

TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

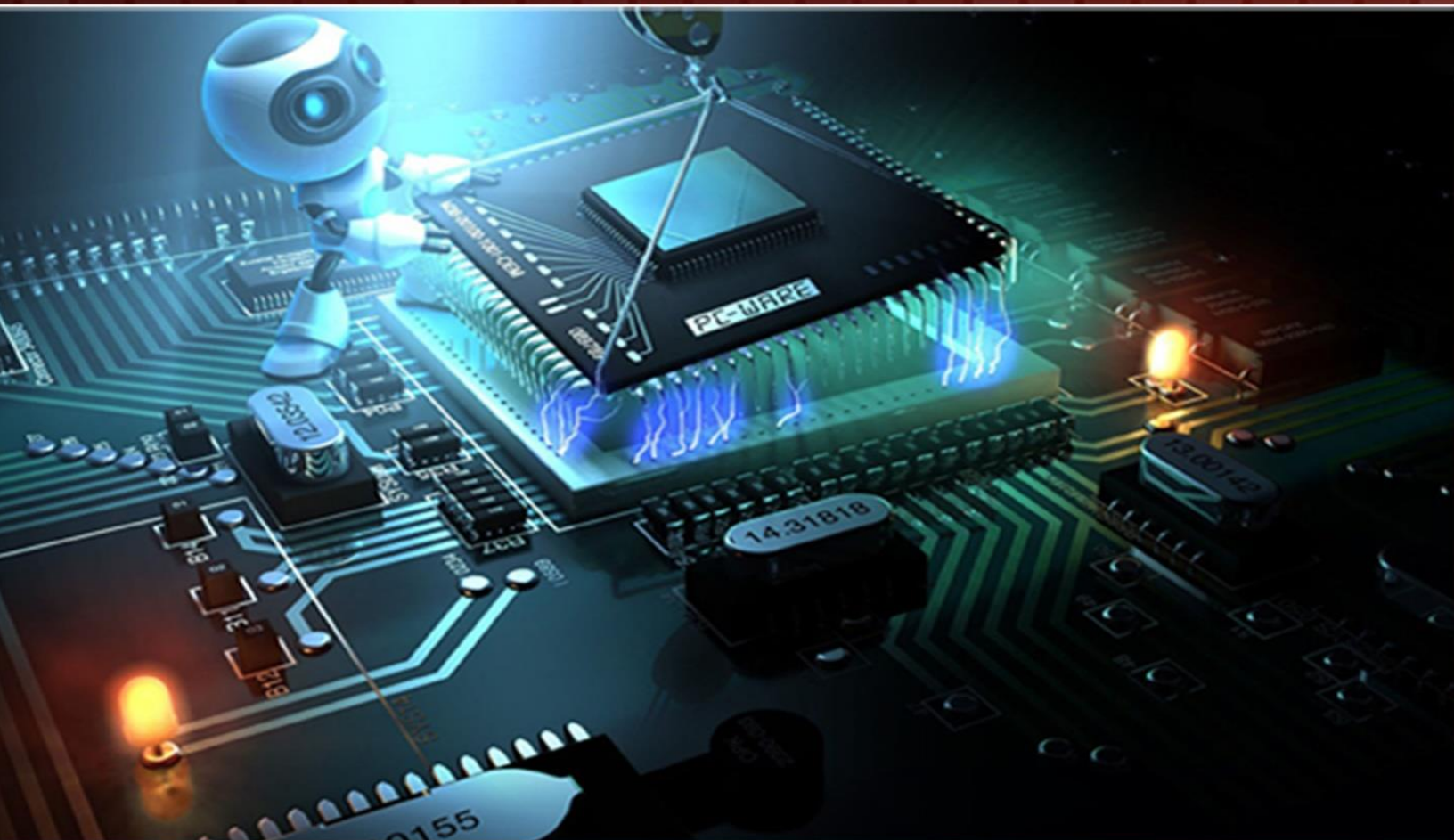


TABLE OF CONTENTS

| | |
|--|-----|
| EDITORIAL ADVISORY BOARD | I |
| DISCLAIMER | II |
| Analysis of Blind Adaptive Multiuser Detection Receivers R. Nirmaladevi and K. Kishan Rao | 1 |
| Determination of Temperature Distribution in a Nuclear Fuel Element Consisting of a Sphere of Fissionable Material and a Spherical Shell of Aluminum Cladding using Finite Element Method J. A. Akpobi and G.O Ariavie | 17 |
| The Modernization of Feed Water Pump Turbine (FWPT) Controller for Nuclear Power Plants Unit 5 and 6 Kwangyoung SOHN, Jayoung LEE and Changhwan CHO | 25 |
| A Novel Method for Calculating Approximate Weight of an Object through its Image Amaresh Chandra Meher, Uddgam Singh and K.K. Senapati | 32 |
| Intelligent Hearing Assistance using Self-Organizing Feature Maps Michael Hodges and Mohamed Zohdy | 40 |
| Identification of Query Forms for Retrieving the Information from Deep Web Nripendra Narayan Das and Ela Kumar | 53 |
| Comparison of Metaheuristic Algorithms for Evolving a Neural Controller for an Autonomous Robot Sergii Zhevzyk and Wilfried Elmenreich | 62 |
| Ontological Support for the Evolution of Future Services Oriented Architectures Bilal Gonen, Xingang Fang, Eman El-Sheikh, Sikha Bagui, Norman Wilde and Alfred Zimmermann | 77 |
| An Analytical Approach of Computational Complexity for the Method of Multifluid Modeling A.K. Borah and P.K. Singh | 91 |
| Impulse Noise Removal Technique Based on Fuzzy Logic Mithilesh Atulkar, A.S. Zadgaonkar and Sanjay Kumar | 99 |
| Development of the ISR3M Model for IS Risk Management Evaluation using the Focus Area Structure according to the MMDPIS Generic Process Mina El maallam and Abdelaziz Kriouile | 106 |
| Identification of Erythemato-Squamous Skin Diseases using Support Vector Machines and Extreme Learning Machines: A Comparative Study towards Effective Diagnosis S. O. Olatunji and Hossain Arif | 124 |

EDITORIAL ADVISORY BOARD

Professor Er Meng Joo

Nanyang Technological University
Singapore

Professor Djamel Bouchaffra

Grambling State University, Louisiana
United States

Prof Bhavani Thuraisingham

The University of Texas at Dallas
United States

Professor Dong-Hee Shin,

Sungkyunkwan University, Seoul
Republic of Korea

Professor Filippo Neri,

Faculty of Information & Communication Technology,
University of Malta,
Malta

Prof Mohamed A Zohdy,

Department of Electrical and Computer Engineering,
Oakland University,
United States

Dr Kyriakos G Vamvoudakis,

Dept of Electrical and Computer Engineering, University of
California Santa Barbara
United States

Dr M. M. Fraz

Kingston University London
United Kingdom

Dr Luis Rodolfo Garcia

College of Science and Engineering, Texas A&M University,
Corpus Christi
United States

Dr Hafiz M. R. Khan

Department of Biostatistics, Florida International
University
United States

Professor Wee SER

Nanyang Technological University
Singapore

Dr Xiacong Fan

The Pennsylvania State University
United States

Dr Julia Johnson

Dept. of Mathematics & Computer Science, Laurentian
University, Ontario,
Canada

Dr Chen Yanover

Machine Learning for Healthcare and Life Sciences
IBM Haifa Research Lab, Israel

Dr Vandana Janeja

University of Maryland, Baltimore
United States

Dr Nikolaos Georgantas

Senior Research Scientist at INRIA, Paris-Rocquencourt
France

Dr Zeyad Al-Zhour

College of Engineering, The University of Dammam
Saudi Arabia

Dr Zdenek Zdrahal

Knowledge Media Institute, The Open University, Milton
Keynes
United Kingdom

Dr Farouk Yalaoui

Institut Charles Dalaunay, University of Technology of
Troyes
France

Dr Jai N Singh

Barry University, Miami Shores, Florida
United States

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

Analysis of Blind Adaptive Multiuser Detection Receivers

¹R. Nirmaladevi and ²K. Kishan Rao

¹Dept. of Electronics & Instrumentation Engineering, KITS Warangal, Andhra Pradesh, INDIA.

²Dept. of Electronics & Communication Engineering, Vaagdevi Group of institutions, Andhra Pradesh, INDIA.

¹nimala123@yahoo.com, ²prof_kkrao@rediffmail.com

ABSTRACT

In this paper, blind adaptive multiuser detection with Mean Output Energy (MOE) are designed in receivers for synchronous MC-DS-CDMA systems in Rayleigh fading channels. Pre and post detection in multiuser detectors are implemented in the system to detect their effects in the outputs. On the basis of the delayed received signals, we proposed Affine Projection Algorithm-link multiuser detector (APA) and its performance is evaluated with the existing multiuser blind detectors (Normalized blind LMS and Kalman Filter). Compared to existing multiuser detectors APA-link detector outperformed the BER performance and provided a tradeoff between performance and computation cost. The analysis is further extended by implementing decorrelating multiuser detector based receivers in SC-DS-CDMA and MC-DS-CDMA to eliminate the MAI caused by other users.

Keywords: MC-DS-CDMA, MOE, APA, BER, MAI, SC-DS-CDMA

1 Introduction

Unlike the adaptive MMSE receivers, blind adaptive interference suppression receivers do not require training sequences. It has been shown in the original work by Honig et al. [1] that, by knowing only the spreading code and the timing of the desired user, the MMSE multiuser receiver can be implemented blindly. They have developed the first blind LMS-based multiuser detection receiver in the absence of multi-path fading. It is based on the minimization of a MAI criterion by using a stochastic gradient approach. However, this method is not convenient in a time-varying environment. For this reason, other approaches have been proposed. (i) On the one hand, blind detectors based on Recursive Least Square RLS [2] and Kalman filter [3] algorithms make it possible to improve the convergence features and tracking capabilities in a dynamic environment and interferers. (ii) On the other hand, Mucchi et al. [4] have proposed a derived version of the pioneering blind LMS-based detector, which makes it possible to operate in a time-varying frequency-selective multi-path fading channels. For this purpose, they first completed channel compensation and time alignment on the signal replicas along each independent path and then combine the resulting signals before or after multiuser detection, resulting in two receiver schemes. The first scheme is called the RAKE blind adaptive multiuser detection receiver where combining is performed after multiuser detection. The second scheme is called pre-detection combining blind adaptive multiuser detection receiver where combining is performed before multiuser detection. The pre-detection combining based receiver has the advantage of using only one detector for the

combined replicas instead of one detector for each signal replica. In addition, according to [4], this yields a remarkable complexity reduction, more reliable decision variable and more robust convergence procedure.

However, the above blind adaptive multiuser detection techniques were only developed for single-carrier DS-CDMA systems. In this work, our purpose is to design blind adaptive multiuser detection receivers for synchronous MC-DS-CDMA systems in Rayleigh fading channels. For this purpose, we first reformulate the ideas presented in [4] to design two blind adaptive receivers for MC-DS-CDMA systems. Namely:

- *The first receiver provides a blind adaptive multiuser detector for each carrier followed by a post-detection combiner.*
- *The second receiver consists of a pre-detection combiner followed by a single blind adaptive multiuser detector.*

To implement them, we have proposed a blind APA-like multiuser detector [5]. The proposed detector can be seen as a generalization of the blind LMS-based detector [1], on the basis of multiple delayed input signal vectors. A comparative study is then carried out with existing blind LMS [1] and Kalman filter [3] based multiuser detectors initially developed for single-carrier DS-CDMA systems.

1.1 Receiver Structure with Post-Detection Combining

In this subsection, we propose a blind adaptive multiuser receiver with post-detection combining for synchronous MC-DS-CDMA systems in time-varying fading channels (see Figure 2.1). Thus, to retrieve the symbol sequence of the first user $d_1(n)$ we first recall the $N \times 1$ discrete-time received vector over the m^{th} carrier given as follows:

$$X_m(n) = \sqrt{P_1} d_1(n) h_m(n) c_1 + \sum_{k=2}^K \sqrt{P_k} d_k(n) h_m(n) c_k + \eta_m(n) \quad (1)$$

Here, as our goal is to suppress the MAI, we assume that the fading processes $\{h_m(n)\} m=1,2,\dots,m$ are available at the receiver. Thus, channel compensation over the m^{th} carrier can be performed in the following manner:

$$\begin{aligned} \underline{X}_m(n) &= \text{Re}(h_m^*(n) X_m(n)) \\ &= \sqrt{P_1} d_1(n) |h_m(n)|^2 c_1 + \sum_{k=2}^K \sqrt{P_k} d_k(n) |h_m(n)|^2 c_k + \text{Re}(h_m^*(n) \eta_m(n)) \end{aligned} \quad (2)$$

where the multiplication with $h_m^*(n) = |h_m(n)| e^{-j\phi_m(n)}$ compensates for the phase and weights the signal amplitude by a positive time-varying factor $|h_m(n)|^2$.

After channel compensation, the resulting vector $\underline{X}_m(n)$ over the m^{th} carrier defines the input to a blind adaptive multiuser detector whose canonical linear representation for user 1 was firstly established in [1], as follows:

$$W_m(n) = c_1 + a_m(n) \quad (3)$$

subject to the constraint:

$$c_1^T a_m(n) = 0 \quad (4)$$

or equivalently, since $c_1^T c_1 = 1$:

$$c_1^T c_1 = 1 \quad (5)$$

Where c_1 is the normalized spreading vector of the first user and $a_m(n)$ is the adaptive part of the detector. Thus, the practical implementation of this detector is achieved by means of two orthogonal filters (see Figure 2): the spreading code of the desired user c_1 and the adaptive part $a_m(n)$ that is used to eliminate the MAI.

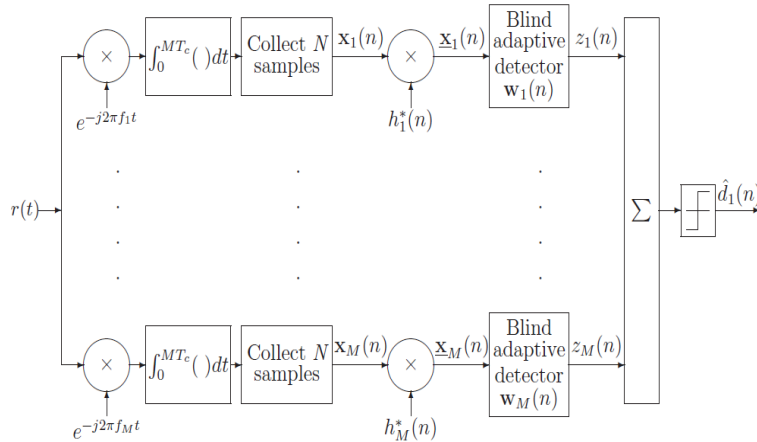


Figure 1: Blind adaptive receiver structure with post-detection combining.

The detector $W_m(n)$ is designed to minimize the Mean Output Energy (MOE) cost function:

$$J_{MOE}[W_m(n)] = E \left[\left| W_m^T(n) \underline{x}_m(n) \right|^2 \right] \quad (6)$$

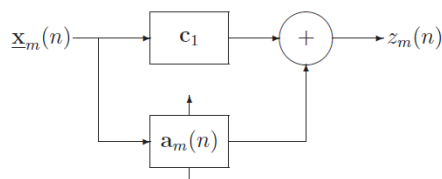


Figure 2: Blind Adaptive Multiuser Detector $W_m(n) = c_1 + a_m(n)$.

It should be noted that the MOE criterion given by (6) is related to a “scaled” version of the MSE criterion given as follows:

$$J_{MSE}[W_m(n)] = E \left[\left| \sqrt{P_1} d_1(n) - W_m^T(n) \underline{X}_m(n) \right|^2 \right] \quad (7)$$

$$= J_{MOE}[W_m(n) - P_1]$$

Where it is assumed that $E \left[|h_m(n)|^2 \right] = \sigma_h^2 = 1$. As P_1 is a constant, minimizing the MOE criterion will also minimize the MSE criterion. Hence, the MOE multiuser detector is equivalent to the MSE based one [1]. Since the MOE criterion does not depend on the data symbol $d_1(n)$, there will be no need for training sequences. Therefore, this leads to the blind adaptive implementations that will be presented in subsection [1.2]

The outputs of the blind adaptive detectors over all carriers are finally combined by a post-detection combiner resulting in the following decision about the desired user data symbol:

$$\hat{d}_1(n) = \text{sgn} \left(\sum_{m=1}^M W_m^T(n) \underline{X}_m(n) \right) \quad (8)$$

In the next subsection, as the post-detection based receiver structure requires M blind adaptive multiuser detectors, we propose a blind adaptive receiver based on pre-detection combining which requires only one multiuser detector.

1.2 Receiver Structure with Pre-Detection Combining

The proposed blind adaptive MOE receiver with pre-detection MRC is shown in Figure 2.3. Thus, after channel compensation and time alignment, the resulting vectors $\{\underline{X}_m(n)\}_{m=1,2,\dots,M}$ given by (2) are combined before detection as follows:

$$\underline{X}_{tot}(n) = \sum_{m=1}^M \underline{X}_m(n) \quad (9)$$

The combined vector $\underline{X}_{tot}(n)$ is then processed with a single blind adaptive multiuser detector $W(n) = c_1 + a(n)$ as in (3), in order to minimize the following MOE criterion:

$$J_{MOE}[W(n)] = E \left[\left| W^T(n) \underline{X}_{tot}(n) \right|^2 \right] \quad (10)$$

Where $a(n)$ is the adaptive part of the detector which satisfies $c_1^T a(n) = 0$. Finally, the symbol decision of the desired user can be obtained as follows:

$$\hat{d}_1(n) \text{sgn} \left(W^T(n) \underline{X}_{tot}(n) \right) \quad (11)$$

Where the detector $W(n)$ will be implemented adaptively in the next subsection.

It should be noted that the pre-detection combining receiver structure has the advantage of reducing greatly the computational cost by using only one detector for the combined vector instead of a detector for each carrier.

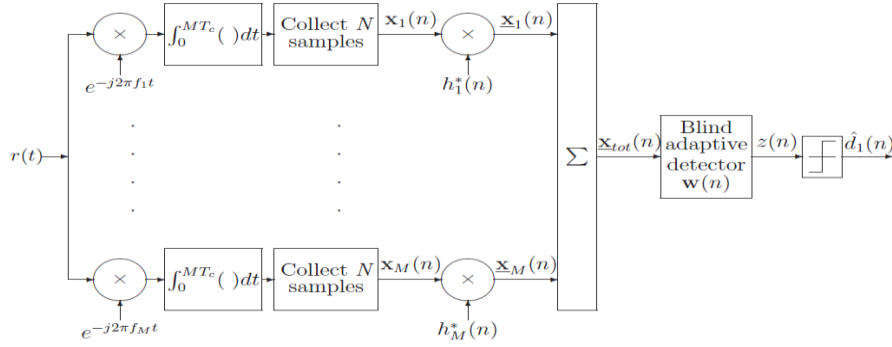


Figure 3: Blind adaptive receiver structure with pre-detection combining.

To implement the post-detection and predetection based blind adaptive receivers, various blind adaptive algorithms are considered in the next subsection.

1.3 Blind Adaptive Implementation

To use unique notations, let $\underline{X}(n)$ and $W(n)$ respectively denote:

- The received vector $\underline{X}_m(n)$ and the filter weights when considering the post-detection receiver structure.
- The combined received vector $\underline{X}_{tot}(n)$ and the filter weights $W(n)$ when using the pre-detection receiver structure.

1.3.1 Normalized Blind LMS Based Multiuser Detector

By using a stochastic gradient approach, Honig et al. [1] have proposed to derive an adaptive implementation for the detector (3) that minimizes the MOE criterion (6). Thus, they first evaluated the unconstrained gradient of the MOE cost function (6) as follows:

$$\nabla J_{MOE}[W(n)] = \nabla E \left[\left| W^T(n) \underline{X}(n) \right|^2 \right] = 2 \left[W^T(n) \underline{X}(n) \right] \underline{X}(n) \quad (12)$$

The projected gradient, orthogonal to c_1 is then obtained:

$$\nabla J_{MOE}[W(n)]_{c_1} = 2 \left[W^T(n) \underline{X}(n) \right] \left[\underline{X}(n) - (c_1^T \underline{X}(n)) c_1 \right] \quad (13)$$

Therefore, a stochastic gradient blind algorithm to update the adaptive part of the detector can be written as follows:

$$\begin{aligned} a(n) &= a(n-1) - \mu \nabla J_{MOE} [W(n-1)]_{c_1} \\ &= a(n-1) - \mu z(n) \left[\underline{X}(n) - z_{MF}(n) c_1 \right] \end{aligned} \quad (14)$$

Where $z(n) = W^T(n-1) \underline{X}(n)$ the output of the detector is, $z_{MF}(n) = c_1^T \underline{X}(n)$ is the output of the conventional matched-filter, and μ is the step-size that controls the adaptation speed.

To avoid the gradient noise amplification problem, a normalized version of this algorithm (as the NLMS algorithm) can be written as follows:

$$a(n) = a(n-1) - \frac{\mu_N}{\delta + \|\underline{X}(n)\|^2} Z(n) [\underline{X}(n) - z_{MF}(n) c_1] \quad (15)$$

Where $\mu_N \in (0, 2)$ is the normalized step-size and δ is a small positive regularization constant that insures stability when $\|\underline{X}(n)\|^2$ is too small.

1.3.2 Blind APA-like Multiuser Detector

To improve the convergence features in high MOE environments and time-varying fading scenarios, we propose to generalize the algorithm in (15) by using L delayed input signal vectors [5]. Toward this end, we first define the following received and code matrices of L column each:

In addition, the MOE cost function in (6) is modified to account for L delayed input signal vectors:

$$X(n) = [\underline{X}(n) \quad \underline{X}(n-1) \quad \dots \quad \underline{X}(n-L+1)] \quad (16)$$

$$C_1 = [c_1 \quad c_1 \quad \dots \quad c_1] \quad (17)$$

$$J_{MOE}[W(n)] = E \left[\left\| X^T(n) W(n) \right\|^2 \right] \quad (18)$$

Taking the unconstraint gradient of the MOE cost function:

$$\nabla J_{MOE}[W(n)] = \nabla E \left[\left\| X^T(n) W(n) \right\|^2 \right] = 2X(n) [X^T(n) W(n)] \quad (19)$$

The projected gradient, orthogonal to C_1 satisfies:

$$\nabla J_{MOE}[W(n)]_{C_1} = 2 \left[X(n) - C_1 (C_1^T X(n)) \right] [X^T(n) W(n)] \quad (20)$$

Then, a stochastic gradient algorithm that updates the adaptive part of the detector (3) can be written as follows:

$$a(n) = a(n-1) - \mu \nabla J_{MOE}[W(n-1)]_{C_1} \quad (21)$$

Where μ is the step-size.

Substituting (20) in (21) and introducing a factor $[\delta I_L + X^T(n) X(n)]^{-1}$ similar to that in the APA, a new blind algorithm (APA-like) to update $a(n)$ can be expressed by [5]:

$$a(n) = a(n-1) - \mu_N \left[X(n) - C_1 Z_{MF}(n) \right] \left[\delta I_L + X^T(n) X(n) \right]^{-1} z(n) \quad (22)$$

Where $z(n) = X^T(n) w(n-1)$, $Z_{MF}(n) = C_1^T X(n)$, $\mu_N \in (0, 2)$ and δ is the regularization constant.

To insure that the orthogonally condition $c_1^T a(n) = 0$ is satisfied at each iteration, we replace $a(n)$ by its orthogonal projection onto c_1 :

$$a(n) = a(n) - c_1^T a(n) c_1 \quad (23)$$

It should be noted that when $L = 1$, the blind APA-like multiuser detector (22) reduces to the normalized blind LMS multiuser detector (15).

1.3.3 Blind Multiuser Detection Based on Kalman Filtering

In [3], Zhang et al. have proposed to use an alternative standard representation for the blind adaptive multiuser detector:

$$W(n) = c_1 - A_1 a(n) \quad (24)$$

Where the columns of the $N \times N - 1$ matrix A_1 span the null space of c_1 , i.e.:

$$c_1^T A_1 = 0 \quad (25)$$

It should be noted that A_1 can be precomputed off-line via one of many orthogonalization procedures such as the Gram-Schmidt orthogonalization. Unlike (3), the adaptive part $a(n)$ in (24) is now of size $(N - 1) \times 1$ and has the advantage of being unconstrained.

Let us define the output of the detector as follows:

$$z(n) = W^T(n) \underline{X}(n) \quad (26)$$

then $z(n)$ has zero-mean and its variance is given by (due to equations (6) and (7)):

$$E\left[\left[Z(n)\right]^2\right] = J_{MOE}[W(n)] = J_{MSE}[W(n)] + P_1 \quad (27)$$

Thus, when the detector is optimal (i.e., $J_{MSE}[W(n)]$ attains its MMSE value), the variance of $z(n)$ corresponds to the minimum MOE and is dominated by the power of the desired user P_1 .

Substituting (24) in (26) results in the following measurement equation:

$$Z_{MF}(n) = \underline{d}^T(n) a(n) + Z(n) \quad (28)$$

Where $Z_{MF}(n) = c_1^T \underline{X}(n)$ and $\underline{d}^T(n) = \underline{X}^T(n) A_1$.

If the detector is assumed to be time-invariant, one can write:

$$a(n) = a(n-1) \quad (29)$$

As (28) and (29) define a state-space representation of the adaptive part of the detector, Kalman filtering makes it possible to recursively update $a(n)$ [3].

1.3.4 Computational Cost of the Various Algorithms

Here, we provide the computational cost of the various blind adaptive multiuser detectors when used to implement the post-detection and pre-detection combining receiver structures. According to Table 1,

the predetection combining receiver structure with the Table 1: Computational complexity of the various blind adaptive multiuser detectors when considering the post-detection and predetection combining receiver structures.

Table: 1

| Blind Adaptive Detector | Post-detection Combining | Pre-detection Combining |
|-------------------------|--------------------------|-------------------------|
| LMS based detector | $O(MN)$ | $O(N)$ |
| APA-like Detector | $o(MNL^2)$ | $o(NL^2)$ |
| Kalman Based Detector | $O(MN^2)$ | $O(N^2)$ |

Various blind adaptive detectors has the advantage of greatly reducing the computational cost when compared with the post-detection combining based one. In addition, the proposed blind APA-like detector has the advantage of providing a scalable complexity by adjusting the parameter L which is usually much less than the filter length (i.e., $L \ll N$). The scalable complexity of the APA-like detector can be traded with the performance as we will see in the simulation results presented in the next subsection. Therefore, the APA-like detector with the pre-detection combining scheme is a priori preferable to design the receiver.

2 Simulation Results

In this subsection, we first carry out a comparative study between the following blind adaptive multiuser detectors:

- The normalized version of the standard blind LMS multiuser detector [1],
- The proposed blind APA-like multiuser detector,
- The blind Kalman filter based multiuser detector [3], when they are used to implement the pre-detection and post-detection combining receiver structures.

In addition, we compare the performances of the proposed blind APA-like detector with the training based APA filter presented in subsection III.

A synchronous MC-DS-CDMA system with K active users and $M = 4$ carriers is considered. The spreading sequences used are gold codes of length $N = 31$. The fading processes $\{h_m(n)\}_{m=1,2,\dots,M}$ are generated according to the complex Gaussian distribution with zero-mean and unit-variance. User 1 is assumed to be the desired user with SNR per transmitted carrier kept constant at 10 dB.

In the first example, a high MAI scenario is assumed with 14 multiple-access interfering users (i.e., $K = 15$), among which five users have ISR (see equation 1.50) of 10 dB each, five users have ISR of 20 dB each, two users have ISR of 30 dB each and two other users have ISR of 40 dB each. According to Figure 2.4, the pre-detection combining receiver slightly outperforms the post-detection receiver with the various blind adaptive detectors, when considering the BER performance. In addition, the proposed blind APA-like detector provides much better BER performance and convergence features than the normalized blind LMS detector. Nevertheless, the blind Kalman detector yields the best BER performance and convergence features, but at the price of increased computational cost (see Table 2.1).

Figure 2.5 demonstrates the average Signal to Interference- plus-Noise Ratio (SINR) performance of the various blind detectors in the pre-detection combining receiver. The average SINR is defined as follows:

$$SINR(n) = \frac{\sum_{l=1}^{loop} \left| W^T(n) \left(\pi \sqrt{P_1} d_1(n) c_{1,tot} \right) \right|^2}{\sum_{l=1}^{loop} \left| W^T(n) \left[X(n) - \sqrt{P_1} d_1(n) c_{1,tot} \right] \right|^2} \quad (30)$$

Where

$$c_{1,tot} = \sum_{m=1}^M \left| h_m(n) \right|^2 c_1 \quad (31)$$

and the average is performed over $loop = 300$. The SINR improvement of the proposed APA-like detector is better than that obtained with the normalized blind LMS detector and approaches that of the blind Kalman detector, when L is getting higher. Therefore, for the various reasons mentioned above, we recommend to use the blind APA- like detector with the pre-detection combining scheme to design the receiver.

To illustrate the advantages of the blind adaptive MOE detectors over the training based adaptive MMSE filters presented in the previous section, we consider a high MAI scenario with 8 interfering users for which the $ISR=20$ dB (this hence corresponds to a severe near-far scenario). Figure 6 shows the SINR performance improvement of the proposed APA-like detector and the training based APA filter in the pre-detection combining receiver. According to this figure, the blind detector yields better SINR performance improvement than the training based filter which suffers from slow convergence in this severe near-far scenario. This is due to the fact that the blind detector uses the code sequence of the desired user in addition to an adaptive part, whereas the training based filter starts adaptation from a zero initial weight vector. Thus, in a high MAI environment, it is recommended to use a blind adaptive detector to efficiently suppress the MAI and to mitigate the near-far problem.

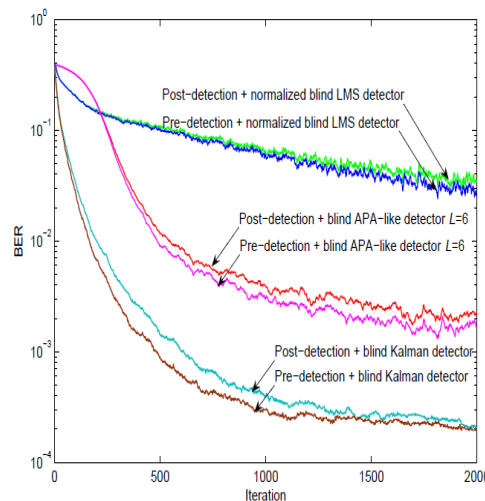


Figure 4: BER performance of the pre-detection and post-detection receivers with the various blind adaptive multiuser detectors.

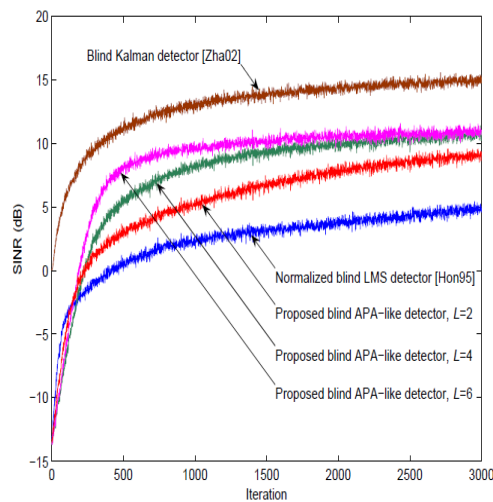


Figure 5: SINR improvement of the various blind adaptive detectors in the predetection combining receiver.

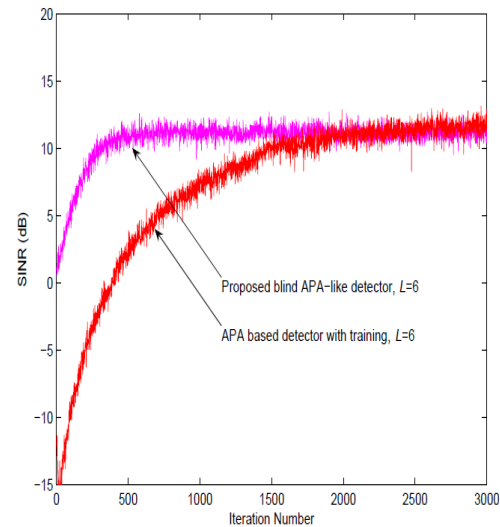


Figure 6: SINR improvement of the proposed blind APA-like detector and the training based APA filter in the pre-detection combining receiver.

3 Receiver Design Based on Decorrelation Detection

As an alternative to the MMSE multiuser receivers, the decorrelating multiuser detector based receivers are designed to completely eliminate the MAI caused by other users, using the spreading sequences of all users [7]. In this section, we first present the state of the art on single-carrier DS-CDMA receiver design based on the decorrelating detector. Then, we propose a MC-DS-CDMA receiver structure consisting of a decorrelating detector, a Kalman filter based channel estimator and a MRC [6].

3.1 Decorrelation detection for single-carrier DS-CDMA

Given the spreading codes of all active users $\{c_k\}_{k=1,\dots,K}$ the decorrelating multiuser detector for user 1 can be written in the following form [8]:

$$W_1 = \sum_{k=1}^K [R^{-1}]_{1k} c_k \quad (32)$$

Where $R = [c_1 \ c_2 \ \dots \ c_k]^T [c_1 \ c_2 \ \dots \ c_k]$ the normalized cross-correlation matrix of the spreading is vectors and $[R^{-1}]_{ij}$ denotes the $(i, j)^{th}$ element of the inverse of the matrix R .

Such decorrelating multiuser detector satisfies:

$$W_1^T c_1 = 1 \quad (33)$$

and

$$W_1^T c_k = 0, \quad k = 2, \dots, K \quad (34)$$

Equations (33) and (34) can be verified as follows:

$$\begin{aligned}
 W_1^T c_k &= \sum_{i=1}^K [R^{-1}]_{ii} c_k \\
 &= \sum_{i=1}^K [R^{-1}]_{ii} [R]_{ik} \\
 &= [R^{-1}R]_{1k} \\
 &= [I]_{1k} \\
 &= \begin{cases} 1 & k=1 \\ 0 & k=2, \dots, K \end{cases}
 \end{aligned} \tag{35}$$

It is obvious from (34) that the decorrelating detector is orthogonal to the subspace spanned by the spreading sequences of all interfering users. Therefore, the decorrelating detector can completely eliminate the MAI and can achieve optimal near-far resistance whatever the ISR of the interfering users. It should be emphasized that the decorrelating detector does not require the knowledge of the power of all users.

The decorrelating detector based receiver for synchronous DS-CDMA systems is first proposed by Lupas and Verdú [9]. The generalization to asynchronous DS-CDMA systems is then reported in [10]. However, only AWGN channels are considered. Since then, the extension of these receivers to operate in fading channels has been extensively studied by several authors. Thus, Zvonar et al. [11] have analyzed the performance of the decorrelating detector in slowly time-varying flat-fading channels. Nevertheless, the fading processes are assumed to be available at the receiver. In [12], Kawahara et al. have proposed to combine decorrelation multiuser detection with channel estimation for asynchronous DS-CDMA systems in multi-path slowly fading channels. The fading processes are estimated by using a training based RLS algorithm. In [13], a decorrelating detector based receiver structure is considered for application in rapidly time-varying Rayleigh fading channels where the fading processes are estimated adaptively using the LMS or the RLS algorithm. Recently, in [14], Wu et al. have compared the performances of a Kalman filter based channel estimator combined with various multiuser detectors, such as the decorrelating detector, the decision-feedback detector, the parallel and successive interference cancellation. According to their study, the decorrelating detector is the most robust detector against the MAI and the near-far problem.

3.2 MC-DS-CDMA receiver based on decorrelation detection

Although decorrelation multiuser detection have been extensively used for single-carrier DS-CDMA systems, few approaches are developed for MC-DS-CDMA systems. Thus, Yang et al. [15] have proposed several decorrelation multiuser detection schemes for TF-domain spreading MC-DS-CDMA systems in AWGN channels. However, they did not investigate these schemes in fading channels. In [16], the authors have developed a multichannel joint detection scheme for MC-DS-CDMA systems in time-invariant frequency-selective fading channels. The scheme consists of a decorrelating detector followed by a RAKE multi-path combiner for each carrier. Nevertheless, perfect channel knowledge is assumed at the receiver.

Here, we propose to extend, to the multi-carrier case, the combination of decorrelation multiuser detection and Kalman channel estimation scheme presented in [14]. In particular, our scheme [6] operates in three steps (see Figure 7):

1. The decorrelating multiuser detector is carried out along each carrier to completely eliminate the MAI,
2. The fading channel responses, modeled by Auto Regressive (AR) processes, are estimated by using Kalman filtering,
3. The fading processes estimates are fed into a frequency diversity MRC rule to obtain the data symbol estimate.

Thus, to retrieve the desired symbol sequence of the first user $d_1(n)$ from the received signal, let us recall the $N \times 1$ discrete time received vector over the m^{th} carrier given as follows:

$$X_m(n) = \sqrt{P_1} d_1(n) h_m(n) c_1 + \sum_{k=2}^K \sqrt{P_k} d_k(n) h_m(n) c_k + \eta_m(n) \quad (36)$$

Where the fading processes $\{h_m(n)\}_{m=1,2,\dots,M}$ are assumed to be rapidly time-varying. At that stage, the received vector at the m^{th} carrier $X_m(n)$ is processed by the decorrelating detector W_1 given by (32).

By taking into account (33) and (34), the decorrelating detector yields the following observation:

$$y_m(n) = W_1^T X_m(n) = \sqrt{P_1} d_1(n) h_m(n) + v_m(n) \quad (37)$$

Where $v_m(n) = W_1^T \eta_m(n)$ is a zero-mean Gaussian noise with variance $\sigma_v^2 = \sigma_\eta^2 [R^{-1}]_{11}$. From equation (37), the decorrelating detector is able to completely eliminate the MAI caused by other users, but at the expense of slightly enhancing the additive noise.

Based on the observations $\{y_m(n)\}_{m=1,\dots,M}$ and by using an AR model for the fading processes, Kalman filtering can be carried out to provide an estimation of the fading processes $\{\hat{h}_m(n)\}_{m=1,2,\dots,M}$ [6].

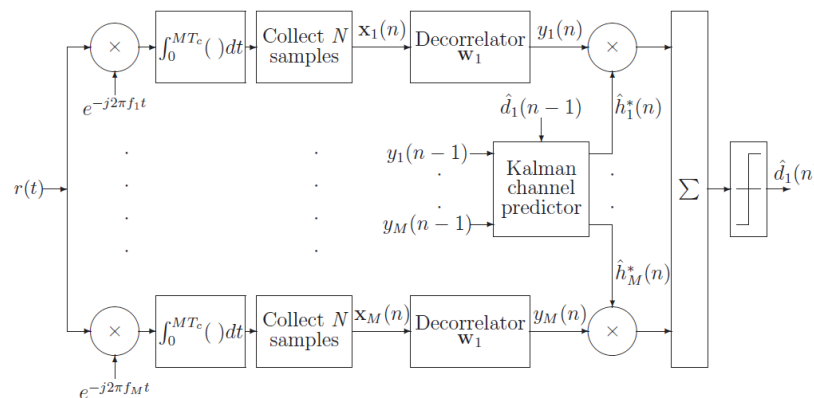


Figure 7: MC-DS-CDMA receiver structure with decorrelation detection for user 1.

The estimation of AR fading channels based on Kalman filtering will be investigated in chapter 3. This also includes the development of several channel estimation techniques that can be directly applied to the proposed receiver in Figure 7.

Finally, MRC makes it possible to provide the estimate of the desired user data symbol as follows:

$$\hat{d}_1(n) = \text{sgn} \left(\text{Re} \left(\sum_{m=1}^M h_m^*(n) y_m(n) \right) \right) \quad (38)$$

3.3 Simulation results

In this subsection, we first carry out a comparative simulation study between the proposed decorrelating detector based receiver and the correlator based one presented in [15].

A synchronous MC-DS-CDMA system is considered with M carriers and K active users, each using a gold code of length $N = 31$ to spread his information. The fading processes $\{h_m(n)\}_{m=1,2,\dots,M}$ are generated according to the modified Jakes model with $L_o = 16$ oscillators and Doppler rate $f_d T_b = 0.05$. Here, to focus our attention on the effect of the MAI on both receivers, the fading processes are assumed to be available at the receiver. The performance of the decorrelating detector based receiver when the fading processes are estimated by various approaches will be presented in the simulation results.

Figure 8 shows the effects of the ISR on the BER performance of both receivers for number of carriers equal to $M = 1$ and $M = 3$. On the one hand, the BER of the decorrelating detector based receiver does not depend on the ISR and, hence, it is near-far resistant. On the other hand, the BER of the correlator detector based receiver is highly dependent on the ISR, where degradation of the BER can be noticed starting at low ISR. In addition, for the decorrelating detector based receiver, a high frequency diversity gain is obtained when increasing the number of carrier from $M = 1$ to $M = 3$. This is not the case for the correlator based receiver when the ISR is high.

According to Figure 9, increasing the number of users greatly increases the BER of the correlator based receiver. On the other hand, the decorrelating detector based receiver is insensitive to the number of users. This is due to the fact that, contrary to the correlator based receiver, the decorrelating detector based receiver can completely eliminate the MAI caused by other users.

Here, we also present a simulation example that illustrates the performance of the proposed decorrelating detector based receiver compared with that of the adaptive receivers proposed in the previous sections. Namely, the training based Separate Detection (SD) receiver with APA and the post-detection combining receiver with blind APA-like detector. To focus on the MAI suppression capabilities of these receivers, we consider only AWGN channels (without the effect of fading) with $K = 10$, $M = 3$, SNR=5 dB and ISR=15 dB.

According to Figure 10, the decorrelating detector based receiver provides the lowest BER results without the need for any training period. This is due to the fact that it uses the spreading codes of all users and, hence, can completely eliminate the MAI. The blind APA based receiver has faster convergence than the training APA based one, but it results in higher steady state BER. To take the advantages of both adaptive receivers, their combination can also be considered. Thus, the blind APA-

like detector is first carried out up to iteration number 500. As the blind APA-like detector uses the spreading code of the desired user, this will ensure the fast suppression of high amount of MAI.

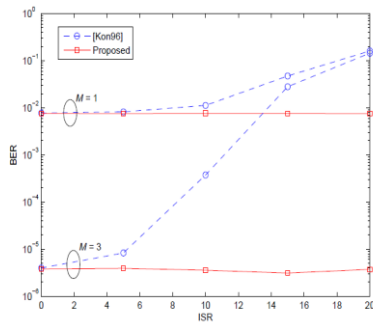


Figure 8: BER performance versus ISR for the proposed receiver and the correlator based one [Kon96]. Number of carriers considered are $M = 1$ and $M = 3$. $K = 10$ and $\text{SNR}=15$ dB.

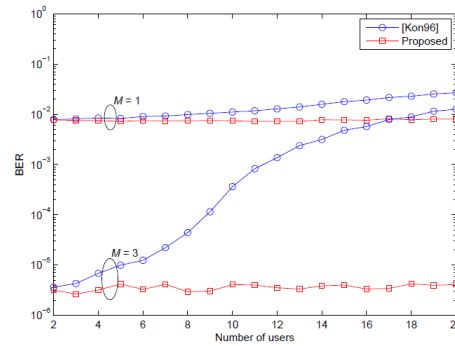


Figure 9: BER performance versus number of active users K for the proposed receiver and the one in [18]. Number of carriers considered are $M = 1$ and $M = 3$. $\text{ISR}=10$ dB and $\text{SNR}=15$ dB.

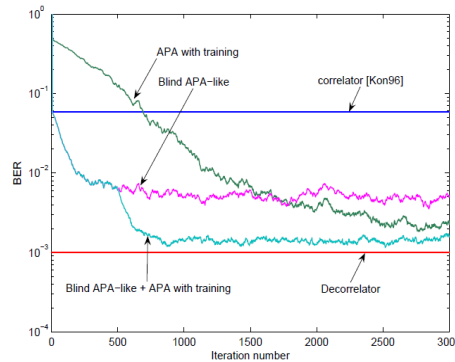


Figure 10: BER performance versus iteration number for the various receivers. $M = 3$, $K = 10$, $\text{ISR}=15$ dB and $\text{SNR}=5$ dB.

At that stage, starting from the final filter weight values provided by the blind APA-like detector (at iteration 500), a training based APA can then be used to continue eliminating the residual MAI and can provide a BER performance close to that of the decorrelating detector based receiver.

4 Conclusions

In this work, we have proposed five MAI suppression receivers for MC-DS-CDMA systems in Rayleigh fading channels.

We first presented two adaptive MMSE receiver structures based on adaptive filters such as NLMS, APA and RLS. The so-called Separate Detection (SD) structure consists in considering a particular adaptive filter for each carrier, whereas the so-called Joint Detection (JD) structure is defined by the concatenation of the adaptive filter weights dedicated to each carrier. Simulation results show that the JD structure provides lower BER than the SD structure. In addition, the order of complexity of both structures are the same when considering the APA, but this is not the case for the RLS filter where the JD structure has higher computational cost than the SD one. Therefore, APA in the JD structure corresponds to a trade-off between performance and computational cost.

However, as adaptive receivers require training sequences, we have proposed two blind adaptive Mean Output Energy (MOE) receivers based on an APA-like multiuser detector, where only the spreading sequence and the timing of the desired user are required. The so-called post-detection combining based receiver structure provides a blind adaptive detector for each carrier where combining is performed after detection. The so-called pre-detection combining based receiver structure uses a single blind adaptive detector after combining the signals of all carriers. Therefore, the computational cost of the pre-detection combining based receiver is much less than that of the post-detection combining based one. Simulation results show that the pre-detection combining based receiver provides slightly lower BER results than the post-detection combining based one. In addition, the comparative study is carried out with existing blind LMS and Kalman filter based detectors shows that the proposed APA-like detector can provide a trade-off between performance and computational cost. Furthermore, the proposed blind APA-like detector is shown to outperform the training based APA in severe near-far scenarios, when considering the BER and SINR improvement.

When the spreading codes of all active users are available, we propose a receiver structure based on the decorrelating detector which includes also a Kalman channel estimator and a MRC [6]. The comparative simulation study we have carried out shows that, while the correlator based receiver is highly sensitive to the MAI and the near-far problem, the proposed receiver can completely eliminate the MAI and is insensitive to the near-far problem. In addition, the proposed receiver can provide approximately the same BER performance as the adaptive receivers without the need of any training sequences for MAI suppression.

Nevertheless, the design of receivers usually requires the explicit estimation of the fading process over each carrier to achieve optimal diversity combining and coherent symbol detection.

REFERENCES

- [1]. M. Honig, U. Madhow, and S. Verdu, "*Blind adaptive multiuser detection*", *IEEE Trans. Inf. Theory*, vol. 41, no. 7, pp.944 -960 1995.
- [2]. Yihai Zhang, Wu-Sheng and T.Aron Gulliver, "*Recursive multi-user detection for DS-WMB system*", *IEEE Conf. Communications, Computers and Signal Processing (PACRM)*, pp- 534-537, Aug., 2005.
- [3]. X. D. Zhang and W. Wei, "*Blind adaptive multiuser detection based on Kalman filtering*", *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp.87 -95 2002.
- [4]. L. Mucchi, S. Morosi, E. Del Re and R. Fantacci, "*A new algorithm for blind adaptive multiuser detection in frequency selective multipath fading channel*," *IEEE Trans. On wireless commun.*, vol. 3, no1, pp. 235-247, January 2004.
- [5]. A. Carini, G.L. Sicuranza, "*Optimal Regularization Parameter of the Multichannel Filtered- x Affine Projection Algorithm*", *IEEE Transactions on Signal Processing*, v.55 n.10, p.4882-4895, October 2007.

- [6]. S. Alireza Banani, Rodney G. Vaughan "Blind channel estimation for MRC systems with maneuvering transmit/receive terminals", Canadian Conference on Electrical and Computer Engineering, (CCECE 2010), Calgary, Alberta, Canada, 2-5 May, 2010.
- [7]. J.Ravindrababu, P.Venumadhav, E.V.Krishna Rao,"BER Performance of Multi-User MC-DS-CDMA using spreading codes in fading Channels," IEEE ICCIC-2011 pp. 301-304.
- [8]. S. Ulukus and R. D. Yates, "A blind adaptive decorrelating detector for CDMA systems," *IEEE J. Select. Areas Commun.* vol. 16, pp. 1530–1541, Oct. 1998.
- [9]. R. Lupas and S. Verd'u, "Linear Multiuser Detectors for Synchronous Code Division Multiple-Access Channels" *IEEE Transactions on Information Theory*, vol. 35, pp. 123--136, January 1989.
- [10]. Lupas R., and S. Verdd, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, pp. 496-508, 1990.
- [11]. Z. Zvonar, "Combined multiuser detection and diversity reception for wireless CDMA systems," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 205–211, Feb. 1996.
- [12]. T. Kawahara and T. Matsumoto, "Joint decorrelating multiuser detection and channel estimation in asynchronous CDMA mobile communications channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 506–515, Aug. 1995.
- [13]. Shakya, F.H. Ali, and E. Stipidis, "Improved successive interference cancellation for DS-CDMA using constant modulus algorithm", In International Symposium on Communications Theory and Applications, Amblseide, UK, 1-6, July 2007.
- [14]. S.-H. Wu, U.Mitra and C.-C.J.Kuo, "Performance of Linear multistage receivers for Ds-CDMA in frequency Selective fading channels", *IEEE Trans. Inf. Theory*, Vol.51, no.10, pp-3493-3517, Oct.2005.
- [15]. Lie-Liang Yang, *Multicarrier Communications*, John Wiley Sons, 2009.

Determination of Temperature Distribution in a Nuclear Fuel Element Consisting of a Sphere of Fissionable Material and a Spherical Shell of Aluminum Cladding using Finite Element Method

¹J. A. Akpobi and ²G.O Ariavie

¹Department of Production Engineering,

^{1,2}Mechanical Engineering Department,

^{1,2}University of Benin, Benin City, Edo State, Nigeria.

²Federal University of Petroleum Resources, Effurun, Delta State, Nigeria.

¹akpobi@uniben.edu, ²ariavie.godfrey@fupre.edu.ng

ABSTRACT

This paper involves the use of Rayleigh-Ritz finite element method to determine the temperature distribution in a nuclear fuel element consisting of a sphere of fissionable material and a spherical shell of aluminum cladding. The differential equation is a one – dimensional second order differential problem. The finite solutions obtained when compared with the exact solutions shows that the accuracy increases as the number of elements increases with decrease in error, and this was shown graphically. It can be stated that finite element method is an accurate method for determining the temperature distribution in a nuclear fuel element consisting of a sphere of fissionable material and a spherical shell of aluminum cladding.

Keywords: Rayleigh – Ritz Finite Element method, temperature distribution, nuclear fuel, aluminum cladding.

1 Introduction

A nuclear fuel element for use in the core of a nuclear reactor is disclosed and has a composite cladding having a substrate and a metal barrier metallurgically bonded on the inside surface of the substrate so that the metal barrier forms a shield between the substrate and a nuclear fuel material held within the cladding.

A number of researchers have used finite element and other method to study temperature distribution. In recent times, the heat generation due to fission within a nuclear fuel rod is not uniform and for a cylindrical fuel rod, the heat generation is given by [1]. The fuel elements are usually long cylindrical rod or rectangular plates of uranium (or thorium) enclosed by cladding. The uranium may be in the pure metallic form, in the form of a compound such as uranium dioxide UO_2 or in the form of an alloy with another metal such as aluminum or zirconium in [2]. Reference [3] analyzed the temperatures and cooling rate that arises during welding. [4] determined temperature distribution on cutting tool in end-milling. ABACUS software based on finite element method was used to study the temperature and heat

flux changes in a nuclear fuel rod by [5]. Reference [6] proposes to analyze in-pile fission gas release from UO₂ fuel. The finite element model describes fission gas transfer from the grain interior to grain boundaries by simultaneous mechanisms of diffusive flow and boundary sweeping considering the effect of irradiation induced resolution when gas amount in grain. [7] used finite element method to determine the temperature distribution on TRISO fuel kernel. In [8] finite element is used to analyze the thermo-structural behaviour of cladding process. Reference [9] used finite element method to solve the problem of stress distribution in a cylindrical nuclear fuel element with a graphite matrix and spherical inclusions. The governing one-dimensional equation for heat transfer in the nuclear fuel element of a fissionable material and aluminum cladding is given in [10].

It is obvious that a number of researchers seem not to have analyzed the temperature distribution in a nuclear fuel element of a spherical form consisting of a sphere of fissionable material surrounded by a spherical shell of aluminum cladding. Hence, this paper using finite element analysis tends to fill this gap.

2 Methodology

2.1 Problem

Consider a nuclear fuel element of spherical form, consisting of a sphere of “fissionable material surrounded by a spherical shell of aluminum “cladding” as shown in the figure below. Nuclear fission is a source of thermal energy, which varies non-uniformly from the center of the sphere to the interface of the fuel element and the cladding. We wish to determine the temperature distribution in the nuclear fuel element and the aluminum cladding.

The governing equations for the two regions are the same, with the exception that there is no heat source term for the aluminum cladding. We have

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 k_1 \frac{dT_1}{dr} \right) = q_1 \quad \text{for} \quad 0 \leq r \leq R_f \quad (1)$$

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 k_2 \frac{dT_2}{dr} \right) = 0 \quad \text{for} \quad R_f \leq r \leq R_c \quad (2)$$

Where subscript 1 and 2 refer to the nuclear fuel element and cladding, respectively. The heat generation in the nuclear fuel element is assumed to be of the form

$$q_1 = q_0 \left[1 + c \left(\frac{r}{R_f} \right)^2 \right] \quad (3)$$

Where q_0 and c are constants depending on the nuclear material. The boundary conditions are;

$$kr^2 \frac{dT_1}{dr} = 0 \quad \text{at} \quad r = 0 \quad (4)$$

$$T_1 = T_2 \quad \text{at} \quad r = R_f, \quad \text{and} \quad T_2 = T_0 \quad \text{at} \quad r = R_c \quad (5)$$

Use eight linear elements to determine the finite element solution for the temperature distribution and compare the nodal temperature with the exact solution.

$$T_1 - T_0 = \frac{q_0 R_f^2}{6k_1} \left\{ \left[1 - \left(\frac{r}{R_f} \right)^2 \right] + \frac{3}{10} c \left[1 - \left(\frac{r}{R_f} \right)^4 \right] \right\} + \frac{q_0 R_f^2}{3k_2} \left(1 + \frac{3}{5} c \right) \left(1 - \frac{R_f}{R_c} \right) \quad (6)$$

$$T_2 - T_0 = \frac{q_0 R_f^2}{3k_2} \left(1 + \frac{3}{5} c \right) \left(\frac{R_f}{r} - \frac{R_f}{R_c} \right) \quad (7)$$

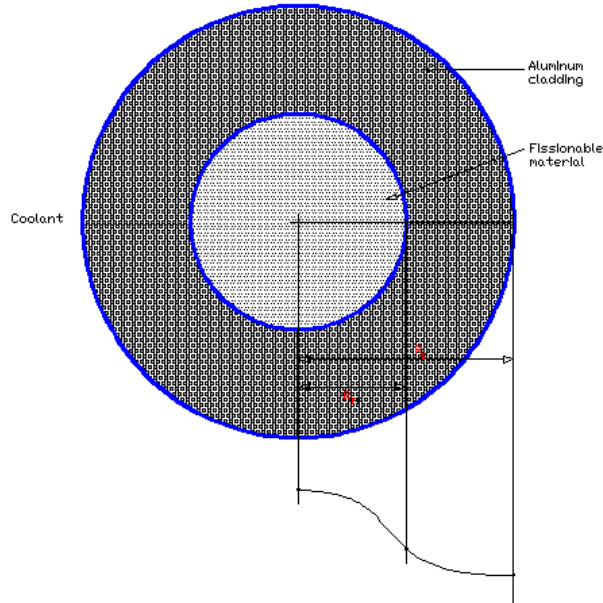


Figure 1: A nuclear fuel elements of spherical form

2.2 Solution

Using Rayleigh – Ritz Finite Element method.

Since the problem is a two in one problem, we consider both governing equations separately;

Considering the governing equation of the **fissionable material** and putting it in residual form as shown below;

$$-\frac{d}{dr} \left(r^2 k_1 \frac{dT_1}{dr} \right) = q_1 r^2 \quad (8)$$

we develop the weak form by multiplying through by the weight function (w), equate to zero and then integrate over the whole area. This gives rise to the equation below;

$$0 = 2\pi \left(\int_{r_a}^{h+r_a} \frac{dw}{dr} r^2 k_1 \frac{dT_1}{dr} - q_1 w r^2 \right) dr - \left(2\pi w k_1 r^2 \frac{dT_1}{dr} \right) \Big|_{r_a}^{h+r_a} \quad (9)$$

The general weak form for this fissionable material is therefore written as;

$$0 = \left(\int_{r_a}^{h+r_a} \frac{dw}{dr} r^2 k_1 \frac{dT_1}{dr} - q_1 w r^2 \right) dr - \left(w k_1 r^2 \frac{dT_1}{dr} \right) \Big|_{r_a}^{h+r_a} \quad (10)$$

Considering the governing equation of the **Aluminum Cladding** and putting it in residual form as shown below;

$$-\frac{d}{dr} \left(r^2 k_2 \frac{dT_2}{dr} \right) = 0 r^2 \quad (11)$$

we develop the weak form by multiplying through by the weight function (w), equate to zero and then integrate over the whole area. This gives rise to the equation below;

$$0 = 2\pi \left(\int_{r_a}^{h+r_a} \frac{dw}{dr} r^2 k_2 \frac{dT_2}{dr} \right) dr - \left(2\pi w k_2 r^2 \frac{dT_2}{dr} \right) \Big|_{r_a}^{h+r_a} \quad (12)$$

Where $(r_a, h+r_a)$ is the domain of the element along its radius.

Since there is no heat loss, the weak form for the Aluminum Cladding is therefore written as;

$$0 = \left(\int_{r_a}^{h+r_a} \frac{dw}{dr} r^2 k_2 \frac{dT_2}{dr} \right) dr - \left(w k_2 r^2 \frac{dT_2}{dr} \right) \Big|_{r_a}^{h+r_a} \quad (13)$$

The approximate solution is of the form;

$$T(r) = \sum_{j=1}^n T_i^e \psi_j^e(r) \quad (14)$$

Where
$$\sum_{j=1}^n T_i^e \psi_j^e(r) = T_1^e \psi_1^e + T_2^e \psi_2^e + T_3^e \psi_3^e + \dots + T_n^e \psi_n^e(r) \quad (15)$$

2.2.1 For the Fissionable material

Substituting equation (14) into the weak form of the fissionable material, and w for ψ (for $i = 1$ to 3), we have;

$$0 = \int_{r_a}^{h+r_a} \left(r^2 k_1 \frac{d\psi_i}{dr} \left(\sum_{j=1}^3 T_i^e \frac{d\psi_j}{dr} \right) - q_1 w r^2 \right) dr - \sum_{j=1}^3 \psi_j^e Q_j^e(r) \quad (16)$$

The above can be written in the form;

$$\sum_{j=1}^n k_{ij} w_j^e = f_i + Q_i^e \quad (17)$$

The finite element model can therefore be represented as;

$$k_{ij} = f_i^e + Q_i^e \quad (18)$$

Where

$$k_{ij} = \int_{r_a}^{h+r_a} \left(r^2 k_1 \frac{d\psi_i}{dr} \cdot \frac{d\psi_j}{dr} \right) dr \quad (19)$$

And

$$f_i^e = \int_{r_a}^{h+r_a} q_1 \psi_i r^2 dr \quad (20)$$

2.2.2 For Aluminium Cladding;

Substituting equation (14) into the weak form of the aluminium cladding, and w for ψ (for $i = 1$ to 3), we have;

$$0 = \int_{r_a}^{h+r_a} \left(r^2 k_2 \frac{d\psi_i}{dr} \left(\sum_{j=1}^3 T_j^e \frac{d\psi_j}{dr} \right) \right) dr - \sum_{j=1}^3 \psi_j^e Q_j^e(r) \quad (21)$$

The above can be written in the form;

$$\sum_{j=1}^n k_{ij} w_j^e = f_i + Q_i^e \quad (22)$$

The finite element model can therefore be represented as;

$$k_{ij} = f_i^e + Q_i^e \quad (23)$$

Where

$$k_{ij} = \int_{r_a}^{h+r_a} \left(r^2 k_2 \frac{d\psi_i}{dr} \cdot \frac{d\psi_j}{dr} \right) dr \quad (24)$$

And $f_i^e = 0$ (25)

Evaluating the coefficient matrix using MathCAD software, results are obtained for k_1 and k_2 . The matrix obtained is thus;

$$k_{ij} = \frac{k}{15h} \begin{pmatrix} 3h^2 + 15hra + 35ra^2 & -(6h^2 + 20hra + 40ra^2) & 3h^2 + 5hra + 5ra^2 \\ -(6h^2 + 20hra + 40ra^2) & 32h^2 + 80hra + 80ra^2 & -(26h^2 + 60hra + 40ra^2) \\ 3h^2 + 5hra + 5ra^2 & -(26h^2 + 60hra + 40ra^2) & 23h^2 + 55hra + 35ra^2 \end{pmatrix}$$

$$f_i = \frac{q_1 h}{15} \begin{pmatrix} \frac{(h^2 + 10ra^2)}{4} \\ 3h^2 + 10hra + 10ra^2 \\ \frac{(9h^2 + 20hra + 10ra^2)}{4} \end{pmatrix}$$

Where we have k , we substitute into it, k_1 and k_2 which are used in solving the equations of fissionable material and aluminum cladding respectively.

2.3 Assembly of the Matrix Using Eight Elements

Next, we assemble the matrix using eight elements with four elements each for the fissionable material and Aluminium Cladding, since they both share the same centre, thereafter; we mesh the elements to get a 17×17 matrix.

An illustration using one element each is shown below;

For fissionable material;

$$\begin{pmatrix} k_{11}^1 & k_{12}^1 & k_{13}^1 \\ k_{21}^1 & k_{22}^1 & k_{23}^1 \\ k_{31}^1 & k_{32}^1 & k_{33}^1 \end{pmatrix} \begin{pmatrix} T_1^1 \\ T_2^1 \\ T_3^1 \end{pmatrix} = \begin{pmatrix} Q_1^1 \\ Q_2^1 \\ Q_3^1 \end{pmatrix}$$

For Aluminium Cladding;

$$\begin{pmatrix} k_{11}^2 & k_{12}^2 & k_{13}^2 \\ k_{21}^2 & k_{22}^2 & k_{23}^2 \\ k_{31}^2 & k_{32}^2 & k_{33}^2 \end{pmatrix} \begin{pmatrix} T_1^2 \\ T_2^2 \\ T_3^2 \end{pmatrix} = \begin{pmatrix} Q_1^2 \\ Q_2^2 \\ Q_3^2 \end{pmatrix}$$

From boundary condition,

$$T_1 = T_2 \text{ at } r = R_F \text{ and } T_2 = T_0 \text{ at } r = R_c$$

The mesh becomes;

$$\begin{bmatrix} k_{11}^1 & k_{12}^1 & k_{13}^1 & 0 & 0 \\ k_{21}^1 & k_{22}^1 & k_{23}^1 & 0 & 0 \\ k_{31}^1 & k_{32}^1 & k_{33}^1 + k_{11}^2 & k_{12}^2 & k_{13}^2 \\ 0 & 0 & k_{21}^2 & k_{22}^2 & k_{23}^2 \\ 0 & 0 & k_{31}^2 & k_{32}^2 & k_{33}^2 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{bmatrix} = \begin{bmatrix} Q_1^1 \\ Q_2^1 \\ Q_3^1 + Q_1^2 \\ Q_2^2 \\ Q_3^2 \end{bmatrix}$$

The eight element mesh follows the same procedure.

3 Results

The results obtained from the finite element method are compared to that of the exact solutions.

Assuming that; $k_1 = 25m^3/h$, $k_2 = 20m^3/h$, $R_F = 8m$, $R_c = 16m$, $q_1 = 30m^3/h$, $q_0 = 11.5385m^3/h$, $c = 0.4$, $T_o = 20^0C$, $h = \frac{L}{n} = \frac{8}{4} = 2$.

Four elements each are used for the fissionable material and the aluminum cladding. The fourth element of the fissionable material is merged with the first element of the aluminum cladding material

to get a total of eight elements. The finite solution is determined after the boundary conditions have been applied. The results are shown in the table below;

Table 1: Values of Exact and Finite Element Method Solutions for Temperatures of Nuclear Fuel for Eight Elements

| RADIUS | FEA SOLUTIONS | EXACT SOLUTIONS | ERROR | PERCENTAGE ERROR (%) | ACCURACY (%) |
|--------|---------------|-----------------|------------|----------------------|--------------|
| 0 | 42.9813 | 33.14464342 | 0.22885898 | 22.885898 | 77.114102 |
| 1 | 41.7413 | 33.06757595 | 0.20779717 | 20.779717 | 79.220283 |
| 2 | 40.9013 | 32.83464276 | 0.19722251 | 19.722251 | 80.277749 |
| 3 | 40.3169 | 32.44065154 | 0.19535848 | 19.535848 | 80.464152 |
| 4 | 39.1347 | 31.87694841 | 0.18545566 | 18.545565 | 81.454435 |
| 5 | 38.0828 | 31.13141797 | 0.18253338 | 18.253338 | 81.746662 |
| 6 | 36.3864 | 30.18848327 | 0.17033608 | 17.033608 | 82.966392 |
| 7 | 34.8207 | 29.0291058 | 0.16632619 | 16.632619 | 83.367381 |
| 8 | 32.5991 | 27.63078551 | 0.15240649 | 15.240649 | 84.759351 |
| 9 | 29.8015 | 25.93505539 | 0.12973993 | 12.973993 | 87.026007 |
| 10 | 27.5597 | 24.5784713 | 0.10817348 | 10.817348 | 89.182652 |
| 11 | 25.7278 | 23.46853887 | 0.08781400 | 8.781400 | 91.218600 |
| 12 | 24.1999 | 22.54359517 | 0.06844263 | 6.844263 | 93.155737 |
| 13 | 22.9080 | 21.76095050 | 0.05007201 | 5.007201 | 94.992799 |
| 14 | 21.8000 | 21.09011222 | 0.03256366 | 3.256366 | 96.743634 |
| 15 | 20.8401 | 20.50871903 | 0.01590112 | 1.590112 | 98.409888 |
| 16 | 20.0000 | 20.00000000 | 0.00000000 | 0.000000 | 100.000000 |

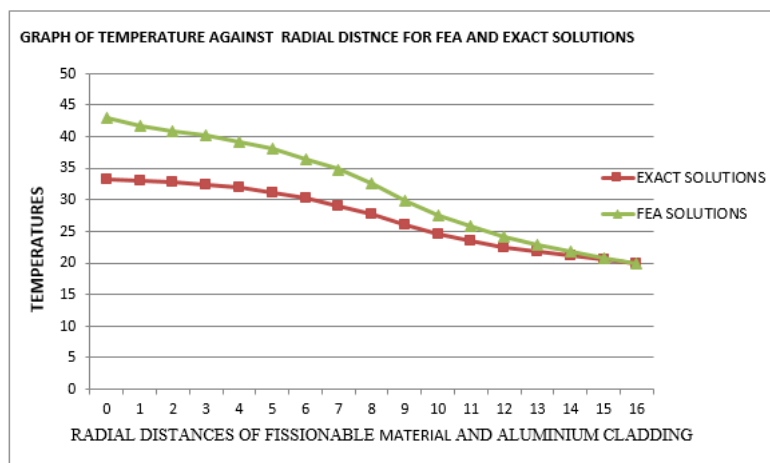


Figure 2: Graph of Temperature against Radial Distance for FEA and Exact Solutions

4 Discussion

The Finite Element solutions obtained in the problem can be used to determine the distribution of temperature distribution in a nuclear fuel element consisting of a sphere of fissionable material and a spherical shell of aluminum cladding. This is as a result of substituting the appropriate values of the domains and boundary conditions into the formulated coefficient matrix equations. The results are represented in the table, and it shows that there is increase in accuracy as the number of element

increases with decrease in error. The graph of the finite element solutions and the exact solutions has separate paths but converges at a point.

5 Conclusion

The results shows that finite element method is a more reliable and accurate method for determining the temperature distribution in a nuclear fuel element consisting of a sphere of fissionable material and a spherical shell of aluminum cladding successfully.

REFERENCES

- [1]. Pandey, K.M and Mahesh, M. (2010), "Determination of temperature distribution in a cylinder nuclear Fuel Rod", a Mathematical Approach. International Journal of Innovation, Management and Technology. Volume 1, no. 5.
- [2]. Gupta, G.P. and Rajendra Prakash (1994), "Engineering Heat Transfer", Nemchand and Brothers, 6th Edition.
- [3]. Elijah, Kannatey – Asibu, Jri, Jalld A.R and Noboru, K.(1989),"Journal of engineering materials and technology", vol. 11.
- [4]. Noor, M. M., Kadirgama, Rahman, M.M, Harun, W.S.M and Haron, C.H.C (2009),"Finite Elwment Analysis and Statistical Method to determine temperature distribution of cutting tool in End-Milling.
- [5]. Eskandari. M.R, Bavandi and Mihandoost (2012), "Studies on Nuclear Fuel Rod Thermal Performance".
- [6]. Kenichi ITO, Ryo Iwasaki and Yoshihiko Iwano (1984), "Finite Element Model for Analysis of Fission gas Release from uranium dioxide fuel", pp 129 – 138.
- [7]. Saragi E. (2012), "Modelling of Temperature distribution in TRISO fuel based on Finite Element Method", vol. 1448, pp 270.
- [8]. Purohit, G.K and Kalakesh Baligar (2012), "Cladding Process Simulation And Residual Stress Estimation Using Finite Element Coupled Field Analysis". International Journal of Engineering Research and Technology. Vol 1, ISSN 2278 – 0181.
- [9]. Isakor, V., Zaznoba, V., and Konopler, E. (2006), "Theoretical Foundation of Chemical Engineering", Volume 40, pp 431.
- [10]. Reddy, J. N (1984), "An introduction to Finite Element Method", Texas, McGraw Hill Second Edition.

The Modernization of Feed Water Pump Turbine (FWPT) Controller for Nuclear Power Plants Unit 5 and 6

Kwangyoung SOHN, Jayoung LEE and Changhwan CHO
Mirae-en, Daejeon, Republic of Korea
kwangyoung.sohn@mirae-en.co.kr

ABSTRACT

For modernization, Feed water Pump Turbine (FWPT) for HANUL Nuclear Power Plants Unit 5 & 6 has been replaced with Triple Modular System including the upgrade of control logic actuating Low Pressure/High Pressure (LP/HP) stop valves and control valves. This modernization includes hardware platform change as well as ladder logic changes. This paper provides the lesson learned from the design support and the verification and validation (V&V) for control logic software in accordance with IEEE 1012[1], which describes the Software Development Life Cycle (SDLC) phase activities for independent verification and validation (V&V). As usually it is necessary to interpret standards by upward and/or downward tailoring, i.e. interpretation, based on the SIL level and application function for practical independent V&V. It has been conducted to list up the inspection viewpoints for software itself as well as architectural design including the hardware interfaces. For successful independent V&V of FWPT, the specific viewpoints and approach are employed according to the functional characteristics and code optimization. The V&V for this project reviews only the requirement, design, implementation, and test phase. This article also provides the difficulty experienced during independent V&V including the design support, and concludes by addressing a couple of lessons learned for FWPT V&V.

Keywords: FWPT (Feed water Pump Turbine), Software Development Life Cycle (SDLC), Verification and Validation (V&V), Nuclear Power Plant (NPP)

1 Introduction

Due to the hardware aging and obsolescence, the upgrade of FWPT for HANUL nuclear power plants unit 5 & 6 was brought up as necessary. In the course of upgrade, independent V&V has been requested to validate the design integrity of the software and its system which is classified to safety-related in accordance with the organization as Figure 1.

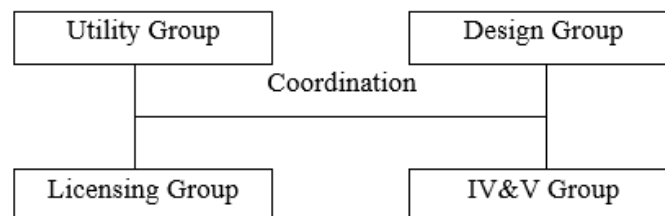


Figure 1: The Organization for FWPT V&V for HANUL NPP unit 5 & 6

For the transparency of V&V activity, the design team and review team is officially separated for managing the independent review of the system and the component design, which is also the requirement from licensing organization of Korea Institute of Nuclear Safety (KINS) shown in Figure 1.

However IEEE 1012 code is generally conceptual that is applicable to all the software of various fields including FWPT, it is necessary to devise application-specific review points to V&V team, which might be enhancing the reliability of the FWPT software system.

2 FWPT Overall

The critical characteristics[8] for FWPT is straight forward, i.e., processing the sensor signals, process engineering con-version, send out the result to monitor and control the Low Pressure/High Pressure (LP/HP) stop valves and control valves.

Figure 2 indicates the interconnection diagram between FWPT and other auxiliary systems which provide the process input and control output. The main function (critical characteristics) of FWPT is to control the turbine output by manipulating the stop and control valves for Feed water Pump.

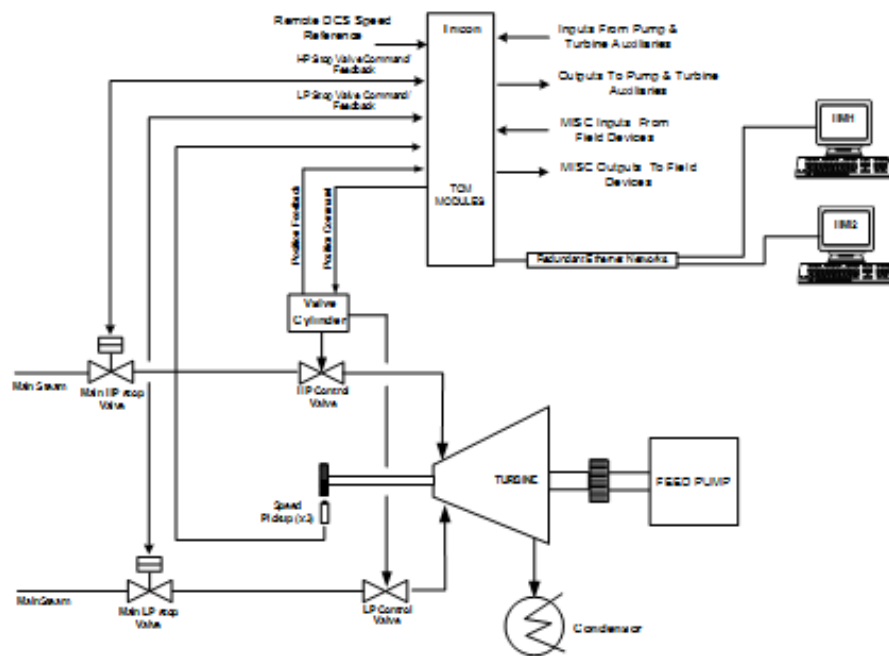


Figure 2: FWPT for HANUL unit 5 & 6

Input and output for FWPT that is controlling LP stop/control valve and HP stop/control valve simply is composed of triple sensors and signal processors. These signals are processed by the algorithm in each triple processor. Also one of the final outputs of triple modules is selected as a result of voting for actual actuation of valves by comparing the speed values from the magnetic speed pick-up devices...

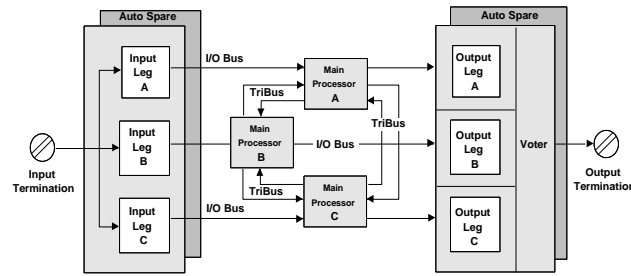


Figure 3: Triple modular software structure

Figure 3 provides the general configuration of the FWPT system including hardware and software. A single input and output is divided into 3 inputs and outputs for FWPT, which are processed in each hardware. And those signals are processed in the triple processor (PLC) for calculation and PID control.

3 Approach for FWPT V&V

3.1 Identification of critical characteristics

FWPT software on HANUL 5&6 is independently verified and validated. Based on the conceptual V&V activities of IEEE 1012, the major viewpoint is selected as below through the system function and performance analysis.

1. Identification of the critical functional characteristics for the FWPT – controlling LP/HP stop and control valves
2. Identification of the interface between the internal and external sub-components like the communication and its transmission frequency (including the serial data links) – interface through HMI (Human-Machine Inter-face)
3. Identification of the performance characteristics – response time for actuating valves
4. Identification of the appropriateness on the functional cohesion and coupling in final implementation [2] – decomposition of logic module
5. Identification of the test coverage – test plan for a type of test
6. Exceptional handling

The following section will address the detail of the several item enumerated above, and difficulty that has been experienced during independent verification and validation and design support.

3.2 Modified verification and validation

The verification and validation team has been organized to meet the independencies of managerial and technical aspect. However the team is not independent from the financial budget, which means there is a decisive weakness for performing a sincere verification and validation. See Table 1 for independency for this project.

3.3 Identification of the interface

For the modernization of FWPT controller, some of the hardwired interconnections between components are connected through a communication network. When communication is used for data exchange, there might be a discontinuity of data when a network failure occurs and is recovered soon

again. Through these communication media, the data pertaining to FWPT control and monitoring is sent out to MCR and relevant human interface.

3.4 Identification for performance characteristic

FWPT could have a malfunction mainly caused not by human intervention error but by internal software malfunction. Because the human intervention is limited to the manipulation of hand switch for MANUAL/AUTO transition, and the determination of a series of set points. This could result in critical hazard of malfunction in controlling the turbine in the type of human errors.

3.5 Cohesion and coupling

It is the design verification to check if the software module is decomposed to be constructed well based on the logical function decomposition, getting rid of implementation complexity and ambiguity resulting from incomplete software design. It is a critical measure to judge the testability and maintainability of software [2].

3.6 Reliability of function and performance

It is a new aspect of verification and validation to check if the function to be implemented is implementable in a new software logic composer or hardware. Recently most of functions, even implemented with hardware in a legacy plant, are reformed as software, targeting for a digital system. Also hardware on which the software is running is different from the legacy hardware, and also the legacy code is transformed into a different algorithm although upgraded algorithm is implemented to be functioning in the same way to legacy code. It is very important check point to verify that the function and performance is equivalent to legacy.

3.7 Exception handling

In any software function, there is an exception of partial function. This partial function shall be clearly designed and implemented, which supports a reliable test plan and procedure in the test phase. But in this project, it is found that the software is functioning by assuming the prefabricated route, which means this is not resilient to the exceptions.

3.8 Test coverage

Practically exhaustive test coverage is not desirable and not recommended for robust software testing. However, when the output of the software is actuating the hardware devices connected to system, the maximum test coverage is recommended in test. For this, systematic and concrete test coverage has been generated by designer as well as independent verifier and validator based on the 6 criteria in Section 2.1. It was very helpful to remove the delicate failure sources, and it becomes the solid basis of test procedure preparation.

4 Lesson learned in FWPT V&V

The following are the patterns found through the analysis of anomaly reports in each phase about FWPT in HANUL nuclear power plant unit 5 & 6.

4.1 Requirement refinement

In case of this project, there has been only the legacy source code and control logic which could be source of SDLC documentation. Based on this limited source, Software Requirement Specification (SRS) and Software Design Specification (SDS) have been prepared. So it omitted the some of the requirements, which have been finally resolved through anomaly report and analysis, and there was feedback for this to design documents. In requirement extraction on legacy plants without complete design data, it is necessary to discuss requirement refinement. As noted in [8], the inception phase of design is very critical for further process of SDLC. Some of the requirement has been identified in the design phase, not in requirement phase, which has been reincorporated into the design by iteration.

4.2 The Code optimization in implementation

The code analysis as well as decomposition analysis such as cohesion and coupling have been conducted for software integrity and completeness. The following is one of the example for this

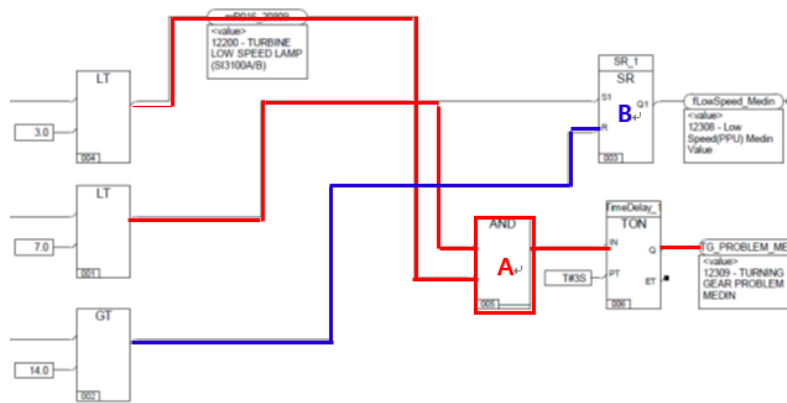


Figure 4: logic diagram for optimization

In figure 4, the input value is less than 3.0, then the block “A” has a meaning to generate the effective output, i.e. 1. Thus without AND gate logic it is possible to generate the trigger condition by comparing that input is less than 3.0. This is one of the examples that the optimization is required as necessary, which means that the developer does not fully understand the requirements.

4.3 The management of V&V

According to Annex C “definition of independent V&V” of IEEE 1012-2004[1], independent V&V is categorized by three parameters: technical independence, managerial independence, and financial independence as described in Table 2.

Table 1: Independent verification and validation form

| Independent V&V Form | Technical | Management | Financial |
|--|-----------|------------|-----------|
| Classical | l | l | l |
| Modified | l | i | l |
| Integrated | i | l | l |
| Internal | i | i | i |
| Embedded | e | e | e |
| FWPT project | i | l | e |
| NOTE l : Rigorous, i : Conditional Independence, e : Minimal Independence | | | |

However FWPT project has the variance deviating from the independency from IEEE 1012. For reference, Korea Hydraulic and Nuclear Power (KHNP) has changed the project verification structures in hardware qualification into CLASSICAL form by reinforcing the internal regulation but still no action for software verification and validation.

4.4 Test environment

Unit testing and integration testing is for verifying its function and performance only. However Factory Acceptance Test (FAT) and Site Acceptance Test (SAT) should integrate the hardware and/or peripheral interfaces with software itself, which is ultimately an architectural system design testing for FWPT for HANUL unit 5 & 6. In this project, verification and validation should have considered the complete integration with hardware interfaces, but it was incomplete in integration.

4.5 Design data freeze

As we're well aware of that the process for verification and validation is to review the SDLC design data in each phase right after the issue of frozen design data of each phase. But the design data such as SRS, SDS, source code and a multiple of test procedure that had not been finished and signed appropriately were transmitted to verification and validation organization, which brought up repetitive review process wasting a time. Thus the verifier has felt the difficulty in conducting the review process, i.e., entailing the repetitive review process for a single design documents like SRS, SDS, multiple test procedure and source code. For verification and validation team it seemed to have been an activity to support the fundamental design work, not verification and validation work. Also there has been case that utility changed the requirement in the implementation and testing phase all of sudden.

4.6 Anomaly resolution process

This is not a technical issue but a managerial issue for verification and validation process. Once verifier finds out the anomaly for target system regarding function, performance and interfaces, these anomaly report should have been justified and resolved by the designer with feedback to FWPT design selectively and appropriately.

5 Conclusion

Through the independent verification and validation for FWPT in HANUL nuclear power plant 5 & 6, a couple of obstacles in conducting the verification and validation have been found practically.

In this project, the selection of critical characteristics of FWPT was very important even though some of the requirements are found in phase later than requirement phase. Also it is found that the requirement refinement, code optimization, the management of independent V&V, test environment, design data freeze, and anomaly resolution process is one of difficulty in performing successful verification and validation.

Among them, the design data freeze, anomaly resolution process and design support due to the lack of understanding of FWPT system have been a tiresome factors deteriorating the smooth and successful independent verification and validation.

Once the completion of independent V&V, issuing the anomaly report, a resolution meeting between independent verifier and validator and designer to obtain the optimal solution should have been held in

every SDLC phases. Unfortunately there is a tendency that the designer will not partly accept the anomaly issued just because the function is any-way performed well even though there is room for optimization and documentation [8].

To be a successful independent verification and validation, it is important to make an effort to find out the technical anomaly as well as to cooperate each other, i.e., designer and validator.

ACKNOWLEDGEMENTS

This work has been conducted with the support of INVENSYS Korea, especially thank to director Myung. Soon. Kim.

REFERENCES

- [1]. IEEE Std 1012™, "IEEE Standard for Software Verification and Validation", 2004
- [2]. Roger S. Pressman, Ph. D, ISBN 0073655783, "Software Engineering, A Practitioner's Approach", Fifth Edition, McGraw-Hill Higher Education.
- [3]. Kyung Ho Cha, Kee Choon Kwon, Chun Se Woo, "The software verification and validation tasks for a safety critical system in nuclear power plants", International journal of safety, vol.3 no.1, pp.38-46, 1598-7302, 2004.
- [4]. C. Ponsard, P. Massonet, J. F. Molderez, A. Rifaut, A. van Lamsweerde, H. Tran Van, "Early verification and validation of mission critical systems", Formal Methods in system Design, Volume 30, Number 3, pp.233-247, 0925-9856.
- [5]. IEEE 829, "IEEE Standard for Software and System Test Documentation", 2008
- [6]. ANSI/IEEE 1008, "IEEE Standard for Software Unit Testing", 1987
- [7]. NUREG/CR-6430, "Software Safety Analysis", 1995
- [8]. Lessons learned from Practical Independent Verification and Validation based on IEEE 1012, A Journal of Software Engineering and Applications, JoonKu LEE, YangMo KIM

A Novel Method for Calculating Approximate Weight of an Object through its Image

¹Amaresh Chandra Meher, ²Uddgam Singh and ³K.K. Senapati,

¹Dept of Computer Science, SUIIT, Jyoti Vihar, Sambalpur-768019, India

^{2,3}Dept of Computer Science, BIT, Mesra, Ranchi-835215, India

¹amaresh.dcs@gmail.com; ²uddgam.singh4386@gmail.com; ³kksenapati@bitmesra.ac.in

ABSTRACT

The field of computer science is capable of calculating approximate weight of an object through its image. It is new and challenging. The idea behind the weight of an object is nothing but approximation of the pixel carries some weight value which is specific for the corresponding specific object. It is a challenge for the computer science people to develop methods through pixel weight. The main focus of this paper is to find the approximate weight of an object through its digital image. Generally weight calculating machine is used to find out the weight of any object(s). To improve the process and to make it somehow digitized we have proposed a novel method that does not require any physical standard to calculate weight.

Initially, a surface image is captured by the system and whenever a newly taken image, i.e., the image of object placed over the same surface is input, it will be subtracted from our pre-existing surface image. Based on the features of Colour attribute the weight of the object is estimated. From the experimental results, it is observed that the proposed model works perfectly on both training data as well as novel data. We believe this paper could form a base for future works on related topics.

Index Terms – Algorithms, Approximation methods, Colour, Surface image, Weight, Weight per pixel, Weight approximation, Image subtraction, Threshold.

1 Introduction

In today's scenario people are becoming more application oriented rather than using physical equipment. In this context, here our work is an approach to meet the goal to some extent which will find the approximate weight of an object from its image in a convincing way.

A digital image is a numeric representation (normally binary) of a two-dimensional image. Depending on whether the image resolution is fixed, it may be of vector or raster type. In a simple way, the term "digital image" usually refers to raster images also called bitmap images. As image consists of millions of pixels, Yao and Odobez described each pixel mode as a combination of intensity and texture [1]. Image features which can be used for retrieval purposes can be in the form of Colour, shape, and texture, as described in [2]. Image retrieval by Colour feature is the most widely used method [2]. Our proposed model also utilizes the Colour feature towards the approximation of weight.

A background generation and background subtraction approach to separate the moving objects from other components in the image is derived by Yufang Zhang, Peijun Shi, Elizabeth G. Jones, and Qiuming Zhu [3]. The term “Surface image” in our proposed method is used to indicate a background image which has captured before the entrance of any moving object. Ruihua MA, Liyuan LI, Weimin HUANG, Qi TIAN proposed a model where they derived the relation for geometric correction for the ground plane and proves formally that it can be directly applied to all the foreground pixels to estimate the crowd density by pixel counting. They explained there how to implement pixel counting with geometric correction efficiently [4]. Saad Choudri, James M. Ferryman, Atta Badii, in their proposed model, tried to increase the robustness of “Pixel-based people counting” methods when background pixels become increasingly difficult to resolve and to reduce the loss of people when they got absorbed into the background after being stationary for a long period by introducing a selective background update method[5]. In the proposed method also surface image is updated to avoid error reading during pixel counting.

Weight is directly proportionate to the number and type of atoms in the object and physically, atom is the smallest unit in an object, likewise, pixel is the smallest unit for that object in an image and to find the weight of object within the image we have to find the number of pixels covering only for the required object within the image and then finding out the weight per unit pixel or weight per pixel (w). Telling in other way, ‘w’ uses the concept of density in spite of the fact that we have kept the word ‘density’ silent in our paper. The density of a substance is its mass per unit volume and since density varies from substance to substance, same is the case for “weight per pixel” (from object to object). Section-II contains proposed method where the whole process of finding weight per pixel for a specific type of object is derived. After finding out weight per pixel we can accumulate the weight for whole object. Section III derives about the flow of processes, section IV and V contains experimental design and conclusions respectively. Finally section VI contains the references.

2 Proposed Method

The proposed method is based on known and similar type of objects. Similar type of objects refer to the objects made up of same materials, e.g., all fishes of the category *Paralichthys olivaceus* are similar, which in turn are different from fishes of category *Oplegnathus fasciatus* or any other, wooden pencils come under one category which are different from pencils having plastic sheath, eggs of pigeons are different from eggs of parrots etc.

This method is basically performed in two phases.

1: Pre-processing phase, 2: Execution phase

In the pre-processing phase some similar type of objects are collected and their weights have been measured manually. Consider the physical weights $W_1, W_2, W_3, W_4, \dots, W_n$ respectively are for objects $O_1, O_2, O_3, O_4, \dots, O_n$. The average physical weight ‘WAvg’ is calculated as given below:

$$W_{Avg} = \frac{\sum_{i=1}^n W_i}{n} \quad (1)$$

In the subsequent step the corresponding pixels of the object in the image is counted. Similarly this process is repeated for rest of the images. Now calculate average number of pixels per object. Consider the total number of pixels covering for object O_1 is $Pix(O_1)$, for object O_2 , it is $Pix(O_2)$ and so on.

Suppose “ Pix_{Avg} ” is the average number of pixels per object. It can be calculated as:

$$Pix_{Avg} = \frac{\sum_{i=1}^n Pix(O_i)}{n} \quad (2)$$

In the next step, the weight per pixel, suppose “ w ” is calculated as follows:

$$\begin{aligned} w &= W_{Avg} / Pix_{Avg} \\ &= \frac{(\sum_{i=1}^n W_i) / n}{(\sum_{i=1}^n Pix(O_i)) / n} \\ &= \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Pix(O_i)} \\ &= \sum_{i=1}^n (W_i / Pix(O_i)) \end{aligned} \quad (3)$$

Now store the value of “ w ” for further use.

In the execution phase the weight of the required object in a novel image can be finding out in two steps:

Step-1: Find out the number of pixels (Count) covering the object (in novel image) which is to be weighted.

Step-2: The weight of object will be calculated by the following formula as:

$$\text{Weight of the object (W)} = \text{Count} * w. \quad (4)$$

So finally “ W ” is our desired result.

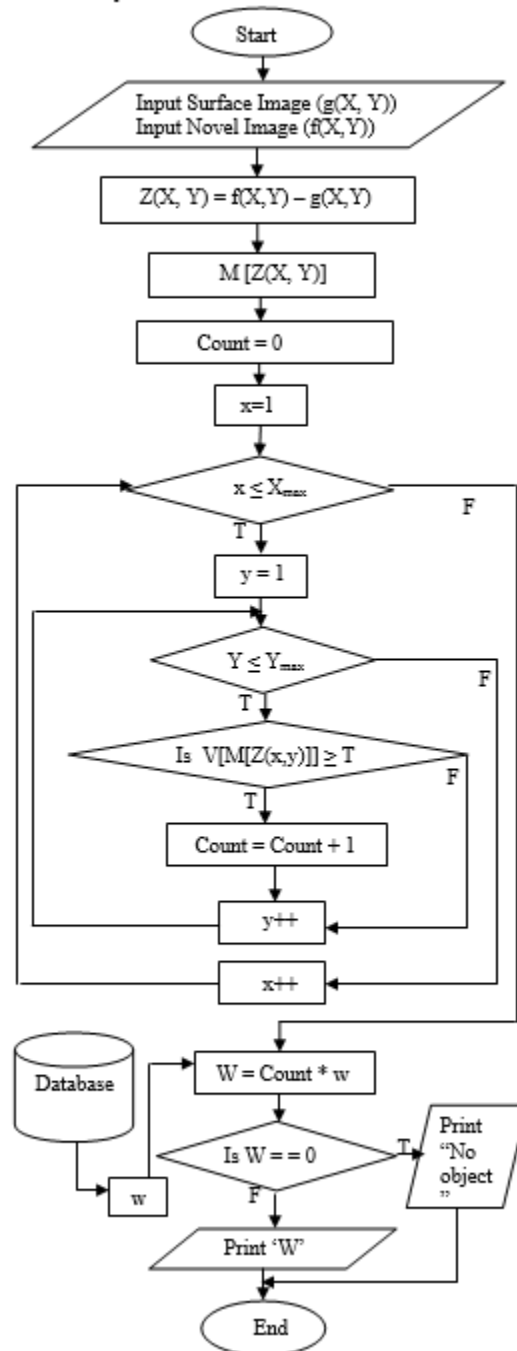


Figure 1: (Flow chart for the proposed method)

3 Flow of Processes

As mentioned in the previous section the general principle of the techniques presented in this paper is based on the similar image database for the object.

The activities done are shown in figure-1 briefly which indicates that the input novel image is subtracted immediately from the surface image stored just before the novel image where the object will be placed. Let $f(X, Y)$ be the image to be input and $g(X, Y)$ is the surface image. Both $f(X, Y)$ and $g(X, Y)$ are 2-D images. $Z(X, Y)$ is the resultant image after subtraction of $g(X, Y)$ from $f(X, Y)$. $M [Z(X, Y)]$ is the matrix form of $Z(X, Y)$. Conversions of the resultant image into matrix form simplify the job of separation of

object from the background. X_{\max} and Y_{\max} represent the last row and column of the matrix respectively. $V [M [Z(x, y)]]$ is the value of each pixel in the matrix. The matrix shows that the values of pixels for background in the resultant image (excluding for the object) are zero or nearly equal to zero (due to slight changes in intensity of light) irrespective of the colours of the background. Taking this matter into consideration a threshold value 'T' for pixels is set. Then all the pixels having value greater than or equal to threshold are counted. These pixels cover for the required object and the rest of the pixels having values below threshold are for background. Initially the counter is set to 0 and is incremented each time for the pixels having values greater than or equal to 'T'. The database already contains 'w' whose values are objects specific. The appropriate value of 'w' corresponding to the type of object is found out carefully from the database. Finally to get the approximate weight 'W' of the object, value of Count is multiplied with the value of 'w'.

4 Experimental Design and Results

The equipments required for the experiment are a good quality camera, conveyor belt, the objects to be tested and a processor.

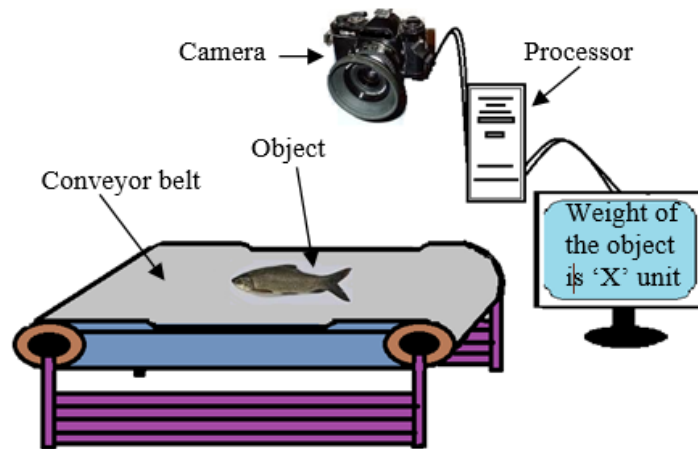


Figure 2

To test the accuracy of the proposed method, thousands of experiments have been performed. Here we have shown the calculated weight of twelve pencils and five no. of "Labeo rohita" fishes by keeping them at the distances 150 cm and 60 cm respectively from the camera. The Image of the object is captured only when the object is in maximum contact to the surface. The results produced by our system are compared with the weight calculated by weighting machine.

The proposed method was experimentally verified using a Nikon camera (model no. D3000) keeping resolutions 1936 x 1296 for pencils and images of fishes are taken by a Canon camera (model no. DIGITAL IXUS 95 IS) keeping resolutions 3648 x 2736. Horizontal resolution and vertical resolution are 300 dpi for pencils and 180 dpi for fishes. Bit depth is 24 and Colour representation is RBG for both.

For the proposed method the position of the camera is static and its resolutions are kept constant. The colours of surface image and object should not match, as well as the distance between the camera and the surface of the conveyor belt is fixed. The conveyor belt used here has an important characteristic, that is, it always keeps the object placed on it at the proper place to be imaged. In the experiment of

discussion we used Classmate Carbon Black Pencil. Also for the purpose of clarity in understanding we have provided a surface image, an object image and the resultant image (after subtraction) for a fish whose experimental discussion is not shown here to avoid the lengthiness.

The physical weight of the pencil is 4.6274 gm. The surface image and image of pencil are shown in Figure 3 and 4 respectively. The resulted figure after subtraction is shown in Figure 5.

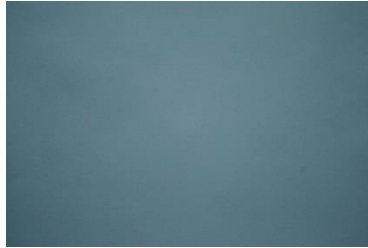


Figure 3



Figure 4



Figure 5

From figure 5, number of pixels (Count) above threshold is 13555 keeping threshold value at 30. As derived in our proposed method, approximate weight of Classmate carbon Black pencil is the product of Count and 'w'. The value of 'w' is obtained as 0.0003877529824 gm. So the approximate weight of the pencil=13555 *0.0003877529824=5.255992gm.

Table 1

| Pencil Name | Physical Weight | Number of pixels covering the pencil |
|-----------------------|-----------------|--------------------------------------|
| Ole-Grip (super dark) | 5.17710 | 13167 |
| Camlin Flora | 4.7574 | 14693 |
| Nataraj 621 (HB) | 4.6552 | 13633 |
| Papyrus (HB) | 5.9298 | 12678 |
| POLO CHECKING | 6.3607 | 15152 |

Taking into consideration for five pencils as shown in table-1, value of 'w' can be calculated where values of W_{avg} and Pix_{avg} are calculated as 5.37604 and 13864.6 respectively.

$$\text{Weight per pixel 'w'} = (W_{Avg} / Pix_{Avg})$$

Physical weight of classmate carbon pencil is 4.6274 gm and with our proposed method it is coming as 5.255992 gm. So the accuracy is of 88.04%.

Similarly many experiments have been performed through the proposed method whose results are presented in table -2 & table-3 with its accuracy. Table-2 is for pencils and table-3 is for fishes.

Table 2

| Name of pencils | Weight calculated by Weighting Machine (in gm) | Results produced by our system (in gm) | Accuracy (%) |
|------------------------------|--|--|--------------|
| Ole-Grip: (super dark) | 5.17710 | 5.80389 | 89.20 |
| Papyrus: (HB) | 5.9298 | 4.915932 | 82.90 |
| Camlin Flora | 4.7574 | 4.00450 | 84.17 |
| Nataraj 621: (HB) | 4.6552 | 5.286236 | 86.44 |
| Classmate:(Carbon black) | 4.6274 | 5.25599 | 88.04 |
| FUN (ULTRA DARK BONDED LEAD) | 5.0082 | 5.765111 | 86.87 |
| REYNOLDS HB | 4.8439 | 5.811642 | 83.35 |
| Cello LEADER HB | 4.9393 | 5.502603 | 89.76 |
| POLO CHECKING | 6.3607 | 5.875233 | 92.37 |
| APSARA Platinum (Extra Dark) | 4.5673 | 5.097013 | 89.61 |
| APSARA DRAWING 701 | 4.8490 | 5.874458 | 82.54 |
| STOR HB | 3.6607 | 3.734061 | 98.04 |

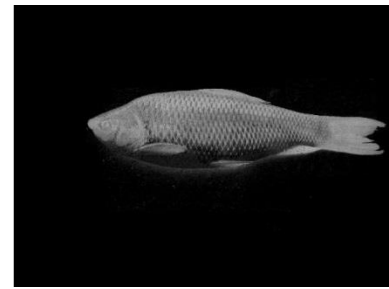


Figure 6: (Surface image for the fish)

Figure 7:(Object image for the fish)

Figure 8: (Resultant image)

Table 3

| Weight calculated by Weighting Machine (in gm) | Results produced by our system (in gm) | Number of pixels covering object | Accuracy (%) |
|--|--|----------------------------------|--------------|
| 1000 | 9979.51988 | 1074969 | 99.98 |
| 1100 | 1185.836874 | 1274999 | 92.83 |
| 1200 | 1256.285024 | 1350744 | 95.52 |
| 1400 | 1341.91173 | 1442809 | 95.85 |
| 1450 | 1472.456197 | 1583169 | 98.51 |
| Threshold = 34 , w = 0.000930068866 | | | |

5 Conclusion

In this paper we have presented a novel method for finding the weight of an object(s) through an image and the result of the method as observed is nearly equals to the actual weight of the object which shows that our method works properly. Again more precise result can be obtained by further prudent work on this.

As author we must accept that this paper contains lot of constraints and it works only for a limited category of objects in the real world but keeping in mind that as this is the first step towards the goal constraints must be there. We also feel strongly that the idea is new and unique.

This method can also be useful in photographic industry, study of celestial objects and in many more areas.

This is an initial attempt; further enhancement on this technique can facilitate capturing images by a portable device. This novel technique can be further upgraded for dissimilar object in a single frame with the aid of object recognition.

REFERENCES

- [1] J.Yao and J.Odobez, "Multi-layer background subtraction based on color and texture," IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.
- [2] A.K. Jain, "Fundamentals of Digital Image Processing", Englewood cliffs, NJ: Prentice-Hall, 1980.
- [3] Yufang Zhang, Peijun Shi, Elizabeth G. Jones, and Qiuming Zhu" Robust Background Image Generation and Vehicle 3D Detection and Tracking". 2004 IEEE Intelligent Transpoltation Systems Conference Washington, D.C., USA, October 36, 2004.
- [4] Ruihua MA, Liyuan LI, Weimin HUANG, QiTIAN, "On Pixel Count Based Crowd DensityEstimation for Visual Surveillance" Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems Singapore, 1-3 December,2004
- [5] Saad Choudri, James M. Ferryman, Atta Badii, "Robust Background Model for Pixel Based People Counting using a Single Uncalibrated Camera", 978-1-4244-5504-1/09/\$25.00 ©2009 IEEE

Intelligent Hearing Assistance using Self-Organizing Feature Maps

¹Michael Hodges and ²Mohamed Zohdy

¹Computer Science and Engineering, Oakland University, Rochester, MI, USA;

²Electrical and Computer Engineering, Oakland University, Rochester, MI, USA;

¹mjhodges@oakland.edu; ²zohdyma@oakland.edu

ABSTRACT

A novel hearing assistance system is proposed which classifies sounds and selectively tunes them according to the needs of the hearing impaired. This differs from the usual hearing aids available today in that it uses computational intelligence to filter and tune sounds based on real life categorical classifications. The system can significantly improve audibility for the hearing impaired, including bringing completely inaudible tones into audibility. For classification a self-organizing feature map is used with a vector of sound features from the joint time-frequency domain. The map is trained with input sounds until a map of neurons is clustered according to these. The resulting map is used to classify new sounds. Based on this classification, a sound can be tuned to improve audibility for the hearing impaired. Techniques proposed for audio output include gamma tone frequency filtering, Fourier compression, low pass filtering, spectral subtraction, and an original algorithm to choose how to best boost the amplitude of deaf frequencies.

Keywords: self-organizing feature maps; Artificial Intelligence; machine learning; neural networks; pattern classification.

1 Introduction

Most common hearing aids today do not include computational intelligence mechanisms such as those which can classify the type of sound. Typically at best, one can find distinction between voice and noise. A deaf person could therefore not be able to separate the hearing aid processing of different categories of sound. For example, one may wish to amplify the sound of a baby while blocking out the sound of background nature sounds.

This research aims to define the development of a system which has the promise of achieving exactly this sort of intelligent classification for hearing aids. By using a self-organizing feature map as unsupervised learning to train a neural network to classify sound categories, the output can be adjusted based on such a classification.

Given a sound classification, this research further studies methods to adjust the output sound to improve hearing ability for a deaf user. Gamma tone filters and amplification can together separate a sound into its frequency components and boost a desired range. Fourier domain compression can recreate a sound with its Fourier signature shape intact while compressing the range of frequencies so

that a high-frequency deafness or low-frequency deafness can be compensated. Additional algorithms can determine how to boost particular frequencies most efficiently without losing other audible tones. Noise reduction can be accomplished by low pass filter blurring, or by spectral subtraction of a target noise source Fourier signal.

Combining the intelligent sound classification with these methods of output tuning produces a robust system which can significantly improve hearing ability. This paper analyzes the amount of hearing improvement possible, which can range from a likely significant amount to occasionally great gains under specific circumstances.

A flow-graph of the proposed system is shown in (Figure 1).

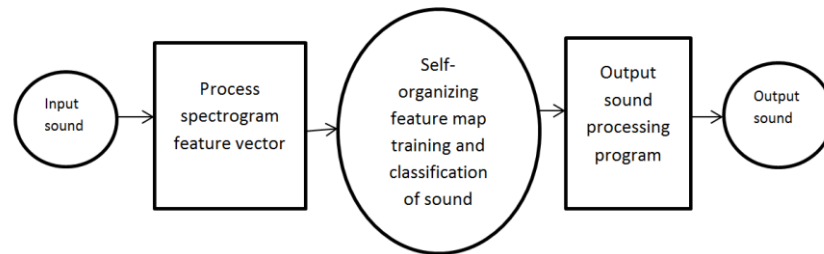


Figure 1: Flow-graph of the proposed system for intelligent sound classifying hearing assistance system.

The system starts with an input sounds, from which features are extracted to uniquely describe the sound. These features are fed into a self-organizing feature map which trains itself to classify sounds. Based on the map's classification, the sounds are fed to an output processing program, which tunes the output sound for the hearing impaired.

2 Modeling the System

Based on the diagram in (Figure 1), a model is created to demonstrate the flow of the system.

The input sound is a time series of amplitude values which can also be represented in frequency domain form. A number of these frequencies may likely have amplitudes below the threshold of hearing for a deaf person. Also the total range of frequencies present may extend beyond the capabilities of the hearing impaired. The output processing techniques described for this system perform amplification, frequency selective tuning, frequency compression, and computed algorithm-based sound adjustment.

The system model can be viewed as a sound input that goes through processing stages until a final output emerges according to the needs of a hearing impaired person. The input is the original sound to be processed. This can be obtained from a microphone and stored in a memory buffer while processing occurs. In a hardware oriented system processing can be much faster as it can be done in hardware.

Parallel to the input is the training of a self-organizing feature map neural network. Given a set of sounds from different categories, a neural map is trained which clusters the sound inputs and enables classification. The inputs to this system are mathematical feature vectors describing several training sounds.

After the input enters the system, it will need to use the trained map to classify the sound into one of a selected number of defined categories. When the category is determined, the input is given such a tag, and it proceeds to the output processing stage.

In the output processing stage, frequencies are tuned and compressed, algorithms search for optimal hearing improvement techniques, and amplification occurs. Based on the category from the previous step, these settings can be controlled to filter certain types of noises while tuning others to be heard. For example, music may be processed for output while nature sounds may be muted. This would give an output of music sounds while tuning out sounds classified as nature sounds.

The output then emerges from a speaker so the hearing impaired person can receive the entire processed sound from all the stages of the model. In (Figure 2) is a class diagram of the system.

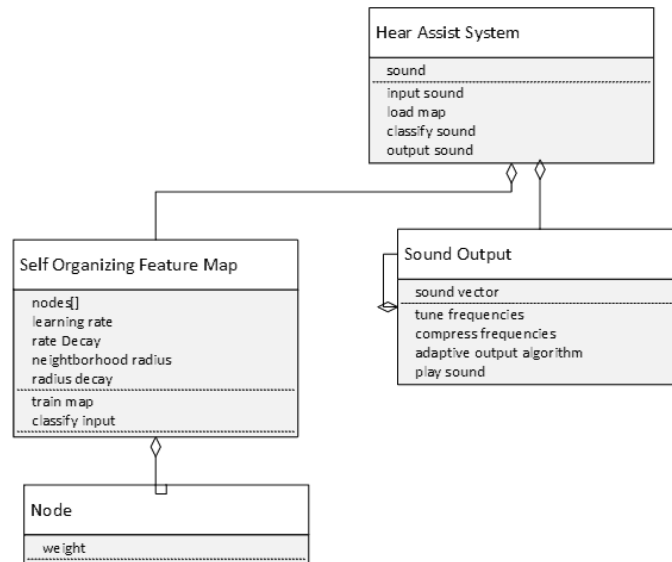


Figure 2: Class diagram model of the proposed system.

It shows that the main system contains subsystems of the self-organizing feature map and the output system. The self-organizing map contains an array of nodes. The map object trains and classifies sounds. The output object tunes these sounds for the hearing impaired. The main system object connects these two parts.

3 Spectrograms

Like the Fourier series a spectrogram divides a signal into its frequency components. However, a difference is that the components of a spectrogram have differing lengths in time duration along with a frequency. Therefore spectrograms can be used to visualize a signal's duration as well as frequency content.

The theory of using spectrograms for this application comes from the human hearing and brain system role as a spectral analyzer [1] [2]. The ear and brain search for dominant components in a spectrum. From this fact it is likely that sounds with similar spectrums should be similar and separable from other sounds.

Because of the logarithmic separation of frequency sensors in the inner ear, low frequencies have high frequency resolution while high frequencies exhibit high time resolution. This makes the dual time and frequency representation of spectrograms useful for such an application. The descriptive features of sounds in our classification system are therefore based on spectrograms, explained here.

A spectrogram is similar to the Fourier transform but differs in important ways. A spectrogram is a plot of a function in the domain of time and frequency. It can be plotted as time vs. frequency with an amplitude and drawn using color intensity. Examples are shown in (Figure 3).

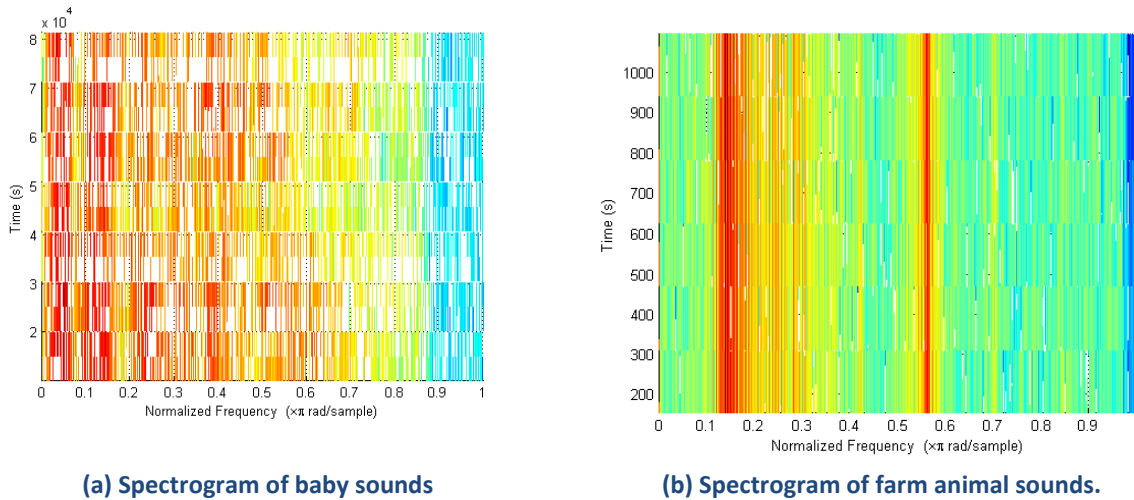


Figure 3

A form of a spectrogram can be created with the short time Fourier transform, defined as [3]:

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t} dt \tag{1}$$

Its discrete form is [9]:

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x(n)w(n-m)e^{-j\omega n} \tag{2}$$

In these two equations w is the window function, a particularly shaped sampling of the data outside the bounds of which the value is 0. The variables τ and m are used for sample shifts.

The spectrogram value used for feature description in our system is the magnitude of this transform squared.

4 Intelligent Unsupervised Learning using Self-Organizing Feature Maps

To classify a sound, we propose an unsupervised application of self-organized feature maps. To classify the sounds, a self-organizing feature map was trained, as explained in this section. The trained map can then classify new sounds into a particular category to control output adjustment accordingly. For example, a deaf person may choose to tune music sounds and block out animal sounds. This system proposes a method to reach this goal.

The method starts with a feature vector describing a sound sample. By dividing the sound's spectrogram values into four quadrants, then dividing the most densely valued quadrant into four more quadrants, a total of seven quadrants is obtained. The mean and range values of each of these quadrants were used to produce a feature vector of size 14. This choice of features captures the bio inspired qualities of a spectrogram, focusing on the most significant regions of this graph as the brain recognizes dominant patterns in the audio spectrum. Patterns which are similar to each other are recognized which can be used to separate different categories of sounds.

Once the feature vectors for a training set are calculated, they are fed into a self-organizing feature map network. This is a two-dimensional matrix of neurons, each with a weight of dimension equal to that of the feature vector. The matrix starts with randomized weight values which are to be adjusted so as to cluster the map into training categories. As a feature vector is fed to the system, the closest weight node and its immediate neighbors are adjusted in weight to be slightly more similar to the feature vector. Through repeated iterations, similar feature qualities will congregate together and separate themselves from differing ones, producing a clustering of the map based on patterns in the feature data.

The equation for the training of a self-organizing feature map is [4]:

$$W_v(t+1) = W_v(t) + \alpha(t) * \beta(t) * [\text{dist}(t) - W_v(t)] \quad (3)$$

For this weight adjustment, increase t , which is a positive constant and repeat from 2 while $t < \lambda$. The parameters are t , the current iteration; λ , the limit on time iteration; W_v , the current weight vector of node v ; $\text{dist}(t)$, the target input data vector; $\alpha(t)$, the learning restraint due to time; and $\beta(t)$, the restraint due to distance from the best matching node, usually called the neighborhood function.

Once a map is trained, it can be used to classify new data. As the self-organizing feature map's signature difference from a typical neural network is its visual interpretation, one can examine the trained maps to see how well the sounds are clustered. The map can be plotted based on vector magnitude, as in (Figure 4a), or color coded based on classifications such as those based on Euclidean distance from a training average, as in (Figure 4b).

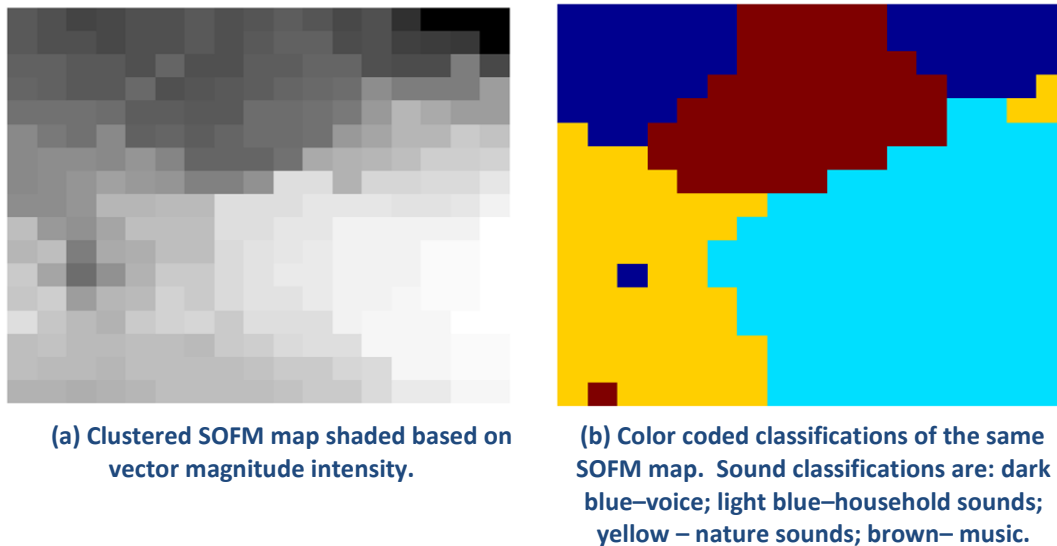


Figure 4

5 Output Processing

To produce the goal of amplifying deaf frequencies for hearing improvement, a number of techniques can be used. Our implementation includes a gamma tone filter deconstruction of a sound to be processed. A gamma tone filter can be defined in the time domain as:

$$g(t) = t^{n-1} e^{-2\pi b t} \cos(2\pi f_0 t + \phi) \quad (4)$$

In this equation, the order of the filter is n which relates to the sharpness of the transition from stop band to pass band and to the gain of each filter. An order 4 filter was used in the system described, which is considered ideal for auditory processing and modeling the human hearing system [5]. The variable b , in Hertz, controls the duration of the impulse response. The variable f_0 is the center frequency. Φ is a phase shift, which represents relative position of the cosine term.

After this filter is applied, a sound will be broken up into a number of samples centered at different frequencies. For quality reconstruction, we used 40 channels, or center frequencies. These were divided logarithmically, in accordance with the logarithmic frequency sensing of the human hearing system.

Once the filter has divided the sounds by their center frequencies, particular deaf frequencies can be boosted as needed, while others are correspondingly lowered. We designed an algorithm to find how to best boost a deaf frequency without sacrificing audible sounds. Given a deaf frequency threshold from 0 to 1 and a hearing frequency threshold from 0 to 1, these define what level of amplitude is required to hear this tone. Next is to define what frequency is deaf.

The algorithm attempts to boost this frequency from below the threshold to above the threshold as required, and measures the loss of the other frequency weights to determine whether they have been lowered beyond audible range. A range of thresholds above and below the input thresholds are tried so the algorithm is adaptive to future changes in hearing ability. The best hearing improvement is used for the output.

Another technique for hearing improvement we used was Fourier compression. After transforming a sound signal into the Fourier domain, we simply compress it by interpolation. This keeps the nature of the frequency signal intact while reducing the range in which it is output. This is useful when a deaf person cannot hear very high or very low frequencies.

Fourier compression in our system is applied using interpolation. In this method, a Fourier spectrum plot is recreated using different frequency axis separations ordered as natural numbers to produce a new plot that is compressed in the x axis direction. For example, a plot 1000 values wide can be compressed to 800 values wide by sampling the plot at intervals of $1000/800$, and numbering them from 100 to 900 for a compressed plot.

Noise reduction was applied in two ways. First, a low pass blurring filter was used in anticipation of high frequency noise. However, the noise may not always originate from such a source. The source may be a fan, vacuum cleaner, traffic, or other such disturbances. For this we take a sample of such a sound and transform it into the Fourier domain. To the output signal we subtract the Fourier mean in the frequency domain. This assumes that the sound was the desired sound with an additional frequency signal from the disturbance. After subtracting this disturbance, the sound can be more faithfully produced for output.

6 Results and Discussion

6.1 Intelligent classification

Based on visual observation and statistical training data, map parameters may need to be adjusted to produce well clustered results. To try to produce an optimal situation for map training, we tested

different parameter values for classification success while holding the others constant. We used a training sample of voice sounds, household sounds, nature sounds, and music sounds.

Learning rate was varied from 0.775 to 1.00 in increments of 0.025. Learning rate decay was varied from 0.775 to 1.00 in increments of 0.025. Neighborhood radius was varied from 0.8 to 8.0 in increments of 0.8. Radius decay was varied from 0.775 to 1.00 in increments of 0.025. Therefore there were ten different values of each of four parameters, for 10 000 combinations for each sample sound set.

The results are shown in (Table 1). These results show the best classifications for each sound set and the parameters in which this classification was made. It is therefore desirable to use these parameters to test the SOFM maps for classification accuracy, described next.

Table 1: Calculated parameters to optimize self-organizing feature map training classification of sounds from voice, household, nature, and music categories.

| Test # | Learn Rate | Learn Rate Decay | Radius | Radius Decay | Average Classification Rate |
|--------|------------|------------------|--------|--------------|-----------------------------|
| 1 | 0.93 | 0.98 | 5.60 | 0.90 | 0.85 |
| 2 | 0.80 | 0.98 | 8.00 | 0.88 | 0.89 |
| 3 | 0.80 | 0.90 | 6.40 | 0.83 | 0.78 |
| 4 | 0.78 | 0.90 | 4.00 | 0.90 | 0.63 |
| 5 | 0.90 | 0.90 | 8.00 | 0.78 | 0.64 |

Based on the optimal parameters for each set of sounds, 100 tests each were done of the same sound samples in which a SOFM map was generated and input sounds were classified as described. Significant results of classification rates are shown in (Table 2).

Table 2: Significant classification results of 24 total voice, household, nature, and music sounds.

| Result # | Voice | Household | Nature | Music |
|----------|-------|-----------|--------|-------|
| 1 | 1.00 | | 0.82 | 0.78 |
| 2 | 1.00 | 0.74 | 0.93 | 0.87 |
| 3 | 0.998 | | 0.80 | |
| 4 | 0.98 | | | 0.82 |
| 5 | 1.00 | | | 0.93 |

Further tests were made to classify different categories of sounds, as shown in (Table 3).

Table 3: Significant classification results of three sets of 24 total sounds each set of different categories.

| Sound Selection | Classification category 1 | Classification category 2 | Classification category 3 | Classification category 4 | Average Classification |
|---|---------------------------|---------------------------|---------------------------|---------------------------|------------------------|
| Babies, Emergency, Fireworks, Livestock | 0.67 | | 0.67 | 0.67 | 0.60 |
| Bells, Crowds, Emergency, Vehicles | 0.83 | 0.83 | 0.67 | | 0.64 |
| Babies, Trains, Crowds, Livestock | 0.67 | | 0.67 | | |

Our results show that intelligent classification is definitely possible. A practical implementation can successfully distinguish certain sounds from each other and process the output accordingly. The weaknesses are that sweeping generalizations cannot easily be made, and perfect results cannot be

expected for what are clearly human defined categories. In the worst cases careful engineering is required to choose training sounds. In any case, some sound samples cluster better than others, and the human intelligence aspect of the system is a subject for further study.

6.2 Output Processing

Our adaptive output algorithm is demonstrated in (Figure 5).

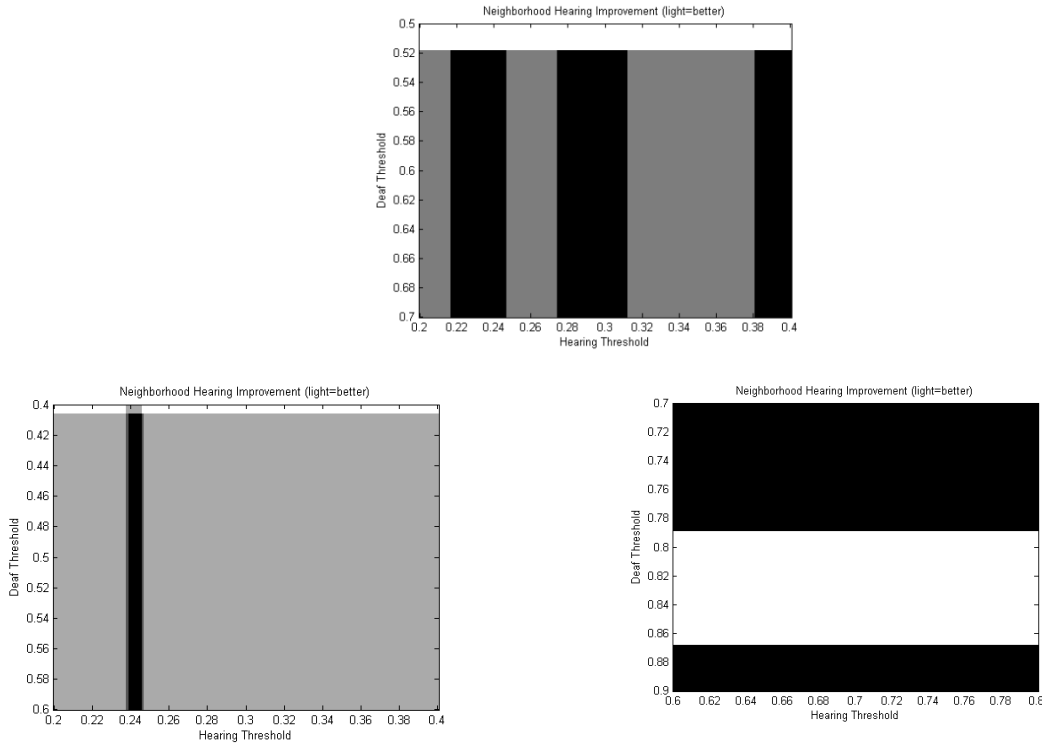


Figure 5: (a) – (c) Example cases of our adaptive output algorithm optimization maps for deaf thresholds and hearing thresholds at optimized improvement ratios.. For a person who needs the two particular threshold levels to hear, and who cannot hear a target frequency, the algorithm calculates hearing improvement for surrounding thresholds based on trying to boost the deaf frequency and considering the loss of audible frequencies in the processing.

In this figure, the thresholds are defined along the axes, and the hearing improvement possible is shaded in the grid. The best improvement is the lightest shades, and the corresponding thresholds are what were used in the calculation.

Results from Fourier compression techniques are shown in (Figure 6).

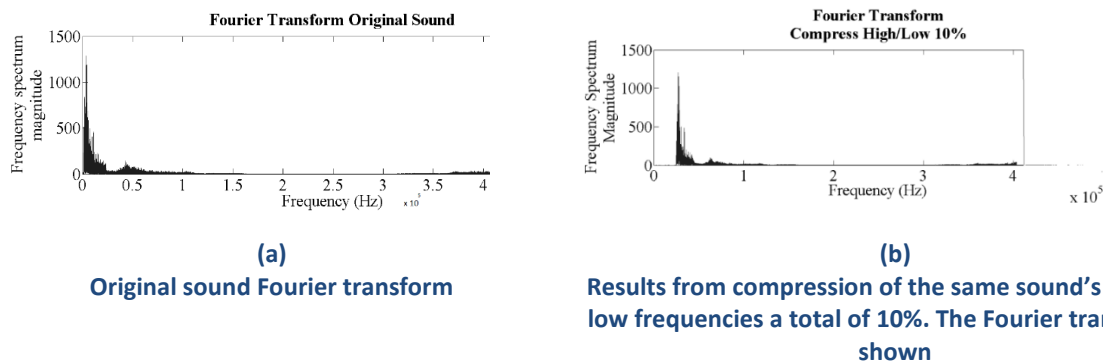


Figure 6

Spectrogram output results are shown in (Figure 7). The figure shows the original spectrogram and the spectrogram from a 10% high and low compression, then a 10% reduction in amplitude of the highest 30% of frequencies.

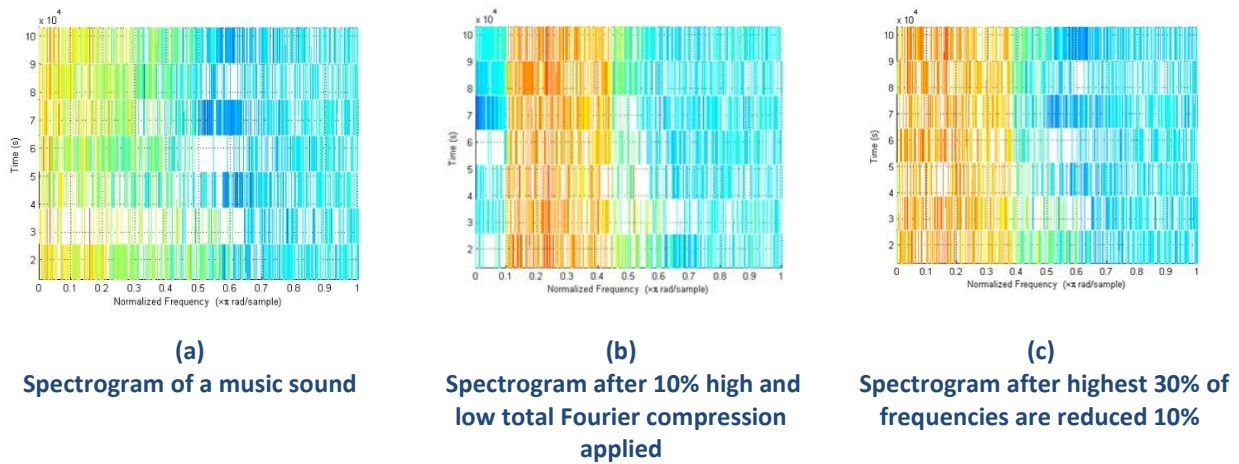


Figure 7

6.3 Hearing metrics

A goal is to describe a metric to measure hearing improvement for the system's adjustable frequencies. Each volume bar in the program controls a range of frequencies f_{low} to f_{high} . In a specific case of the sliding bars, a particular frequency is raised from level x_i to x_f , while the other bars are lowered from y_i to y_f for each bar.

The total proportion weight of the initial state for a target frequency, is:

$$W_i = \frac{x_i}{\sum_y y_i + x_i} \quad (5)$$

The proportion weight of the final state for the target frequency is:

$$W_f = \frac{x_f}{\sum_y y_f + x_f} \quad (6)$$

For a deaf person who cannot hear the target frequency normally, the total improvement in proportion of the target frequency weight from initial to final can be obtained by dividing equation (6) by equation (5):

$$W_{imp} = \frac{W_f}{W_i} \quad (7)$$

The proportional amount of frequency range adjusted by changing a particular sliding bar is:

$$f_{adj} = \frac{\int_{f_{low}}^{f_{high}} f}{\int_0^{f_{max}} f} \quad (8)$$

Using the preceding definitions, the amount of hearing gain proportion obtained by adjusting a particular deaf person's target frequency range, is:

$$\text{Hearing improvement} = W_{imp} f_{adj} \quad (9)$$

This means that for f_{adj} percent of the overall frequency content improved by w_{imp} , the resulting overall hearing improvement can be calculated. This calculation must consider that the y frequencies are not lowered below an audible amplitude when adjusting f .

To describe a metric for the compression in the output program, consider the following. The amount of frequency of the initial sound can be called f . The adjusted frequency is $c*f$, where c is the compression proportion factor.

The total frequency range of the sound sample is:

$$f_{total} = \int_0^{f_{max}} f_{sample} \quad (10)$$

A target frequency range, based on the deaf person's needs, can be defined from its low to high frequency as:

$$f_{target} = \int_{f_{low}}^{f_{high}} f_{sample} \quad (11)$$

For compression, consider when the target frequency of the sound is brought into audible range. The amount brought into audible range is the hearing improvement proportion (multiply by 100 for percent):

$$\text{hearing improvement} = \frac{f_{target}}{f_{total}} \quad (12)$$

In this case, an example is bringing frequencies 3600 – 4000 in a maximum 4000 frequency sample into the audible range by high compression. The improvement would be 400/4000, or 10%, meaning that 10% of the sample has now been changed and brought into the audible range.

The adaptive output algorithm will generally improve hearing ability a minimum of 10%. This is in the case that a deaf frequency can be boosted into hearing range without sacrificing audible frequency ranges. For ten different frequency ranges, if 1 of 10 is deaf, boosting it will improve audibility 10%. In the case where less than 10 of the bands are represented in the output, the hearing improvement will be even greater.

Fourier compression of x percent can bring x percent of the sound into audible range. Therefore, a 10% compression can improve hearing ability 10%, while a 30% compression can improve 30%. However, care must be taken because large amounts of compression can tend to distort the quality of the signal, and even create a ringing effect.

While the hearing improvements may seem somewhat modest, it is very significant that the system allows entirely unheard frequencies to be brought audible. This is an immeasurable improvement because it performs like a miracle to allow a deaf person to hear a brand new part of a sound sample.

7 Conclusions

The project we designed contributes a unique implementation that is not currently readily available in hearing aids. It incorporates intelligent computing to expand the domain of hearing aid technology. Its results are successful to a significant extent, although not perfect. The concept is novel and practical at the same time.

The output processing system we described can produce significant and sometimes drastic improvements in hearing ability. These methods are shown to be practical and effective in the goal of hearing improvement.

The techniques we used were successfully chosen to produce positive results. The feature vector uncovered patterns in sounds to classify them into categories. The self-organizing feature map was a useful tool to cluster and classify sounds. The output techniques each contribute in their own way to an improvement in hearing ability for the hearing impaired.

Further study and tuning of these types of systems can provide more robust and finely engineered solutions capable of practical use in real world applications.

ACKNOWLEDGEMENT

The authors would like to thank Thomas Bryant at Oakland University for his work in developing preliminary models and results of a sound classification system that helped lead to the development of ours.

REFERENCES

- [1] Loizou, P. C., *Mimicking the human ear*. IEEE Signal Processing Magazine, 1998. 15(5): p. 101-130.
- [2] Sottek, R., & Genuit, K., *Models of signal processing in human hearing*. AEUE - International Journal of Electronics and Communications, 2005. 59(3): p. 157-165.
- [3] Allen, J., *Short term spectral analysis, synthesis, and modification by discrete Fourier transform*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1977. 25(3): p. 235-238.
- [4] Bryant, T., Hodges, M., & Zohdy, M. *Self-organizing maps applied to engine health diagnostics*. International Journal of Computer and Information Technology, 2014 3(2): p. 205-211.
- [5] Abdel-Aty-Zohdy, H. S., & Zohdy, M. A., *Neural networks for pattern discovery and optimization in signal processing and applications*. , 1995. 1: p. 202-206 vol.1.
- [6] Bryant, T., & Zohdy, M., *Noise signal identification by modified self-organizing maps*. International Journal of Computer and Information Technology, 2014. 3(1): p. 48-53.
- [7] Chung, K., Killion, M. C., & Christensen, L. A., *Ranking hearing aid input-output functions for understanding low-, conversational-, and high-level speech in multitalker babble*. Journal of Speech, Language, and Hearing Research, 2007. 50(2): p. 304-322.
- [8] Cosentino, S., Falk, T., McAlpine, D., & Marquardt, T.,. *Cochlear implant filterbank design and optimization: A simulation study*. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 2014. 22(2): p. 347-353.

- [9] Couturier, M. M. *Disambiguating words with self-organizing maps*. Master's Thesis, Massachusetts Institute of Technology, 2011.
- [10] Gan, R. Z., Wood, M. W., & Dormer, K. J., *Human middle ear transfer function measured by double laser interferometry system*. *Otology & Neurotology* : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 2004. 25(4): p. 423-435.
- [11] Ghaffari, R., Page, S. L., Farrahi, S., Sellon, J. B., & Freeman, D. M., *Electrokinetic properties of the mammalian tectorial membrane*. *Proceedings of the National Academy of Sciences of the United States of America*, 2013. 110(11): p. 4279-4284.
- [12] Gladston, A. R., Vijayalakshmi, P., & Thangavelu, N.; *Improving speech intelligibility in cochlear implants using vocoder-centric acoustic models*, 2012: p. 66-71.
- [13] Gonzalez, R., & Woods, R.; In McDonald M., Dworkin A. (Eds.), *Digital image processing*, 3rd edition ed2008, New Jersey: Pearson Education.
- [14] Kelly, F., Kelly, F., Harte, N., & Harte, N.; *Auditory features revisited for robust speech recognition*. 2010: p. 4456-4459.
- [15] Mayer, R., Aziz, T. A., & Rauber, A.; *Visualising class distribution on self-organising maps*. 2007: p. 359-368. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [16] Ngamkham, W., Ngamkham, W., Sawigun, C., Sawigun, C., Hiseni, S., Hiseni, S., et al., *Analog complex gammatone filter for cochlear implant channels*. 2010: p. 969-972.
- [17] Nsour, A., & Zohdy, M., *Self-organized learning applied to global positioning system (GPS) data*. *Proceedings of the 6th WSEAS International Conference on Signal, Speech, and Image Processing*, Lisbon, Portugal. 2006.
- [18] Puengue, L., Liu, D., & Zohdy, M.; *Modified self-organizing feature maps for classification of mathematical curves*. *International Journal of Computer and Information Technology*, 2013. 2(5), 972-973-980.
- [19] Purves, D., Augustine, G., Fitzpatrick, D., Katz, L., LaMantia, A., McNamara, J., et al., *Neuroscience* ed2001, Sunderland: Sinauer Associates.
- [20] Stekh, Y., Sardieh, F. M. E., & Lobur, M., *Neural network based clustering algorithm*. 2009: p. 168-169.
- [21] Su, M. C., & Chang, H. T., *Fast self-organizing feature map algorithm*. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, 2000. 11(3): p. 721-733.
- [22] Vesanto, J., & Alhoniemi, E., *Clustering of the self-organizing map*. , 2000. 11 (3): p. 586-600.

- [23] Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C.,. *Machine learning in medical imaging*. IEEE Signal Processing Magazine, 2010. 27(4): p. 25-38.
- [24] Wheeler, D. E. *Measurement of industrial hearing loss*. Noise Control, 1955. 1(4), 9.
- [25] Yan, P., Suzuki, K., Wang, F., & Shen, D.,. *Machine learning in medical imaging*. Machine Vision and Applications, 2013. 24(7): p. 1327-1329.
- [26] Zhu, X., Zhang, L., Wei, J., & Zhou, S.,. *Application of wavelets and principal component analysis to process quantitative feature extraction*, 2008.

Identification of Query Forms for Retrieving the Information from Deep Web

¹Nripendra Narayan Das and ²Ela Kumar

¹Rawal Institute of Engineering and Technology, India

²IGDT University for Women, India

¹nripendradas@gmail.com; ²ela_kumar@rediffmail.com

ABSTRACT

Web databases are now present everywhere. The data from the Deep Web cannot be accessed by Search engine and web crawlers directly. The only way to access the hidden database is through query interfaces and filling up number of HTML forms for a specific domain [18]. In this paper a technique called QFORT (QUERY FORM RETRIEVAL TECHNIQUE) has been developed for identifying the relevant query form is presented. Retrieving information from deep web pages using wrappers is a fundamental problem arising in a huge range of web pages of vast practical interests. In this paper, we propose a novel technique to the problem of identifying the query forms from Web pages, which is one of the key problems in automatic extraction approach. The problem is resolved by many authors by using different technique Intensive experiments on real web sites show that the proposed technique can effectively help extracting desired data with high accuracies in most of the cases.

Keywords: Crawler, deep web, wrapper generation, attribute.

1 Introduction

Information searching has been becoming one of the most important and popular activities on the Web. Today's web based crawler retrieve content only from a portion of the web called the publicly indexable web (PIW) [1] i.e. the set of web pages which can be seen purely by following hypertext links, by ignoring search forms and the pages which require user authorization or registration. During studies it was observed that significant fraction of web content lies outside the PIW [18]. It is estimated that there are several million hidden-web sites [2] with various important information whose results are hidden behind the search form. These types of web are called the hidden web or deep web [2]. Pages in the hidden web are dynamically extracted after submitting the query through different search forms.

The great challenge with the researchers are to identify and automatically fill these search forms. As all the pages, which contain the search forms, include at least one HTML form which is used to submit the query parameters. In this paper a technique called QFORT (QUERY FORM RETRIEVAL TECHNIQUE) has been developed for identifying the relevant query form is presented.

Extracting information from the deep web can be categorized as follows:

- (1) There must be detail description of search method.
- (2) Searching of relevant query forms which are relevant to the task.

(3) Search form must be filled up and examine the results of each relevant and useful information.

(4) Query must be formulized.

It is very much challenging for user to retrieve and analyze relevant as well as important information from the deep web automatically. Now a days, the users manually fills input values to web forms and extract data from the returned web pages. In case of user wish to fill complex queries, filling out forms manually is not practical, but these queries are required for many web based applications [18]. In this paper we are concentrating only on one specific domain i.e. Automobile. While browsing different types of automobile web site, it was observed that the web sites are designed in same pattern and users need to fill up number of consecutive forms to retrieve the information. It is explained with an example.

Example 1: As explained above it was observed that all car search related web sites are designed using the same model. As shown in figure-1 the first web page consists of the these fields (i) Make field (ii) Radio/Checkbox button with “New car” and “Used cars (iii) Name of state / city / pin code and (iv) Search button etc. etc.

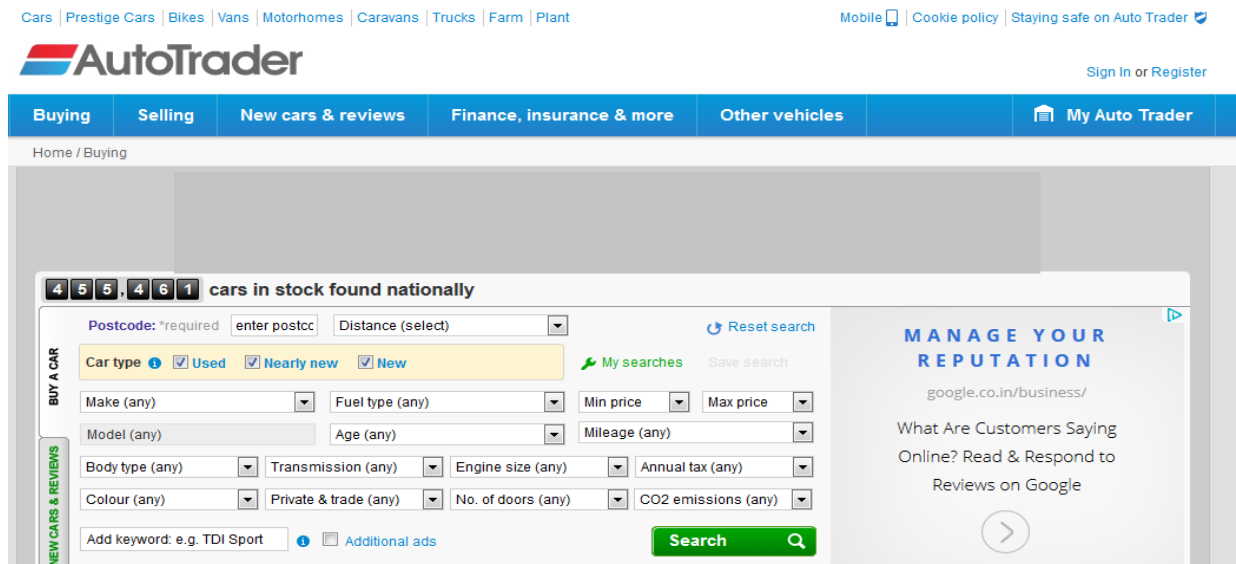


Figure 1: Car Search Form

1.1 Site Form Analysis:

It has been assumed that the given HTML page has a form that applies to our chosen application. In the case when more than one form is on the page, we consider only the largest form—the one with the largest number of characters between the open and closing form tags. Its content are then parsed into a DOM tree.

Form designers create many fields with input tags. For example:

```
<input type="text" size="10" name="zip" maxlength="5" value="">
```

They use select and textarea to create other fields. An example is:

```
<form method="post" class="searchForm" action="http://www.autotrader.co.uk/search/form">
```

```
<select name="radius" id="radius" class="searchFilter">
```

```

<option value="1500">Distance (select) </option>
<option value="1">within 1 mile</option>
-----
-----
-----
<option value="200">within 200 miles</option>
<option value="1501">National</option>
</select>
    </p>
<option class="" value="" >Make (any) </option><option class="" value="abarth" >ABARTH (42)
</option>
-----
-----
-----
<option class="" value="ac" >AC (24) </option>
</select>
<option class="" value="" >Min price </option><option class="" value="0" >Rs. 0
</option><option class="" value="500" >Rs. 5,00,000 </option>
<option class="" value="1000" >Rs. 1,00,000 </option>
-----
-----
-----
<option class="" value="1500" >Rs. 15,00,000 </option>
<option class="" value="2000" >Rs. 20,00,000 </option>
<option class="" value="" >Mileage (any) </option>
<option class="" value="up_to_100_miles" >up to 100 miles (17037) </option>
-----
-----
-----
<option class="" value="up_to_1000_miles" >up to 1,000 miles (22049) </option>
-----
-----
-----
</form>

```

Figure 2: Equivalent codes of Figure-1

Although there are many attributes for the input tag, we are only interested in the type, name, and value attributes. After parsing the input tag, we store the field name, field type, and field value for fields with type text, hidden, checkbox, radio, and submit. For the text area tag, we store the field name with

field type text area. For the select tag, we analyze the content between the opening and closing tags to extract and store the field name, the option values (values inside the option tags), and the displayed values (values displayed in the selection list on the Web page).

User may give any type of unexpected query in search engine text box e.g. suppose the user wants to find information about the “used” Japanese red cars of 2006 made within Rs. 15 Lakh and available in “Delhi”. User will have to fill up the form shown in figure 1 to formulate such types of query by selecting the choice in make field (e.g. Honda, Toyota etc.). More importantly, User has to fill up and submits the form repeatedly for all different Japanese car. As most of the web pages do not contain colour option, user has to browse each and every result for a particular colour car.

The above process can be efficiently completed by using an automatic form querying system, but it is not an easy task to design this type of automated query processing method due to several challenges.

The challenges are as follows:

- (a) Automatic filling of forms: As web pages provides different types of interfaces, automatic filling of forms is a challenging task. Moreover, the user may not be aware of some of the important required field which may be mandatory field for some web site. (E.g. Filling of PIN code to find out the city name is a difficult task for user).
- (b) Extraction of results: As most of the data displayed in result pages of web site are embedded in HTML code and this is another difficult problem to extract the result from the web pages. The search and the extraction of required data from such pages are very much complicated task since each web form interface is designed for user’s convenient and every web page format are always different from each other.
- (c) Navigational complexity: The pages which are generated after submission of query form may contain link to another web pages consists of relevant informations and therefore, it is necessary to navigate these links to see the detail record. It was also observed that during navigation of such web sites repeated filling of web forms are required which are dynamically generated by the server side programs due to submission of pervious query form. These forms are collectively called consecutive forms.

In this paper, a query technique is discussed for the deep web called QFORT (QUERY FORM RETRIEVAL TECHNIQUE) and resolve above mentioned challenges.

To solve the above issues following steps are required:

- (a) *Identification of relevant forms for specific domain database*
- (b) *Automatic filling up of forms with the keywords and display the results.*

The rest of the paper is organized as follows: Section 2 discusses related research in this area. Section 3 and 4 present model for representing normalized forms.

2 Related Work

The huge growth of the deep web has motivated interest in the study of web crawlers [3, 4]. Currently, many research works are going on for the extraction of information from deep web in better way and many solution have been proposed by the researchers. Some of important types are:

- Manual Approach
- Wrapper generation
- Automatic extraction.

Manual Approach: In this approach programmer tries to extract the information from web pages after identifying it's schema by writing equivalent source code.

Wrapper generation [5]: In this approach, set of rules are designed to extract the information from web pages.

Automatic Extraction [6]: Data items have different meaning and role in web pages. In this approach authors have proposed a method to identify the mapping between data items having same role in different pages.

3 Searching of Single Html Form

In this section, the attribute-value representation is discussed in a single HTML form. It was observed that the nature and type of the layout markup are not same in HTML forms. The web application must follow the uniformity in designing the form.

3.1 HTML forms:

An HTML form consists of various parameters as per the requirement e.g. textboxes, dropdown menus, radio buttons, check boxes including banners, advertisement, diagram etc... After filling up all the required information available in forms user submit the form to web server or mail server for further processing. HTML form are designed with the HTML code and a form is embedded inside the <FORM> tag. Each HTML form contains a set of form field and the URL address of server side program so that whenever user submits the query form, server returns a set of result pages.

To process a FORM in server side mainly three essential attributes are required:

- Action attribute – It contains the URL address of the form.
- Method attribute – HTTP method is used to submit the form.
- Enctype attribute – It specify the content type used for form processing.

3.2 Searching of relevant query forms [18]:

In the previous section it was discussed how to design a single form. Sometime the desired results were retrieved by using only a single form called root form. But it was observed in some forms(e.g. auto-search form, book-search form etc..) that user need to fill up more than one form known as the child form(s) or descendant form(s) to retrieve the desired result as shown in figure 1 because the root form has dependency on child form.

Example : As shown in figure-1, if the user wants to search for different model of Maruti make, in this case user has to select the Maruti make in choice make field and after submitting this root form the second form will display all models of Maruti make as shown in figure 1(b).

3.3 Issues in modelling of consecutive forms [18]

Some of the issues were observed in modelling consecutive forms. Most consecutive forms have dependencies between the main and descendant form.

Example: In search interface form as shown in figure-1, user requires two consecutive forms to be fill out, if user searched for “Used Car” & “New Car”. In all existing search forms, it was also observed that there is form dependency exists between form fields. For example, if user selects the “Maruti” in make menu, the all models of “Maruti” make will be displayed in its child form in drop down menu option and if the user wants to see the details of all model one by one he has to select one model at a time and submit the query. It means if there are 10 models available for “Maruti” make (in our example), user has to select 10 times and submit the query to retrieve the results of different model, which is tedious as well as time consuming job.

4 Comparison of PIW Crawler and HW Crawler

In the traditional PIW crawler the basic idea is to follow different web pages from one to another using links. Another way to think it is that the Web is a large directed graph and that the PIW Crawler is simply exploring the graph using a graph traversal algorithm.

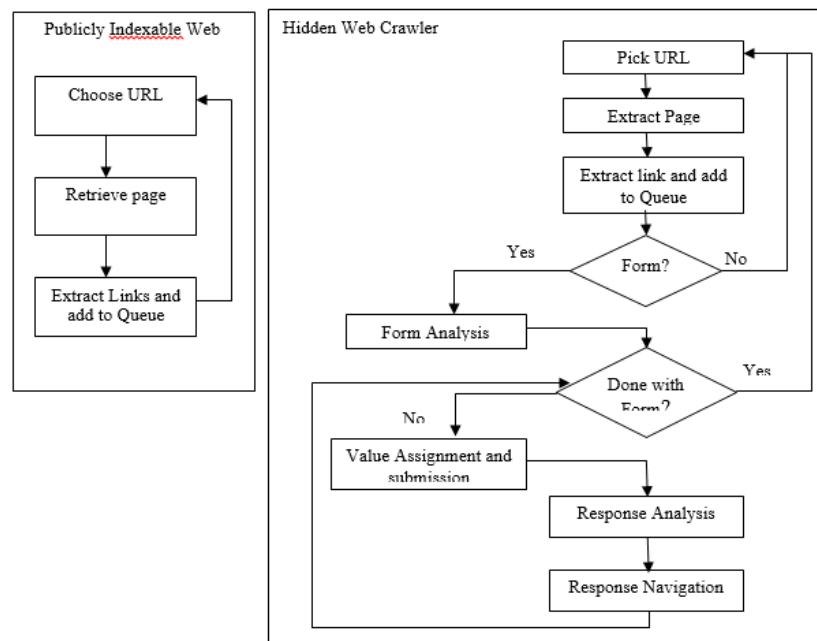


Figure 3: Flow chart comparison of PIW crawler and HW crawler

4.1 Algorithm for Data Extraction from Web pages:

BEGIN

- (i) Insert the keyword in search box (e.g. HONDA city car price 5 lac onwards).
- (ii) Extract the one key word from the number of keywords user entered in the search box.
- (iii) Search the web addresses stored in database (The database store the web addresses of different domain).
- (iv) if keyword matches with the keyword of web addresses, the web page will be opened internally and the corresponding results will be extracted from the that particular web page and display the result. Likewise software will start searching on remaining web addresses and extract the results accordingly.

END

5 Observation

It was observed that in existing form query interface, user fill up the keyword in search box and after clicking the submit/search button server sends the results as web addresses. And suppose user is not satisfied with the result of that web page he/she has to open another web link and compare the results as per his/her requirement and he/she will keep on browsing the link unless and until he/she does not get satisfied with the results. But in our approach user need to fill the query in search box in the beginning only and after that crawler will search in the database equivalent web address, open it internally and extract the desired result from that web site and it follow the same procedure for other web addresses also.

6 Experimental Results and Analysis

In this paper, an experimented was conducted on ten Web sites for each of two applications: car ads and Hotel Booking ads. The approach, however, is not limited to the two applications on which experiment was conducted. It can work with other applications as long as those applications have Web sites with forms, and we have ontologies for those applications. The process of rewriting queries in terms of site forms is the same.

6.1 Experimental Results

We are interested in three kinds of measurements: field-matching efficiency, query submission efficiency, and post-processing efficiency.

To know if we properly matched the fields in a user query with the fields in a site query, we measured the ratio of the number of correctly matched fields to the total number of fields that could have been matched (a recall ratio for field matching (f_m)), and we measure the ratio of the number of correctly matched fields to the number of correctly matched fields plus the number of incorrectly matched fields (a precision ratio for field matching (f_m)):

$$\text{Recall}_{f_m} = \frac{\text{Number of correctly matched fields}}{\text{Total number of fields that should have been matched}} \quad \text{Precision}_{f_m} = \frac{\text{Total number of matched fields}}{\text{Number of correctly matched fields}}$$

To know if we submitted the query effectively, we measure the ratio of the number of correct queries submitted to the number of queries that should have been submitted (a recall ratio R_{qs} for query submission, qs), and we measure the ratio of the number of correct system queries submitted to the number of correct queries submitted plus the number of incorrectly submitted queries (a precision ratio P_{qs}):

$$\text{Recall}_{qs} = \frac{\text{Number of correct queries submitted}}{\text{Total number of queries that should have been submitted}} \quad \text{Precision}_{qs} = \frac{\text{Number of correct queries submitted}}{\text{Total number of queries submitted}}$$

An overall efficiency measurement was also conducted which was obtain by multiplying the three recall measurements and the three precision measurements together:

$$R_{\text{overall}} = R_{f_m} * R_{qs}$$

$$P_{\text{overall}} = P_{f_m} * P_{qs}$$

Because the two kinds of metrics measure two stages of one single process, we used the products to calculate the overall performance of the process with respect to our extraction ontology.

Table 1: Experiment result of used car search

| | | | |
|--|----------------|------------------|---------|
| Number of Forms : 10 Number of Fields Forms: 42 Number of Fields applicable to the ontology: 37(78.5%) | | | |
| | Field Matching | Query Submission | Overall |
| Recall | 100% (37/37) | 100% (325/325) | 100% |
| Precision | 100% (37/37) | 85.5% (325/380) | 85.5% |

Table 2: Experiment result of Hotel Booking search

| | | | |
|--|----------------|------------------|---------|
| Number of Forms : 10 Number of Fields Forms: 25 Number of Fields applicable to the ontology: 18(72%) | | | |
| | Field Matching | Query Submission | Overall |
| Recall | 94% (17/18) | 100% (212/212) | 94% |
| Precision | 100% (18/18) | 212% (212/212) | 100% |

7 Conclusion

In this research, a system is designed and implemented that can fill out and submit Web forms automatically according to a given user query against a corresponding application extraction ontology. From the returned results, the system extracts information from the pages, puts the extracted records in a database, and queries the database with the original user query to get, to the extent possible, just the relevant data behind these Web forms.

Our system has been tested on two applications: car advertisements and Hotel Booking advertisements. In average, there were 78.9% fields in the site forms that were applicable to the extraction ontologies. The system correctly matched 95.7% of them. Considering only the fields that were applicable to the extraction ontologies and were correctly matched, the system correctly sent out all queries that should have been submitted to the Web sites we tested. It, however, also sent out some additional queries that are not necessary according to the original user query. Among all queries our system submitted for our experiments, only 91.4% of them are necessary. Further, for the Web sites we tested, our Output Analyzer correctly gathered all linked pages. Finally, of the records correctly extracted by Ontos, our system always correctly returned just those records that satisfied the user-specified search criteria.

REFERENCES

- [1]. S. Lawrence, C. L. Giles, Searching the world wide web, Science 280(5360)(1998) 98-100.
M. K. Bergman, The Deep Web: Surfacing Hidden Value, September 2001, <http://www.brightplanet.com/deepcontent/tutorials/deepwebwhitepaper.pdf>.
- [2]. S. Chakrabarti, M. van den Berg, B. Dom, Focused Crawling: A new approach to topic specific web resource discovery, in 8th World Wide Web conference may 1999.
- [3]. J. Cho, H. Gracia-Molina. The evolution of the web and implication for an incremental crawler, in: Proc, 26th Int. Conf. Very Large Data Bases. VLDB, 2000.

- [4]. Yanhong Zhai and Bing Liu, "Extracting Web Data Using Instances Based Learning." WISE conference, 2005.
- [5]. Hu D, Meng X, "Automatic Data Extraction from Data-Rich Web Pages", The 10th Data System for Advanced Applications (DASFAA), Beijing, 2005.
- [6]. LIU Wei, MENG Xiao-Feng, MENG Wei-Yi, A Survey of Deep Web Data Integration, Chinese Journal of Computers, Vol. 30, No. 9, Sept 2007, pp : 1475-1489.
- [7]. B. He, Patel M, Zhang Z, C Chang, Accessing the deep web, Communication of the ACM, Col. 50 (5), May 2007; 95-101.
- [8]. Yoo Jung An, James Geller, Yi-Ta Wu, Soon Ae Chun: Semantic deep web: automatic attribute extraction from the deep web data source. In proceeding of the 2007 ACM symposium on Applied Computing (SAC2007), Seoul, Korea, March, 2007, pp: 1667-1672.
- [9]. Longzhuang Li, Y onghuai Liu, Abel Obregon, and Matt A.Weatherston "Visual Segmentation-Based Data Record Extraction from Web Documents," IEEE IRJ 2007.
- [10]. Wei Liu, Xiaofeng Meng, and Weiyi Meng, "ViDE: A Vision-based Approach for Deep Web Data Extraction," IEEE TKDE, 2009.
- [11]. Hongkun Zhao, Weiyi Meng, Zonghuan Wu, Vijay Raghavan, and Clement Yu, "Fully automatic wrapper generation for search engines," in ACM WWW, 2005.
- [12]. Wu, W., Doan, A., Yu, C.: WebIQ: Learning from the Web to match Deep-Web query interfaces. In: Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE 2006), p. 44 (2006).
- [13]. B. He, Z. Zhang, and K. C.-C. Chang. Knocking the door to the deep web: Integrating web query interfaces. In SIGMOD Conference, System Demonstration, 2004.
- [14]. L. Barbosa and J. Freire. Siphoning hidden-web data through keyword-based interfaces. In SBBD, 2004.
- [15]. A. Ntoulas, P. Zerfos, and J. Cho. Downloading hidden web content. Technical report, UCLA, 2004.
- [16]. J. Cope, N. Craswell, and D. Hawking. Automated Discovery of Search Interfaces on the Web. In Proc. of ADC, pages 181–189, 2003.
- [17]. Das N. N., Kumar Ela, Hidden Web Query Technique for Extracting the Data from Deep Web Data Base, WCECS2012_PP410-414, in proceeding of the world congress on engineering and computer science 2012 vol-1 WCECS October 24-26 2012

Comparison of Metaheuristic Algorithms for Evolving a Neural Controller for an Autonomous Robot

¹Sergii Zhevzyk and ²Wilfried Elmenreich

Institute of Networked and Embedded Systems, Alpen-Adria Universität Klagenfurt, Klagenfurt, Austria;
¹sergii.zhevzyk@aau.at; ²wilfried.elmenreich@aau.at

ABSTRACT

Evolutionary algorithms are a possible way to automatically design the behavior of autonomous robots. In this paper we compare different evolutionary algorithms (EA), namely simple EA, two dimensional cellular EA, and random search, according to their performance in a simple simulation, where a phototaxis robot with two sensors of limited range has to find a light source in a closed area. In our experiments we studied the effects on performance of EA parameters, such as population size and number of generation. The results explain how the choice of the neural network (three-layered or fully-connected) may influence the quality of a final solution.

Our findings indicate that acceptable results can be achieved using all EAs but not with random search. The utilization of a fully-connected neural network allows achieving better results for all EAs as compared to a three-layered neural network. Two dimensional cellular EA and simple EA evolve the best strategies for a robot's behavior which allow the robot to reach the light source in almost all cases.

Keywords: Evolutionary algorithm; neural network; robot simulation.

1 Introduction

One way to create a control system for an autonomous robot is to apply an evolutionary approach for evolving a neural controller. EAs can be used to solve different problems especially in machine learning and function optimization domains [13], [17], [20]. A lot of performed experiments with evolving of adaptive behavior confirm that evolutionary algorithms can generate a number of successful system controllers [17], [9], [10]. In this work we compare three metaheuristic algorithms (simple EA, two dimensional cellular EA, and random search) in their ability to discover the most apt neural controller for a phototaxis robot.

The idea of using natural selection for evolving of control systems for robots was proposed by Turing in 1950s. In the next decades a set of metaheuristic algorithms have been developed [12], [15], but the actual usage of these algorithms started after 1990 by the continuous improvement of computer technology and the advent of evolutionary robotics [5]. A large number of EAs causes a need for a generally accepted methodology that allows comparing different EAs and exploring which parameters significantly affect performance [6]. Comparison of EAs on theoretical level was carried out by He and Yao applying Markov chains [14]. Empirical studies in this field were done by De Jong [7] and by Schaffer, et al. [19]. Their works are mostly applied to genetic programming. In our study we compare

metaheuristic algorithms for evolving a neural controller. Czarn mentioned that the results of theoretical work may not comply with practical outcomes [6]. Therefore, here we focus on practical experiments and analysis of these results.

Programming a controller for a robot can become a challenging task because of algorithm complexity, processing of results from sensors, and actuator modeling [9]. A possible way to overcome this issue is to use an evolutionary algorithm to design a neural controller. Applying this approach raises the issue of selecting an appropriate metaheuristic algorithm and its settings since it can have a significant impact on the quality of the evolved behavior.

The purpose of this paper is to compare metaheuristic algorithms applicable to designing the behavior for an autonomous robot. Robot simulation provides a convenient test bed to compare the performance of differently evolved control algorithms. We used a computer simulation as application of metaheuristic algorithms. We defined a simple computer simulation of an autonomous robot, its neural controller and the environment. The simulation of the robot is based on the former experiment we used in [18] for comparing the evolvability of ANNs and Finite State Machines. The main objective is to compare different metaheuristic algorithms and determine dependencies (e.g., population size, number of generations) which may have an impact on performance. This knowledge can be helpful in selection of an appropriate evolutionary approach in future research. Section 2 of this paper provides background about evolutionary algorithms and briefly describes simple EA, two dimensional EA, and random search. Section 3 defines the problem definition and specifies configurations of the conducted experiments. In Section 4 we examine the results of EAs evaluations. Section 5 concludes the paper.

2 Evolutionary Algorithms

In the past decades a large number of evolutionary algorithms were presented, but in general they are united by the same idea: a population with limited resources competing for resources, therefore activating natural selection. Algorithm 1 shows a generalized functioning scheme of evolutionary algorithms. They work with a pool of candidates, each described by a candidate's genotype. The initial population is filled either with randomly generated representations or with candidates which are specifically adapted for this problem. A parameter N denotes the population size in the evolutionary algorithms.

```

initialize population;
evaluate candidates;
while not (termination criterion) do
    parent selection;
    recombination;
    mutation;
    evaluate candidates;
    survivor selection;
end

```

Algorithm 1: General algorithm of EA.

Operators of variation (recombination, mutation) and selection are the two forces that drive evolution forward. The main role of variation operators is the generation of new candidates for the next evolutionary steps. Selection is used to choose individual genomes from the population for later breeding. The fitness function is a result of the candidates' evaluation in one or multiple simulation runs

and allows to measure a quality of a genotype [8]. The quality of a final solution for an evolutionary algorithm strongly depends on matching of representations, variation operators, and fitness function [11].

2.1 Simple EA

A simple EA is the reflection of the generalized scheme of evolutionary algorithms. Appropriate solutions are achieved through application of variation operators (crossover, mutation) and selection of the most fitted individuals. To use this algorithm, the parameters ρ_e , ρ_r , ρ_m , ρ_c , respectively representing the rate of elite candidates, the rate of randomly selected individuals, the rate of representations for mutation, and the rate of candidates as results of crossover, should be defined. Based on these parameters and known size of the population N , we calculate the number of elite candidates, the number of randomly selected representations, the number of individuals for mutation, the number of candidates for breeding, which are stored in the parameters n_e , n_r , n_m , and n_c , respectively.

```
X ← randomly generated population;
while not (termination criterion) do
  foreach candidate  $x_i$  of X do
    | run experiment for  $x_i$ ;
    | compute the fitness  $f(x)$  of  $x_i$ ;
  end
  descending sort of X based on  $f(x)$ ;
   $X_{new}$  ← empty population ;
  for  $i$  ← 1 to  $n_e$  do
    | add  $x_i$  to  $X_{new}$ ;
  end
   $k$  ←  $n_e$ ;
  for  $i$  ←  $k$  to  $k + n_r$  do
    |  $x_r$  ← randomly selected individual from X;
    | add  $x_r$  to  $X_{new}$ ;
  end
   $k$  ←  $k + n_r$ ;
  for  $i$  ←  $k$  to  $k + n_m$  do
    |  $x_e$  ← randomly selected elite candidate;
    |  $x'_e$  ← mutate  $x_e$ ;
    | add  $x'_e$  to  $X_{new}$ ;
  end
   $k$  ←  $k + n_m$ ;
  for  $i$  ←  $k$  to  $k + n_c$  do
    |  $x_{e1}$  ← randomly selected elite candidate;
    |  $x_{e2}$  ← randomly selected elite candidate;
    |  $x_c$  ← mate  $x_{e1}$  and  $x_{e2}$ ;
  end
   $k$  ←  $k + n_c$ ;
  for  $i$  ←  $k$  to  $N$  do
    |  $x_n$  ← randomly generated individual;
    | add  $x_n$  to  $X_{new}$ ;
  end
   $X$  ←  $X_{new}$ ;
end
```

Algorithm 2: The above pseudo-code outlines the algorithm of simple EA.

Typically, the elite candidates in a population having the highest fitness value are selected for the next generation. Accordingly, their offspring, as results of crossover and mutation, take the places of the less fit representations. However, a small percentage of individuals not belonging to the elite can also be selected to the next generation, because of their property or characteristic in their structure might be useful in the next generation. Moreover, non-elite candidates allow increase diversity of the population and thus increment a number of different unique solutions. Algorithm 2 is describing the implementation of simple EA.

2.2 Cellular EA

A cellular EA (cEA) [23] is a kind of evolutionary algorithms, in which the population is placed in a toroidal two dimensional space. Candidates can only communicate with their neighbors, what corresponds to the behavior of individuals in nature. There are many models of neighborhoods for cEA, such as Von Neumann (linear) neighborhood, Moore (compact) neighborhood, diamond neighborhood and others [16]. Usage of different models can lead to completely different strategies. In our experiments we use only Moore neighborhood with radius one, which means that only the closest neighbors are taken into consideration. The neighborhood R also includes the central candidate for which we calculate the neighborhood. In this evolutionary algorithm the parameters ρ_e , ρ_m , ρ_c , which respectively denote the rate of elite individuals, the probability of mutation, and the probability of crossover, are applied for the neighborhoods. The number of elite candidates n_e is calculated from ρ_e and the neighborhood size. Algorithm 3 shows the pseudo-code for a cellular EA.

```

X ← randomly generated population;
while not (termination criterion) do
  Xnew ← empty population;
  foreach candidate xi of X do
    run experiment for the neighborhood Ri of xi;
    compute the fitness f(x) of Ri;
    descending sort of Ri based on f(x);
    if number of xi in Ri ≤ ne then
      | add xi to Xnew;
    else
      | generate random number r ∈ [0, 1];
      | if r < ρm then
      |   | xe ← randomly selected elite candidate;
      |   | x'e ← mutate xe;
      |   | add x'e to Xnew;
      | else if r < ρm + ρc then
      |   | xe ← randomly selected elite candidate;
      |   | ci ← mate xi and xe;
      |   | add ci to Xnew;
      | else
      |   | xn ← randomly generated individual;
      |   | add xn to Xnew;
      | end
    end
  end
end
X ← Xnew;
end

```

Algorithm 3: The above pseudo-code outlines the algorithm of cellular EA

2.3 Random search

Random search finds a solution using an undirected search (see Algorithm 4). In each generation, all candidates of the population are replaced with randomly generated candidates, which are subsequently evaluated. A single candidate that has the highest fitness value is kept for the next generation. If the search space is small and the number of evaluations, i.e. generations times population size, is comparably high, then this algorithm has a chance to pick an acceptable solution. In case of large search space this chance goes down. Compared to other algorithms, random search does not try to improve candidates via mutation or crossover, therefore it can be treated as an undirected search. The random search approach gives a reference for the size of the search space.

```

X ← randomly generated population;
while not (termination criterion) do
  foreach candidate  $x_i$  of X do
    run experiment for  $x_i$ ;
    compute the fitness  $f(x)$  of  $x_i$ ;
  end
  descending sort of X based on  $f(x)$ ;
  for  $i \leftarrow 2$  to  $N$  do
     $x_r \leftarrow$  randomly generated candidate;
    replace  $i$ -th individual of X by  $x_r$ ;
  end
end
end

```

Algorithm 4: The above pseudo-code outlines the algorithm of random search.

3 Experiment Setup

3.1 Physical setup

Figure 1 sketches the simulation setup of our phototaxis robot searching for a light source. The testbed for our robot is a closed quadratic room. The start position is in the center of this room. The position of the light source is outside a restriction circle with the central point in the center of the room. The restriction circle prevents a finding of the light on the first steps. All environment settings are shown in Table 1.

For our experiments we used a differential wheeled robot with configuration described in Table 2. It has 2 sensors to detect whether the distance to the light source is within their sensing range.

Table 1: Configuration of the environment.

| Parameter name | Parameter value |
|----------------------------------|-----------------|
| Width of the field | 200 cm |
| Diameter of the light source | 10 cm |
| Radius of the restricting circle | 80 cm |

Table 2: Robot configuration parameters.

| Parameter name | Parameter value |
|----------------------------|-----------------|
| Diameter of the robot | 10 cm |
| Diameter of the wheels | 5 cm |
| Range of the sensors | 70 cm |
| Angle of the sensor vision | 45° |
| Maximum speed | 12 cm/s |

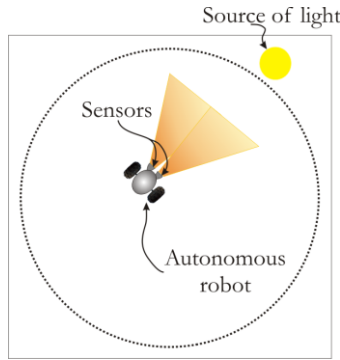


Figure 1: An autonomous robot is looking for the light source in the closed area.

3.2 Fitness function

The main task of the robot is to reach its target in a minimal amount of movements – on this basis we implemented the fitness function as in Equation 1.

$$F = k_t P_t + k_s P_s + k_l P_l \tag{1}$$

$$P_t = \begin{cases} 1 & \text{if robot reached the target} \\ 0 & \text{if robot did not reach the target} \end{cases} \tag{2}$$

$$P_s = \begin{cases} e^{\frac{r-d}{r}-1} & \text{if robot senses the target} \\ 0 & \text{if does not sense the target} \end{cases} \tag{3}$$

$$P_l = e^{\frac{m-l}{m}-1} \tag{4}$$

P_t is the reward for a successful strategy allowing to reach the target (see Equation 2). The value P_s shows how close the robot is to the target at the end of simulation (see Equation 3). This value is especially important in the beginning of an evolution to teach the robot to come closer to the light and finally reach it. The maximum range of the robot’s sensors is represented as parameter r . The distance is encoded in parameter d in case the robot senses the light. Finally, Equation 4 represents how fast the robot can reach the target. The value m represents the maximum amount of time steps in the simulation. The number of time steps that is required to reach the target for the selected strategy is defined as l . Coefficients k_t , k_s and k_l describe the influence of P_t , P_s , and P_l on the fitness value. In our work they have been set to the following values:

$$k_t = 0:3; k_s = 0:3; k_l = 0:4$$

The fitness function is designed in a way that all possible fitness values should lay in the range [0;1]. The maximum number of time steps for our experiments is 300.

3.3 Evolvable control system

The robot was controlled by an ANN. In our simulations we used two different representations: a fully-connected ANN and a three-layered ANN.

The three-layered neural network is a time-discrete ANN which has a feed-forward structure. It means that each neuron of the input layer is connected to each neuron of the hidden layer which at the same

time is connected to each neuron of the output layer. The fully-connected neural network is a discrete-time and recurrent ANN. Instead of feed-forward structure of three-layered neural network, each neuron of the fully-connected neural network is connected to every other neuron and itself, thus making it a recurrent artificial neural network [24]. A recurrent network can retain information about the past, but in general is hard to train [25]. In our case, the training of the two network types follows the same approach of mutating and recombining a genome consisting of weights and biases of the ANN. A fully-connected neural network has a larger search-space whereas it employs more connections between neurons. At the same time this feature and presence of recurrent connections might help to achieve more sophisticated behavior.

The number of inputs and the number of outputs are the same for both candidates. Two inputs which represent distances measured by sensors are connected to the input neurons. From two output neurons we receive information about the speed of robot's wheels. With regard to the number of neurons in the hidden layer, there is no straightforward way to determine the optimal number of hidden neurons analytically. The optimal number depends on the complexity of the function to be approximated, and, therefore, indirectly on the number of input and output nodes. Besides a trial and error approach, there are some empirically derived rules-of-thumb, of these, the most commonly relied on is the optimal size of the hidden layer is usually between the size of the input and size of the output layers [2]. Swingler [22] and Berry [1] propose a maximum of two times the number of input nodes for the hidden nodes. Boger and Guterman [3] suggest that the number of hidden nodes should be 70%-90% of the number of input nodes. Caudill and Butler [4] recommend that the number of hidden nodes should be two third of the sum of input and output nodes. Since determining the optimal number of hidden nodes for a problem is outside the scope of this paper, we have chosen two hidden nodes in accordance with most of the rules of thumb given above.

In our experiments we apply metaheuristic algorithms to train these networks. The main idea of this training is to optimize the weights w_{ji} , where j represents the neurons which have incoming connection to i , and the bias b_i of each neuron i in the ANNs. We calculate the output of the neuron for step k by applying an activation function F :

$$o_i(k) = F\left(\sum_{j=0}^n w_{ji}o_j(k-1) + b_i\right) \quad (5)$$

where the sigmoid function is employed as activation function F :

$$F(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

3.4 Experiment parameters

All experiments are developed using the FREVO tool [21] which has a workflow for the selection of building blocks (problems, representations, evolutionary algorithms and ranking systems) and provides an easy setup for all necessary settings.

Settings of evolutionary algorithms have a huge impact on results of experiments. Information about used configurations is specified in tables 3 and 4.

We conducted a set of experiments with 2000, 5000, 10000, and 100000 evaluations. For each of these values we run experiments 100 times with different initial seeds in order to get sufficient statistical data. The results obtained from these experiments allow to watch an evolutionary process in detail. To check how the number of candidates in the population influences the results of evolutionary algorithms, we used the following population sizes: 25, 36, 49, 64, 81, 100, 121, 144, 169, 196, 225, 256, which are the squares of natural numbers. This is the requirement of cellular EA that builds a toroidal two dimensional space. The number of evaluations equals the population size multiplied by the number of generations.

Table 3: Settings of simple EA.

| Name | Value |
|-----------------------|-------|
| Elite rate | 0.1 |
| Mutation rate | 0.6 |
| Crossover rate | 0.1 |
| Renew rate | 0.1 |
| Random selection rate | 0.1 |
| Mutation severity | 0.3 |
| Mutation probability | 0.3 |

Table 4: Settings of cellular EA.

| Name | Value |
|--------------------------------|-------|
| Elite rate | 0.1 |
| Probability of elite mutation | 0.6 |
| Probability of elite crossover | 0.1 |
| Renew probability | 0.2 |
| Mutation severity | 0.3 |
| Mutation probability | 0.3 |

4 Experiment Setup

We have conducted a set of experiments on evolving the autonomous robot controller using different evolutionary algorithms. Since the runtime of the simulation accounts for the majority of time spent for evaluating solutions, we specified a given number of evaluations for each experiment. Figure 2 depicts the results after 2000 evaluations, which corresponds to a rather short time of evolution. Thus, this figure indicates which algorithm and parameter setting is preferable if there is no possibility for extensive simulation, e.g., there is a limit on run time. The fitness values (ranging from 0 to 1, according to the definition in Section III) show a large dispersion of results. The values for random search mark an inefficient algorithm, while cellular EA and simple EA show comparable good results for short time of evolution.

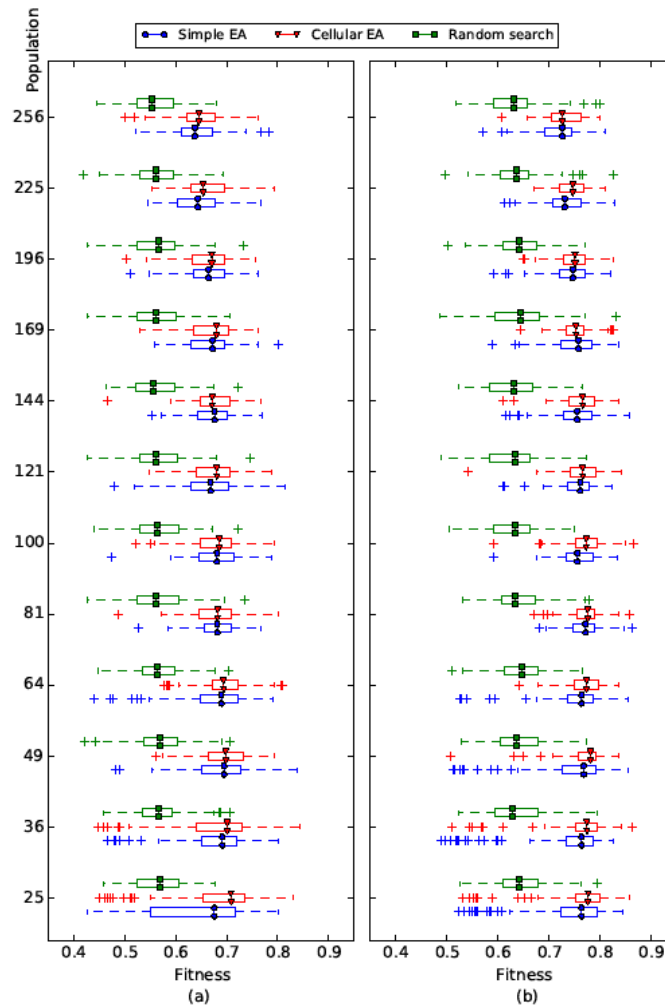


Figure 2: Box-and-whisker diagrams of the fitness values after 2000 evaluations for (a) three-layered ANN and (b) fully-connected ANN.

Figure 3 shows the results after 5000 evaluations which yield better fitness values than after 2000 evaluations. The relative effectiveness of the algorithms stayed the same.

Figures 4 and 5 extend the number of evaluations towards 10000 and 100000, respectively. The latter corresponds to a case where sufficient simulation time is available and the question shifts from *which algorithm provides good results the fastest?* to *which algorithm provides the best results if we wait long enough?*. The fitness values are more gathered after 100000 evaluations, but the performance of evolved controllers is good enough. The difference in terms of efficiency of neural networks after 10000 and 100000 evaluations is negligible compared to waiting time.

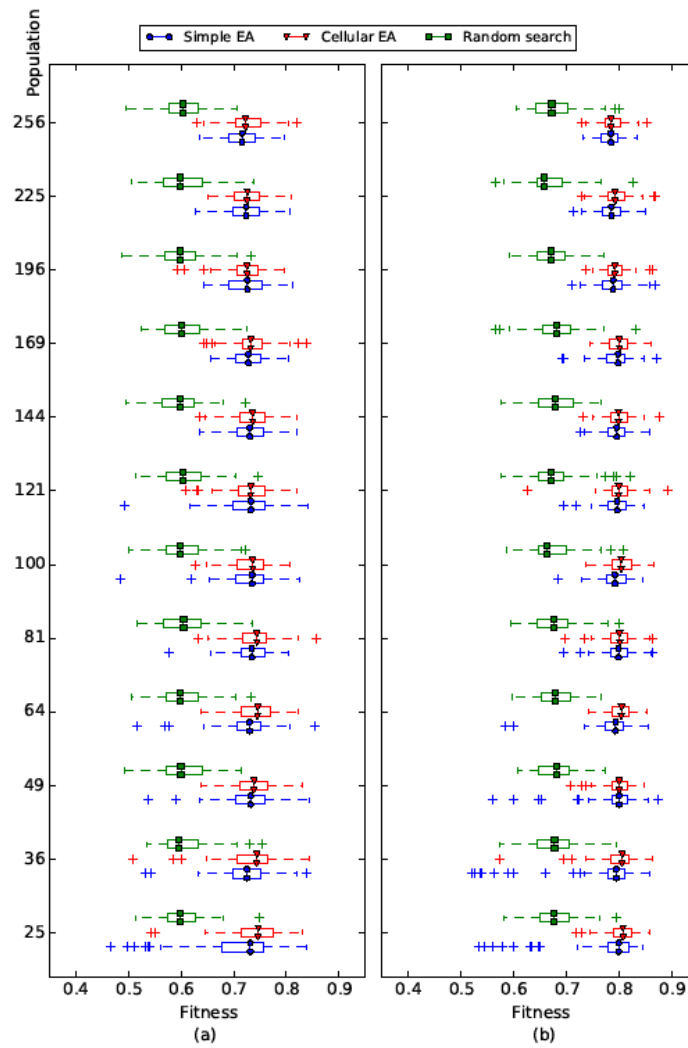


Figure 3: Box-and-whisker diagrams of the fitness values after 5000 evaluations for (a) three-layered ANN and (b) fully-connected ANN.

We can see that fully-connected ANN performs better than three-layered ANN employing all evolutionary algorithms, but with increasing number of evaluations this difference becomes insignificant.

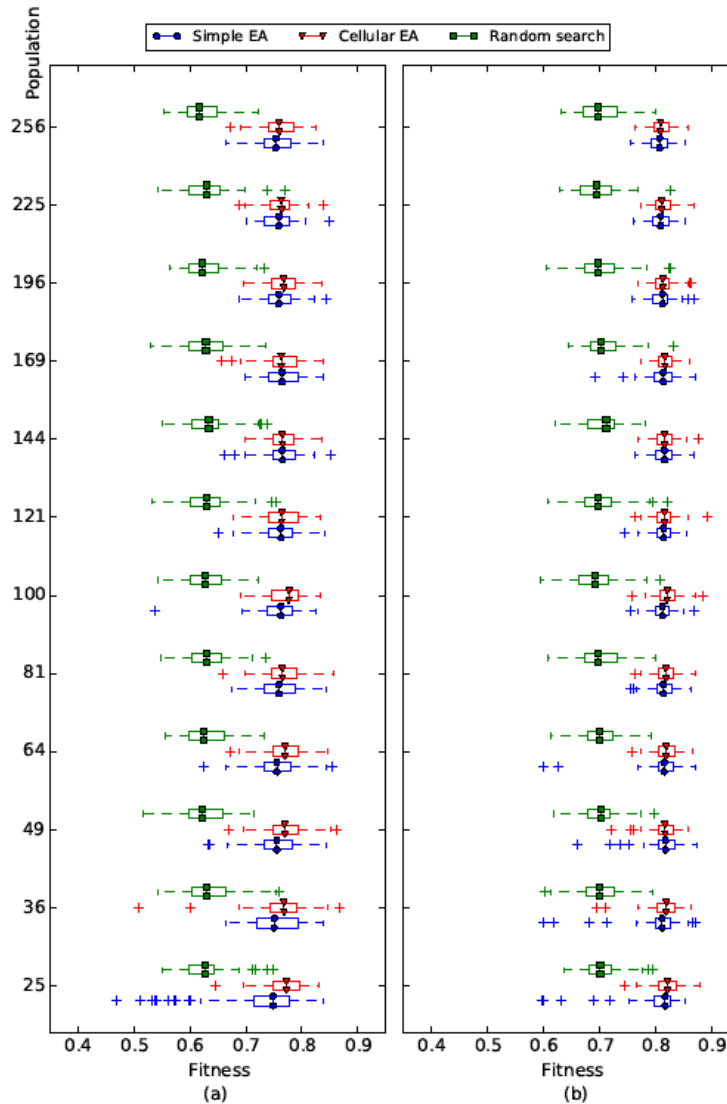


Figure 4: Box-and-whisker diagrams of the fitness values after 10000 evaluations for (a) three-layered ANN and (b) fully-connected ANN.

4.1 Evaluation of Significance

Considering that the results from the simulations are affected by random factors it is not so easy to affirmatively define which algorithm and settings work better and which show similar performance. To answer this question we model the fitness values for the two algorithms as two independent events – X for cEA and Y for simple EA:

$$X \sim N(\mu_X, \sigma_X^2),$$

$$Y \sim N(\mu_Y, \sigma_Y^2),$$

where μ_X , μ_Y are means and σ_X^2 , σ_Y^2 are estimated variances of results measured using multiple simulation runs with different random seeds.

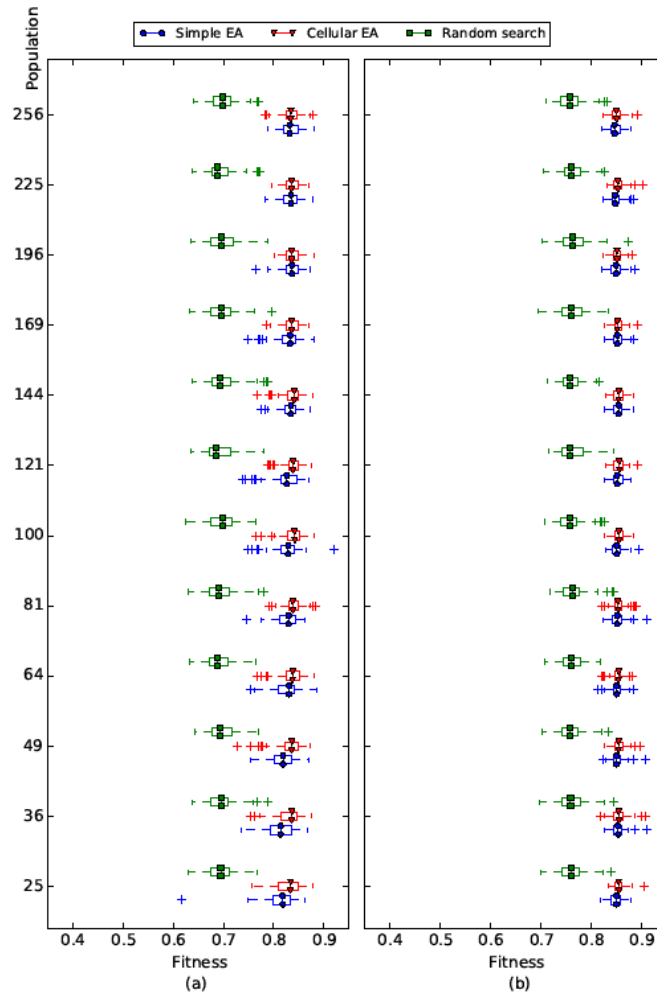


Figure 5: Box-and-whisker diagrams of the fitness values after 100000 evaluations for (a) three-layered ANN and (b) fully-connected ANN.

In the next step, we calculate the difference between two events:

$$Z \sim N(\mu_Z, \sigma_Z^2) \sim X - Y \sim N(\mu_Y, \sigma_Y^2) - N(\mu_X, \sigma_X^2) \sim N(\mu_Y - \mu_X, \sigma_Y^2 - \sigma_X^2) \quad (7)$$

In order to compare cEA and simple EA, we calculate a chance, that probability of Z is less than 0:

$$P(Z \leq 0) = \frac{1}{2} \operatorname{erfc}\left(\frac{\mu_Z}{\sqrt{2}\sigma_Z}\right) \quad (8)$$

The probability $P(Z \leq 0)$ corresponds to the probability that cEA is better than simple EA. The probability that simple EA is better than cEA can be obtained using Equation 9.

$$P(Z > 0) = 1 - P(Z \leq 0) \quad (9)$$

Figure 6 shows the difference between calculated probabilities $P(Z \leq 0)$ and $P(Z > 0)$, which at the same time allows observing how cellular EA is better than simple EA. The trends for different neural networks vary. Cellular EA employing three-layered ANN works better than simple EA for larger number of evaluations. For fully-connected ANN simple EA works better for small number of evaluations and with increasing number of evaluations this difference becomes insignificant. Figure 6 points that for

fully-connected ANN cellular EA provides better results than simple EA. If we employ three-layered ANNs, the cellular EA also dominates over simple EA with a few exceptions.

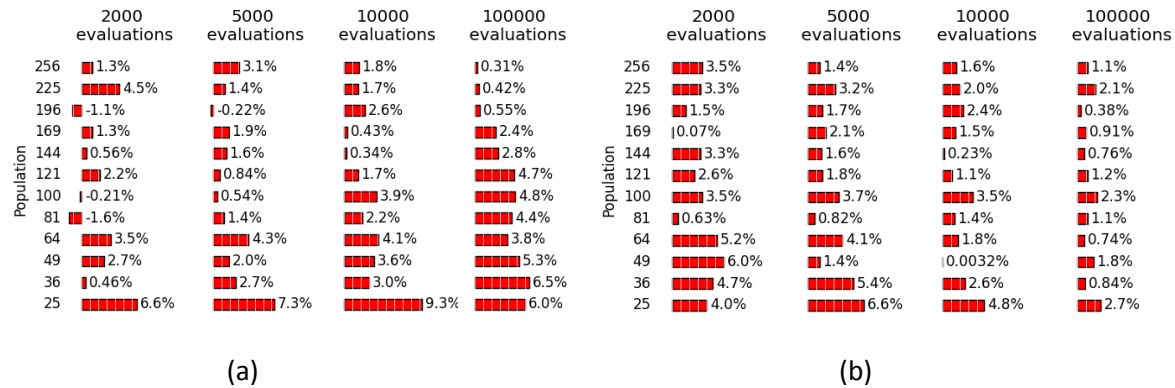


Figure 6: Prevalence of cellular EA compared to simple EA for different neural networks: a) three-layered ANN; b) fully-connected ANN.

5 Conclusion

Two dimensional cellular EA and simple EA show acceptable results in evolving behavioral designs of an autonomous robot. Examination of outcome robot strategies using these algorithms shows that the light source can be found in the vast majority of experiments. Achieved performance results using different evolutionary algorithms demonstrate efficiency of metaheuristic approach for evolving of an autonomous robot.

The results of the experiments help to determine, that cEA and simple EA are the most applicable for evolving a neural controller. A fully-connected ANN outperforms three-layered ANN in all conducted experiments. Based on our findings, we recommend to use cEA and fully-connected ANN for problems that require short evaluation phase. For a large number of generations and population size the efficiency of both algorithms are approximately the same. In the experiments we measured an influence of population size and number of generations on performance of metaheuristic algorithms. The dependencies on these parameters are negligible. This information is important for the conduction of experiments. To accelerate a simulation, the population size should be the same as the number of cores on the server, where these experiments will be performed.

In future work we are planning to validate our results for different application scenarios and to extend our analysis to further parameters, for instance, mutation and crossover rate.

ACKNOWLEDGEMENT

This work was performed in the research cluster Lakeside Labs funded by the European Regional Development Fund, the Carinthian Economic Promotion Fund (KWF), and the state of Austria under grant 20214—21532—32604 (Project MESON) and 20214—22935—34445 (Project Smart Microgrid Lab). We would like to thank A. Monacchi for the feedback.

REFERENCES

- [1]. M. Berry and G. Linoff. *Data Mining Techniques: For Marketing, Sales, and Customer Support*. Database management / Wiley. Wiley, 1997.
- [2]. A. Blum. *Neural Networks in C++: An Object-Oriented Framework for Building Connectionist Systems*. Wiley, 1 edition, 5 1992.
- [3]. Z. Boger and H. Guterman. Knowledge extraction from artificial neural network models. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference*, volume 4, pages 3030–3035 vol.4, 1997.
- [4]. M. Caudill and C. Butler. *Understanding Neural Networks; Computer Explorations*. MIT Press, Cambridge, MA, USA, 1992.
- [5]. D. Cliff, P. Husbands, and I. Harvey. Explorations in evolutionary robotics. *Adaptive Behavior*, 2(1):73–110, 1993.
- [6]. A. Czarn, C. MacNish, K. Vijayan, B. Turlach, and R. Gupta. Statistical exploratory analysis of genetic algorithms. *Evolutionary Computation, IEEE Transactions*, 8(4):405–421, 2004.
- [7]. K. A. De Jong. Analysis of the behavior of a class of genetic adaptive systems. 1975.
- [8]. A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Springer, 2007.
- [9]. W. Elmenreich and G. Klingler. Genetic evolution of a neural network for the autonomous control of a four-wheeled robot. In A. Gelbukh and A. F. Kuri Morales, editors, *Sixth Mexican International Conference on Artificial Intelligence*, pages 396–406. IEEE Computer Society, 2007.
- [10]. D. Floreano and L. Keller. Evolution of adaptive behaviour in robots by means of darwinian selection. *PLoS biology*, 2010.
- [11]. D. Fogel. What is evolutionary computation? *Spectrum, IEEE*, 37(2):26–28, 2000.
- [12]. L. Fogel, A. Owens, and M. Walsh. *Artificial Intelligence Through Simulated Evolution*. John Wiley & Sons, 1966.
- [13]. D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [14]. J. He and X. Yao. From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms. *Evolutionary Computation, IEEE Transactions*, 6(5):495–511, 2002.

- [15]. J. H. Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press, 1975.
- [16]. J. Kari. Theory of cellular automata: a survey. *Theoretical Computer Science*, 334(1):3–33, 2005.
- [17]. A. L. Nelson, E. Grant, and T. C. Henderson. Evolution of neural controllers for competitive game playing with teams of mobile robots. *Robotics and Autonomous Systems*, (46):135–150, 2004.
- [18]. A. Pinter-Bartha, A. Sobe, and W. Elmenreich. Towards the light – Comparing evolved neural network controllers and finite state machine controllers. In *Proceedings of the Tenth International Workshop on Intelligent Solutions in Embedded Systems*, pages 83–87, Klagenfurt, Austria, jul 2012.
- [19]. J. D. Schaffer, R. A. Caruana, L. J. Eshelman, and R. Das. A study of control parameters affecting online performance of genetic algorithms for function optimization. In *Proceedings of the third international conference on Genetic algorithms*, pages 51–60. Morgan Kaufmann Publishers Inc., 1989.
- [20]. M. Sipper, Y. Azaria, A. Hauptman, and Y. Shichel. Designing an evolutionary strategizing machine for game playing and beyond. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, 37(4):583–593, 2007.
- [21]. A. Sobe, I. Fehervari, and W. Elmenreich. FREVO: A tool for evolving and evaluating self-organizing systems. In *Proceedings of the 1st International Workshop on Evaluation for Self-Adaptive and Self-Organizing Systems*, Lyon, France, Sept. 2012.
- [22]. K. Swingler. *Applying Neural Networks: A Practical Guide*. Morgan Kaufmann, pap/dsk edition, 5 1996.
- [23]. M. Tomassini. *Spatially structured evolutionary algorithms: artificial evolution in space and time (natural computing series)*. Springer-Verlag New York, Inc., 2005.
- [24]. M. Schuster, and K. Paliwal. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on* 45, no. 11: 2673-2681, 1997.
- [25]. R. Pascanu, T. Mikolov, Y. Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30 th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013

Ontological Support for the Evolution of Future Services Oriented Architectures

¹Bilal Gonen, ¹Xingang Fang, ¹Eman El-Sheikh, ¹Sikha Bagui, ¹Norman Wilde and ²Alfred Zimmermann

¹Department of Computer Science, University of West Florida, Pensacola, FL, USA;

² Department of Informatics, Reutlingen University, Reutlingen, Germany;

bgonen@uwf.edu; xfang@uwf.edu; eelsheikh@uwf.edu; bagui@uwf.edu; nwilde@uwf.edu;
alfred.zimmermann@reutlingen-university.de;

ABSTRACT

Services Oriented Architectures (SOA) have emerged as a useful framework for developing interoperable, large-scale systems, typically implemented using the Web Services (WS) standards. However, the maintenance and evolution of SOA systems present many challenges. SmartLife applications are intelligent user-centered systems and a special class of SOA systems that present even greater challenges for a software maintainer. Ontologies and ontological modeling can be used to support the evolution of SOA systems. This paper describes the development of a SOA evolution ontology and its use to develop an ontological model of a SOA system. The ontology is based on a standard SOA ontology. The ontological model can be used to provide semantic and visual support for software maintainers during routine maintenance tasks. We discuss a case study to illustrate this approach, as well as the strengths and limitations.

Keywords: Ontology; ontological modeling; services oriented architecture; software evolution; software maintenance; semantic support.

1 Introduction

Services Oriented Architectures (SOA) have emerged as a useful framework for developing interoperable, large-scale systems, typically implemented using the Web Services (WS) standards [1]. SOA typically refers to large systems-of-systems in which composite applications are created by orchestrating loosely coupled service components that run on different nodes and communicate via message passing [2]. Often an infrastructure layer, sometimes called an Enterprise Service Bus (ESB), mediates the communication, providing features such as routing, security, and data transformation. Such systems present several software engineering challenges because they need to orchestrate diverse services having different owners, and have complex reliability requirements. While developing SOA applications presents many software engineering challenges, managing the evolution of such systems presents even greater challenges [3][4].

SmartLife applications are emerging as intelligent user-centered systems that will shape future trends in technology and communication, including social networks, smart devices, and intelligent cars. The development of such applications integrates web services, service-oriented enterprise architectures, cloud computing and Big Data management, among other frameworks and methods [5][6]. SmartLife

systems present even greater software evolution challenges. Such applications exist in complex and dynamic real-world environments with frequent new functional and interface requirements, rapidly emerging security issues, partner services that may be withdrawn or modified at short notice, etc. Software maintainers will need to implement changes rapidly to minimize downtime, yet they will be dealing with a complex combination of external services and internal legacy code, probably not of their own writing.

Although ontologies and ontological modeling have been used to support program comprehension, their integration and application to support maintenance and evolution of complex SOA systems, including SmartLife systems, is novel. Our research integrates multiple research directions to support the evolution of services oriented architectures and address the challenges of future SOA systems [7]. First, we extend a standard SOA ontology to develop a SOA evolution ontology that better supports typical maintenance tasks. We then use this ontology to develop a model of a SOA system and demonstrate how such ontological models can be used to provide semantic support for the maintenance of SOA systems.

In this paper, we present the development and use of ontological modeling methods to address the emerging challenges of SOA evolution. The next section describes Services Oriented Architectures and issues involved in SOA maintenance, including background and related work. Section 3 describes the use of ontologies for modeling SOA systems and specifically for maintenance of such systems. We then show how a standard SOA ontology can be extended to support maintenance in Section 4, followed by a case study that demonstrates this approach on an existing SOA application in Section 5. Finally, we show how such an ontological model can be used in a semantic browser to provide visual support for a maintainer in Section 6, followed by a discussion of conclusions and future work in Section 7.

2 SOA and SOA Maintenance

Services Oriented Architectures (SOA) is not a specific architecture, but rather an architectural style for constructing large software systems. Though definitions of SOA vary, these composite applications are constructed by orchestrating collections of services that run on different nodes and communicate by message passing. Often the messages are mediated by an Enterprise Service Bus (ESB) that provides loose coupling by handling routing, fault tolerance, data translation, etc. (See Figure 1). The services themselves may be heterogeneous, in that they may be implemented using different programming languages, operating systems and vendor infrastructure. Ideally each service represents some specific business functionality but services may also be created by wrapping existing legacy software.

To achieve interoperability, implementers of SOA often adopt some of the Web Services standards [1]. In this case each service will publish an interface in Web Services Description Language (WSDL) and the data contained in messages will be defined in XML Schema Definitions (XSD). In some cases the code will be written using Business Process Execution Language (BPEL) that is essentially a programming language encoded in XML for orchestrating interactions among services.

Large corporations have been adopting SOA because it allows them to create components that can be used across the organization with relatively little inter-divisional coordination. An example is provided by a major multinational insurance group, which has grown by acquisitions to encompass diverse

divisions each with their own IT infrastructure. To control costs and improve agility, the group is developing reusable services for common functions such as claims processing [8].

A great strength of SOA is the flexibility provided by the loose coupling, both between the services themselves and between the IT organizations that manage them. But this flexibility may create significant problems when the composite application needs to evolve. Several authors have pointed out characteristics of SOA that may make such maintenance difficult [3][4][9]. An important part of any maintenance task is acquiring the knowledge needed to make changes to the software; changes made based on incomplete understanding may fail with disastrous results. For full comprehension a maintainer may need to understand many of the technologies that the system uses, including some that may be not be common within his own branch of the organization. Even if he has the necessary technical knowledge, not all of the code or documentation may be available to him.

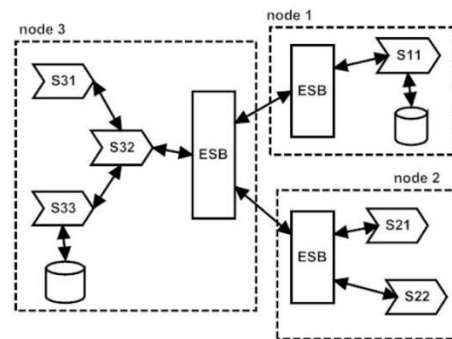


Figure 1: A SOA Composite Application

As to the evolution of future SOA systems with still weaker organizational links, problems of governance may become increasingly challenging. Changes to requirements may need to be negotiated between partners who are not accustomed to close collaboration. Security issues may arise at short notice and it may be difficult to evaluate their impact without knowing details of partners' infrastructure. The mix of partners may also change quickly as new services are offered and old ones withdrawn.

There has been a moderate amount of research on tools and methods to help the maintainer of a SOA application. Papazoglou, Andrikopoulos, and Benbernou categorize changes into "deep" and "shallow" and describe how to perform a "gap analysis" between the existing software and the desired improvement [10]. To provide a view of how an application operates across multiple nodes, several authors propose using dynamic analysis, generally starting from some form of log of inter-process messages. One early proposal was IBM's Web Services Navigator, which provides several visualizations of message logs [11]. A later paper from the same group identifies correlations between different messages by looking at values in specific data fields [12]. An interesting problem sometimes addressed by dynamic analysis is the location of the code that implements one specific "feature" or functionality of a composite application. One technique computes a component relevance index for each message and then displays the messages using a Feature Sequence Viewer [13]. In another approach, Yousefi and Sartipi propose identifying features using an analysis of dynamic call trees from distributed execution traces [14].

Dynamic analysis techniques are very powerful, but it can be difficult to collect the necessary data since usually a crafted set of tests needs to be run on a system which is instrumented to allow the collection

of correlated message logs from the different nodes. Static techniques instead rely only on interface descriptions, documentation, and possibly some source code. However they often cannot give as complete results without considerable human input. For example one method recovers concept maps from the interface descriptions as a starting point for knowledge engineering interviews with system experts [15]. Another approach uses search methods, enhanced by a rule-based system for extracting abstractions from the service interfaces [16].

3 Ontological Modeling

In philosophy the concept of ontology refers to the study of things that exist [17][18]. Ontologies define content theories and representations of concepts as objects, their properties and their relationships. Ontologies specify terms for describing specific domain knowledge by representing domain facts using a structured vocabulary of concepts [17]. The body of knowledge describes a specific domain using that vocabulary as a collection of facts. One most common definition of the term “ontology – as a specification of a conceptualization” is from [19] and [20]. Current mechanisms for representing ontologies are closely related and seem to be similar to the object-oriented terminology. A comparison of ontology terms and meanings in different contexts is in [21]. The ontological terms have object-oriented correspondences with similar meanings: concept relates to class, individual to instance, property of object to relationship, property of data-type to attribute, schema or ontology to class model, and knowledge base to object model.

In our understanding architecture evolution ontologies represent a common vocabulary for software engineers [21] who need to share their information based on explicitly defined concepts. Ontologies include the ability to infer automatically transitive knowledge. Our ontological approach has some practical reasons: share the common understanding of service-oriented architecture domains and their structures; reuse the architectural knowledge; make architectural requirements, structures, building blocks explicit; promote reusability of architectural artifacts; separate the architectural knowledge according to orthogonal architectural domains; and classify, analyze, and diagnose enterprise systems according to the service-oriented reference architecture.

For our purpose, as in [17], [18], the ontology is a formal and explicit description of shareable and automatically navigable concepts of our architectural domain. For modeling purposes we use UML class diagrams to represent concepts, and we describe the attributes as properties (sometimes called roles) and role or property restrictions as facets. This ontology structure, together with the instances of these concepts, constitutes the knowledge base. Practically the knowledge base is a growing structure, which starts with the basic concept structures and is enlarged by more or less an amount of growing number of instances.

We have developed exemplarily metamodels and related ontologies seeded by a student research project [22] for the following main architectural domains from the Enterprise Services Architecture Reference Cube (ESARC) as a starting and extendable set of work results: Business & Information Reference Architecture, Information Systems Reference Architecture, and the Technology Reference Architecture. Metamodels are used, as standardized in [23], to define architecture model elements and their relationships for the reference architectures of ESARC. Metamodels define models of models. In

our approach for architectural modeling we use metamodels as an abstraction for architectural elements and relate them to architecture ontologies.

The Reference Model for Service Oriented Architecture of OASIS [24] is an abstract framework, which defines basic generic elements and their relationships of a service-oriented architecture. This reference model is not a standard, but provides a common semantic for different specialized implementations. Reference models are, as in [24], abstract conceptual models of a functional decomposition of model elements together with the data flows between them. Reference architectures, in [25] and [26], are specialized models of a reference model. It is a composition of related architectural elements, which are built from typed building blocks as the result of a pattern-based mapping of reference models to software elements.

The technical standard of Service Oriented Architecture Ontology from [27] defines core concepts, terminology, and semantics of a service-oriented architecture in order to improve the alignment between the business and IT communities. Following stakeholders are potential users of the SOA ontology, related architecture metamodels, as well as concrete architectural building blocks: business people and business architects, architects for the information systems and software architecture, architects for the technological infrastructure, cloud services architects and security architects.

The SOA Ontology in [27] is represented in the Web Ontology Language (OWL) [18]. The ontology models the core concepts of SOA as classes and properties. The SOA ontology includes additional natural language description of the main concepts and relationships in UML diagrams, which show graphically the semantic concepts as classes and the properties as UML associations. The intent of the UML diagrams are for explanations only, but are helpful constructs for understanding the modeled domain of SOA architecture and more concise than the more spacious formal descriptions in OWL. The SOA ontology defines the relationships between semantic concepts, without mentioning the exact usage of these architectural concepts. To illustrate the SOA ontology the standard uses examples and descriptions of these in natural language.

The two core concepts of the SOA ontology, as in [27], are: System and Element. These two core concepts are generic and often used concepts to define a composite structure of systems that have elements. These abstract meaning of systems and elements is used in different specific architectural modeling situations. Using the concept "*Element*" the technical standard associates following core properties: *uses* and *usedBy* as well as the properties *representedBy* and *represents*. The technical standard of SOA Ontology defines additional concepts of the SOA Ontology like *HumanActor*, *Task*, *Service*, *ServiceContract*, *Effect*, *ServiceInterface*, *InformationType*, *Composition*, *ServiceComposition*, *Process*, *Policy*, and *Event*.

4 A SOA Evolution Ontology

The starting point for our SOA Evolution Ontology is the Open Group SOA ontology technical standard [27], developed by the Open Group in order to provide a common understanding of services oriented architectures and help improve alignment between the business and information technology community. Compared to other service-related ontologies such as OWL-S, WSDL-S and WSMO that focus mostly on the application of service discovery and invocation and composition, the Open Group SOA ontology seems a more appropriate point of departure given the maintainer's need to comprehend the system at multiple levels. The Open Group Ontology is defined in the web ontology language (OWL)

and is ready for extension and population for specific applications. In this ontology, 15 classes and 30 object properties are defined. The class hierarchy is shown in Figure 2.

To develop the SOA Evolution Ontology, the Open Group SOA ontology was extended to improve support for software maintainers. Specifically, the Service class was sub-classed into InternalService and ExternalService classes to distinguish services owned by the maintainer's organization from those owned by partners. A ProcessingModule class was added as a subclass of Element to model code components that perform the InternalServices. The software artifacts that perform ExternalServices are out of the SOA system boundary. A subclass of the top-level Thing class DataItem was added to contain information on every data item appearing in service input or response messages (InformationType in SOA ontology). Detailed data items would add a lot of information to help a maintainer trace problems located in message transmission. The extended class hierarchy is shown in Figure 3. The corresponding object properties hasDataItem and isDataItemOf were defined between InformationType and DataItem classes.



Figure 2: The Open Group SOA Ontology Class Hierarchy

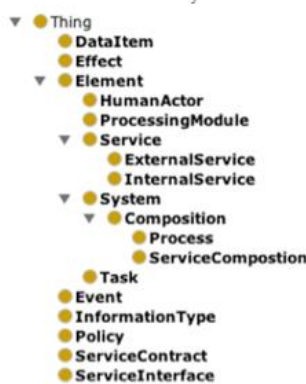


Figure 2: The Extended Open Group SOA Ontology Class Hierarchy

These extensions enhance the ability of the ontology to support software maintenance queries. With the added classes, much more relevant information for maintenance purposes may be stored as individuals (the OWL term for class instances) in the ontology and thus be available for future queries.

In a typical populated ontology for a SOA system, the basic building blocks are service clusters with InternalService or ExternalService individual in the centre with ServiceInterface, InformationType and DataItem individuals attached directly or indirectly. A service in SOA ontology is different from a service in a SOA system. In a SOA system a service (as defined in a single WSDL file) may consist of multiple operations while a service in SOA ontology only represents a single operation to simplify the relationship between service class (InternalService or ExternalService) individuals and ServiceInterface class individuals. With service defined as a single operation, only one ServiceInterface individual is attached to a service, which reduces the complexity and confusion. Thus, each InternalService or ExternalService individual has a ServiceInterface individual attached to define its interface and each ServiceInterface has two InformationType individuals correspond to its input and output messages. Each InformationType individual contains multiple DataItem individuals that represent fields in a message. Service clusters are

connected by System and ProcessingModule individuals. A System may use multiple services while a ProcessingModule may perform an InternalService and use multiple services. In this way, all service clusters are connected to form the complete SOA ontology.

Due to the complexity of a SOA system, the number of instantiated individuals in the extended SOA ontology is expected to be beyond the capability of manual population. Fortunately, the design of an ontology is capable of accommodating automatic or semi-automatic populations of ontologies, depending on the programming language that the SOA system is coded in. In a typical SOA system, most of information need for the ontology population is defined in standard WSDL, XSD, and optional BPEL files. Everything in a service cluster can be found in a single WSDL file and its related XSD files. With these, Services (belonging to InternalService or ExternalService classes) are instantiated to generate ontology individuals with their related ServiceInterface, InformationType and DataItem individuals. The individuals of these five classes can be populated from the WSDL and XSD files automatically or semi-automatically. Individuals belonging to these five classes consist of the biggest portion of the whole populated ontology. The rest, which is a small portion of individuals, belong to the System and ProcessingModule classes, will be populated to connect the service clusters. Here, the automatic population is possible under specific circumstances (for example, a system all written in BPEL), but it will not be easy because of the relationships that exist in programs written in different programming languages. So the invocation and dependence of relationships between services have to be manually extracted from source code to complete the ontology population. Fortunately, it is likely to be only a small portion of the total number of individuals and relationships.

5 Case Study

To illustrate the use of the SOA Evolution Ontology, we populated it for a small publicly available SOA composite application called WebAutoParts that has been used in earlier studies [17]. WebAutoParts is a hypothetical online automobile parts startup that uses an agile development strategy. Instead of writing large amounts of specialized code, commercial SOA services from well-known vendors are orchestrated using a relatively small amount of in-house BPEL. This application provides an order processing workflow in which incoming orders are first checked to confirm that inventory is available, then sales tax and shipping are computed, and finally the order is stored and a note placed in a message queue to trigger order fulfillment (packing and shipping).

Vendor services are represented by their real WSDLs and include Amazon's SimpleDB and SQS simple queue services, a state sales tax service from strikeiron.com, and a shipping cost service from ecocoma.com. In-house services have XSDs for data types and WSDLs that were generated from BPEL code. The BPEL itself is stubbed since runnable code would require setting up paying accounts with each of the vendors. Table 1 lists the sizes of the artifacts that describe WebAutoParts.

Table 1: Artifacts for WebAutoParts

| File | Lines |
|-----------------------------------|-------|
| InventoryRepository.bpel | 71 |
| OrderProcessing.bpel | 118 |
| InventoryRepositoryArtifacts.wsdl | 77 |
| OrderProcessingArtifacts.wsdl | 69 |
| AmazonSimpleDB.wsdl | 611 |
| QueueService.wsdl | 1043 |
| TaxDataBasic5.wsdl | 436 |
| usps.wsdl | 197 |
| InventoryQuery.xsd | 28 |
| PurchaseOrder.xsd | 36 |
| Total | 2686 |

Most classes of the SOA Maintenance Ontology could be populated automatically by parsing the BPEL, WSDL and XSD artifacts. However for this small test system the individuals were generated by hand, with the help of a few special-purpose scripts. The resulting individuals are as shown in Table 2.

Table 2: Populated SOA Maintenance Ontology

| Class | Count |
|-------------------|-------|
| System | 1 |
| Internal Service | 2 |
| External Service | 5 |
| Processing Module | 2 |
| Information Type | 14 |
| Data Item | 105 |
| Total | 129 |

One way to think about the contribution of the ontology is that it easily provides the maintainer with deep knowledge about the individuals in the system, as opposed to the shallow knowledge provided by simple textual search. For example, while the maintainer could use the find command in an editor to locate an XML tag containing tax Rate in an XSD, it would be much more informative to locate the corresponding individual in the ontology. The class of the individual immediately tells the maintainer he is dealing with a DataItem, and simple transitive queries can show him what Information Types it belongs to, the interfaces that use them, and the services that have those interfaces.

For each class in the ontology we defined a small packet of deep knowledge that we think would be of most use to display to a maintainer. We also defined a parameterized query to collect that deep knowledge (see Table 3).

Table 3: Suggested deep knowledge for each class in the ontology

| Class of individual | Desired knowledge |
|---------------------|---|
| InternalService | What ProcessingModule performs this service? What are this service's input and output Information Types? What other services does its ProcessingModule use? |
| ExternalService | What are this service's input and output Information Types? What Processing Module use this service and what are the corresponding Internal Services? |
| Service Interface | What services is this an interface for? What Information Types and Data Items are in its input and output? |
| InformationType | In what Service Interfaces does this InformationType appear as input or output? What services use those Service Interfaces? What Data Items are part of this InformationType? |
| DataItem | What Information Types contain this DataItem? What Service Interfaces use those Information Types? What services use those Service Interfaces? |
| ProcessingModule | What Internal Services does this ProcessingModule perform and what are their Service Interfaces? What services does this ProcessingModule use and what are their Service Interfaces? |

To see how the ontology and the deep knowledge queries would be useful, consider three plausible maintenance scenarios for WebAutoParts:

Scenario A: Some errors have been encountered in the shipping costs computed when WebAutoParts processes orders. A Software Engineer is trying to understand how WebAutoParts gets shipping costs and what data is exchanged as part of that operation. Thus he needs to locate where the "shipping" concept is located in the system. Correct conclusion: Shipping costs are calculated by external service GetUSPSRate. If there are errors, this would be the service to check.

Scenario B: WebAutoParts has been dealing exclusively within the United States but now plans to enter other countries. One of the many changes requires understanding how the application uses American zip codes since Software Engineers will now have to modify the software to handle other kinds of postal code. The Software Engineer needs to locate where the "zip code" concept is located in the system. Correct conclusion: WebAutoParts will need to make changes to the way it computes shipping costs (currently done by external service GetUSPSRate) and to how it calculates local taxes (currently done by external service GetTaxRateUS).

Scenario C: Amazon is retiring their SimpleDB and replacing it with a new database offering. The Software Engineer needs to understand how the SimpleDB services are used in WebAutoParts. Correct conclusion: SimpleDB is used in two places. Its GetAttributes service is used by the InventoryRespository.bpel processing module while its PutAttributes service is used by the OrderProcessing.bpel processing module.

For each scenario the relevant individuals in the ontology can be found by a simple text search on "shipping", "zip" or "SimpleDB" (See Table 4). Then the packaged queries can be applied to each individual and the correct conclusions can be reached immediately.

For comparison, we asked a Software Engineer with Web Services experience to check how the same conclusions could be reached by a sequence of conventional searches on the WebAutoParts artifacts. For each search the Software Engineer used either a Unix "grep" search when he had to scan multiple files, an editor's search command when he had the problem narrowed down to a single file, or he scrolled multiple pages in a file when he needed to understand the nesting of XML elements. The starting points were the same text searches as used on the ontology. The Software Engineer tried to find

the shortest sequence of searches that would rigorously arrive at a correct conclusion. For each search he counted the number of places in the artifacts that matched his query and categorized them as:

- Useful, if they either provided him with the information he wanted or with a good starting point for the next search, or
- Non-useful, if they had to be examined and discarded.

Table 4: Comparison of Ontology and Conventional Searches

| | Ontology Search | --- | Conventional Search | --- |
|----------|--|--------------------|---------------------|-----------------------|
| Scenario | Relevant Ontology Individuals (one ontology query on each) | Searches Needed | Useful Matches | Non-useful Matches |
| A | DataItem: 1- USPS.GetUSPSRate.Out.Shipping | 10 | 12 | 39 |
| B | DataItems: 1- USPS.GetUSPSRate.In.ZipOrigination 2- USPS.GetUSPSRate.In.ZipDestinatio 3- TaxDataBasic5.GetTaxRateUS.Out.ZIPCode | 25 | 36 | 143 |
| C | External Services: 1- AmazonSimpleDB.PutAttribute 2- AmazonSimpleDB.GetAttributes | 8 | 9 | 6 |

The results of the comparison are shown in Table 4. As can be seen, using the ontology drastically reduces the amount of work needed to arrive at correct conclusions about the software.

6 Ontology-Based Visual Interface for SOA Maintenance

The Semantic Web enables machines to comprehend semantic documents and data, not human speech and writings [28]. The Semantic Web provides an infrastructure that enables not just web pages, but databases, services, programs, sensors, personal devices, and even household appliances to both consume and produce data on the web [29]. Knowledge is represented in ontologies as triples. Triples consist of Subject, Predicate, and Object. For instance, John Smith likes Florida. “Likes” is a relationship between the subject John Smith, and the object Florida.

There are 267 triples in our ontology. Some of them are:

AmazonSimpleDB.GetAttributes.Interface – isInterfaceOf - AmazonSimpleDB.GetAttributes.

AmazonSimpleDB.PutAttributes - hasInterface - AmazonSimpleDB.PutAttributes.Interface.

We developed a tool “Semantic Browser” that uses ontologies for annotating documents with semantic information. The semantic browser tool has three parts: document indexing, annotation, and traversing documents. For the document indexing, we use some Java classes, i.e., Map, HashMap, HashSet classes in the Java API. We put all the documents we want to index in a folder. We select this folder from the Semantic Browser tool. We index these documents with the keywords retrieved from the ontology. There are 114 individuals in the ontology, e.g., “AmazonSimpleDB.PutAttributes”, “AmazonSimpleDB.PutAttributes.Interface”, “USPS.GetUSPSRate.In.ZipDestination”. These individuals are used to index the documents. After the documents are indexed, for a given keyword, the Semantic Browser tool returns the list of documents in which the keyword occurs.

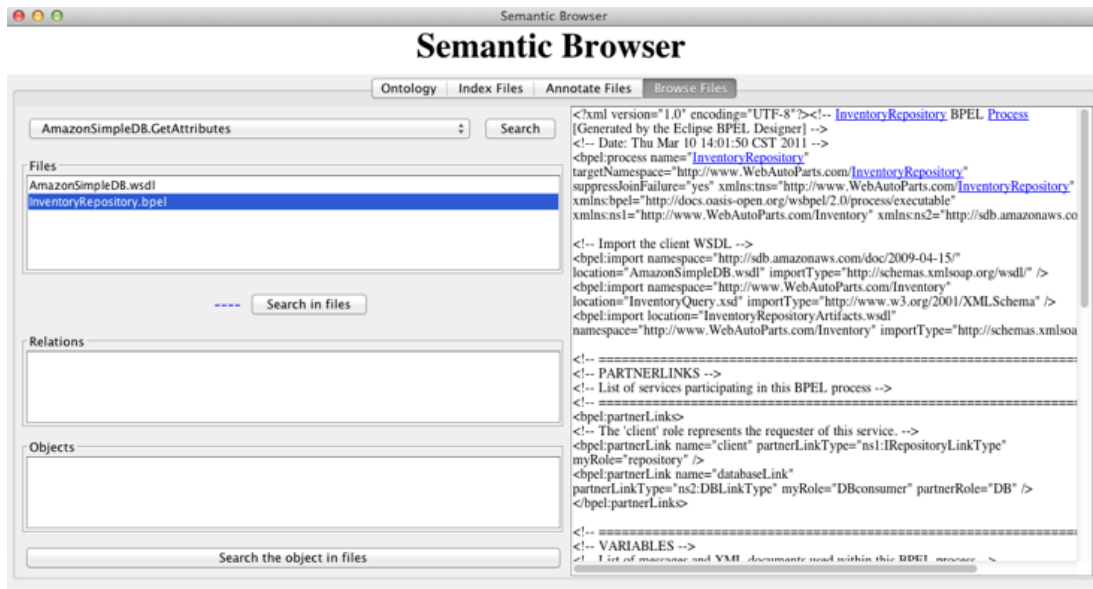


Figure 4: A Screenshot of Semantic Browser Tool

The second part of this Semantic Browser tool does the annotation automatically by processing all of the files in a folder and annotating them. The tool then looks for the individuals from the ontology in the content of the document. These terms are tagged with an html tag. For instance, the term "InventoryRepository" in the document becomes:

```
<a href="InventoryRepository.Module">InventoryRepository</a>.
```

The third part of the Semantic Browser tool does traversing. Figure 4 shows the traversing component of the tool. It allows users to search for related information based on knowledge captured by the ontology. Let's say as a Software Engineer, you are looking for documents containing "GetUSPSRate" and what interfaces it has.

You do an initial query on "GetUSPSRate" and are offered several files containing the term "GetUSPSRate" in them. Assume the "OrderProcessing.bpel" file contains "GetUSPSRate", and the user selects this file. The content of the file appears in the Semantic Browser. The named entities, which we have in our ontology, appear highlighted and underlined. You click on the "GetUSPSRate" in the text, and are offered some relationships, such as: "is a", "has interface". You select "has interface" relationship from the list, and are offered a list of interfaces, which come from the ontology. You select "USPS.GetUSPSRate.Interface" from the list, and are offered all of the files which contain the term "USPS.GetUSPSRate.Interface". After clicking one of the file names from the list, the content of the file appears in the browser.

Consider Scenario C explained in Section 5 let's see how we reach the same conclusion by using the ontological model. By a simple text search on "SimpleDB", we find the relevant individuals in the ontology. We start with the "AmazonSimpleDB.GetAttributes". By querying the ontology, we retrieve two triples from the ontology, which have the "AmazonSimpleDB.GetAttributes" as the subject of the triples. These two triples are:

```
AmazonSimpleDB.GetAttributes -> usedBy -> InventoryRepository.Module
```

```
AmazonSimpleDB.GetAttributes -> hasInterface -> AmazonSimpleDB.GetAttributes.Interface
```

We make another query at the document indexing part of the Semantic Browser tool to find in which document the “InventoryRepository.Module” occurs. As a result, we find that the “InventoryRepository.Module” occurs in the InventoryRepository.bpel file.

One of the benefits of the Semantic Browser tool is to provide semantic links as opposed to physical hyperlinks. Although there was no physical links between the documents, after processing the documents, the documents become linked through the semantic relationships that come from the ontology.

7 Conclusions

In conclusion, maintenance of SOA systems presents many challenges. In this paper, we argued that an ontology-based approach could ease the difficulties of maintenance and introduced an extended SOA Ontology that can be populated with information about any specific SOA-based system. The extensions provide the maintainer with more knowledge of the system, hence more control over the services, traceable steps, and the ability to support software maintenance queries. This ontology extends the Open Group SOA Ontology so it should be compatible with other tools based on this standard. As a first application of the SOA Evolution Ontology we proposed a Semantic Browser tool to aid a maintainer in navigating the many artifacts that describe a SOA system. We illustrated the approach by populating the extended SOA Ontology to model WebAutoParts, an example SOA system.

ACKNOWLEDGMENTS

Work described in this paper was partially supported by the University of West Florida Foundation under the Nystul Eminent Scholar Endowment, and the SOA Innovation Lab Germany.

REFERENCES

- [1]. Nicolai Josuttis, *SOA in Practice: The Art of Distributed System Design*, O'Reilly, 2007, ISBN: 0-596-52955-4.
- [2]. G. Lewis, E. Morris, S. Simanta and D. Smith, "Service Orientation and Systems of Systems," *IEEE Software*, Vol. 28, No. 1, 2011, pp. 58-63. doi:10.1109/MS.2011.15
- [3]. Nicolas Gold, Claire Knight, Andrew Mohan, Malcolm Munro, "Understanding service-oriented software", *IEEE Software*, Vol. 21, March/April 2004, pp. 71-77, doi: 10.1109/ms.2004.1270766.
- [4]. Grace Lewis and Dennis Smith, "Service-Oriented Architecture and its implications for software maintenance and evolution", *Frontiers of Software Maintenance*, 2008. FoSM 2008, pp. 1-10, doi: 10.1109/fosm.2008.4659243.
- [5]. El-Sheikh, E., Bagui, S., Firesmith, D. G., Petrov, I., Wilde, N., Zimmermann, A., 2013. Towards Semantic-Supported SmartLife System Architectures for Big Data Services in the Cloud,

Proceedings of the Fifth International Conference on Advanced Service Computing (Service Computation 2013), May 27 – June 1, Valencia, Spain, pp. 59-64.

- [6]. Zimmermann, A., Pretz, M., Zimmermann, G., Firesmith, D. G., Petrov, I., El-Sheikh, E., 2013. Towards Service-oriented Enterprise Architectures for Big Data Applications in the Cloud, *17th IEEE International EDOC Conference (EDOCW 2013): The Enterprise Computing Conference with SoEA4EE*, 9-13 September 2013, Vancouver, BC, Canada, pp. 130-135, 2013.
- [7]. G. Lewis, "Is SOA Being Pushed Beyond its Limits?" *Advances in Computer Science: an International Journal*, Vol. 2, Issue 1, No. 2, 2013.
- [8]. Deb Ayres, Dave Berry, Babic Hosseinzadeh, Conference Presentation, Oracle OpenWorld, Topic: "Service lifecycle governance", San Francisco, Sept. 21, 2010.
- [9]. Nicolas Gold and Keith Bennett, "Program comprehension for web services, *International Conference on Program Comprehension*, 2004, doi: 10.1109/wpc.2004.1311057.
- [10]. M. P. Papazoglou, V. Andrikopoulos, and S. Benbernou, "Managing Evolving Services," *IEEE Software*, Vol. 28, No. 3, May/June 2011, pp. 49-55, doi: 10.1109/MS.2011.26.
- [11]. W. De Pauw, et al., "Web services navigator: visualizing the execution of web services", *IBM Systems Journal*, Vol. 44, No. 4, Oct. 2005, pp. 821-845, doi: 10.1147/sj.444.0821.
- [12]. W. De Pauw, R. Hoch, and Y. Huang, "Discovering Conversations in Web Services Using Semantic Correlation Analysis", *IEEE 20th International Conference on Web Services, ICWS'2007*, July 2007, pp. 639-646, doi: 10.1109/ICWS.2007.200.
- [13]. John Coffey, Laura White, Norman Wilde, Sharon Simmons, "Locating Software Features in a SOA Composite Application," *ECOWS*, pp.99-106, 2010 Eighth IEEE European Conference on Web Services, 2010.
- [14]. A. Yousefi and K. Sartipi, "Identifying distributed features in SOA by mining dynamic call trees", *IEEE International Conference on Software Maintenance (ICSM)*, Sept. 2011, pp. 73-82, doi: 10.1109/ICSM.2011.6080774.
- [15]. J. Coffey, T. Reichherzer, B. Owsnick-Klewe, and N. Wilde, "Automated Concept Map Generation from Service-Oriented Architecture Artifacts", *Proc. of the Fifth Int. Conference on Concept Mapping CMC2012*, Sept. 2012, pp. 49-56.
- [16]. Eman El-Sheikh, Thomas Reichherzer, Laura White, Norman Wilde, John Coffey, Sikha Bagui, George Goehring, Arthur Baskin, Towards Enhanced Program Comprehension for Service Oriented Architecture (SOA) Systems, *Journal of Software Engineering and Applications*, Volume 6, Number 9, September 2013, pp. 435-445, doi: <http://dx.doi.org/10.4236/jsea.2013.69054>.
- [17]. B. Chandrasekaran, J. R. Josephson, V. R. Benjamins, "What Are Ontologies, and Why Do We Need Them?", *IEEE Intelligent Systems*, pp. 20-26, 1999.

- [18]. D. Gasevic, D. Djuric, V. Devedzic, *Model Driven Engineering and Ontology Development*, Springer 2009.
- [19]. T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, Vol. 5, no. 2, pp. 199-220, 1993.
- [20]. S. Staab, R. Studer (eds.), *Handbook on Ontologies*, Springer, 2004.
- [21]. K. Schneider, *Experience and Knowledge Management in Software Engineering*, Springer 2009.
- [22]. A. Zimmermann, G. Zimmermann, "Enterprise Architecture Ontology for Services Computing," *Service Computation 2012 Nice, France*, pp. 64-69, 2012.
- [23]. OMG, "Meta Object Facility (MOF) Core Specification", Version 2.0, Object Management Group, 2006.
- [24]. C. M. MacKenzie, K. Laskey, F. McCabe, P. F. Brown, and R. Metz, OASIS "Reference Model for Service Oriented Architecture" 1.0, OASIS Standard, 12 October 2006.
- [25]. J. A. Estefan, K. Laskey, F. G. McCabe, and D. Thornton, OASIS "Reference Architecture for Service Oriented Architecture" Version 1.0, OASIS Public Review Draft 1, 23 April, 2008.
- [26]. The Open Group, "SOA Reference Architecture", Technical Standard, 2011, <https://www2.opengroup.org/ogsys/jsp/publications/PublicationDetails.jsp?catalogno=c119>
- [27]. The Open Group, "Service-Oriented Architecture Ontology", Technical Standard, 2010, <https://www2.opengroup.org/ogsys/jsp/publications/PublicationDetails.jsp?catalogno=c104>
- [28]. Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web", *Scientific American*, May 2001, p. 29-37.
- [29]. James Hendler, Tim Berners-Lee and Eric Miller, 'Integrating Applications on the Semantic Web', *Journal of the Institute of Electrical Engineers of Japan*, Vol 122(10), October, 2002, p. 676-680

An Analytical Approach of Computational Complexity for the Method of Multifluid Modeling

¹A.K. Borah and ²P.K. Singh

¹Department of Mathematics, R G Baruah College, Gauhati University, Guwahati-India,

²Department of Mathematics, Allahabad University, Allahabad UP, India,

¹borah.arup@yahoo.com, ²pramod_ksingh@rediffmail.com

ABSTRACT

In this paper we deal the building blocks of the computer simulation of the multiphase flows. Whole simulation procedure can be viewed as two super procedures, the implementation of VOF method and the solution of Navier Stoke’s Equation, Moreover, a sequential code for a Navier Stoke’s solver has been studied.

Keywords: Bi-Conjugate Gradient Stabilized (Bi-CGSTAB), SIMPLE algorithm, Krylov Subspace, ILUT function, Preconditioner, multifluid flows

1 Introduction

The solution of multifluid model have been developed in this studies. Implementation of both steps of VOF (PLIC) method have been derived. The SIMPLE algorithm is used for solving the momentum equation generates large sparse linear systems of equations. These methods are solved by iterative method such as K-S methods. On the other hand convergence rate of this method can be accelerated by preconditioning techniques. Development and improvement of numerical schemes have encouraged researchers to investigate almost every branch of fluid dynamics and its application to real life problem.

Multiphase flows occur in mant industrial and natural phenomena, petroleum refining [1], biological flows [2] and interaction of air with the sea surface [3]. The simulation of multiphase fluid flows in one of the most challenging problem in CFD as it involves in modeling of sharp interfaces separating multiple fluids. The numerical simulation—into fluid flow modeling and multiphase modeling in Fig. (1)

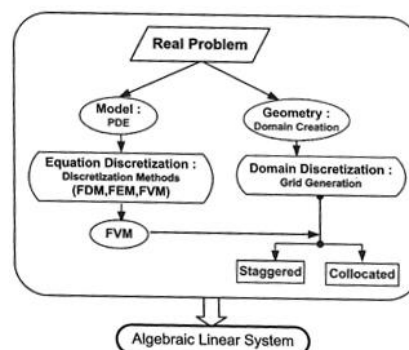


Figure 1: PDE to Linear System

The fluid flow properties (velocity, pressure, etc.) can be represented by partial differential equations such as N-S equation. Furthermore, numerical solutions of these equations constitute fluid flow modeling. Multiphase modeling involves defining the interface between various phases and then calculating the flux in all directions by using the solution from fluid flow model.

2 Multiphase modeling

Maintaining sharp interface during fluid transportation is a difficult task in the modeling of interfacial flows [4]. The interface between two phases can be modeled by scalar transport equation [5]. The modelling involves the construction and movement of the interface. An effective approach for interface modeling is interface capturing, we study a VOF method based on interface capturing approach has been applied. This method has two steps reconstruction and advection of the interface between two fluids [6-7]. Moreover, the interface is established by calculating the volume fractions of each fluid. The transportation algorithm is employed for the movement of the interface. The main challenging of the modeling of the interfacial fluid flow is the implementation of a method which can efficiently move the sharp interface without stretching and wrinkling [6] [8]. In the latter step the interface is approximated by a straight line (a plan 3 dimension) [9]. On the other hand, the area of the geometrical shape represented by triangles or rectangle below the lines is calculated to evaluate the volume of the fluid based on the position of the interface. The interface is approximated in the subsequent time steps by using the volume of fluid of the previous time step and this is where time difficulty maintaining the sharpness of the 8interface in the interface approximation arises [10].

3. *Interface Reconstruction* In the volume of fluid (PLIC) method, the interface between two fluids in a grid cell is approximated by a line segment which intersects the cell's faces. But the line segment divides the cell into two parts- each of them containing one of the two fluids as shown in polygon ABFGD in Fig.(2), the notations are as flows,

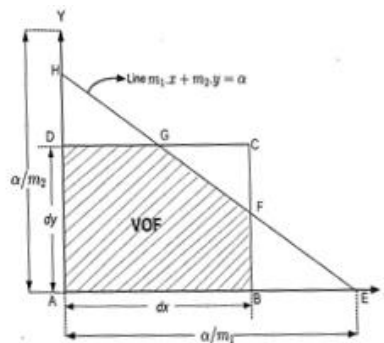


Figure 2: Cell ABCD is cut by the straight line EH and contains fluid 2 in region ABFGD

Rectangle ABCD, represents a grid cell, dx the length of the cell in X direction, dy the length of the cell in Y direction, line EH approximation of the interface, polygon ABFGD: volume of one fluid in the cell and α/m_1 , projection of the segment EH on the X axis.

The general equation of straight line L1 with normal \tilde{m} may be represented as;

$$m_1x + m_2y = 0 \quad (1)$$

Where m_1 and m_2 are the components of the normal vector in the x and y directions respectively and α being the constant which is related to the distance of the line from the origin. The coordinate at the

points at which the line intersects the axes X and Y are respectively $(\alpha/m_1, 0)$ $(0, \alpha/m_2)$ to the points E and H in Fig. (2).

However, in the simulation the values of volume fractions are provided initially for all the cells. But at the next time step the fluid mixture moves and the interface changes its position and hence new values of the volume fractions have to be calculated. In order to evaluate the volume fraction of one fluid in a cell, one has to calculate the area below the line L1, polygon ABFGD in Fig. (2).

2.1 The Estimation of the Normal Vector

In the first part of reconstruction, a normal vector is estimated by a nine-point finite difference formula [6]

$$m = \nabla C \tag{2}$$

Eq. (2) represents the gradient of the colour function C in the direction of coordinates axes. The discrete approximation to Eq. (2) is given by

$$\nabla C = \begin{pmatrix} \nabla^x C \\ \nabla^y C \end{pmatrix} \equiv \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \tag{3}$$

Where ∇^x and ∇^y denotes the gradient in the x and y direction respectively. For approximating the values of these gradient terms, we choose eight nearest neighbors in 2 dimension all the neighbors sharing the vertex, [11]. For a uniform mesh, in a grid cell at location (i, j), the gradient terms of Eq.(3) in the coordinate form can be expressed [6]

$$(m_1)_{i,j} = \frac{1}{\Delta x} (C_{i+1,j+1} + 2 C_{i+1,j} + C_{i+1,j-1} - C_{i-1,j+1} - 2C_{i-1,j} - C_{i-1,j-1}) \tag{4}$$

$$(m_2)_{i,j} = \frac{1}{\Delta y} (C_{i+1,j+1} + 2 C_{i,j+1} + C_{i-1,j+1} - C_{i+1,j-1} - 2C_{i,j-1} - C_{i-1,j-1}) \tag{5}$$

Eq.(3)-(4) represents the normal vector estimation formula for x, y directions [7], this scheme produces a good estimation of the normal vector. An investigation of accuracy test of different normal estimations schemes have been demonstrated [11] that the linear fit (using the nine-point) stencil produce the same order error as other methods such as quadratic fit which require more computations.

2.2 The calculation of VOF from the normal vector and the line constant

$$\square ABFGD = \underbrace{\Delta AEH}_{\Delta_{big}} - \underbrace{\Delta BEH}_{\Delta_1} - \underbrace{\Delta BEH}_{\Delta_2} \tag{6}$$

$$\text{Area} = \frac{\alpha^2}{2m_1m_2\Delta_{big}} \left\{ 1 - H(\alpha - m_1 dx) \left(\frac{\alpha - m_1 dx}{\alpha} \right)^2 - \underbrace{H(\alpha - m_2 dy) \left(\frac{\alpha - m_2 dy}{\alpha} \right)^2}_{\Delta_2} \right\} \tag{7}$$

where H(x) be the Heaviside step function

$$H(x) = \begin{cases} 0 & x < 0; \\ 1 & x > 0 \end{cases} \tag{8}$$

Eq.(6)-(7) calculates the area below the line in Fig.(2) , the procedure for calculating area below the line in two cases for the x direction using Eq.(7) the case of a line with positive slope intersecting the grid cell at the axes as shown in Fig.(3) represents the three triangles

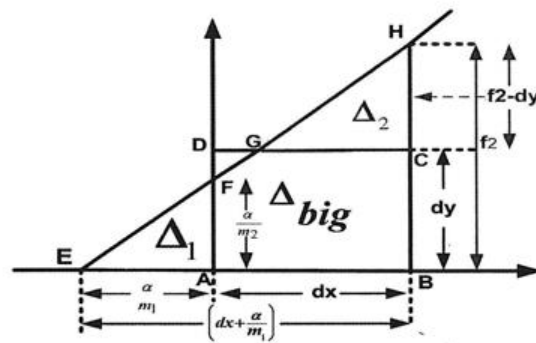


Figure 3: Line of Positive slop intersects the cell faces

Δ_{big} - triangle BEH, Δ_1 - triangle AEF, Δ_2 - triangle CGH

Now the area of the polygon \square is required to calculate the actual volume of the fluid contained in the cell.

$$\Delta_{big} = \left| 0.5 \times \left(dx + \frac{\alpha}{m_1} \right) f_2 \right|$$

$$\Delta_1 = \left| 0.5 \times \frac{\alpha}{m_1} \frac{\alpha}{m_2} \right|$$

$$\Delta_2 = \Delta_{big} \frac{(CH)^2}{(BH)^2}$$

$$\Delta_2 = \Delta_{big} \frac{\{ (f_2 - dy)^2 \}}{f_2^2} = \Delta_{big} \left(1 - \frac{dy}{f_2} \right)^2$$

$$\left(\frac{\alpha}{m_1} < 0 \right) \Rightarrow h_1 = 1 \text{ and } (f_2 - dy) > 0 \Rightarrow h_2 = 1$$

$$\text{Area} = \Delta_{big} \{ 1 - h_1 \cdot \Delta_1 - h_2 \cdot \Delta_2 \} \quad (9)$$

In Eq.(9) h1 and h2 represents the Heaviside step-function $H(\alpha - m_1 dx)$,

$H(\alpha - m_2 dx)$ respectively. A clearer picture of this case is depicted in Fig.(4).

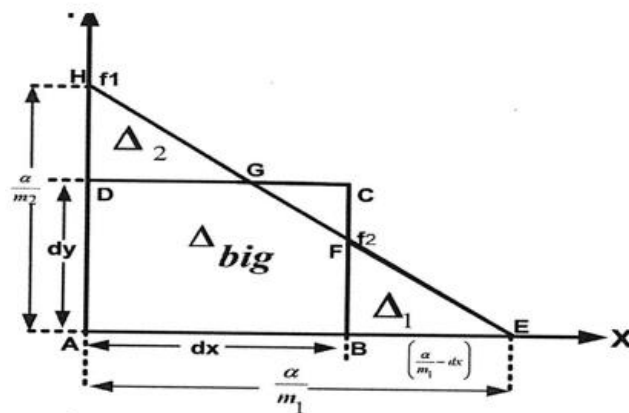


Figure 4: Line of negative slop intersects the cell faces

$$\Delta_{big} = 0.5 \times \frac{\alpha \times \alpha}{m_1 m_2}$$

$$\Delta_1 = \Delta_{big} \left(\frac{BE}{AE} \right)^2 = \Delta_{big} \times \left(\frac{\frac{\alpha}{m_1} - dx}{\frac{\alpha_1}{m_1}} \right)^2$$

$$\Delta_2 = \Delta_{big} \left(\frac{DH}{AH} \right)^2 = \Delta_{big} \times \left(\frac{\frac{\alpha}{m_2} - dy}{\frac{\alpha_2}{m_2}} \right)^2$$

$$\Delta_2 = \Delta_{big} \times \left(\frac{\alpha - m_2 dy}{\alpha} \right)^2$$

$$(\alpha - m_1 dx) > 0 \Rightarrow h1 = 1 \text{ and } (\alpha - m_2 dy) > 0 \Rightarrow h2 = 1$$

$$\text{Area} = \Delta_{big} \{1 - h1. \Delta_1 - h2. \Delta_2\}$$

From the Eqs. (9)-(10) it has been demonstrated that when the line intersects the cell such that it generates two small triangles. But other cases may arise when there is only one triangle $h1=0$ or $h2=0$ or there is no triangle present at all as such $h1=h2=0$.

2.3 Performance of Preconditioners

The preconditioners accelerates the convergence rate of the Krylov-Subspace method. During simulation the matrix is being generated at each time step. Thus the matrix entries changes at each time step which changes the matrix properties. In this study, the Bi-CGSTAB method has been employed because it has been found in literature to provide small convergence for non-symmetric matrices. Now the effects of preconditioners applied to the Bi-CGSTAB method implemented on different problem are demonstrated.

To simulate multifluid flow, the VOF method has been implemented. This method treats the mixture of two fluids as one fluid which is determined by the interface. During the advection the mesh is kept fixed and the interface is reconstructed from the values of the colour function and its gradient in a grid cell. These values contribute to the calculation of the coefficients matrix entries.

Since the fluid moves at each time step, it has been observed that the magnitude of the matrix entries change because of the changes in the interface position. Due to this change the condition number- the ratio of the highest to the lowest eigenvalues of the matrix may vary.

A matrix with high condition number makes Krylov-Subspace solvers converge slowly and so, in order to increase the convergence rate preconditioning techniques are applied [12-13]. Incomplete to L-U Symmetric Successive Over Relaxation (SSOR), Diagonal Scaling (DS) are the are the most appropriate preconditioners for the solvers.

3 Computational Complexity

In ILU factorization, the original matrix A is decomposed into two matrices L and U. Its Algorithm in dense format is shown in 4.1. In this Algorithm, there three nested for loops required (Steps -1, 2 and 9), these nested loops

```

1   for ( i = 0 to n)
2   for (j = i+1 to n) do
3   Uj,i = A j,i ;
    
```

```

4   if (Uj,i = 0.0) then
5   Lj,i = 0.0 ;
6   else
7   L_(j,i) = U_(j,i)/U_(i,i) ;
8   end if
9   for (k = i+1 to n) do
10  if (Uj, k ≠ 0.0) then
11  Uj, k = L j, i X Ui, k
12  end if
13  end for
14  end for
15  L i,, i = 1.0 ;
16  end for

```

Algorithm 4.1 Dense ILUT algorithm [14] generates the data depending of the matrix elements of L and U. This depending implies that for calculating the elements of the (i+1)th , (i+2)th rows, the elements from the ith or previous row are required . Further, this data dependency is a hindrance to parallelizing the Algorithm [15]. In the parallel version the matrix is divided into different parts which are available on different processors of the parallel computer [16]. Therefore, the elements of previous rows may not available on the same processor and those elements have to be brought from other processors.

In Algorithm 4.1 there are three for loops, its computational complexity can be calculated by observing the number of counts in each loop. The outer loop runs from 0 to (n—1) and other loop run from (i+1) to (n—1). The total number of counts can be expressed

$$\sum_{i=0}^n (n-i)^2$$

Which has complexity has order $\mathcal{O}(n^3 - n^2)$ but in algorithm ILUT algorithm in diagonal format, there is one for loop so its complexity is given as $\mathcal{O}(n)$. Moreover, the matrix vector product in Algorithm requires only one for loop. Therefore, its computational complexity can also be given as n.

The computational complexity of the matrix-vector products in other sparse formats [17-18] are 5n.

4 Discussion and Conclusion

The main conclusions can be summarized from the above studies:

- (i) the diagonal format occupies less memory storing penta diagonal matrices.
- (ii) solving the linear system consumes most of the computational time of the simulation.
- (iii) a short description of the iterative methods has been investigated.
- (iv) the Bi-CGSTAB method requires four inner products and two matrix vectors products.
- (v) these products are developed a diagonal format reducing the computational complexity of the solver.
- (vi) the need for preconditioners has been highlighted.
- (vii) the main computational steps of Krylov-Subspace methods have been demonstrated.
- (viii) the parallel Bi-CGSTAB method has been integrated into Navier Stokes solver.

Furthermore, the all the parallel computational tools (C++ code with MPI) for simulating the multiphase flow phenomena have been investigated. In the future work, we will further improve the algorithm.

5 Acknowledgment

The author would like to acknowledge the financial support from the University Grants Commission, Bahadurshah Zafar Marg, New Delhi, India under F. No 40-236/2011 (SR).

REFERENCES

- [1] Mayer, A and Lenhard, R., Special Issue on Multiphase Flows and Chemical Transport, *Advances in Water Resource*, 1998. 21: p.75-76.
- [2] Christopher, E., *Fundamentals of Multiphase Flows* Cambridge University, Press U.K. 2005
- [3] Melville, W., The Role of Surface Wave Breaking in Air-Sea Interaction, *Annual Review of Fluid Mechanics*, 1996. 28: p.279-321.
- [4] Rider, W and Kothe, D., Reconstructing Volume Tracking, *Journal of Computational Physics*, 1998.141: p.112-152.
- [5] Greaves, D., A Quadrative Adaptive method for Simulating Fluid flow with moving Interfaces, *Journal of Computational Physics*, 2004. 194: p. 35-56.
- [6] Rudman, M., Volume Tracking Methods for Interfacial Flow Calculations, *International Journal of Numerical methods in Fluids*, 1997. 24: p. 671-691.
- [7] Denis, G ., Ali, N., Scardovelli, R., and Zaleski, S. Volume-of-fluid interface tracking with Smoothed Surface Stress Methods for Three Dimensional Flows, *Journal of Computational Physics*, 1999. 152: p. 423-456.
- [8] Alibadi, S and Shujaee, K., Free surface flow simulations using parallel finite element method, *SIMULATION*, 2001. 76(5): p. 257-262.
- [9] Scardovelli, R and Zaleski, S., Analytical relations connecting Linear Interfaces and Volume Fractions in Rectangular Grids, *Journal of Computational Physics*, 2000. 164(6): p. 228-237.
- [10] Ruben, S and Zaleski, S., Interface Reconstruction with Least-Square Fit and Split Eulerian Lagragian Advection, *International Journal for Numerical Methods in Fluids*, 2003. 41: 251-274.
- [11] Scardovelli, R and Zaleski, S., Interface Reconstruction with Least Square fit and split Eulerian – Lagragian advection. *International Journal of Numerical Methods in Fluids*, 2003. 41: p. 251-274.
- [12] Saad, Y., *Preconditioning Techniques for Nonsymmetric Indefinite Linear systems*. Technical Report, Centre for Supercomputing Research and Development, 1992. University of Illinois at Urbana Champaign.

- [13] Sun, J, Cao, J and Yang., Paralled Preconditioners for large scale partial differential equations systems, *Journal of Computational and Applied Mathematics*, C. 2009. 226: p. 125-135.

- [14] Saad, Y., *Iterative methods for Sparse Linear System*. PWS Publishing Company, 1996. International Thompson Publishing Inc. Boston.

- [15] Basermann, A., *Parallel Block ILUT/ILDT Preconditioning for Sparse Eigen Problems and Sparse Linear system*. *Numerical Linear Algebra with Applications*, 2000. 7: p. 635-648.

- [16] Li, K, *First and Scalable Parallel Matrix Computations on Distributed memory Systems*. 2005. In 19th IEEE International Parallel and Distributed Processing Symposium.

- [17] Shahnaz, R., Usman, A., and Chugati, I *Implementation and Evaluation of parallel sparse Matrix-Vector products on Distributed Memory Parallel Computers*, Barcelona, Spain. In IEEE International Conference on Cluster Computing (CLUSTER), 2006. Barcelona, Spain, pages 1-6.

- [18] Straubhaar, J. *Parallel Preconditioners for the conjugate Gradient Algorithm using Gram Scimidt and least square methods*, *Parallel Computing*, 2008. 34(10): p. 551-569.

Impulse Noise Removal Technique Based on Fuzzy Logic

¹Mithilesh Atulkar, ²A.S. Zadgaonkar and ³Sanjay Kumar

C V Raman University, Kota, Bilaspur, India

¹m.atulkar@gmail.com, ²arunzad28@hotmail.com, ³sanraipur@rediffmail.com

ABSTRACT

This paper presents an algorithm to remove random impulse noise in a digital grayscale image based on fuzzy logic approach. Numbers of filters are used to remove noise using fuzzy rules. Filter window used the local statistic to determine local weighted mean. Pixel that are detected noisy are filtered, other pixels are remain unchanged. Experimental results show that this method has better performance than other state of the art filters. The effectiveness of this method is 44.95 in terms of peak-signal-to-noise-ratio.

Key words: fuzzy logic, image denoising, impulse noise, filter.

1 Introduction

In today's information world image act as an information carrier like navigation system traffic observation. Images are captured by use of satellite system video camera, mobile and other acquisition devices. It may likely possible images are corrupted, blurred or degraded due to bad acquisition of images, transmission of images in electromagnetic fields or bad recording[1]. Image consists information and since image is corrupted hence some useful information of in images are removed or image is less informative. Different types of noise are identified; some of them are impulse noise, additive (Gaussian) noise and multiplicative noise. In impulse noise pixel value is replaced by a fixed value either maximum value or minimum value sometimes it may be a random value using random distribution [2]. In additive noise a fixed value is added to each pixel to the image, added value is generated using Gaussian distribution hence it is commonly referred as Gaussian noise. In literature many of them deal with Gaussian noise. In this paper we have to concentrate on impulse noise. During literature many filter found that deal with still image corrupted with noise [3-5]. Several works have been done on fuzzy logic which has better performance compared to non-fuzzy approaches [6-8]. Many filter found that describe video image corrupted by Gaussian (additive) noise [9-12]. Analogously it is necessary to made distinction between filters for color image and gray scale image, filters for gray scale can be used in RGB color image by applying filter on each of the color band separately. Here we present a method for noise removal on a gray scale image.

The paper is structured as follows: in Section 2 describe the methodology and different steps in the actual filtering process are detailed in proposed method .Section 3 gives the experimental results. Finally, in Section 4 paper is concluded.

2 Proposed Denoising Algorithm

Consider is a grayscale image to be processed; w is a filtering window of pixel. Filtering window as in figure 2.1 below centered at is surrounded with no of neighbor points of different layer. Points at vertical and horizontal direction of outer layer 1 is stored in and points at diagonals are stored in, similarly points of layer 2 are stored in & and points of layer 3 are stored in &. In each no of noisy points are find out and stored in an array named.

| | | | | | | |
|-----|-----|-----|-------|-----|-----|-----|
| NW3 | | | N3 | | | NE3 |
| | NW2 | | N2 | | NE2 | |
| | | NW1 | N1 | NE1 | | |
| W3 | W2 | W1 | (X,Y) | E1 | E2 | E3 |
| | | SW | S1 | SE1 | | |
| | SW2 | | S2 | | SE2 | |
| SW3 | | | S3 | | | SE3 |

Figure 1: Filter window of size 7*7.

$$vec_1 = [N_1, S_1, W_1, E_1] \quad (1)$$

$$vec_2 = [NW_1, NE_1, SW_1, SE_1] \quad (2)$$

$$vec_3 = [N_2, S_2, W_2, E_2] \quad (3)$$

$$vec_4 = [NW_2, NE_2, SW_2, SE_2] \quad (4)$$

$$vec_5 = [N_3, S_3, W_3, E_3] \quad (5)$$

$$vec_6 = [NW_3, NE_3, SW_3, SE_3] \quad (6)$$

$$vec = [val_1, val_2, val_3, val_4, val_5, val_6] \quad (7)$$

Where, val_i is no of noisy points in any vector (vec_i) is calculated by using count () method as described below for $i=1$ to 6.

$$val_i = count(vec_i) \quad (8)$$

vec is provided as input to the fuzzy inference system to evaluate fuzzy rules. If $F(x, y)$ is corrupted and points of vec_1 is noiseless than mean value of vec_1 is used to assist the restoration of noisy pixel, but if vec_1 is corrupted than vec_2 is used to assist new value to noisy point. It is not necessary for any vec_i that all points are noiseless or corrupted. To solve this problem fuzzy logic is used. Fuzzy rules are used to find the vec_i , which are responsible to restore the noisy pixel. Filter window is 7×7 here but for pixel at corner a limited filter of 5×5 and 3×3 size is used.



Figure 2: Experimental result with Lena image (a) standard Lena image, (b) noisy image with 30% impulse noise, Restoration result by corresponding (c) FIRE[16], (d) BDND[17], (e) DWM, (f) LUO, (g) FMEM, (h) SAM, (i) PROPOSED METHOD filters.



Figure 3: Experimental result with Gold Hill image (a) standard Gold Hill image, (b) noisy image with 30% impulse noise, Restoration result by corresponding (c) FIRE[16], (d)BDND[17], (e) DWM, (f) LUO, (g) FMEM, (h)SAM, (i) PROPOSED METHOD filters.

Figure 2 presents visual quality of resultant denoised image of various methods. Figure 2 (a) shows the original lena image. Figure 2 (b) shows 30% noisy lena image. Figure 2 (i) shows the resultant image of proposed method. Remaining figures are output image of described methods. Figure 3 shows the corresponding original noisy and output image of standard gold hill image.

3 Result & Discussion

Experiment is performed on two standard image of Lena and gold hill image. Value of peak-signal-to-noise-ratio (PSNR) and mean square error (MSE) in images is calculated on the basis of below equation (1) and equation (2).

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) (db) \quad (1)$$

$$MSE = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [y(i, j) - o(i, j)]^2}{M \cdot N} \quad (2)$$

PSNR values of denoised image of various methods at different noise density are detailed in table 1 using standard lena image. It is clear from the results that lower the noise density, higher PSNR value. PSNR values are linearly decreasing order with respect to increasing noise level. PSNR value of all methods in table 1 is also calculated in another image gold hill image. In both table 1 and table 2 FIRE and DWM have greater signal to noise ratio corresponding to all other methods.

Table 1: Performance Comparisons of PSNR Value in Lena Image

| Method | 10% | 20% | 30% | 40% | 50% |
|----------------|-------|-------|-------|-------|-------|
| FIRE [14] | 35.16 | 31.49 | 28.96 | 26.63 | 24.49 |
| BDND [15] | 23.64 | 20.41 | 18.55 | 17.06 | 15.78 |
| DWM [16] | 34.82 | 31.30 | 29.24 | 27.05 | 25.33 |
| LUO [17] | 30.71 | 24.98 | 21.73 | 19.37 | 17.46 |
| FMEM [18] | 34.19 | 28.68 | 23.73 | 20.96 | 17.99 |
| SAM [19] | 22.25 | 19.24 | 17.47 | 16.26 | 15.29 |
| CURRENT METHOD | 44.95 | 42.41 | 41.18 | 39.78 | 38.99 |

Mean Square Error MSE of resultant denoised image by use of different filtering methods in standard lena image is tabulated in table 3. Unlike PSNR, MSE values are increasing linearly for corresponding noise density. DWM and current methods have better MSE at higher noise level. Table 4 also describes the MSE values for all method in previous table in gold hill image. In gold hill image LUO and current methods have better result.

Table 2: Performance Comparisons of PSNR Value in Gold Hill Image

| Method | 10% | 20% | 30% | 40% | 50% |
|-----------|-------|-------|-------|-------|-------|
| FIRE [14] | 34.43 | 31.13 | 28.51 | 26.43 | 24.39 |
| BDND [15] | 23.07 | 19.95 | 18.10 | 16.54 | 15.38 |
| DWM [16] | 34.20 | 30.66 | 28.83 | 26.42 | 24.95 |
| LUO [17] | 30.02 | 24.68 | 21.36 | 19.00 | 17.15 |
| FMEM [18] | 31.90 | 27.52 | 23.32 | 20.57 | 17.71 |
| SAM [19] | 22.02 | 19.08 | 17.27 | 16.04 | 15.05 |
| TLIDE | 33.95 | 29.06 | 23.67 | 20.77 | 16.49 |
| PROPOSED | 44.69 | 42.87 | 40.96 | 39.79 | 38.84 |

Table 3: Comparison of MSE Value in Lena Image

| Method | 10% | 20% | 30% | 40% | 50% |
|-----------|------|------|-------|-------|-------|
| FIRE [14] | 1.01 | 2.11 | 3.38 | 4.93 | 6.89 |
| BDND [15] | 3.03 | 6.36 | 9.85 | 13.99 | 18.82 |
| DWM [16] | 0.80 | 1.69 | 2.68 | 3.91 | 5.53 |
| LUO [17] | 0.98 | 2.83 | 5.48 | 8.99 | 13.55 |
| FMEM [18] | 1.42 | 2.22 | 3.86 | 6.57 | 10.73 |
| SAM [19] | 3.82 | 7.76 | 11.56 | 13.49 | 19.48 |
| TLIDE | 0.80 | 1.92 | 4.27 | 7.89 | 16.06 |
| PROPOSED | 1.61 | 2.90 | 3.84 | 5.31 | 6.37 |

Table 4: Comparison of MSE Value in Gold Hill Image

| Method | 10% | 20% | 30% | 40% | 50% |
|-----------|------|------|-------|-------|-------|
| FIRE [14] | 1.10 | 2.28 | 3.61 | 5.22 | 7.27 |
| BDND [15] | 3.38 | 6.98 | 10.80 | 15.36 | 20.27 |
| DWM [16] | 0.98 | 2.09 | 3.37 | 5.01 | 7.06 |
| LUO [17] | 1.26 | 3.26 | 6.12 | 9.88 | 1.46 |
| FMEM [18] | 2.24 | 3.19 | 4.93 | 8.05 | 12.36 |
| SAM [19] | 3.99 | 7.95 | 12.03 | 16.10 | 20.34 |
| TLIDE | 0.98 | 2.27 | 4.78 | 8.43 | 16.60 |
| PROPOSED | 1.71 | 2.61 | 4.05 | 5.31 | 6.60 |

PSNR, MSE and run time (RT) value of proposed method is described in table 5(a) in standard lena image. Highest PSNR value is 44.95 which is better compare to all other methods as in literature. MSE value should be as possible as smaller, is 1.61 at 10% noise in lena image. Column 4, RT shows time required in mile seconds to remove impulse noise from the noisy image. Proposed method requires 135 msec at 40% noise density. Table 5(b) presents corresponding PSNR, MSE and RT values in gold hill image. In order to noise density, PSNR is linearly decreasing and MSE is in increasing order.

Table 5: Determination of the PSNR, MSE & RT Values in (a) Lena Image, (b) Gold Hill Image.

| Noise level | PSNR | MSE | RT(msec) |
|-------------|-------|-------|----------|
| 10 | 44.95 | 1.616 | 237.7 |
| 20 | 42.41 | 2.90 | 137.57 |
| 30 | 41.18 | 3.84 | 261.51 |
| 40 | 39.78 | 5.31 | 135.70 |
| 50 | 38.99 | 6.37 | 191.15 |

| Noise level | PSNR | MSE | RT(msec) |
|-------------|-------|------|----------|
| 10 | 44.69 | 1.71 | 171.85 |
| 20 | 42.87 | 2.61 | 219.59 |
| 30 | 40.96 | 4.05 | 160.77 |
| 40 | 39.79 | 5.31 | 160.24 |
| 50 | 38.84 | 6.60 | 213.75 |

4 Conclusion

In this paper we present fuzzy logic based noise removal, detail preserving restoration method. This is the ultimate filter for removal of impulse noise. Even at very high noise density image details texture and edges are preserved. The proposed method for image enhancement is sufficient but the future work is to develop a method for image enhancement in RGB color image and video sequenced image.

REFERENCES

- [1]. A. Bovik, Handbook of Image and Video Processing. New York: Academic, 2000.
- [2]. Raymond H. Chan, Chung-Wa Ho and Mila Nikolova, "Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Detail-Preserving Regularization", IEEE Trans. Image Process. October 2005. 14(10): p. 1479-1485. DOI: 10.1109/TIP.2005.852196.
- [3]. Tuan-Anh Nguyen, Won-Seon Song, and Min-Cheol Hong, "Spatially Adaptive Denoising Algorithm for a Single Image Corrupted by Gaussian Noise", IEEE Trans. Consumer Electron. August 2010. 56(3): p. 1610-1615.
- [4]. R.C. Hardie, C.G. Boncelet, "LUM filters: a class of rank-order-based filters for smoothing and sharpening", IEEE Trans. Signal Process., 1993. 41: p. 1061-1076.
- [5]. S.J. Ko, Y.H. Lee, "Center weighted median filters and their applications to image enhancement", IEEE Trans. Circuits Syst., vol. 38, pp. 984-993, 1991.
- [6]. S. Schulte, V. De Witte, M. Nachtegael, D. Van der Weken, E.E. Kerre, "Fuzzy random impulse noise reduction method", Fuzzy Sets Syst., 2007. 158: p. 270-283.
- [7]. S. Schulte, M. Nachtegael, V. De Witte, D. Van der Weken, E.E. Kerre, "A fuzzy impulse noise detection and reduction method", IEEE Trans. Image Process., 2006. 15(5): p. 1153-1162.
- [8]. S. Schulte, S. Morillas, V. Gregori, E.E. Kerre, "A new fuzzy color correlated impulse noise reduction method", IEEE Trans. Image Process., 2007. 16(10): p. 2565-2575.
- [9]. S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Video denoising based on inter-frame statistical modelling of wavelet coefficients", IEEE Trans. Circuits Syst. Video Technol., 2007. 17(2): p. 187-198.

- [10]. L. Jovanov, A. Pizurica, V. Zlokolica, S. Schulte, P. Schelkens, A. Munteanu, E. E. Kerre, and W. Philips, "Combined wavelet-domain and motion-compensated video denoising based on video codec motion estimation methods", IEEE Trans. Circuits Syst. Video Technol., 2009. 19(3): p. 417-421.
- [11]. H. B. Yin, X. Z. Fang, Z. Wei, and X. K. Yang, "An improved motion-compensated 3-D LLMMSE filter with spatio-temporal adaptive filtering support," IEEE Trans. Circuits Syst. Video Technol., 2007 17(12): p. 1714-1727.
- [12]. T. Mélangé, M. Nachtegael, E. E. Kerre, V. Zlokolica, S. Schulte, V. De Witte, A. Pizurica, and W. Philips, "Video denoising by fuzzy motion and detail adaptive averaging", J. Electron. Imaging, 2008. 17(4).
- [13]. Tom Mélangé, Mike Nachtegael and Etienne E. Kerre, "Fuzzy Random Impulse Noise Removal From Color Image Sequences", IEEE Trans. Image Process., 2011. 20(4): p. 959-970, DOI: 10.1109/TIP.2010.2077305.
- [14]. F. Russo and G. Ramponi, "A fuzzy filter for images corrupted by impulse noise," IEEE Signal Process. Lett., June 1996.3(6): p. 168-170.
- [15]. P. E. Ng and K. K. Ma, "A switching median filter with boundary discriminative noise detection for extremely corrupted images," IEEE Trans. Image Process., June 2006. 15(6): p. 1506-1516.
- [16]. Y. Dong and S. Xu, "A new directional weighted median filter for removal of random-valued impulse noise," IEEE Signal Process. Lett., Mar 2007. 14(3): p. 193-196.
- [17]. W. Luo, "An efficient algorithm for the removal of impulse noise from corrupted images," AEU-Int. J. Electron. Commun, 2007. 61: p. 551-555.
- [18]. J. Zhang, "An efficient median filter based method for removing random valued impulse noise," Digit. Signal Process, July 2010. 20(4): p. 1010-1018.
- [19]. H. Ibrahim, N. S. P. Kong, and T. F. Ng, "Simple adaptive median filter for the removal of impulse noise from highly corrupted images," IEEE Trans. Consumer Electron., Nov. 2008. 54(4): p. 1920-1927.

Development of the ISR3M Model for IS Risk Management Evaluation using the Focus Area Structure according to the MMDPIS Generic Process

¹Mina El maallam and ²Abdelaziz Kriouile

IMS Team, SIME Lab., ENSIAS, University Mohammed V-Souissi, Rabat, Morocco;

¹elmaallam@gmail.com; ²kriouile@ensias.ma

ABSTRACT

Risk management (RM) is one of the main IS governance pillars. However, to remain a center of profit and cost optimization for the company, this activity must be evaluated, monitored and improved continuously. Hence the interest to develop an IS risk management maturity model. This paper aims to address this need by providing the ISR3M (Information System Risk Management Maturity Model) model. After a summary of literature review, it presents the design approach, then describes the model and evaluates it.

Keywords: Information system, risk management, Maturity, Maturity model, Focus Area structure.

1 Introduction

Information System (IS) risk management contributes to the protection of the IS assets. It saves the organism from the losses caused by the emergence of unwanted events having an incidence on the IS objectives and consequently on its strategy. It also has an important role in making good decisions about entering new opportunities. In addition, it promises an optimal allocation of resources [1]. However, it presents a set of challenges for both professionals and researchers.

Indeed, a primary mission of Risk Managers is to help companies maximize profit through minimizing the cost of risk [2]. The latter being a combination of the cost of risk management and loss due to their eventual realization [3].

But "we can control only what we can measure" and we can measure only what we know. The first challenge is then raised to know the status of the IS risk management. The second challenge is to improve this activity to ensure efficiency and continuity of its implementation.

The present article aims to answer these challenges by proposing the ISR3M (Information System Risk Management Maturity Model) maturity model for IS risk management. This model has for objective to evaluate the maturity of risk management and ensure its continuous improvement through the implementation of small changes, frequent, incremental, quickly obtained [4] and guided by kaizen philosophy of "better than yesterday, less than tomorrow".

The second section presents a summary of literature review. It evokes the problems covered by section as well as a comparative study of existing solutions. The third section describes the approach adopted

for the ISR3M design. This is the MMDPIS process [5]. The proposed model is presented in fourth and fifth sections and evaluated in sixth section. The paper is concluded in section seven.

2 Literature review

The interest of IS maturity models increases both for the researchers and for the professionals. However, their development and adoption still face several problems [6] especially for IS risk management discipline. There are two natures of problems.

The first is related to the definition of the studied field. Indeed, one of the problems hindering the IS risk management is the ambiguity that prevails. In IS, the risks interpretations differ from community to community [1]. In addition, IS risk management does not cover all aspects of the IS and is, in most cases, related to IT aspects.

The second problem is related to the model design. The latter presents numerous points of improvement such us:

- Lack of satisfactory answers to the implementation of improvement actions: the objective of the maturity assessment models is to identify gaps that could be filled by improvement actions [7]. However, most of these models do not describe how to effectively carry out these actions to deal with identified gaps ([8], [9], [10]).
- Falsified Certainty given to the decision-makers: another one critical, often shown, is that maturity models can give in the decision-makers a "falsified certainty" regarding the diagnosis and the evaluation of maturity. This report can be due to the carelessness of certain important aspects of the studied domain. It is also due to the very limited understanding of the reality ([8], [9]). For example, ([11]), [12], [13], [14]) qualify the models of maturity as "recipe stage by stage" which simplify excessively the reality.
- Poor theoretical basis: The third point considered by [15] as the most important criticisms of the existing maturity models are their poor theoretical basis. Indeed, most models are based on "best practices" or "success factors" from projects that have favorable results to a company or an industry [8]. Thus, to be consistent with the maturity model would not necessarily guarantee that the company would be successful [8]. There is no agreement on a "right way" to ensure a positive result [16].
- Lack in tests of model validity: According to [12], the ambiguity noted in results of maturity models lies in the insufficient emphasis on models testing in terms of validity, reliability and genericity. [8] confirms this and thinks the maturity models are a foundation that lack of experimentation.
- Not adequacy to the specific needs: other criticisms refer to the multitude of almost identical maturity models and to the not thoughtful adoption of the CMM main plan ([17], [18], [19]). The carelessness of the specific needs can hinder the achievement of the objectives of the maturity model. In the same line, [10] think that the maturity models must be customizable, because the internal and external characteristics can limit the applicability of a maturity model in its normalized version [19]. On the other hand, the design of the model in question must consider the context in which it will be deployed. The context is "all information that can be used to characterize the situation of an entity" [20].

- The high level of formalism: [21] think that too much emphasis on the formalization of improvement activities accompanied by extensive bureaucracy can block people innovation. [13] postulate that maturity models should not focus on one level sequence to an "end state" predefined, but on factors driving the evolution and change. The exaggerated formalism can also urge to remain motionless on a single path of improvement. In such cases, the models of maturity tend to neglect the potential existence of several paths so advantageous [22] and being able to better answer the possible evolutions in studied IS.

The review of the literature also concerned a comparative study of the maturity models in risk management and IS risk management. In the light of both natures of problems expressed previously, the proposed criteria of comparison are:

- C1: Genericity: the proposed solution should be generic viewpoint processes and IS risk management concepts.
- C2: independence of the context of application: the solution must be applicable in all the contexts and the business sectors.
- C3: Adaptability: The solution must take into account the specificities of the studied area.
- C4: transparency: the solution has to insure the documentation and the traceability of the maturity measures.
- C5: plan improvement: does the model assist its users in the definition of an improvement plan?
- C6: theoretical Basis: does the model based on the theoretical aspect of the domain studied for the measure of the maturity?
- C7: adequacy to needs (IS RM): is the model adapted to the IS risk management?

Table 1 shows the result of this study.

Table 1: Comparative study of RM maturity models.

| Model | C1 | C2 | C3 | C4 | C5 | C6 | C7 |
|------------------------------------|----|----|----|----|----|----|----|
| RMM [23] | - | + | - | - | - | - | - |
| Project RMM [24] | - | + | - | + | - | - | - |
| COPS [25] | - | + | - | + | - | - | - |
| J-RMMM [26] | - | - | - | + | - | - | - |
| ERMM-Level Assessment Tool [27] | + | + | - | + | - | - | - |
| CMMI [28] | - | + | - | + | - | - | - |
| MMGRSeg [29] | + | + | - | - | - | - | - |
| Modèle de maturité de Risk IT [30] | - | + | - | + | - | - | - |

The natures of the problems being so defined in a rather clear and precise way, the proposed answer is the design of a new model of maturity of the IS risk management: ISR3M. This model is developed using the MMDPIS process [5] described in section 3.

3 MMDPIS Process Description

The MMDPIS process is structured in three blocks: (1) design, (2) implementation, and (3) continuous improvement (Figure 1).

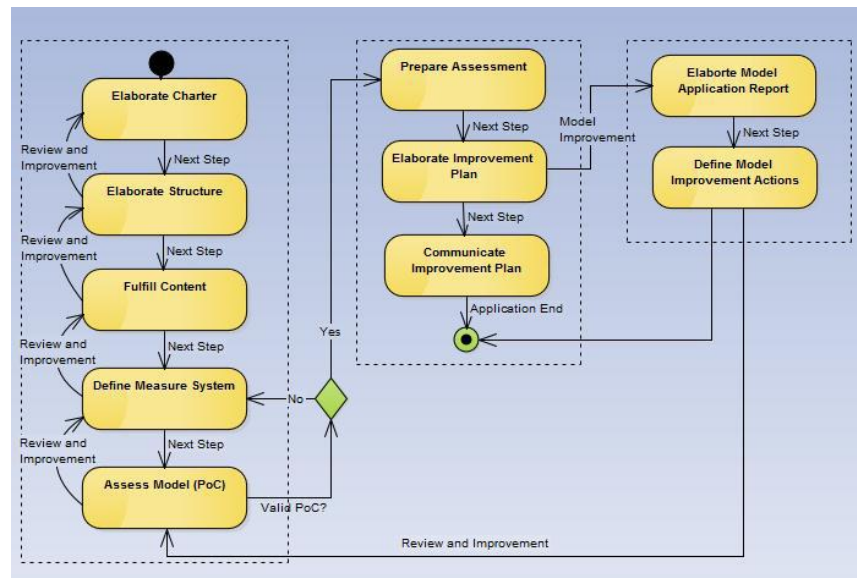


Figure 1: MMDPSI process description

3.1 Block 1: Design

The first block presents the design stages.

3.1.1 Establish charter

The establishment of the charter is the first step in developing the maturity model. The Charter may mainly include: model denomination, studied area description, scope, model purpose, development strategy, success factors, requirements, and stakeholders.

3.1.2 Establish the structure

The aim is to structuring the concepts of maturity depending on the purpose and scope of the model to develop and also defining the architecture or the representation according to which the evaluation will be made. In contrast to the other approaches cited in the literature review the definition of architecture at the MMDPIS process takes into consideration the purpose and requirements of the model to be developed.

This architecture can be of type: (1) staged, (2) continuous or (3) Focus Area. The choice must be justified and documented.

The staged architecture is adapted when it is a question of estimating the global maturity of the organization with regard to a given activity. This choice is recommended for example in case of benchmarking study.

The objective of the continuous architecture is to evaluate domains or process according to predefined levels of evolution, called "levels of capacity". This choice is recommended when it is question of making a comparative study between these domains in a perspective to prioritize improvement axes.

As for the third architecture "Focus Area", it defines control objectives specifically for each area of studied activity depending on its life cycle phase. It enables to take into account the interdependencies between these control objectives.

3.1.3 Fulfill the structure

This step begins by identifying the elements of the model according to the adopted structure.

The explanation of these elements can be based on different approaches. The choice of method depends on the context of the model development.

Indeed, the most common and recognized approach of the methods of maturity models development is the literature review along with interviews of experts in the field. This approach is complete for some items when needed, especially for complex or new areas by exploratory methods such as Delphi technique, Focus group or case study.

The elements to be defined for all the types of structures are:

- Axes and elements of evaluation.
- Domains and groups of domain: the identified domains must be mutually exclusive and collectively exhaustive. The link Domain/Group of domain can be established according to a Top-Down or Bottom-up approach according to the context of model development.
- Objectives of control: for every domain, define the objectives of control.

Specific elements of each structure are:

- Maturity levels for staged architecture: define the maturity levels of the model and control objectives required for each domain to reach a given level.
- The capacity levels for continue architecture: define the capacity levels and control objectives required for each domain to reach a given capacity level.

The interdependencies and ranks of control objectives (CO) for FA architecture: define for each control objective a rank and dependencies on other control objectives (1).

If the CO is independent of all others CO, then rank (CO) = 1.

*If the CO depends on a number of other CO: {CO1, CO2, ..., CO_n},
then the rank is calculated as rank (CO) = Max (CO_i) +1, 1 <= i <= n.* (1)

- These two elements are used to determine the maturity matrix and define maturity levels.

3.1.4 Defining the measurement system

In this stage are defined the elements of the measurement system:

- Elements of control: for every control objective define the corresponding control elements,
- Method and evaluation tools: it is necessary to define the method according to which the evaluation will be made and which tools to use for the collection of the measures and their exploitation,
- Evaluation Team: the human element is central to the measurement system. It is important at this stage to identify the requirements for this element or detail them if they are already identified in the model charter. These requirements may be under three aspects: skill, function and behavior. The latter can result in motivation, commitment and adherence to continuous improvement project.

3.1.5 Evaluate the model (PoC: Proof of Concept)

The purpose of this evaluation is to check whether the model designed meets the predefined requirements. Evaluation can be done through the progress of a case study example.

If the evaluation is satisfactory, the maturity model developed is implemented. Otherwise, a second iteration is started. The recovery was made from the stage representing the source of the problem or dissatisfaction.

3.2 Block 2: Implementation

The second block guides the implementation of the designed maturity model.

3.2.1 Prepare evaluation

The proposed evaluation process involves the following steps:

- Constitution of the evaluation team: It consists on constituting the team of evaluation according to the requirements predefined in the charter and the measurement system. The participation in this exercise must be accepted without additional or hidden responsibility for the evaluation team. The communication and the raising awareness of the team are important. A Quiz can help to do this and also to measure the adhesion of the team before beginning the evaluation.
- Define IS to evaluate: At this stage, we have to define the list of IS to evaluate. For each IS, it's important to define the corresponding life cycle phases and calculate the weight reflecting its importance in organism. We propose to calculate the weight based on three elements: (1) consumption of the cost, (2) consumption of the load, and (3) contribution to the strategy. Table 2 provides a calculation example of this weight. Column 1 provides the name of IS. Column 2 gives the rate of annual consumption compared to the total annual load. Column 3 shows the annual rate of consumption relative to the overall annual cost. Columns 4 to n describe the IS contributions in organization strategic objectives (SO). This qualification is made on a qualitative scale to which corresponds a quantitative scale: F: Strong 3, M: Medium 2 and Fb: Low 1. The first part of the column n+1 calculates the contribution according to predefined quantitative scale (2). The second part of this column gives the contribution in the form of rate with regard to the global contribution (for example for IS 1 is equal to $67\% = 6 / (3 + 6)$). The last column n + 2 is dedicated to the calculation of the overall weight of the IS based on the three elements previously defined: load consumption, consumption cost and contribution in the strategy. This weight is given by the formula (3).

$$\text{Contribution} = [(3 * F \text{ value}) + (2 * \text{number } M) + (1 * \text{number } Fb)] \quad (2)$$

(for example for IS 1 is equal to $3 * 2 = 6$)

$$\text{Weight} = (\text{Val_col } 2 + \text{Cal_col } 3 + \text{Val_part } 2(\text{Col } n+1)) (\text{IS}) / \text{Total} (\text{Val_col } 2 + \text{Cal_col } 3 + \text{Val_part } 2(\text{Col } n+1)) (\text{all IS}). \quad (3)$$

Table 2: Example of calculating of the IS weight

| IS | load Cons. | cost Cons. | Col.4S.O | ... | Col.n S.O. | Col.n+1 Cont. S.O. | Col.n+2 weight IS |
|------|------------|------------|----------|-----|---------------|-----------------------|----------------------|
| IS 1 | 25% | 40% | F | | F | 6 67% | 44% |
| IS 2 | 75% | 60% | | | F | 3 33% | 56% |

3.2.2 Establish an improvement plan

An evaluation plan depends on the adopted evaluation system. However, it usually contains: sessions of presentation and explanation of the model/ system evaluation tools, evaluation Workshops and outcomes discussion sessions.

The development of the improvement plan first requires an analysis of assessment results.

Once this analysis made, a list of the scenarios of improvement is established. These scenarios represent the possible paths of improvement. The person responsible for implementing the maturity model should select the most optimal path answering the objectives and constraints of the organism. Optimization paths for improvement may require the definition of the elements of calculating the improvement effort. He can involve in particular the knowledge of the cost and the load estimated as well as of the impact of implementation of the improvement actions.

3.2.3 Communicate the improvement plan

The improvement plan should be communicated to all stakeholders in an objective of validation, implementation and monitoring of improvement actions.

3.3 Block 3: continuous improvement

The third block is dedicated to the model continuous improvement. It is a question of defining the improvement actions of the developed model following its implementation.

- Develop the application report: This report serves to register reports and remarks resulting from the phase of implementation.
- Define the model improvement actions: it can give rise to a new iteration for its development.

4 Design of ISR3M Model

4.1 Elaboration of the model charter

Charter is constituted of following elements:

- **Denomination and studied area:** the model is called ISR3M (Information Systems Risk Management maturity model) and it is developed for the evaluation of IS risk management.
- **Scope:** domains concerned by the ISR3M model both in research and in practice are the risk management and the information systems. The definition adopted for the IS is that of a special case of work system [31]. The elements to be considered in the study of such systems are: participants, information, technology, processes, products and services, customers, infrastructure, environment and strategy. As for the risk management device, we adopt the ISO 31000 Framework [32] with the generic management cycle proposed by Sienu [33]. This cycle resumes the stages of the process proposed by ISO 31000 with a restructuring of its phases. Indeed: (1) communication is considered as an activity inherent to every phase of the process [33], (2) the cycle of management preserves its iterative character, but no longer requires synchronization of all stages with a monitoring phase [33], and (3) Treatment may be the cause of a new iteration process [33].

- **Model purpose:** the objective of the ISR3M model is assessing IS risk management. The development of this model should provide answers to the above problem from two perspectives. The first perspective is academic. The model must address a problem not sufficiently addressed in IS research: the assessment of IS risk management. The proposed solution must also be able to open new avenues and opportunities in scientific research in this area. The second perspective relates to the practical side. The proposed model should be easy to implement and comply with the best practices of risk management.
- **Development strategy:** it is a question of developing a new IS risk management maturity model using the MMDPIS process [5].
- **Success factors:** success factors of ISR3M model concern mainly: (1) taking into account the theoretical aspects of the concerned areas especially information and risk management systems, (2) involvement of stakeholders, and (3) simplicity and ease of implementation.
- **Application scope:** given the nature of the area studied, the application of ISR3M model concerns the whole organism.
- **Requirements:** the proposed requirements are the comparison criteria of maturity models presented in second section: c1: Genericity, c2: Independence of application context, c3: adaptability, c4: transparency, c5: plan improvement, c6: Theoretical basis and c7: need Adequacy (IS RM).
- **Stakeholders:** it involves target audience (IS risk management device evaluators, auditors...), implementation responsible (IS risk management managers), and respondents and interviewees (risk managers, IS project managers..., etc).

4.2 Establishment of the structure

The ISR3M model is structured along two dimensions. The first dimension includes evaluated activities. It is a question of risk management activities. The second dimension represents the aspects under which these activities are evaluated. It is a question of evaluation axes and elements related to an IS defined as a WS.

The maturity assessment is made according to the architecture "Focus Area". The choice is justified by the fact that this architecture provides a more sophisticated approach than the other two architectures in relation to the purpose and scope of the ISR3M model. Indeed, it defines small evolutionary steps thus making improvement easier, less risky and less costly. The road improvement is clearer.

The choice is also justified by taking into account the interdependence of the control objectives which is an important characteristic of the risk management business.

4.3 Populate the structure

4.3.1 Determination of the axes and elements of evaluation

The areas of assessment are the elements of the IS defined as WS [31]. These are: (1) infrastructure, (2) strategy, (3) environment, (4) technology (5) Information (6) participants, (7) process (8) products, and (9) customers.

The evaluation elements of each axis are identified through (1) the missions and requirements of WSF as defined in the literature [31], (2) the application of the theory RBV (Resource Based-view) [34] on IS

defined as WS considering both dynamic resources such as skills, as static as the technical infrastructure, (3) the IS risk factors [35], and (4) interviews with IS experts.

Table 3 lists the evaluation elements for each component.

Table 3: IS evaluation elements

| Axis | Evaluation element | Axis | Evaluation element |
|----------------|------------------------------|--------------|--------------------|
| Infrastructure | Technical infrastructure | Information | Security |
| | Human infrastructure | | Reliability |
| | Informational infrastructure | | Relevance |
| Strategy | Alignment | Process | Agility |
| | Contribution | | Formalization |
| Environment | Organism | | Updating |
| | Culture | | Interaction |
| | Intra Enterprise regulations | | Coherence |
| IT | competitive importance | | Product |
| | complexity | Quality | |
| | Codifiability | Exploitation | |
| | Potential of credibility | Customer | Needs |
| | Strategic profile | | Satisfaction |
| Competence | Competence | | |
| Participants | Cooperation | Customer | Cooperation |
| | stability | | |

4.3.2 Definition of domains and domains groups

The areas adopted for ISR3M model are the risk management activities. It is deduced from Risk Management Framework proposed by ISO 31000. The domain groups are the three pillars of the ISO 31000 Framework. We have added the "Recording" section.

The areas of maturity are then listed in Table 4. For the domain group "Process", the maturity domains of selected areas are sub-activities of each of its activities. This is justified by the fact that this level reflects more the operational component of the process.

Table 4: ISR3M domains

| Domain group | Domain |
|---|--------------------------------------|
| RM principle | RM principle |
| Organizational framework of risk management | Mandate and commitment |
| | Design of framework |
| | Implement risk management |
| | Monitor and review framework |
| | Improve framework |
| Process | External context |
| | Internal context |
| | Process context |
| | RM criteria |
| | Risk identification |
| | Risk analysis |
| | Risk evaluation |
| | Selection of treatment options |
| Development of the treatment plan | Development of the treatment plan |
| | Implementation of the treatment plan |
| | Monitoring and review |
| Recording | Recording |

4.3.3 Determination of control objectives

The objectives of control describe the way of progressive improvement of a maturity domain. To define them, we used: (1) description of Framework of risk management given by the standard ISO 31000, (2) study of the literature, and (3) focuses group and interview with risk management experts.

A control objective is identified via the following elements: (1) code, (2) name, (3) target (4) actions needed for its implementation, (5) prerequisite control objectives on which it depends, (6) estimated load, (7) estimated cost of implementation, and (8) implementation of impact.

For example, the control objectives defined for the RM principles domain are:

- A: Reminded principles
- B: Principles formalized and communicated
- C: Principles evaluated in terms of understanding and adherence

4.3.4 Determining the position of the objectives control in the maturity matrix

The position of the control objectives in the maturity matrix is defined by calculating their ranks according to the rule (1).

The application of this rule allows obtaining the matrix of maturity illustrated in figure 2.

| N° | Domain | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----|--|---|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | RM principles (PRM) | | A | B | C | | | | | | | | | |
| | Organizational framework | | | | | | | | | | | | | |
| 2 | Mandate and commitment (ME) | | A | B | | | | | | | | | | |
| 3 | Design of framework (CCO) | | | A | B | C | D | E | | | | | | |
| 4 | Implement risk management (MOE) | | | | | A | | B | C | | | | | |
| 5 | Monitor and review framework (SRC) | | | | | | A | | | B | C | | | |
| 6 | Improve framework (ACC) | | | | | | | A | | | B | C | | |
| | Process | | | | | | | | | | | | | |
| | Establish context | | | | | | | | | | | | | |
| 7 | External context (ECX) | | A | B | C | | | | | | | | | |
| 8 | Internal Context (ECI) | | A | B | C | | | | | | | | | |
| 9 | Process context (ECP) | | A | B | C | D | | | | | | | | |
| 10 | RM criteria (ECC) | | | A | | | B | C | | | | | | |
| | Risk assessment | | | | | | | | | | | | | |
| 11 | Risk identification (API) | | | | A | | B | C | D | E | | | | |
| 12 | Risk analysis (APA) | | | | | A | | B | C | | | | | |
| 13 | Risk evaluation (APV) | | | | | A | | | B | C | | | | |
| | Treatment | | | | | | | | | | | | | |
| 14 | Selection of treatment options (TSD) | | | | | A | | | | B | C | | | |
| 15 | Development of the treatment plan (TEP) | | | | | A | | | | | B | C | | |
| 16 | Implementation of the treatment plan (TMP) | | | | | | A | | | | | B | C | |
| 17 | Monitoring and review (SR) | | | | | | | A | | | | | B | C |
| 18 | Recording | | A | B | C | | | | | | | | | |

Figure 2: Positioning matrix

4.4 Definition of evaluation system

4.4.1 Definition of control elements

In order to build this system, we use an approach based on the principle of the method GQM (Goal-Question-Metric) [36]. This process involves the following steps:

- Determine the objectives of the evaluation system (Goal): The system of evaluation has for objective to estimate the domains of maturity through the evaluation of the realization of the corresponding control objectives. This allows fulfilling the matrix of maturity pre-established according to the verified OC and to define the level of maturity of the risk management of IS studied. There are 18 goals. They can be so formulated as: G_i : " estimate the domain of maturity D_i ", $1 \leq i \leq 18$,
- Formulate questions (Question) to identify aspects to be measured to assess the achievement of defined objectives. In light of defined objectives, questions can be formulated in the following way: Q_i : "In what stage of development is the domain D_i ?" ($1 \leq i \leq 18$),
- Define metrics (Metric) to evaluate these aspects: Metric responding to the question Q_i D_i for each domain are related control objectives,
- Define the elements to measure these metrics: these elements, called in the ISR3M model control elements, are specifically defined for each control objective from its requirements.

Table 5 shows an example of the definition of the control elements: PRM.C: "Principles evaluated in terms of understanding and adherence" of domain "RM principles".

Table 5: Example of control element

| OC | OC goal | Aspect to be verified | Elements of control of the OC |
|---|---|--|---|
| Principles evaluated in terms of understanding and adherence. | Assess the understanding and application of risk management principles. | -Understanding of risk management principles -Application And adherence to risk management principles | -Does the organism measure the level of understanding of IS risk management principles (surveys, quizzes ...)? - Does the organism measures the degree of adherence to IS risk management principles (surveys, quizzes ...)? |

A control objective is checked whether all control elements have a favorable response ie equal to "yes."

4.4.2 Method and Assessment Tool

The evaluation is done through a self-assessment questionnaire. The latter is formed from the control elements previously defined. It is divided into three categories according to three categories of evaluation: (1) category 1: organism, (2) category 2: IS, and (3) category 3: IS with considering axes and elements evaluation.

This questionnaire according to its three categories is implemented in Excel.

For purposes of consolidation measures made by control elements, the answers are translated into quantitative values: "Yes" = 1 and "no" = 0.

The first category concerns the management of the IS risks at the level of the organism in a global way. it concerns in particular domains belonging to the groups of domain " RM principles" and " organizational framework". A control objective is considered achieved if all the answers to the relevant questions (control elements) are "yes".

The second category concerns studied IS. However, the questions are not declined in axes and elements of evaluation. An objective of control is considered reached if all the answers to the corresponding questions (elements of control) are in "yes". The activity "recording" and "the treatment" are examples.

The third category requires checking elements of control of a domain at the level of every evaluation element of the studied IS. This questionnaire allows to consider the specificities of every organism through a configuration variable called "Applicable" indicating the applicability or not of an element of evaluation and if it presents an eliminatory characteristic. The Measure of a control element (EC) relative to an evaluation axis (IS elements) is the rounded value of the arithmetic mean of the measurements of its evaluation elements taking into account the settings of variable values (4).

$$\text{Measure_EC (Axe_Eval)} = \text{Round} [\text{mean (Mesure_EC (Elt_Eval } i) \times \text{Valeur_Applicable (Elt_Eval } i))]; 1 \leq i \leq \text{number of assessment items in Axe_Eval} \quad (4)$$

The questionnaire also takes into consideration the IS lifecycle. Indeed, the measure of a control element is the rounding of the weighted average of its measurements by axis evaluation (5). The weight (Axe_Eval) of this weighting is the importance of each axis evaluation in the phase of the life cycle of the information system at the time of the evaluation. In the absence of studies dedicated to the calculation of this weight, we propose to hold focus groups to define for each type of SI as the context of the business.

$$\text{Measure (EC)} = \text{Round} [\text{Sum (weight (Axe_Eval } j) \times \text{Mesure_EC (Axe_Eval } j))]; 1 \leq j \leq 9 \quad (5)$$

(number of evaluation axes = 9)

The measure of a control objective is "yes" if all the corresponding questions are "yes" (value 1) and "no" (value 0) otherwise (6).

$$\text{Measure (OC)} = \text{product (Measure (EC } i)); 1 \leq i \leq \text{number of EC in OC} \quad (6)$$

Once the control objectives evaluated, the matrix is populated. The lines of each domain are marked with a different color until the corresponding cell to the maximum value of the ranks of control verified objectives. The maturity level of IS risk management for each domain group is the one corresponding to the right column which all the cells harboring the required control objectives are colored. Figure 3 gives an example of the filled matrix.

The company may have a global view of the maturity of its IS risk management through the consolidation of its various IS measures. Indeed, for each maturity domain, the overall rank is the rounding of the weighted average of the ranks in each IS. Quantitative values for each control objective are the corresponding ranks at the maturity scale. For example in the matrix shown in Figure 3, the value corresponding to the control objective "D" verified by the CCO field is "5". The consolidation weight is the weight of the IS reflecting its importance in the organism strategy.

| N° | Domain | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----|--|---|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | RM principles (PRM) | | A | B | C | | | | | | | | | |
| | Organizational framework | | | | | | | | | | | | | |
| 2 | Mandate and commitment (ME) | | A | B | | | | | | | | | | |
| 3 | Design of framework (CCO) | | | A | B | C | D | E | | | | | | |
| 4 | Implement risk management (MOE) | | | | | A | B | C | | | | | | |
| 5 | Monitor and review framework (SRC) | | | | | | A | | | B | C | | | |
| 6 | Improve framework (ACC) | | | | | | | A | | | B | C | | |
| | Process | | | | | | | | | | | | | |
| | Establish context | | | | | | | | | | | | | |
| 7 | External context (ECX) | | A | B | C | | | | | | | | | |
| 8 | Internal Context (ECI) | | A | B | C | | | | | | | | | |
| 9 | Process context (ECP) | | A | B | C | D | | | | | | | | |
| 10 | RM criteria (ECC) | | | A | | | B | C | | | | | | |
| | Risk assessment | | | | | | | | | | | | | |
| 11 | Risk identification (API) | | | | A | B | C | D | E | | | | | |
| 12 | Risk analysis (APA) | | | | | A | B | C | | | | | | |
| 13 | Risk evaluation (APV) | | | | | A | | B | C | | | | | |
| | Treatment | | | | | | | | | | | | | |
| 14 | Selection of treatment options (TSO) | | | | | A | | | | B | C | | | |
| 15 | Development of the treatment plan (TEP) | | | | | A | | | | | B | C | | |
| 16 | Implementation of the treatment plan (TMP) | | | | | | A | | | | | B | C | |
| 17 | Monitoring and review (SR) | | | | | | | A | | | | | B | C |
| 18 | Recording | | A | B | C | | | | | | | | | |

Figure 3: Example of maturity matrix

4.4.3 Definition of the evaluation team

The evaluation team should be constituted mainly from:

- Responsible of implementation: he can involve for example auditors or external consultants.
- People asked: it is mainly about risk managers, IS project managers and business managers.

4.4.4 Model evaluation

The evaluation can be realized through the application of the designed model on a pilot information system. It has for objective to test the applicability of the proposed version.

We applied the model in three iterations on three different IS. Adjustments and adaptations concern including control elements, dependencies and positioning control objectives of the various fields in the maturity matrix.

5 Implementation

5.1 Evaluation approach

The implementation begins with the definition of the team of evaluation and their raising awareness with regard the importance and to the added value of the work to be made.

A working session was then held with the constituted team. Its primary purpose is to explain to members the main concepts used in the ISR3M model. These include the information system definition and the risk management process. A second objective is to present the evaluation system.

The evaluation team has the task of describing the IS object of the valuation. This description requires the definition of four main elements for each SI: (1) the nine elements of the WSF, (2) the weight of these elements in relation to the different phases of the IS life cycle, (3) IS current phase over its life cycle, and (4) the IS weight in relation to the management of global IS risks. This weight can be calculated using the method proposed by the MMDPIS process (3).

The questionnaire is informed by the evaluation team. A workshop is provided later to discuss the answers and finalize the results. This workshop allows the administration of the questionnaire via the "face-to-face." This approach allows for more reliable and less confused answers.

5.2 Elaboration of improvement plan

5.2.1 Analysis of evaluation outcome

The results of the evaluation are maturity matrix for each IS and an overall maturity matrix. The latter is established and analyzed in three visions. The first is a "pessimistic" vision involving the minimum control objectives per domain. Through this vision are located the most defective areas in risk management. The second is an "optimistic" vision. It brings maximum control objectives per domain. This allows to know the most advanced areas in risk management and to create a positive synergy between all the IS. The third vision provide the overall maturity of IS risk management. It is obtained through a weighted average of different IS maturity matrices.

The design of a pivot table allows to conduct an analysis of the evaluation results (by IS, control objective, level of maturity, etc.) according to the need for the decision-makers.

5.2.2 Determination of the improvement strategy

To define the improvement strategy, the organism defines at first the approach of improvement. The latter can be made according to two approaches. The first one is an approach "top - down". In this case, the target is defined at the level of the global maturity matrix and then declined on the maturity matrices of the various IS.

The second approach is an approach "button-up". This approach allows the organism to define the target maturities at the level of every IS. The improvement of the global maturity is the consolidation of the improvements of the various IS.

Secondly the organism defines the desired way of improvement: by maturity level target or control objective target.

If the organism decides to improve his IS risk management maturity towards a given maturity level then all the objectives of control placed before this level must be reached. The actions to be realized are clear.

If the organism decides to improve the maturity by control objectives target, it must ensure that the prerequisites of target control objectives are checked. These are mainly related to dependency constraints.

To guide the selection and implementation of the treatment strategy, we propose two algorithms: the first, called "prerequisite-way", is dedicated to the definition of the prerequisites to achieve a given target.

The second called "improvement declination» define from a global target control objective the control objectives by IS to achieve with optimizing effort.

Table 6 presents a summary of the proposed improvement strategies.

Table 6: Improvement strategies.

| Improvement strategy | Improvement approach | Target definition | Proposed algorithm |
|--------------------------------|----------------------|----------------------|---|
| Top-Down by maturity level | Top-Down | By maturity level | Prerequisite-way Improvement declination |
| Top-Down by control objective | | By control objective | Prerequisite-way Improvement declination |
| Bottom-up by maturity level | Bottom-Up | By maturity level | Prerequisite-way |
| Bottom-up by control objective | | By control objective | Prerequisite-way |

5.2.3 Elaboration of improvement plan

Once the improvement strategy elaborated, the plan of improvement is established. The latter is formed by the necessary actions for the achievement of the target control objectives. These actions arise from those proposed for the implementation of the control objectives. The treatment plan has to specify the responsible and the schedule of implementation as well as monitoring milestones.

5.2.4 Communicate the improvement plan

Improvement plan is communicated to all stakeholders for monitoring and implementation.

6 ISR3M Evaluation

The evaluation of the model is realized according to two axes. The first one concerns the criteria proposed in second section (table 7). The second axis concerns a naturalistic evaluation of the model. It will be presented in an upcoming paper.

Table 7: Evaluation of ISR3M

| Criteria | Evaluation |
|---|--|
| C1: Genericity: the proposed solution should be generic viewpoint processes and IS risk management concepts, | The process on which is based the proposed model is the one of the standard ISO 31000 which is generic viewpoint process and RM concepts. |
| C2: independence of the context of application: the solution must be applicable in all the contexts and the business sectors, | The model can be applied to any IS in all contexts and sectors. |
| C3: Adaptability: The solution must take into account the specificities of the area studied, | A "Applicable" parameter is introduced during the evaluation at the level of every element of evaluation. This parameter allows to specify if each of the latter is considered or not according to the context of IS or of the organism. |
| C4: transparency: the solution has to insure the documentation and the traceability of the measures of maturity, | Evaluation system as well as the documents registered for every version of the model insures the traceability of the measures. |
| C5: plan improvement: does the model assist its | The model assists its users in the definition of the strategy |

| Criteria | Evaluation |
|--|---|
| users in the definition of a improvement plan? | and the improvement ways of both proposed algorithms. |
| C6: theoretical Basis: does the model base itself on the theoretical aspect of the domain studied for the measure of the maturity? | The model ISR3M is effectively based on the theoretical aspect of the various domains of its scope through the detailed study and the effort dedicated to the corresponding literature specially in information system and risk management. |
| C7: adequacy to needs (IS risk management): is the model adapted to the IS risk management? | The model is developed for the evaluation of the IS risk management. |

7 Conclusion

IS risk management is becoming increasingly widespread. The evaluation of this discipline through maturity models in a perspective of continuous improvement is a guarantee of the preservation of its value as a profit center. A series of works related to the assessment of this discipline exist but have limitations in scope and design. Indeed, on the one hand, the existing maturity models deal only with aspects related to IT. On the other hand, these models have conceptual limitations such as: (1) lack of satisfactory answers to the implementation of the improvement actions, (2) the certainty falsified given to decision makers, (3) poor base theoretical, (4) inadequate model validity tests, (5) failure to fit the specific needs of the areas studied, and (6) the high level of formality. The absence of documentation and transparency as for valuation methods are added to these limits.

The solution proposed in this article is the ISR3M model. This maturity model is dedicated to the evaluation of the IS risk management. It is developed according to the MMDPIS process [5] so as to address the problem stated above. Its evaluation shows compliance with pre-established requirements.

The presentation of the implementation results of ISR3M model on an actual case study is planned in future work.

REFERENCES

- [1] Salvati, D. (2008). *Management of Information System Risks*. Zurich: University of Zurich.
- [2] Lei, Y. (2011). Minimizing the Cost of Risk with Simulation Optimization Technique. *Risk Management and Insurance Review*, 14(1), 121-144.
- [3] Zhang, Y. (2009, May). A Study on Risk Cost Management. *International Journal of Business and Management*, 4(5), 145-148.
- [4] Bronet, V. (2006, Septembre). *Amélioration de la performance industrielle à partir d'un processus Référent Déploiement inter entreprises de bonnes pratiques*. Savoie: Université de Savoie.
- [5] El maallam, M and Kriouile, A, (2014). A generic process for the development and the implementation of IS maturity models. *International Journal of Computer Science Issues (IJCSI)*, 11(6), pp. 34-42.
- [6] Poepplbuss, J., Niehaves, B., Simons, A., and Becker, J. (2011). Maturity Models in Information Systems Research: Literature Search and Analysis. *Communications of the Association for Information Systems (AIS)*, 29(27), 505-532.

- [7] Pfeffer, J., and Sutton, R. I. (1999). Knowing what to do is not enough: turning knowledge into action. *California Management Review*, 42(1), 83-108.
- [8] Mettler, T. (2010). Thinking in Terms of Design Decisions When Developing Maturity Models. *International Journal of Strategic Decision Sciences (IJSDS)*, 1(4), 76-87.
- [9] Mettler, T. (2011). Maturity assessment models: a design science research approach. *International Journal of Society Systems Science*, 3(1/2), 81-98.
- [10] Mettler, T., and Rohner, P. (2009). Situational maturity models as instrumental artifacts for organizational design. *4th International Conference on Design Science Research in Information Systems and Technology DESRIST'09*. 22, pp. 1-9. New York, NY, USA: ACM.
- [11] Benbasat, I., Dexter, A. S., Drury, D. H., and Goldstein, R. C. (1984, May). A critique of the stage hypothesis: theory and empirical evidence. *Communications of the ACM*, 27(5), 476-485.
- [12] De bruin, T., Freeze, R., Kulkami, U., and Rosemann, M. (2005). Understanding the Main Phases of Developing a Maturity Assessment Model. *Australasian (ACIS)*. Australie, Sydney.
- [13] King, J. L., and Kraemer, K. L. (1984, May). Evolution and organizational information systems: an assessment of Nolan's stage model. *Communications of the ACM*, 27(5), 466-475.
- [14] McCormack, K., Willems, J., Bergh, v. d., Deschoolmeester, D., Willaert, P., Stemberger, M. I., et al. (2009). A global investigation of key turning points in business process maturity. *Business Process Management Journal*, 15(5), 792-815.
- [15] Biberoglu, E., and Haddad, H. (2002, Decembre). A survey of industrial experiences with CMM and the teaching of CMM practices. *Journal of Computing Sciences in Colleges*, 18(2), 143-152.
- [16] Montoya-Weiss, M. M., and Calantone, R. (1994). Determinants of New Product Performance: A Review and Meta-Analysis. *Journal of Product Innovation Management*, 11(5), 397-417.
- [17] Becker, J., Knackstedt, R., and Pöppelbuß, J. (2009). Developing Maturity Models for IT Management – A Procedure Model and its Application. *Business & Information Systems Engineering (BISE)*, 1(3), 213-222.
- [18] Becker, J., Niehaves, B., Pöppelbuß, J., and Simons, A. (2010). Maturity Models in IS Research. *18th European Conference on Information Systems (ECIS 2010)*. Pretoria, South Africa.
- [19] Iversen, J. H., Nielsen, P. A., and Norbjerg, J. (1999). Situated Assessment of Problems in Software Development. *DATA BASE*, 30(2), 66-81.
- [20] Dey, A. K. (2000). *Providing architectural support for building context-aware applications*. Atlanta, GA, USA: Georgia Institute of Technology.
- [21] Herbsleb, J. D., and Goldenson, D. R. (1996). A systematic survey of CMM experience and results. *Proceedings of the 18th international conference on Software engineering* (pp. 323-330). Washington, DC, USA: IEEE Computer Society.

- [22] Teo, T. S., and King, W. R. (1997). Integration between Business Planning and Information Systems Planning: An Evolutionary-Contingency Perspective. *Journal of Management Information Systems*, 14(1), 185-214.
- [23] Hillson, D. A. (1997). Towards a risk maturity model. *The International Journal of Project and Business Risk Management*, 1(1), 35-45.
- [24] Hopkinson, M. (2011). *The Project Risk Maturity Model: Measuring and improving risk management capability*. Gower.
- [25] Ren, Y. T., and Yeo, K. T. (2009). Risk management capability maturity model for complex product systems (CoPS) projects. *Systems Engineering*, 12(1), 275-294.
- [26] Saito, O., Matsui, T., and Morioka, T. (2007). Organizational Risk Management Maturity Model and Assessment Tool Designed for High-hazard Industries. *International Symposium on Symbiotic Nuclear Power Systems for 21st Century (ISSNP)*, 42-47.
- [27] COSO. (2004). *The Committee of Sponsoring Organizations of the Treadway Commission - Enterprise Risk Management - Integrated Framework - Executive Summary*. New York: AICPA.
- [28] Basque, R. (2011). *CMMI 1.3 - Guide complet de CMMI-DEV et traduction de toutes les pratiques CMMI-ACQ et CMMI-SVC*. Dunod.
- [29] Mayer, J., and Fagundes, L. L. (2009). A Model to Assess the MaturityLevel of the Risk Management Process in Information Security. *4rd IFIP/IEEE International Workshop on BDIM*. New York.
- [30] ISACA. (2010). *RISK IT Framework*.
- [31] Alter, S., and Sherer, S. A. (2004). A General but Readily Adaptable Model of Information System Risk. *Communications of the Association for Information Systems (ACM)*, 14, 1-28.
- [32] ISO. *ISO 31000:2009 Risk Management. Principles and Guidelines on Implementation*. Tech. rep.
- [33] Sienou, A. (2009). *Proposition d'un cadre méthodologique pour le management intégré des risques et des processus d'entreprise*. Thèse doctotale, Institut National Polytechnique de Toulouse, Toulouse.
- [34] Wade, M., and Hulland, J. (2004). Review: The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly*, 28(1), 107-142.
- [35] Alter, S., and Sherer, S. A. (2004). A General but Readily Adaptable Model of Information System Risk. *Communications of the Association for Information Systems (ACM)*, 14, 1-28.
- [36] Basili, R. V., Caldiera, G., and Rombach, H. D. (1994). Goal/Question /Metric Paradigm. *Encyclopedia of Software Engineering*, 1, 528-532.

Identification of Erythematous-Squamous Skin Diseases using Support Vector Machines and Extreme Learning Machines: A Comparative Study towards Effective Diagnosis

S. O. Olatunji¹ and Hossain Arif²

¹Computer Science Department, College of Computer Science and Information Technology,
University of Dammam, Saudi Arabia

²School of Engineering and Computer Science, BRAC University, Dhaka–Bangladesh

¹oluolunji.aadam@gmail.com; ¹osunday@uod.edu.sa

²hossain.arif@gmail.com; ²hossain.arif@bracu.ac.bd

ABSTRACT

Extreme Learning Machine (ELM) is a recently introduced learning algorithm for single hidden-layer feedforward neural network. Compared with classical learning algorithms in neural network, e.g. Back Propagation, ELM can achieve better performance with much shorter learning time. In the existing literature its better performance and comparison with Support Vector Machine (SVM), over regression and general classification problems, have caught the attention of many researchers. In this work, a comparison between ELM and SVM on identification of Erythematous-Squamous skin diseases is investigated. Detailed comparative studies were carried out through adequate experimentation. Experimental results indicated that ELM outperformed SVM. The effect of varying the size of training and testing sets on the performance of classifiers was also investigated in this study. The two techniques compared proved to be viable tools in this germane field of medical diagnosis.

Keywords: Extreme Learning Machine; Support Vector Machine; Erythematous-squamous skin diseases.

1 Introduction

The differential diagnosis of erythematous-squamous diseases is a real and difficult problem in dermatology. There are six groups in erythematous-squamous diseases. They are psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. These diseases are frequently seen in the outpatient dermatology departments. Since all erythematous-squamous diseases share the clinical features of erythema and scaling with very little differences, it is very challenging to identify which particular variant of the diseases is present in a particular patient.

With the recent advancement in Artificial Intelligence (AI) based techniques and their successful application in real life engineering and medical diagnosis, an interesting research problem is to look at how these diseases can be identified easily based on such AI techniques. Although the diagnosis of erythematous-squamous diseases have recently caught the attention of AI researchers leading to the use of some AI techniques [1-4], no one has yet implemented extreme learning machines in the identification of these diseases and compared its performance with support vector machines.

SVM is acclaimed by many researchers for its leading performance [3, 4]. Therefore, it has been widely used for different classification purposes. Most recently, a new learning algorithm, extreme learning machine (ELM), is available for the training of single layer feedforward neural network. The inventors of ELM have done a set of comprehensive experiments in regression and general classification to compare its performance with SVM [4-7]. The experimental results show that compared with classical learning algorithms in neural network, e.g. Back Propagation, ELM can achieve better performance with much shorter learning time [8]. Compared with SVM, ELM is sometimes better than SVM in terms of accuracy, though not always. But since the number of neurons available for each ELM machine is the only parameter to be determined, ELM is much simpler for parameter tuning compared with SVMs whose kernel functions are nonlinear, e.g. RBF functions, thus saving tremendous time in searching for optimal parameters.

Therefore, this work is set to present a comparative study of the performance of ELM and SVM on the identification of erythemato-squamous skin diseases. The effect of varying size of training and testing sets on the performance of classifiers was also investigated in this study. To do this, we partition the available dataset into training and testing sets based on four different ratios, which include 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively. Each of these ratio partitioning was used to carry out experiments involving training and testing the proposed classifiers to be compared.

Results from experiments carried out indicated that ELM outperformed SVM in both training and testing results. The testing set is the most crucial, as it is where the ability to identify unseen sample is tested.

The remaining part of this paper is organized as follows: section two contains the literature review; section three describes the proposed models, while empirical works and discussion of results are presented in section four. Conclusion and recommendations are presented in section five.

2 Literature Review

Juanying Xie et al. proposed hybrid feature selection algorithms to build efficient diagnostic models based on a new accuracy criterion, generalized F-score (GF), and SVM [1]. The hybrid algorithms adopt Sequential Forward Search (SFS), and Sequential Forward Floating Search (SFFS), and Sequential Backward Floating Search (SBFS), respectively, with SVM to accomplish hybrid feature selection with the new accuracy criterion to guide the procedure. These hybrid methods are called modified GFSFS, GFSFFS and GFSBFS, respectively. These combine the advantages of filters and wrappers to select the optimal feature subset from the original feature set to build efficient classifiers. Xie et al. conducted 10-fold cross validation experiments on training subset as well as on the whole erythemato-squamous diseases datasets to obtain the best and statistically meaningful classifiers. Experimental results showed that their proposed hybrid methods construct efficient diagnosis classifiers with high average accuracy when compared with traditional algorithms. The results of 10-fold cross validation experiments on erythemato-squamous diseases dataset show that the proposed hybrid feature selection algorithms, modified GFSFS, GFSFFS, and GFSBFS, have obtained the average classification accuracies of 99.17%, 98.33%, and 95.28% with the average size of selected feature subsets of 22, 13.2, and 17.1 respectively. While the classification accuracy of GFSFS, GFSFFS, and GFSBFS are 98.89%, 98.06%, and 95.81%, and the size of selected feature subset of them are 22.2, 13.3, and 19.3 respectively.

In [2], Xie and Wang developed a diagnosis model based on support vector machines (SVM) with a novel hybrid feature selection method to diagnose erythemato-squamous diseases. They proposed an

improved hybrid feature selection method, named improved F-score and Sequential Forward Search (IFSFS), which is a combination of filter and wrapper methods to select the optimal feature subset from the original feature set. The news IFSFS method improved the original F-score from measuring the discrimination of two sets of real numbers to measuring the discrimination between more than two sets of real numbers. The improved F-score and Sequential Forward Search (SFS) are combined to find the optimal feature subset in the process of feature selection. The best parameters of kernel function of SVM are found out by grid search technique. Xie and Wang then conducted experiments on different training-test partitions of the erythematous-squamous diseases dataset taken from UCI (University of California Irvine) machine learning database. Their experimental results show that the proposed SVM-based model with IFSFS achieves 98.61% classification accuracy and contains 21 features. The authors conclude that their method is very promising compared to the previously reported results.

Kecman and Kikec present the results of using Support Vector Machines (SVMs) and Radial Basis Function Neural Networks (RBF NNs) for diagnosing erythematous-squamous diseases [3]. The paper shows an application of RBF NN and SVM for diagnosing skin diseases by transforming a K-class problem into K two class problems (one-vs-all approach for multiclass problems). The data set contains 358 data pairs of 34 dimensional input records of patients with six known diagnosis (outputs). Thus, the data set is sparse and fairly unbalanced. The paper also discusses the strategies for training SVMs. Both networks design six different one-against-other classifier models which show extremely good performance on previously unseen test data. The training and the test sets are obtained by randomly splitting the dataset into two groups ensuring that each group contains at least one patient for each disease. 100 random split trials (equivalent to performing 10-fold-crossvalidation 10 times independently) were carried out for estimating the tests error rates. SVM models perform better than RBF NN ones, and the SVM models using both and polynomial kernels can perfectly classify, during the training unseen, test data. For a given data set SVMs using the polynomials of second order were particularly efficient and accurate. They use only between 5% and 10% of training data as the support vectors achieving perfect, error-less, diagnosis. Kecman and Kikec claim that the results shown are the best known to date for diagnosing erythematous-squamous diseases which represent difficult dermatological problems.

In their research paper, Ubeyli and Guler presented a new approach for detection of erythematous-squamous diseases based on adaptive neuro-fuzzy inference system (ANFIS) [4]. Neuro-fuzzy systems are fuzzy systems which use artificial neural networks (ANNs) theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples. Neuro-fuzzy systems harness the power of two paradigms: fuzzy logic and ANNs, by utilizing the mathematical properties of ANNs in tuning rule-based fuzzy systems that approximate the way man processes information. The ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion. The six ANFIS classifiers were used to detect the six erythematous-squamous diseases when 34 features defining six disease indications were used as inputs. Each of the ANFIS classifiers was trained so that they are likely to be more accurate for one type of erythematous-squamous disease than the other diseases. The predictions of the six ANFIS classifiers were combined by the seventh ANFIS classifier. To improve diagnostic accuracy, the seventh ANFIS classifier (combining ANFIS) was trained using the outputs of the six ANFIS classifiers as input data. The performances of the ANFIS model were evaluated in terms of training performances and classification accuracies, and the results confirmed that the proposed ANFIS

model has good potential in detecting erythemato-squamous diseases. The total classification accuracy of the ANFIS model was 95.5%. The ANFIS model achieved accuracy rates which were higher than that of the stand-alone neural network model.

3 Proposed Modeling Techniques

In order to facilitate comparing the results of ELM experiments with those of SVM, SVM and ELM techniques will be briefly described.

3.1 Support Vector Machines

SVMs are modern learning systems that deliver state-of-the-art performance in real world Pattern Recognition and data mining applications such as Text Categorization, Hand-Written Character Recognition, Image Classification, Material Identification and Bioinformatics, to mention but a few. Support Vector Machines have been recently proposed as a new intelligence framework for both prediction and classification based on both structure risk minimization criterion and soft margin hyperplane. This new framework deals with kernel neuron functions instead of sigmoid-like ones, which allows projection to higher planes and solves more complex nonlinear problems. It has featured in a wide range of medical and business journals, often with promising results.

Generally, in prediction and classification problems, the purpose is to determine the relationship among the set of input and output variables of a given dataset $D = \{Y, X\}$ where $X \in R^p$ represents the n-by-p matrix of p input variables. It may be noted that $Y \in R$ for forecasting problems and $Y \subseteq R$ for classification problems. Suppose $D = \{y_i, x_{i1}, \dots, x_{ip}\}$ is a training set for all $i = 1, \dots, n$ of input variables, X_j where $[X_j = (x_{1j}, \dots, x_{nj})^T]$ for $j = 1, \dots, p$, and the output variables, $Y = (y_1 \dots y_n)^T$. The lower case letters $x_{i1}, x_{i2}, \dots, x_{ip}$ for all $i = 1, \dots, n$ refer to the values of each observation of the input variables, and $y = k$ to the response variable Y to refer to class A_k for all $k = 1, 2, \dots, c$, where $c \geq 2$.

Here we briefly describe the basic ideas behind SVM for pattern recognition, especially for the two-class classification problem, and refer readers to [13, 14] for a full description of the technique.

The goal is to construct a binary classifier or derive a decision function from the available samples which has a small probability of misclassifying a future sample. SVM implements the following idea: it maps the input vectors $\vec{x} \in \mathbb{R}^d$ into a high dimensional feature space $\Phi(\vec{x}) \in \mathbf{H}$ and constructs an Optimal Separating Hyperplane (OSH), which maximizes the margin, the distance between the hyper plane and the nearest data points of each class in the space \mathbf{H} . Different mappings construct different SVMs. The mapping $\Phi(\cdot)$ is performed by a kernel function $K(\vec{x}_i, \vec{x}_j)$ which defines an inner product in the space \mathbf{H} . The decision function implemented by SVM can be written as:

$$f(\vec{x}) = \text{sgn} \left(\sum_{i=1}^N y_i \alpha_i \cdot K(\vec{x}, \vec{x}_i) + b \right) \quad (1)$$

Where the coefficients α_i are obtained by solving the following convex Quadratic Programming (QP) problem:

$$\text{Maximize} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot y_i y_j \cdot K(\vec{x}_i, \vec{x}_j)$$

Subject to

$$0 \leq \alpha_i \leq C \quad (2)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad i = 1, 2, \dots, N.$$

In the equation (2), C is a regularization parameter which controls the tradeoff between margin and misclassification error. These x_j are called Support Vectors only if the corresponding $\alpha_i > 0$.

Several typical kernel functions are:

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d, \quad (3)$$

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \|\vec{x}_i - \vec{x}_j\|^d\right) \quad (4)$$

Equation (3) is the polynomial kernel function of degree d which will revert to the linear function when d = 1. Equation (4) is the Radial Basis Function (RBF) kernel with one parameter γ .

Other kernel functions are:

$$\text{Linear: } K(x_i, x_j) = x_i^T x_j$$

And

$$\text{Sigmoid: } K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r).$$

Here $\gamma, r, \text{ and } d$ are kernel parameters.

3.2 Extreme Learning Machines

Extreme learning machine was introduced not-long-ago as a learning algorithm for single-hidden layer feed-forward neural networks (SLFNs), which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. In general, the learning rate of feed-forward neural networks (FFNN) is time-consuming than required and this has become bottleneck in their applications. According to [5], there are two main reasons behind this behavior; one is slow gradient based learning algorithms used to train neural network (NN) and the other is the iterative tuning of the parameters of the networks by these learning algorithms. To overcome these problems, [5, 14-15] proposed a learning algorithm called extreme learning machine (ELM) for single hidden layer feed-forward neural networks (SLFNs). It is stated that "In theory, this algorithm tends to provide the best generalization performance at extremely fast learning speed since it is a simple tuning-free algorithm" [15]. Therefore, it is an

interesting option to be considered in predictive modeling applications, particularly in reservoir characterization applications, where better generalization ability is often sought for.

The introduction of ELM has been considered as a good and welcome development because, in the past, it seems that there exists an unbreakable virtual speed barrier which classic learning algorithms cannot break through and therefore feed-forward network implementation then take a very long time to train itself, irrespective of the application type whether simple or complex. Also ELM tends to reach the minimum training error as well as it considers magnitude of weights which is opposite to the classic gradient-based learning algorithms which only intend to reach minimum training error but do not consider the magnitude of weights[5]. Also unlike the classical gradient-based learning algorithms which only work for differentiable activation functions, ELM learning algorithm can be used to train SLFNs with non-differentiable activation functions [14]. According to [5], “Unlike the traditional classic gradient-based learning algorithms, like back-propagation method, facing several issues like local minimum, improper learning rate and over-fitting, etc, the ELM is a simple tuning-free three-step algorithm that tends to reach the solutions straightforward without such trivial issues”.

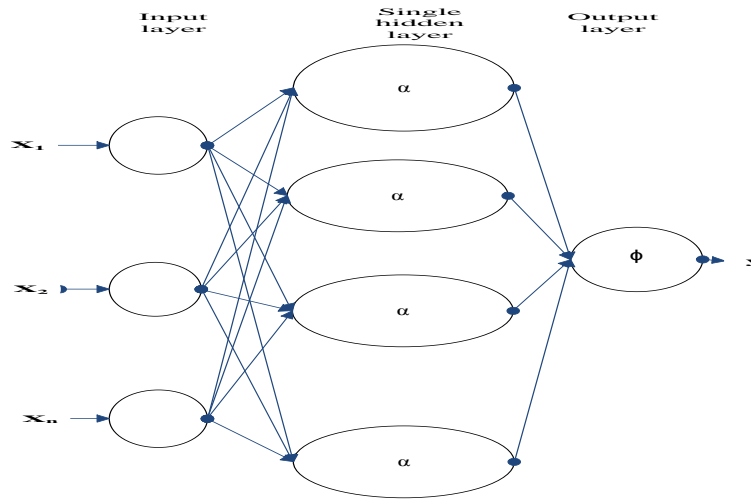


Figure 1: Schematic Representation of ELM ($x_1...x_n$ are the input values, α is the activation function, and y are the output values).

3.2.1 The learning process for the proposed model of Erythemato-Squamous skin diseases based on ELM Framework

Let us first define the standard SLFN (single-hidden layer feed-forward neural networks). If we have N samples (x_i, t_i) , where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, then the standard SLFN with \tilde{N} hidden neurons and activation function $g(x)$ is defined as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, \dots, N \tag{5}$$

Where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight vector that connects the i^{th} hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector that connects the i^{th} neuron and the output neurons, and b_i is the threshold of the i^{th} hidden neuron. The “.” in $w_i \cdot x_j$ means the inner product of w_i and x_j .

SLFN aims to minimize the difference between o_j and t_j . This can be expressed mathematically as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \quad (6)$$

Or, more compactly, as:

$$H \beta = T \quad (7)$$

where

$$H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_{\tilde{N}} \cdot x_{\tilde{N}} + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad T = \begin{bmatrix} T_1^T \\ \vdots \\ T_{\tilde{N}}^T \end{bmatrix}_{N \times m}$$

As proposed by Huang and Babri [16], H is called the neural network output matrix.

With the above SLFN specification background, thus the training procedures for the proposed ELM based model for erythematous-squamous disease identification can be summarized in the following algorithmic steps. See [5, 15] for further details on the workings of ELM algorithm.

Input - The inputs to the system, include clinical attribute obtained from patients, which represent the inputs parameters (input $x_i \in \mathbf{R}^n$ and target $t_i \in \mathbf{R}^m$), activation function, and the number of hidden neuron, \tilde{N} .

Output - The outputs of the ELM system is the target classes identified as the skin diseases type presented and then the weights of the layer.

Mathematically, given a training set

$$N = \{(x_i, t_i) | x_i \in \mathbf{R}^n, t_i \in \mathbf{R}^m, i=1, \dots, N\}$$

activation function $g(x)$, and the number of hidden neuron = \tilde{N} , then, do the following:

Step 0: **Initialization**. Assign random values to the input weight w_j and the bias b_j , $j = 1, \dots, \tilde{N}$

Step 1: Find the hidden layer output matrix H.

Step 2: Find the output weight β as follows:

$$\beta = H^+ T$$

Where β , H and T are defined in the same way they were defined in the SLFN specification above (equations 5, 6, and 7).

4 Empirical Studies, Results and Discussions

4.1 Dataset Description

In this work, the University of California Irvine (UCI) machine learning database containing erythemato-squamous diseases dataset was used and analyzed. There are 366 samples in this database and each sample has 34 attributes. Each sample contains 12 clinical features and 22 histopathological features. Patients were first evaluated clinically with 12 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features. The values of the histopathological features are determined by an analysis of the samples under a microscope. These attributes are listed in Table 1. The family history feature has the value '1' if any of these diseases has been observed in the family and '0' otherwise. The age feature represents the age of the patient. It has been missed in some samples, so it is removed in the experiments. Every other feature (clinical and histopathological) was given a degree in the range of '0' to '3' so that '0' indicates that the feature was not present, '3' indicates the largest amount possible, and '1', '2' indicate the relative intermediate values.

The dataset contains 34 attributes, 33 of which are linear valued and one of them is nominal. The differential diagnosis of erythemato-squamous diseases is a real problem in dermatology. They all share the clinical features of erythema and scaling, with very little differences. The six diseases in this group are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. The class distribution of the diseases is shown in Table 2.

Table 1: Features in the UCI erythemato-squamous diseases dataset

| Clinical Features (take values 0, 1, 2, 3, unless otherwise indicated) | Histopathological Features (take values 0, 1, 2, 3) |
|--|---|
| Feature 1: Erythema | Feature 12: Melanin incontinence |
| Feature 2: Scaling | Feature 13: Eosinophils in the infiltrate |
| Feature 3: Definite borders | Feature 14: PNL infiltrate |
| Feature 4: Itching | Feature 15: Fibrosis of the papillary dermis |
| Feature 5: Koebner phenomenon | Feature 16: Exocytosis |
| Feature 6: Polygonal papules | Feature 17: Acanthosis |
| Feature 7: Follicular papules | Feature 18: Hyperkeratosis |
| Feature 8: Oral mucosal involvement | Feature 19: Parakeratosis |
| Feature 9: Knee and elbow involvement | Feature 20: Clubbing of the rete ridges |
| Feature 10: Scalp involvement | Feature 21: Elongation of the rete ridges |
| Feature 11: Family history (0 or 1) | Feature 22: Thinning of the suprapapillary epidermis |
| Feature 34: Age (linear) | Feature 23: Spongiform pustule |
| | Feature 24: Munro microabscess |
| | Feature 25: Focal hypergranulosis |
| | Feature 26: Disappearance of the granular layer |
| | Feature 27: Vacuolisation and damage of basal layer |
| | Feature 28: Spongiosis |
| | Feature 29: Saw-tooth appearance of retes |
| | Feature 30: Follicular horn plug |
| | Feature 31: Perifollicularparakeratosis |
| | Feature 32: Inflammatory mononuclearinfiltrate |
| | Feature 33: Band-like infiltrate |

Table 2: Class distribution of erythematous-squamous diseases.

| Class code | Class | Number of instances |
|------------|--------------------------|---------------------|
| 1 | Psoriasis | 112 |
| 2 | Seboric dermatitis | 61 |
| 3 | Lichen planus | 72 |
| 4 | Pityriasis rosea | 49 |
| 5 | Chronic dermatitis | 52 |
| 6 | Pityriasis rubra pilaris | 20 |

4.2 Experimental setup and implementation process

As for the implementation of the two methods, we made use of MATLAB environment for the coding, while also calling some functions made available online, relating to ELM and SVM.

To evaluate performance of the two classifiers based on SVM and ELM modeling schemes, the acquired dataset described earlier is divided, using the stratified sampling approach, into 80% training set and 20% testing set for estimating how the investigated model performed on new unseen data. For testing and evaluation of both SVM and ELM framework and to carry out effective comparisons, the classification accuracy was calculated using Percent Correct measure, which is a measure of the percentage of correctly classified target classes.

As for the SVM, parameter settings include kernel function set to be "Gaussian", C set to be 450 and labda set to be 1e-7, based on parameters search outcomes. As for the ELM, hardlim activation function was chosen based on parameter search while the hidden neuron was set to 500.

To further investigate the effect of varying the size of training and testing sets on the performance of the two classifiers, we partition the available dataset into training and testing sets based on four different ratios, which include 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively. Each of these ratio partitioning was used to carry out experiments involving training and testing the two compared techniques.

4.3 Results and Discussions

Experimental results are presented in tables 3 through 6 representing the performance measure for both SVM and ELM classifiers. The results for the four different cases investigated, based on four different percentages of training and testing sets, are presented. These include training and testing sets in ratios 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively.

Table 3: Experimental Results with Training-Testing partition ratio 80:20

| | Training results | Testing Results |
|------------|------------------|-----------------|
| SVM | 100 | 97.26 |
| ELM | 100 | 98.36 |

Table 4: Experimental Results with Training-Testing partition ratio 70:30

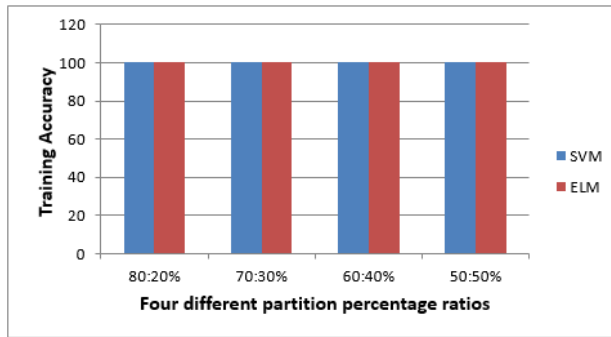
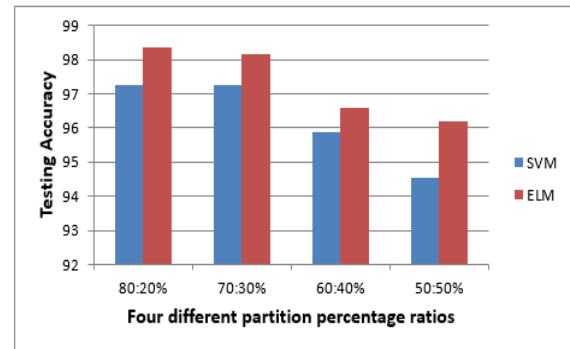
| | Training results | Testing Results |
|------------|------------------|-----------------|
| SVM | 100 | 97.25 |
| ELM | 100 | 98.17 |

Table 5: Experimental Results with Training-Testing partition ratio 60:40

| | Training results | Testing Results |
|-----|------------------|-----------------|
| SVM | 100 | 95.89 |
| ELM | 100 | 96.58 |

Table 6: Experimental Results with Training-Testing partition ratio 50:50

| | Training results | Testing Results |
|-----|------------------|-----------------|
| SVM | 100 | 94.54 |
| ELM | 100 | 96.18 |

**Figure 2: Summary of Training Results for all the Training-Testing partition ratios****Figure 3: Summary of Testing Results for all the Training-Testing partition ratios**

Figures 2 and 3 summarize the training and testing results, respectively.

It can be easily noticed from the tables and figures that ELM outperformed SVM in all the cases considered, for the testing set. For the training set, both ELM and SVM achieved the same high performance. As for the testing set, ELM performed better than SVM for all the four cases by a small margin.

Another important trend that is worth noticing and discussing is the pattern of difference in the performances displayed by each of the two methods for different data partition ratios used. The performance of both ELM and SVM for the testing set slightly decreases as the training set size decreases while testing set size increases. This is only applicable to the testing set. As for the training set, the performance of both ELM and SVM remain constant irrespective of the partition ratio used.

5 Conclusion and Recommendations

The popular SVM and ELM techniques have been proposed and compared as two new identification models to better identify erythemato-squamous skin diseases. The two models based on ELM and SVM have been compared looking into the effect of varying the ratio of training set to testing set. Experimental results indicated that ELM outperformed SVM in all fronts for the testing set case. Both the proposed classifiers have proven to be viable tools in this germane field as demonstrated by high accuracy and consistency of results even in the face of varying training-testing dataset percentages. Thus, SVM and ELM have been shown to demonstrate their unique ability to achieve excellent results in the field of biomedical diagnosis.

Another unique discovery made in this work is the ability of both SVM and ELM to maintain consistent results in the face of varying percentages of training and testing set partition ratio for the training set.

Although there was minimal decrease in the testing set performance as the training set decreases while the testing set increases in size, the reduction in performance is very minimal and acceptable as excellent results were still achieved. In addition, these two models based on each of the SVM and ELM models can be applied to other relevant biomedical diagnosis problems in the future.

REFERENCES

- [1] Xie, J., et al., *Novel Hybrid Feature Selection Algorithms for Diagnosing Erythematous-Squamous Diseases*, in *Health Information Science*, J. He, et al., Editors. 2012, Springer Berlin Heidelberg. p. 173-185.
- [2] Xie, J. and C. Wang, *Using support vector machines with a novel hybrid feature selection method for diagnosis of erythematous-squamous diseases*. *Expert Systems with Applications*, 2011. **38**(5): p. 5809-5815.
- [3] Vojislav Kecman and Mirna Kikec, *Erythematous-squamous diseases diagnosis by support vector machines and RBF NN*, *Proceedings of the 10th International Conference on Artificial intelligence and Soft Computing: Part I*, pp. 613-620, 2010, in *Springer -Verlag Berlin Heidelberg 2010*, e.a. L. Tutkowski, Editor. 2010. p. 613-620.
- [4] Übeyli, E.D. and İ. Güler, *Automatic detection of erythematous-squamous diseases using adaptive neuro-fuzzy inference systems*. *Computers in Biology and Medicine*, 2005. **35**(5): p. 421-433.
- [5] Huang, G.B., Q.Y. Zhu, and C.K. Siew, *Extreme learning machine: a new learning scheme of feedforward neural networks*, in *International Joint Conference on Neural Networks (IJCNN2004)*. 2004: Budapest, Hungary. p. 985 - 990.
- [6] Liu, H., J. Li, and L. Wong, *Use of Extreme Patient Samples for Outcome Prediction from Gene Expression Data*. *Bioinformatics*, 2005. **21**(16): p. 3377-3384.
- [7] Mahmoud, S.A. and S.O. Olatunji, *Automatic Recognition of Off-line Handwritten Arabic (Indian) Numerals Using Support Vector and Extreme Learning Machines*. *International Journal of Imaging*, 2009. **2**(A09): p. 34-53.
- [8] Olatunji, S.O., *Comparison of Extreme Learning Machines and Support Vector Machines on Premium and Regular Gasoline Classification for Arson and Oil Spill Investigation*. *Asian Journal Of Engineering, Sciences & Technology*, 2011. **1**(1): p. 1-7.
- [9] Olatunji, S.O., I.A. Adeleke, and A. Akingbesote, *Data Mining Based on Extreme Learning Machines for the Classification of Premium and Regular Gasoline in Arson and Fuel Spill Investigation*. *Journal Of Computing*, 2011. **3**(3): p. 130-136.

- [10] Olatunji, S.O., S. Ali, and A. Abdul Azeez. *Modeling Permeability Prediction Using Extreme Learning Machines*. in *Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, AMS2010*. 2010. Kota- Kinabalu, Malaysia: IEEE.
- [11] Olatunji, S.O., et al., *Extreme Learning Machine as Maintainability Prediction model for Object-Oriented Software Systems*. *Journal of Computing*, Volume 2, Issue 8, August 2010, 2010. **2**(8): p. 42-56.
- [12] Teddy Mantoro, et al., *Extreme learning machine for user location prediction in mobile environment*. *International Journal of Pervasive Computing and Communications*, 2011. **7**(2): p. 162 - 180.
- [13] Duda, R.O., P.E. Hart, and D.G. Stock, *Pattern Classification*. 2001, New York: John Wiley and Sons. 654.
- [14] Huang, G.B., et al., *Can threshold networks be trained directly?* *IEEE Trans. Circuits Syst. II*, 2006. **53**(3): p. 187-191.
- [15] Huang, G.B., Q.Y. Zhu, and C.K. Siew, *Extreme learning machine: Theory and applications*. *Neurocomputing*, Elsevier, 2006. **70**(1-3): p. 489-501.
- [16] Huang, G.B. and H.A. Babri, *Feedforward neural networks with arbitrary bounded nonlinear activation functions*. *9(1):224–229*. *IEEE Trans Neural Network*, 1998. **9**(1): p. 224-229.