Learning Generalized Weighted Relevance Aggregation Operators Using Levenberg-Marquardt Method

B. S. U. Mendis
Dept. of Computer Science
The Australian National University, Canberra, ACT 0200,Australia
sumudu@cs.anu.edu.au

T. D. Gedeon
Dept. of Computer Science
The Australian National University
Canberra, Australia
tom@cs.anu.edu.au

L. T. Kóczy
Dept. of Telec. & Media Informatics, Budapest University of Tech. & Econ., Budapest, Hungary.
Institute of Inf. Tech. and Elec. Eng. Széchenyi István University, Győr, Hungary
koczy@tmit.bme.hu

Abstract

We previously introduced the generalized Weighted Relevance Aggregation Operators (WRAO) for hierarchical fuzzy signatures. WRAO enhances the ability of the fuzzy signature model to adapt to different applications and simplifies the learning of fuzzy signature models from data. In this paper we overcome the practical issues which occur when learning WRAO from data. This paper discuss an algorithm for learning WRAO using the Levenberg-Marquardt (LM) method, which is one of the most sophisticated and widely used gradient based optimization method. Also, this paper shows the successful results of applying the proposed algorithm to extract WRAO for two real world problems namely High Salary Selection and SARS Patient Classification.

1. Introduction

Fuzzy signatures can model complex structured problems with the help of hierarchically structured vector valued fuzzy sets, and a set of aggregation functions [3]. The hierarchically structured vector valued fuzzy sets represent the degree of relationship of the attributes of the object. The set of aggregation functions map, from lower branches to the higher branches, the different universes of discourse of the hierarchical fuzzy signature structure. Further, these aggregation functions are non-homogenous. We argue that these properties help fuzzy signatures to model problems similar to the nature of human approaches to problem solving [3].

In [2] we further enhanced the inference in fuzzy signatures, by introducing the Weighted Relevance aggregation method. The concept behind the Weighted Relevance aggregation method is that the weights in each branch of the fuzzy signature are the observations of the relevance of that branch to its higher level branches in the hierarchical fuzzy signature structure. Thus, this method introduces additional expert knowledge to the fuzzy signature structure to classify vague data. In addition, it enhances the adaptability of hierarchical fuzzy signatures to different problem domains. In [2], we used simple aggregation functions: minimum, average, maximum-average, and maximum, with these weights. In [4] we showed the methodology of learning Weighted Relevancies from real world data. In [3] steepest gradient descent optimization was successfully used for learning Weighted Relevancies for hierarchical fuzzy signatures.

In [5] we further generalized these Weighted Relevancies and aggregation functions, in hierarchical fuzzy signatures, into one operator called Weighted Relevance Aggregation Operator (WRAO). This allows us to learn both aggregation function and Weighted Relevance at the same time for one node in the hierarchical fuzzy signatures structure. Thus, WRAO simplifies the learning of hierarchical fuzzy signature models from data. In this paper we discuss a Levenberg-Marquardt (LM) optimization method [6, 7] for extracting WRAO from real world data.

In section 2, we briefly discuss the theory of fuzzy signatures and WRAO. In section 3 we discuss the methodology of learning this new operator from data, using the LM optimization method. Finally, in section 4, we apply the proposed method for extracting WRAO for two real world hierarchical fuzzy signature structures namely High Salary Selection and SARS Patient Classification.

Proceedings of the Sixth International Conference on Hybrid Intelligent Systems (HIS'06)
0-7695-2662-4/06 $20.00 © 2006 IEEE

2. Fuzzy Signatures and Generalized WRAO

In this section we discuss the theoretical background of hierarchical fuzzy signature structures and the generalized WRAO.

2.1. Hierarchical Fuzzy Signatures and Weighted Relevance Aggregation

Fuzzy signatures are vector valued fuzzy sets, where each vector component can be a further vector valued fuzzy set [11]. A fuzzy signature s can be defined as [2]:

$$X \to [a_i]_{i=1}^{k}, \quad where \quad a_i = \begin{cases} [0,1]; & if \quad leafe \\ [a_{ij}]_{j=1}^{k_i}; & if \quad branch \end{cases}$$

Fig.1 shows an example of a hierarchical fuzzy signature structure with two arbitrary levels g and $(g+1)$.

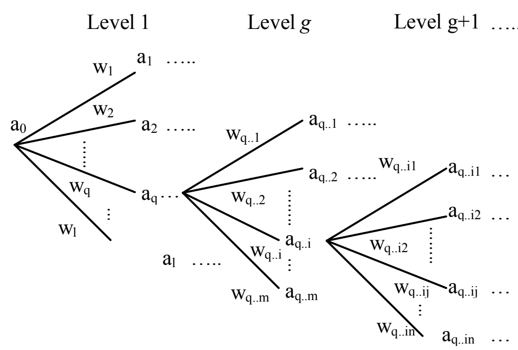

Fig. 1. Fuzzy Signature Structure with two arbitrary levels g and (g+1)

The Weighted Relevance aggregation concept was proposed in [2] to provide additional expert knowledge to the fuzzy signature structure by weighting the relevance of sub branches to its parent branch of the fuzzy signature structure. Thus, the Weighted Relevance will give an additional ability to fuzzy signatures for decision making in situations where input data are vague and complex.

The Weighted Relevance aggregation of an arbitrary branch $a_{q\ldots i}$ in a fuzzy signature (fig.1) can be defined as, $a_{q\ldots i} = @_{q\ldots i}\{w_{q\ldots ij} \bullet a_{q\ldots ij}\}$, where $@_{q..i}$ is an arbitrary aggregation function, $i=1\ldots n$, and $a_{q\ldots ij} \in [0,1]$. The following properties hold for arbitrary weighted relevance w_q,

i) $w_q \in [0,1]$ ii) $\sum_{q=1}^{l} w_q$ is not necessarily equal to 1.

In [1], [2] and [3] we have shown that the Weighted Relevance aggregation can be used to enhance the accuracy of the final results. Also, it helps to find a more abstract hierarchical structure, which represent large numbers of data points, for fuzzy signatures.

2.2. Generalized WRAO

In this section we briefly discuss the theory of Weighted Relevance Aggregation Operator (WRAO) [2] for hierarchical fuzzy signatures. The advantage of WRAO is that it enhances the adaptability of the hierarchical fuzzy signature structure by giving more flexibility to learn better values, which are weighted relevance and aggregation functions, for inference. Also, it simplifies the learning of weighted relevancies and aggregation function by generalizing them into one function. The following definition for generalized WRAO can be found in [2]:

Definition: The generalized WRAO of n branches $s_1, s_2, \ldots, s_n \in [0,1]$ with n weighted relevancies $w_1, w_2, \ldots, w_n \in [0,1]$, in a fuzzy signature, is a function $g : [0,1]^{2n} \to [0,1]$ such that,

$$g(s_1, s_2, \ldots, s_n; w_1, w_2, \ldots, w_n) = \left(\frac{1}{n} \sum_{i=1}^{n} (w_i s_i)^p \right)^{\frac{1}{p}} \qquad (1)$$

where $p \in \Re$, $p \neq 0$, $i \in [1,n]$ and $\sum_{i=1}^{n} w_i$ is not necessarily equal to 1. We call p the aggregation factor of the above function.

Further, in [2] we showed that WRAO generates a class of aggregations which spans the universe of all aggregations which fall between Zadeh's classical conjunction and disjunctions operators, including these two operators. Thus, we argue that WRAO enhances the adaptability of hierarchical fuzzy signature structures to different applications by allowing more flexibility in aggregation.

2.3. Partial Derivatives of WRAO

Weighted Relevance aggregation of an arbitrary branch $a_{q\ldots i}$ in a fuzzy signature (fig.1) using WRAO (1) can be written as,

$$a_{q\ldots i} = \left[\frac{1}{n} \sum_{j=1}^{n} \left(w_{q\ldots ij} a_{q\ldots ij} \right)^{p_{q\ldots i}} \right]^{\frac{1}{p_{q\ldots i}}} \qquad (2)$$

where n is the number of leaves in the branch $q\ldots i$, $w_{q\ldots ij}$ is the Weighted Relevance of the leaf $a_{q\ldots ij}$, where $j \in [1,n]$, and $p_{q\ldots i}$ is the aggregation factor of the branch $q\ldots i$.

To avoid the constraints on the weighted relevance factor w_i, we replaced it by the following sigmoid function [5], $w_i = \frac{1}{1+e^{-\lambda_i}}$, where $\lambda_i \in \Re$. After the above transformations it becomes clear that for any values of the parameter λ_i the weighted relevancies are $w_i \in [0,1]$, and $\sum w_i$ is not necessarily equal to 1. Therefore the constrained optimization problem has been transformed into an unconstrained optimization problem. Now equation (2) can be rewritten as follows,

$$a_{q...i} = \left[\frac{1}{n} \sum_{j=1}^{n} \left(\left\{ \frac{1}{1+e^{-\lambda_{q...ij}}} \right\} a_{q...ij} \right)^{p_{q...i}} \right]^{\frac{1}{p_{q...i}}} \quad (3)$$

We called λ the weighted relevance factor of the above function. Also, as noted before p is the aggregation factor of the above function.

This form of WRAO equation can be used for gradient based learning. The parameters we need to learn are the aggregation factor $p_{q...i}$, and the Weighted Relevance factor $\lambda_{q...i}$. First, we can obtain the partial derivatives of the equation (2) w.r.t. $p_{q...i}$ (fig. 1).

$$\frac{\partial a_{q...i}}{\partial p_{q...i}} = \left[\frac{a_{q...i}^{(1-p_{q...i})}}{np_{q...i}^2} \right] \left\{ \sum_{j=1}^{n} t \ln(t) - n a_{q...i}^{p_{q...i}} \ln\left(a_{q...i}^{p_{q...i}}\right) \right\} \quad (4)$$

where $t = \left(a_{q...ij}\left[w_{q...ij}\right]\right)^{p_{q...i}}$. Similarly, we can obtain the partial derivatives of the equation (4) w.r.t. $\lambda_{q...ik}$.

$$\frac{\partial a_{q..i}}{\partial \lambda_{q..ik}} = \left[\frac{1}{n} \sum_{j=1}^{n} \left(w_{q..ij} a_{q..ij}\right)^{p_{q..i}} \right]^{\left(\frac{1}{p_{q..i}}-1\right)} \frac{d\left(\left[\frac{w_{q..ik} a_{q..ik}}{n}\right]^{p_{q..i}}\right)}{d\lambda_{q..ik}} \quad (5)$$

where $k \in [1,n]$. But we faced the following practical issues with equation (4). The first issue arises when $\lim_{p_{q..i} \to 0} \left(\frac{\partial a_{q..i}}{\partial p_{q..i}}\right) \to +\infty$. The second problem is that theoretically $p_{q..i} \in (+\infty, -\infty)$, but practically it is a problem if we exceed $p_{q..i} = \pm 100$. The program will reach $\pm INF$ of its real values after a few iterations of learning. Therefore, we introduced 4 boundaries to the aggregation factor, $p \in [99, 0.001] \& [-0.001, -99]$, to realize this method. The values +99 and -99 are practically high enough to reach the maximum and minimum for given values of vectors respectively.

3. Levenberg-Marquardt (LM) Optimization for WRAO Learning

The LM algorithm is a widely used advanced optimization algorithm that outperforms simple gradient descent and other gradient methods when applied in a wide variety of problems. The major drawback with the steepest gradient descent method is that if there is no line search method combined with it, there is no guarantee of convergence. LM is a pseudo-second order method in which the Hessian matrix is estimated using the gradients [7].

The LM algorithm is a Sum of Square Errors (SSE) based minimization method. This implies that the function to be minimized is of the following special form [6]:

$$f(s) = \frac{1}{2} \sum_{i=1}^{n} (t_i - s_i)^2 = \frac{1}{2} \left\| \underline{t} - \underline{s} \right\|^2 \quad (6)$$

where $\underline{t}$ stands for the target vector, $\underline{s}$ for the actual output vector of the fuzzy signature, and $\| \ \|$ denotes the 2-norm. Also, it will be assumed that there are m parameters to be learned and there are n records in the training data set, such that $n > m$.

The next update of the LM can be written as:

$$\underline{u}[k] = \underline{par}[k] - \underline{par}[k-1] \quad (7)$$

where the vector $\underline{par}[k]$ contains the all parameters, which are all aggregation factors (p) and all weighted relevance factors (λ) in the fuzzy signature needing to be optimized by the learning algorithm in the k[th] iteration. The LM defines the next update $\underline{u}(k)$ in following manner:

$$\left(J^T[k] J[k] + \alpha I\right) \underline{u}(k) = -J^T[k] \underline{e}[k] \quad (8)$$

, where J is the Jacobian matrix of (6), I is the identity matrix of J, and α is a regularization parameter, which controls both the search direction and the magnitude of the next update $\underline{u}[k]$.

The LM algorithm can be also described as a blend of the steepest gradient descent and the Gauss-Newton method, namely, when $\alpha \to 0$, the LM method converges to the Gauss-Newton method and when $\alpha \to \infty$, the LM method works in parallel with the steepest gradient descent approach.

Equation (8) also can be written as in the following form:

$$\underline{u}[k] = -\left(J^T[k] J[k] + \alpha I\right)^{-1} J^T[k] \underline{e}[k] \quad (9)$$

The Jacobian matrix of (6) can be written as $J = \left[\frac{\partial\left(f(s^k)\right)}{\partial\left(\underline{par}[k]\right)} \right]$, where k is the iteration number of the training algorithm. The chain rule can be used to find the partial derivatives of the above Jacobian matrix, with a more detailed discussion with an example found in [2]. For an example, the partial derivative of $\frac{\partial a_0}{\partial p_{12}}$ (fig.1) can be written as $\frac{\partial a_0}{\partial p_{12}} = \frac{\partial a_0}{\partial a_1} \frac{\partial a_1}{\partial a_{12}} \frac{\partial a_{12}}{\partial p_{12}}$. Now, using (4) we can obtain $\frac{\partial a_{12}}{\partial p_{12}}$. Similarly, we can write $\frac{\partial a_0}{\partial \lambda_{12}} = \frac{\partial a_0}{\partial a_1} \frac{\partial a_1}{\partial \lambda_{12}}$ and (5) can be used to calculate $\frac{\partial a_1}{\partial \lambda_{12}}$.

The LM algorithm uses the restricted step size method to find a good optimal solution. The following algorithm has been given in [9] for calculating the next update $\alpha[k]$ using trust region $r[k]$, which is given in equation (10), to achieve a good convergence:

I. Given $\underline{par}[k]$ and $\alpha[k]$, use (9) to find $\underline{u}[k]$.

II. Calculate $r[k]$ use (10).
If $r[k]$ <0.25 set $\alpha[k+1]=4\alpha[k]$.
If $r[k]$ >0.75 set $\alpha[k+1]=\alpha[k]/2$.
Otherwise $\alpha[k+1]=\alpha[k]$.

III. If $r[k]\le 0$ set $\underline{par}[k+1]=\underline{par}[k]$
Else $\underline{par}[k+1]=\underline{par}[k]+\underline{u}[k]$.

IV. Find new error $\underline{e}[k+1]$, use (6).
If $\underline{e}[k+1]$ > *threshold* then go to I.
Else stop learning.

where k is the current iteration number. The trust region calculation for above algorithm is given by:

$$r[k]=\left[\frac{f(par[k-1])-f([par[k]])}{f(par[k-1]-q(par[k]))}\right] \qquad (10)$$

where approximation of the error [9], $q(par[k])$, is given by $q(par[k])=\{J^T[k]\cdot \underline{u}[k]+\underline{e}[k]\}$. Initially, the algorithm starts by choosing an arbitrary $\alpha[1]>0$ and arbitrary values for $\underline{par}[1]$.

4. Experiments

In this section two real world problems, which are SARS Patients Classification problem and High Salary Selection of employees [3], have been used for the experiments. In [4] we have investigated the results of these problems using automatically constructed weighted relevancies and using manually set aggregation functions, using human experts' suggestions, for fuzzy signatures. The scope of these experiments is to show the proposed WRAO is able to learn good aggregation functions and weighted relevancies simultaneously. Further, the results of experiments should work similar to or better than the results of the experiments in [4]. The same data sets were used for experiments in [4] have been used for the experiments in this paper. We used separate training and test data sets, and only repeat test results from our previous work.

4.1. Experiment 1: High Salary Selection.

The High Salary selection fuzzy signature structure (fig.2) gives the degree of relevance of having a high salary according to contacts, age, and work experience of a particular employee. The scope of the first experiment is to extract the WRAOs for each node in the High Salary Selection fuzzy signature structure automatically. Also, it was expected that the result of the High Salary Selection fuzzy signature with learnt WRAOs should perform at least as well as the results of the similar experiment of the same fuzzy signature with automatically extracted weighted relevancies and manually set aggregation functions [4] (fig. 4).

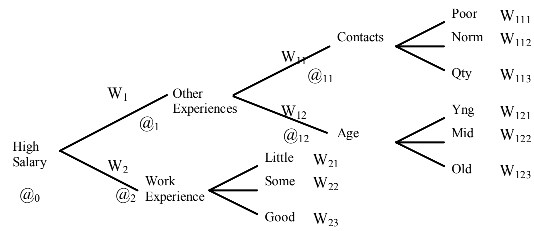

Fig. 2. High Salary Selection Fuzzy Signature

Note that in fig.2 $@_i$ and w_i represent the aggregation function and weighted relevance of the node *i*, respectively. Also, $@_i$ and w_i are functions of aggregation factor p_i and weighted relevance factor λ_i. Fig.3 shows the same fuzzy signature in fig.2 with all aggregation and weighted relevance factors. In fig.3. a_i represents the input of node *i*. Further, the Jacobian matrix, for learning all above factors, is of the following form,

$$J^k=\begin{bmatrix} \frac{\partial a_0^1}{\partial p_0} & \frac{\partial a_0^1}{\partial p_1} & \cdot & \cdot & \frac{\partial a_0^1}{\partial p_{12}} & \frac{\partial a_0^1}{\partial \lambda_0} & \cdot & \cdot & \frac{\partial a_0^1}{\partial \lambda_{123}} \\ \cdot & \cdot & & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & & & \cdot \\ \frac{\partial a_0^r}{\partial p_0} & \frac{\partial a_0^r}{\partial p_1} & & & \frac{\partial a_0^r}{\partial p_{12}} & \frac{\partial a_0^r}{\partial \lambda_0} & & & \frac{\partial a_0^r}{\partial \lambda_{123}} \end{bmatrix}$$

where r is the number of records in the training data set.

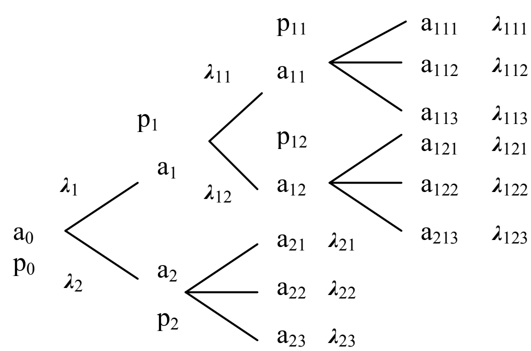

Fig.3 High Salary Selection Fuzzy Signature with all aggregation and weighted relevance factors.

The test data set contain 200 records and the training data set contains 135 records. Each record of test and training data contains an employee's following details: contacts, age, and work experience. Those employees' actual salary scale is divided into 3 soft categories namely low_salary, mid_salary, and high_salary. The *desired values* plot in the fig.4 shows the degree of relevance of each employee's actual salary category to high salary category. For an example, the membership of employees records from 1-90 to the high salary category are less than 0.4. Thus they belong to the low_salary category. The *Actual Results* plot (fig.4) shows the results of the High Salary Selection fuzzy signature with automatically extracted weighted relevancies and manually set aggregation function [4].

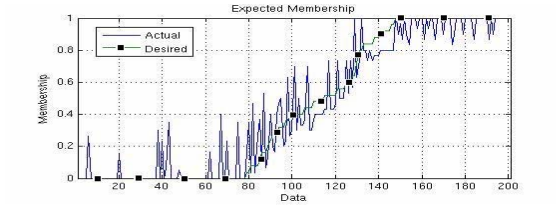

Fig. 4. Results of Similar Experiment in [4]

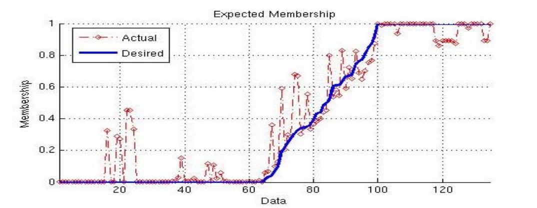

Fig.5. Training Results of Experiment 1

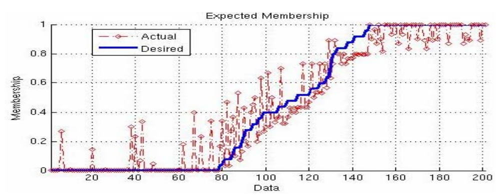

Fig.6. Test Results of Experiment 1

Fig.5 and fig.6 show the results of experiment 1 with automatically learnt WRAOs, using the proposed method, at the training phase and test phase respectively. According to fig.6 it is clear that the *Actual* test results shown in fig.6 is similar to that of the fig.4. Also, table 1 shows the comparison of the mean square error of the previous experiment in [4] versus this experiment for both training and testing phases. These learnt WRAOs can classify the results of salary data into the expected 3 categories accurately. Thus, we can argue that our new WRAO and LM have the capability to extract good aggregation functions and their weighted relevancies for hierarchical fuzzy signatures. The normalized mean square error of our automated technique is relatively higher than our previous technique.

Table 1 Comparison of Mean Square Error for Exp. 1

	Train	Test
Salary Experiment in [4]	0.0122	0.0123
Experiment 2	0.0141	0.0137

4.2. Experiment2: SARS Patients Classification

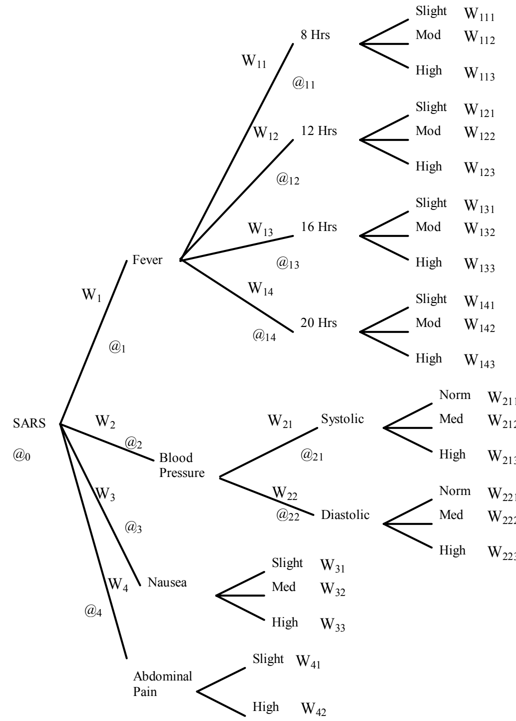

Fig.7. SARS Patients Classification Fuzzy Signature (levels 1 to 3)

The SARS Patients Selection fuzzy signature structure in fig.7 has been used for the second experiment. Both test and training data sets contain 4,000 records patient data. Each set of test and training data contains 4 different categories of patients, namely SARS, normal, pneumonia, and hypertension, and have been used for the second experiment. Each of these categories mentioned above has 1,000 records of data, which were derived from real world values using some randomization.

The experiment with SARS Patients Classification fuzzy signature was carried out to extract the WRAOs, for each node of the hierarchical fuzzy signature (fig.7), to classify the patients according to input data into two different categories of having SARS or not having SARS. We used 1.0 degree of relevance for SARS patient data and 0.0 degree of relevance for non-SARS patient data. Also, as in the first experiment, it was expected that the result of the SARS Patients Classification fuzzy signature with learnt WRAOs should perform at least as well as the results of the similar experiment of the same fuzzy signature with automatically extracted weighted relevancies and manually set aggregation functions in [4].

Fig.6 shows the test results of the similar experiment in [4]. According to this figure all SARS patients data (range 3,000 – 4,000) were kept above 0.5 and all non-

SARS patients data (range 1,000 – 3,000) were kept to 0.0 degree of relevance to SARS condition.

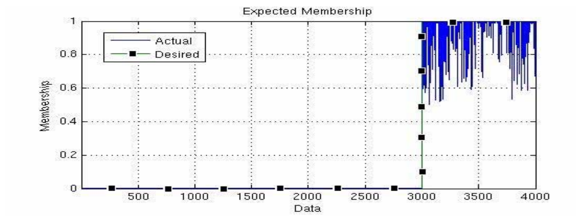

Fig. 8. Results of Similar Experiment in [4]

Fig.9 shows the test results of experiment 1 with automatically learnt WRAOs. The figure of the training phase is very much identical to the test phase figure. Therefore we omit that figure from the paper.

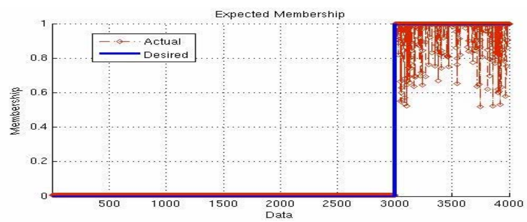

Fig.9. Testing Results of the Experiment 2

According to fig.9 it is clear that the test result of experiment 2 is very similar to that of experiment 2 in [4] (fig.8). Also, table 2 shows the comparison of the mean square error of the similar experiment in [4] versus this experiment for both training and testing phases. Thus, we can argue that our new WRAO and LM have the capability to extract good aggregation functions and their weighted relevancies for different real world hierarchical fuzzy signatures successfully.

Table 2 Comparison of Mean Square Error for Exp. 2

	Train	Test
SARS Experiment in [4]	0.0020	0.0018
Experiment 2	0.0020	0.0017

5. Conclusion

The WRAO for hierarchical fuzzy signatures has been reviewed. Also, some practical issues faced when realizing this method have been solved. Next, a practical algorithm which uses the Levenberg-Marquardt (LM) method has been proposed for realizing the learning of these WRAOs from data for hierarchical fuzzy signatures. Two experiments were carried out to find the flexibility of WRAO for learning, and the applicability of the LM for WRAO learning method. Both experiments concluded the success of the above assumptions we made. Thus, we have made progress towards a fully automated version, which will be essential on the high volume medical data we will process.

References

[1] B.S.U. Mendis, T.D. Gedeon, L.T. Kóczy, "Flexibility and Robustness of Hierarchical Fuzzy Signature Structures with Perturbed Input Data," *International Conf. of Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU)*, Paris, 2006.

[2] B.S.U. Mendis, T.D. Gedeon, J. Botzheim, L.T. Kóczy, "Generalized Weighted Relevance Aggregation Operators for Hierarchical Fuzzy Signatures," *International Conference on Computational for Modeling, Control and Automation (CIMCA), Sydney, Australia*, 2006.

[3] B.S.U. Mendis, T.D. Gedeon, L.T. Kóczy, "Investigation of Aggregation in Fuzzy Signatures", *3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems, Singapore*, December 2005.

[4] B.S.U. Mendis, T.D. Gedeon, L.T. Kóczy, "Learning Weights from Observations for Hierarchical Fuzzy Signatures ," *Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on Advanced Intelligent Systems (SCIS/ISIS),* Japan, 2006.

[5] B.S.U. Mendis, T.D. Gedeon, L.T. Kóczy, "On the Issue of Learning Weights from Observations for Fuzzy Signatures," *World Automation Congress (WAC),* Hungary, 2006.

[6] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math.*, 1 I , 1963, pp. 431-441.

[7] J. Botzheim, C. Cabrita, L. T. K6czy, A. E. Ruano, "Estimating Fuzzy Membership Functions Parameters by the Levenberg-Marquardt Algorithm," *FUZZ-IEEE*, Hungary, 2004.

[8] L.T. Kóczy, "Vector Valued Fuzzy Set," *Busefal, ete* 1980, pp. 41-57.

[9] R. Fletcher, "Practical Methods of Optimization," vol. 1. John Wiley and Sons, 1980.