Investigation of Aggregation in Fuzzy Signatures

B. Sumudu U. Mendis[1], Tamás D. Gedeon[1], László T. Kóczy[2]

[1]Department of Computer Science
The Australian National University
Canberra, ACT 0200
Australia

[2]Department of Telecommunication and Telematics
Budapest University of Technology and Economics
Budapest, Hungary

Abstract

The hierarchical fuzzy signatures structure is a novel concept that can be used to find the degree of similarity or dissimilarity of objects which contain complex structured data, for classification or decision making. Fuzzy signatures are vector valued fuzzy sets, where each vector component can be a further vector valued fuzzy set. Thus, it differs from sparse hierarchical fuzzy rule based systems. Medical and economic diagnoses are the obvious applications of the fuzzy signatures. In this report we present results of three experiments, which were carried out to find the applicability of different aggregation functions, the relationship between the fuzzy signature structure and aggregation functions, and applicability of the fuzzy signatures method for different real world problems. Also, a new method of aggregating fuzzy signatures using weights called the weighted aggregation method has been proposed. Experiments show that the weighted aggregation method provides better results for fuzzy signatures.

I. Introduction

Soft computing research focuses mainly on identifying approximate models for decision support or classification where analytically unknown systems exist. Mostly, those systems consist of very complex structured, high dimensional data, and sometimes with interdependent features. Fuzzy logic approaches have become ideal for soft computing research because of the ability to assign linguistic labels [1] and to model uncertainty in most decision making and classification problems. But conventional fuzzy rule based systems suffer from high computational time complexity. Thus, their applicability still remains on real time control systems with few dimensions of input variables and simply structured data.

As a concept the fuzzy signatures have been discussed in Vámos [2]. In Kóczy [3] further combined the fuzzy signatures with vector valued fuzzy sets to develop the theoretical aspects of fuzzy signatures. In Wong [4] & [5] the construction of hierarchical fuzzy signature structure from data has been discussed. Fuzzy signatures describe objects with the help of a set of interpretation qualitative measures (not necessarily homogeneous), which are also arranged in a hierarchical structure expressing interconnectedness and interdependence in a way modeling the human approach to problems. Thus, fuzzy signatures are capable of handling systems with complex structured data and sometimes with missing values. In addition, it models systems that are analytically unknown and decision making or classification process based on uncertainty.

The ability of the fuzzy signature to process decisions or classifications based on uncertainty is obvious. The specialty of the fuzzy signature approach is firstly based on their hierarchically structured vector valued fuzzy sets, which contain the interconnectedness of features of the object by their hierarchical structure and describe the interdependence of components of that object by using vector valued fuzzy sets. Secondly, the interconnected relationship between higher and lower levels of the fuzzy signature structure is derived by a set of qualitative measures, which are not necessarily homogeneous. That is, different types of aggregations can be used to find the final atomic result by taking different aggregations at different levels as well as at the same level for different sets of vector valued fuzzy sets. This is an ability which is not normal to conventional rule based systems, even for sparse hierarchical fuzzy systems.

Now, it is understandable that the hierarchically structured fuzzy signatures are different from the sparse hierarchical fuzzy systems by their signature structure, the way of aggregating the signature, and the relationship between the signature structure and aggregations. Here, our research was primarily focused on evaluating the ways of aggregating the fuzzy signature structure, and finding the relationship between the fuzzy signature structure and the aggregations. Further, the research focused on finding the applicability of the aggregation in fuzzy signatures for different real world problems.

In order to evaluate the above three objectives, two different example problems have been undertaken, from medical diagnosis, and personnel management. Extracting suitable fuzzy signature structures are beyond the scope of this experiment. Therefore, the SARS fuzzy signature structure in [4] and simple High Salary classification signature in [6] has been used as examples. Three different experiments were carried out with the above two examples.

The first experiment has been set up to evaluate results of the different aggregations with the same fuzzy signature structure. In the second experiment the same set of aggregations has been applied to a different structure of the fuzzy signature used for experiment one. Artificially generated data with 1000 records for each case were used for these experiments. Finally, for the third experiment, a different fuzzy signature has been used with the same set of aggregations.

In the literature two different methods have been proposed by Wong [4] and Kóczy [3] for aggregation of the fuzzy signatures. Wong [4] has proposed to aggregate the fuzzy signature in a straightforward manner, whereas Kóczy [1] proposed to first find the similarity between a pre-identified signature and the input signature before aggregation to a final result. Furthermore, the following two similarity measures have been proposed in Kóczy [3],

$$s = s1 \wedge s2 \qquad \text{eqn1}$$

$$s = (s1 \wedge s2) \vee (\hat{s}1 \wedge \hat{s}2) \qquad \text{eqn2}$$

where *s, s1* and *s2* are fuzzy signatures. Kóczy's method 1 (eqn1) has been modified during the research as follows,

$$s = 1 - |s1 - s2| \qquad \text{eqn3}$$

where the degree of similarity between *s1* and *s2* is high when the difference of *s1* and *s2* are low and vice versa. Also, the weighted aggregation method (explained in section II) for fuzzy signature has been proposed during the research in order to achieve accurate final results. Therefore, eight different methods of taking aggregation have been tested in each experiment, which are the non-weighted and weighted versions of above four methods.

II. The use of different aggregations at different levels of the fuzzy signature.

The scope of the first experiment was to find the applicability of the different aggregation functions to aggregate different levels of fuzzy signature structure and combinations of aggregations which give the best results. Therefore, the first experiment has been set up to evaluate the results of the different aggregations with the same fuzzy signature structure.

The simple aggregations *average* (AV), *minimum* (MN), *maximum* (MX), and a new aggregation called *average maximum* (AM) given by equation 4 were used with SARS Fuzzy signature structure [4] (fig 1).

$$AM = (AV+MX)/2 \qquad \text{eqn4}$$

During experiment 1 it was observed that the results of aggregations *average* and *maximum* give the expected pattern for the final results. But aggregation *average* is always below the expected vales for the final results. And aggregation *maximum* is always well above the expected values for the final results. Therefore, the new aggregation *average maximum*

No	Agg. Name	No	Agg. Name	No	Agg. Name	No	Agg. Name
1	AV,AV,AV	17	AV,AV,MN	33	AV,AV,MX	49	AV,AV,MA
2	MN,AV,AV	18	MN,AV,MN	34	MN,AV,MX	50	MN,AV,MA
3	MX,AV,AV	19	MX,AV,MN	35	MX,AV,MX	51	MX,AV,MA
4	MA,AV,AV	20	MA,AV,MN	36	MA,AV,MX	52	MA,AV,MA
5	AV,MN,AV	21	AV,MN,MN	37	AV,MN,MX	53	AV,MN,MA
6	MN,MN,AV	22	MN,MN,MN	38	MN,MN,MX	54	MN,MN,MA
7	MX,MN,AV	23	MX,MN,MN	39	MX,MN,MX	55	MX,MN,MA
8	MA,MN,AV	24	MA,MN,MN	40	MA,MN,MX	56	MA,MN,MA
9	AV,MX,AV	25	AV,MX,MN	41	AV,MX,MX	57	AV,MX,MA
10	MN,MX,AV	26	MN,MX,MN	42	MN,MX,MX	58	MN,MX,MA
11	MX,MX,AV	27	MX,MX,MN	43	MX,MX,MX	59	MX,MX,MA
12	MA,MX,AV	28	MA,MX,MN	44	MA,MX,MX	60	MA,MX,MA
13	AV,MA,AV	29	AV,MA,MN	45	AV,MA,MX	61	AV,MA,MA
14	MN,MA,AV	30	MN,MA,MN	46	MN,MA,MX	62	MN,MA,MA
15	MX,MA,AV	31	MX,MA,MN	47	MX,MA,MX	63	MX,MA,MA
16	MA,MA,AV	32	MA,MA,MN	48	MA,MA,MX	64	MA,MA,MA

Table 1. Table of Combinations of Aggregations

(eqn4) was introduced to get expected values while preserving the expected pattern. The idea behind that is to increase the final results above the results of the average and decrease the final results from that of the maximum. Table 1 show all the possible combinations of three of the four available aggregations which were used for the experiments.

The doctors know that for certain symptoms, such as SARS, they need to check the patient for possible fever, blood pressure, conditions of nausea, and abnormal pains. In addition, it is fairly important to monitor the fever four times a day. The figure (fig. 1) below show the SARS fuzzy signature [4] used for the experiment 1. Each symptom check has been divided into a number of fuzzy sets, such as "*slight*", "*moderate*", and "*high*" for fever, "*low*", "*normal*", and "*high*" for both blood pressure types, "*slight*", "*medium*", and "*high*" for nausea, and "*yes*", and "*no*" for abnormal pain. Also, the SARS signature contains three levels of hierarchies, which can be aggregated using different aggregations.

Figure 1 also shows weights that are applied to the fuzzy signature, e.g. SARS fuzzy signature, when using our proposed weighted aggregation method. According to notations of fig.1 weight w_{ij} represents the weight j for the level i. Weights are applied having an understanding that some membership values may contribute more to the final result than the others in the same group or level. As an example, contribution of slight fever, moderate fever, and high fever to the final SARS condition can be expressed linguistically as "*less*", "*somewhat*", and "*more*". Therefore, the weights w_{31}, w_{32}, and w_{33} in figure 1 have been configured according to these linguistic expressions. Thus, our attempt can be expressed as an approach which uses linguistic expressions to add further domain knowledge to the fuzzy signatures. Our experiments show that this new weighted aggregation method for fuzzy signatures generally gives better results than the un-weighted aggregation method.

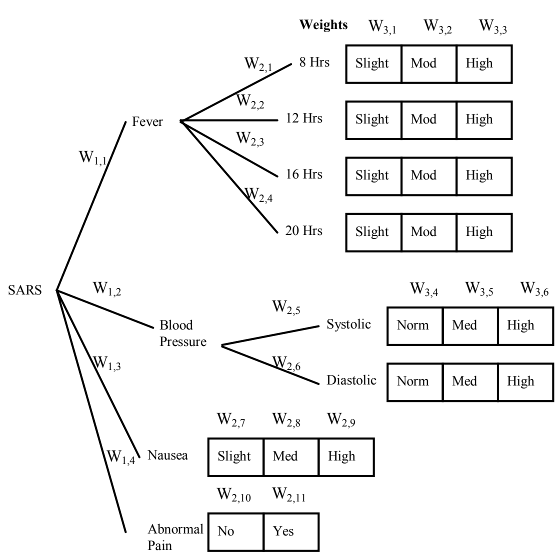

Fig. 1 SARS Fuzzy Signature (levels 1 to 3)

Figure 2 below shows the result of the experiment 1 using weighted Wong's method. All the graphs in figure 2 contain different combinations of aggregations against the degree of membership to the abnormal condition. Four different conditions namely SARS patients, normal persons, pneumonia patients and high blood pressure patients, with 1000 records of data per each condition, have been considered for the evaluation. In order to check the correctness of the SARS signature, as it contain values for fever and blood pressure, the pneumonia and high blood pressure patients data has been considered in addition to normal person data. The figures 2.1, 2.2, 2.3, and 2.4 show the results of the SARS fuzzy signature applying SARS patient data, normal person data, pneumonia data, and high blood pressure patient data respectively. Also, graph 1 in figure 2 has been sorted into descending order of the maximum results of aggregations and the other graphs in figure 2 are ordered according to the order of the aggregations in graph 1. Finally, aggregation numbers which are listed in X axis of every graph, in this paper, are references to table 1. The aggregations in table 1 should read from left to right and they have been applied to the level 1, level 2, and level 3 of the SARS signature respectively.

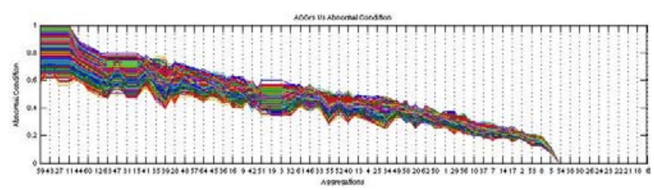

Fig. 2.1. SARS patient data with weighted Wong's method.

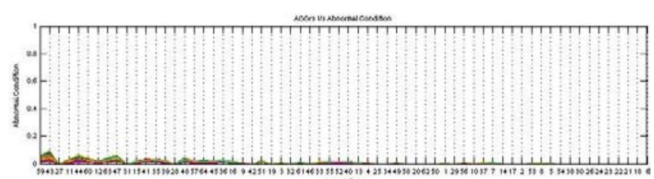

Fig. 2.2. Normal person data with weighted Wong's method.

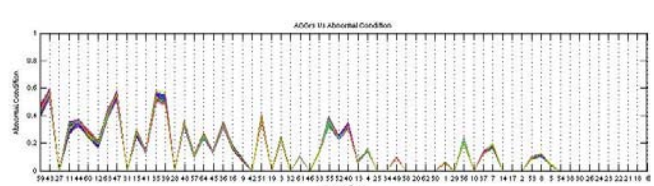

Fig. 2.3. Pneumonia patient data with weighted Wong's method.

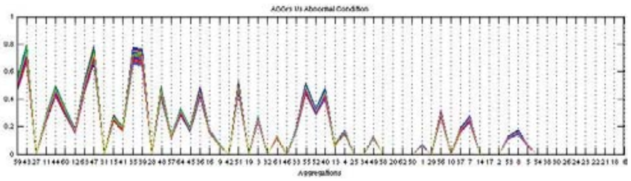

Fig. 2.4. High Blood pressure patient data with weighted Wong's method.

Figure 3 contain the same data as figure 2, where the graphs 3.1, 3.2, 3.3, and 3.4 shows the results of the SARS fuzzy signature after applying SARS patient data, normal person data, pneumonia data, and high blood pressure patient data respectively, but it has been sorted differently. That is, the graph 1 (fig 3.1) is sorted in the same manner as fig 2.1 within 4 sub ranges (3 non empty) being categories of aggregations, which are grouped according to the maximum and minimum results of the aggregations, that are bounded by [1, 0.8], (0.8, 0.5], (0.5, 0.3], and (0.3, 0]. The other 3 graphs in the figure 3 are each sorted according to the ascending order of the minimum result of the aggregations.

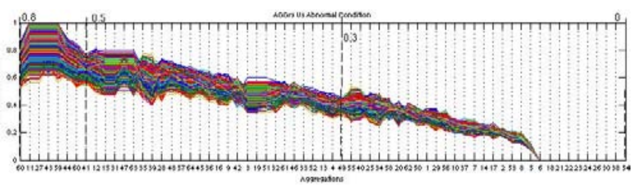

Fig. 3.1. SARS patient data with weighted Wong's method.

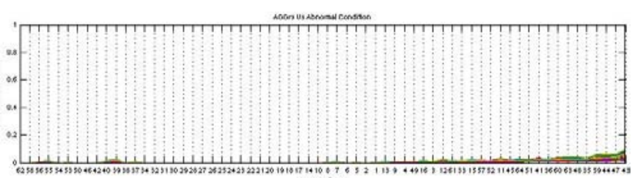

Fig. 3.2. Normal person data with weighted Wong's method.

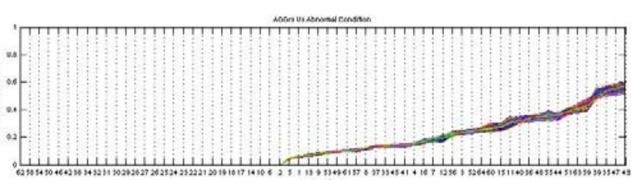

Fig. 3.3. Pneumonia patient data with weighted Wong's method.

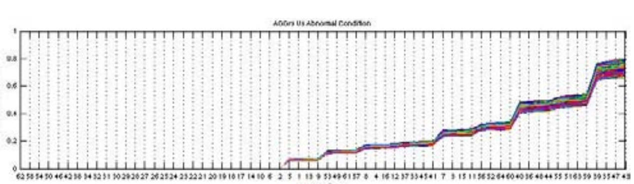

Fig. 3.4. High Blood pressure patient data with weighted Wong's method.

Now, using figures 2 and 3, the best performance aggregations for the SARS fuzzy signature with weighted Wong's method can be identified. Here the best performance aggregation means the aggregation which keeps the results for the SARS patients data at least above 0.5 and keeps the results for other patient data and normal person data below 0.5. The aggregation (MX, MX, MN) can be seemed to be the best according to this method. Also, aggregations (MX, MX, MA), (MX, MX, AV), (MA, MX, MX), (MA, MX, MA), (MA, MX, AV), and (MX, MX, MA) give good results. In addition to that, our new aggregation MA will contribute positively to most of above aggregation combinations.

The figures 4, 5, 6, 7, 8, 9, and 10 show results of Wong's method, Koczy's method 1 (eqn1), Koczy's method 2 (eqn2), modified Koczy's method 1 (eqn3), weighted Koczy's method 1, weighted Koczy's method 2, and weighted & modified Koczy's method 1. Figures 4, 5, 6, 7, 8, 9, and 10 are sorted in the same manner as figure 3, which differ only by the method used for aggregation. For an example, the

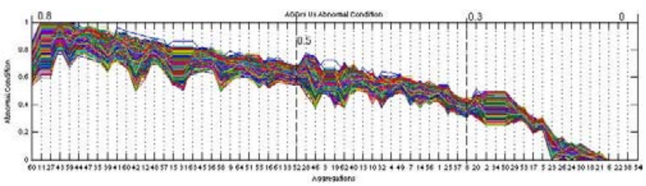

Fig. 4.1. SARS patient data with Wong's method.

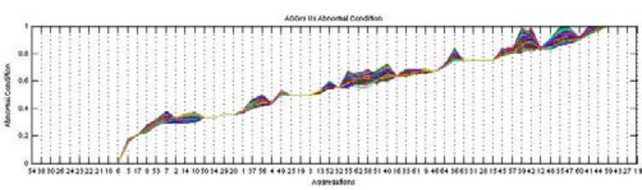

Fig. 4.2. Normal persons data with Wong's method.

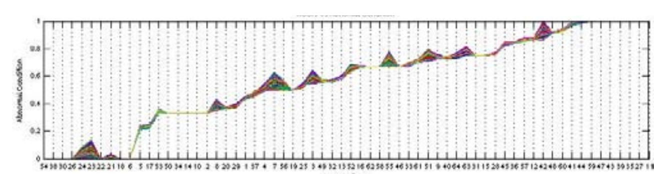

Fig. 4.3. Pneumonia patient data with Wong's method.

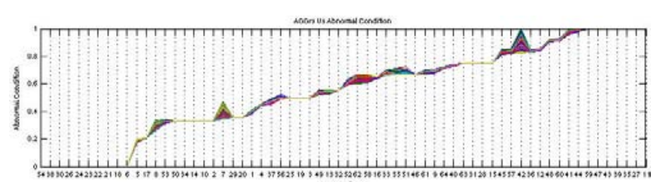

Fig. 4.4. High Blood pressure patient data with Wong's method.

figures 4.1, 4.2, 4.3, and 4.4 shows the results of using Wong's method as aggregation with the SARS fuzzy signature after applying SARS patient data, normal person data, pneumonia data, and high blood pressure patient data respectively.

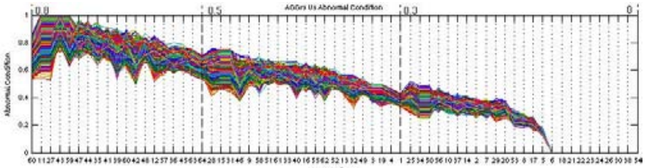

Fig. 5.1. SARS patient data with Koczy's method 1.

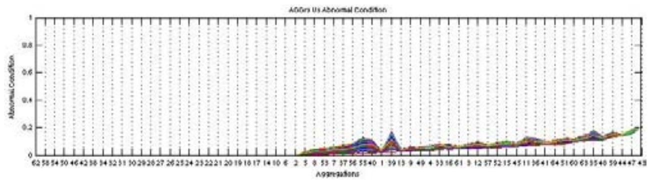

Fig. 5.2. Normal person data with Koczy's method 1.

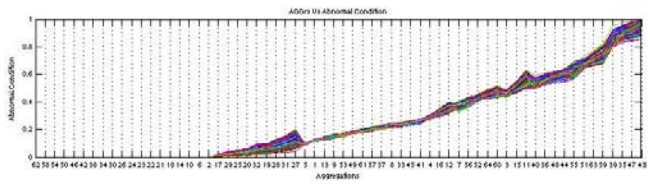

Fig. 5.3. Pneumonia patient data with Koczy's method 1.

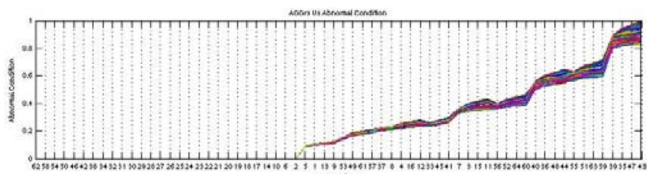

Fig. 5.4. High Blood pressure patient data with Koczy's method 1.

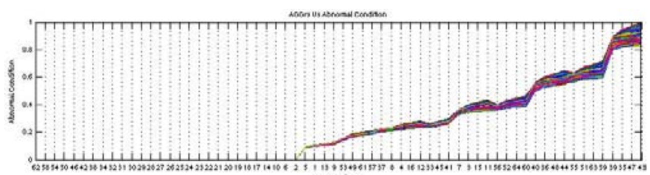

Fig. 6.1. SARS patient data with Koczy's method 2.

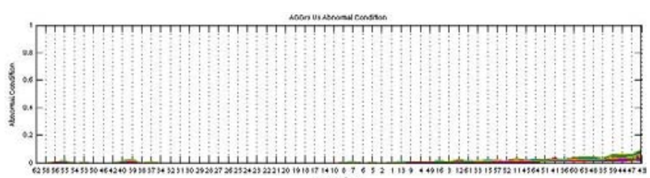

Fig. 6.2. Normal person data with Koczy's method 2.

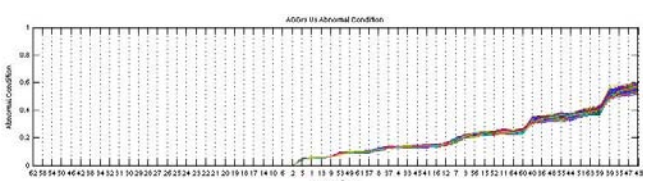

Fig. 6.3. Pneumonia patient data with Koczy's method 2

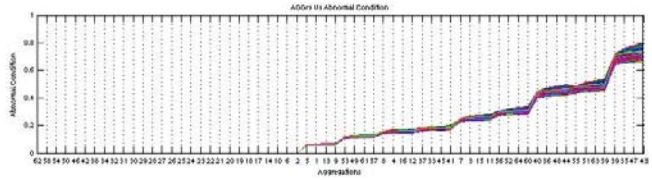

Fig. 6.4. High Blood pressure patient data with Koczy's method 2.

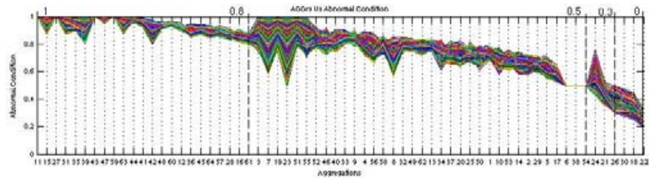

Fig. 7.1. SARS patient data with weighted Koczy's method 1.

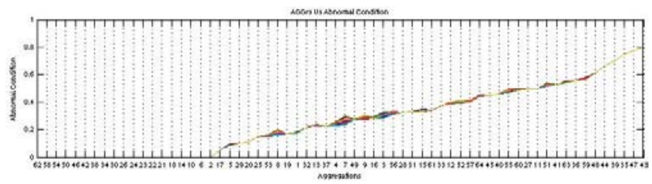

Fig. 7.2. Normal person data with weighted Koczy's method 1.

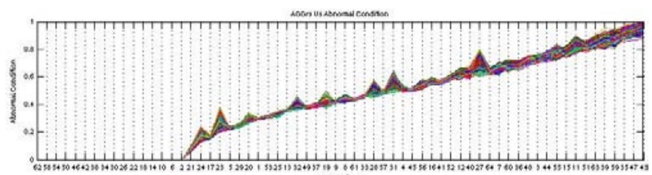

Fig. 7.3. Pneumonia patient data with weighted Koczy's method 1.

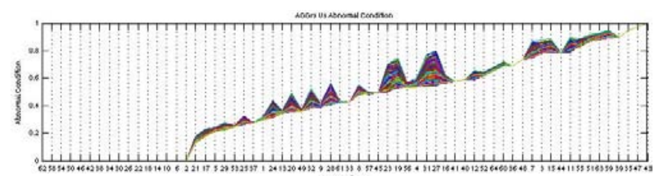

Fig. 7.4. High Blood pressure patient data with weighted Koczy's method 1.

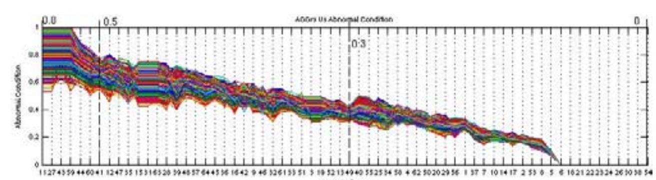

Fig. 8.1. SARS patient data with weighted Koczy's method 2.

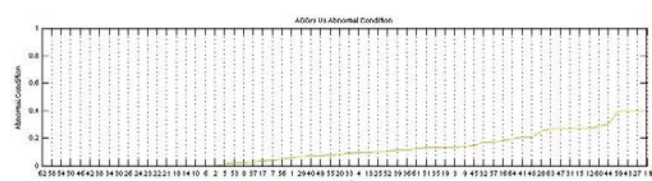

Fig. 8.2. Normal person data with weighted Koczy's method 2.

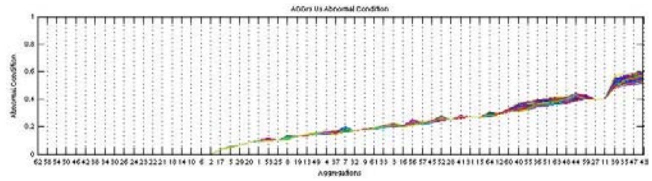

Fig. 8.3. Pneumonia patient data with weighted Koczy's method 2.

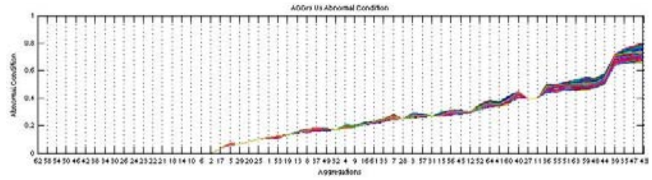

Fig. 8.4. High Blood pressure patient data with weighted Koczy's method 2.

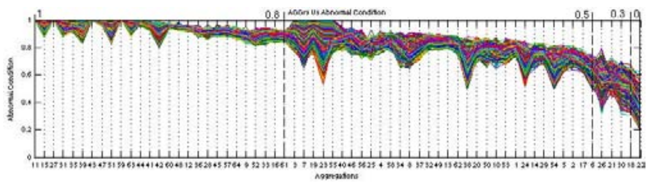

Fig. 9.1. SARS patient data with modified Koczy's method 1.

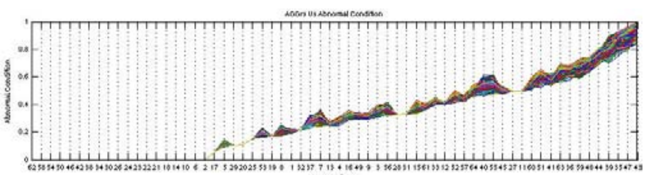

Fig. 9.2. Normal person data with modified Koczy's method 1.

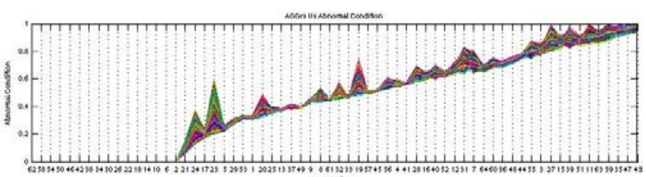

Fig. 9.3. Pneumonia patient data with modified Koczy's method 1.

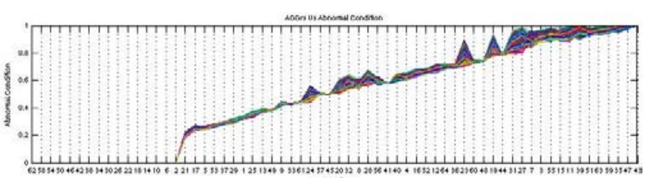

Fig. 9.4. High Blood pressure patient data with modified Koczy's method 1.

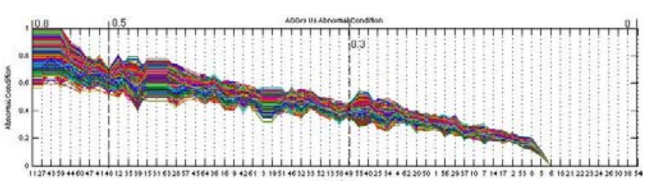

Fig. 10.1.SARS patient data with weighted & modified Koczy's method 1.

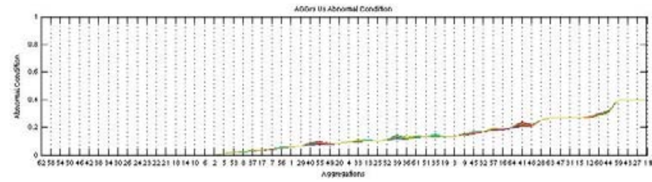

Fig. 10.2. Normal person data with weighted & modified Koczy's method 1

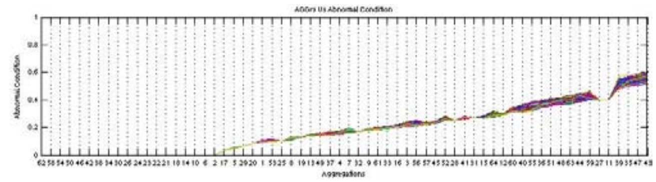

Fig. 10.3. Pneumonia patient data weighted & modified Koczy's method 1

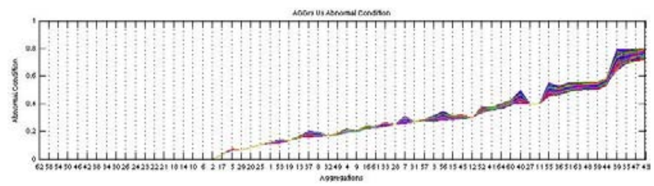

Fig. 10.4. High Blood pressure patient data weighted & modified Koczy's method 1.

Overall, the weighted Wong's method looks stable and gives better results than the other methods discussed above. Kóczy's method 1 works well but all Kóczy's methods are computationally expensive as they have additional comparison with a pre-identified fuzzy signature [3]. Moreover, it needs an extra effort to identify this pre-signature and difficult to deal with when values are missing or removed from inputs.

III. The relationship between the fuzzy signatures structure and the aggregations.

The scope of the second experiment was to find the relationship between the fuzzy signatures structure and the aggregation functions. Therefore, the second experiment evaluates the results of the different aggregations with same SARS data but using a different fuzzy signature structure. Figure 11 shows the new SARS fuzzy signature structure. In order to form a new structure the blood pressure node in level 1 has been removed from the original SARS fuzzy signature structure (fig. 1). Systolic and diastolic pressures are now two branches of level 1 (fig. 11) of new SARS fuzzy signature and labeled measure S and measure D. This is equivalent to considering systolic and diastolic values not being connected to each other.

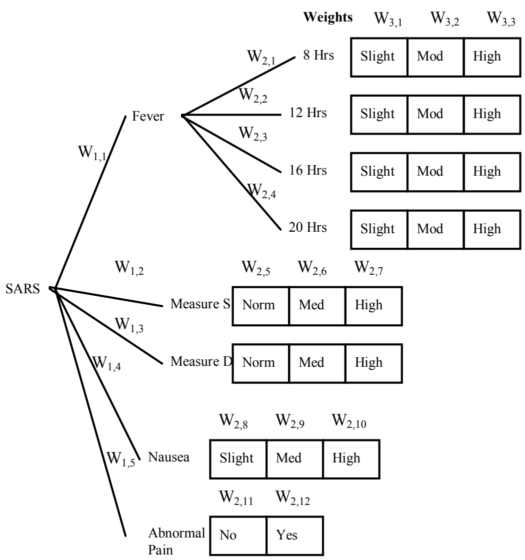

Fig. 11 Modified SARS Fuzzy Signature (levels 1 to 3)

Figure 12 below shows the results of experiment 2 using the weighted Wong's method. All the graphs in figure 12 show different combinations of aggregations against the degree of membership to the abnormal condition. As in experiment 1, all four different conditions namely SARS patients, normal persons, pneumonia patients and high blood pressure patients, with 1000 records of data per condition, have been considered for the evaluation. The figures 12.1, 12.2, 12.3, and 12.4 show the results of the weighted Wong's method with SARS fuzzy signature applying SARS patient data, normal person data, pneumonia data, and high blood pressure patient data respectively. Also, all the graphs in figure 12 have been sorted as the same way as figure 2.

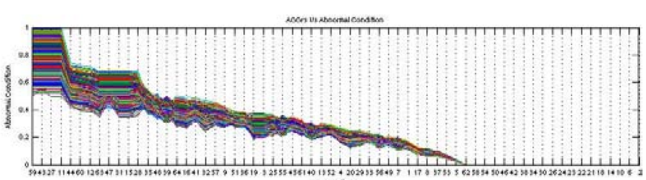

Fig. 12.1. SARS patient data with weighted Wong's method.

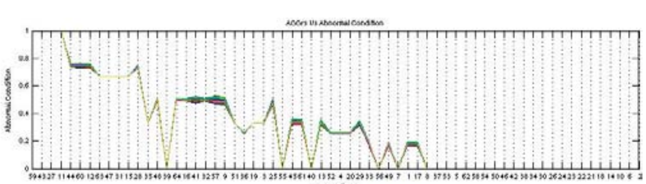

Fig. 12.2. Normal person data with weighted Wong's method.

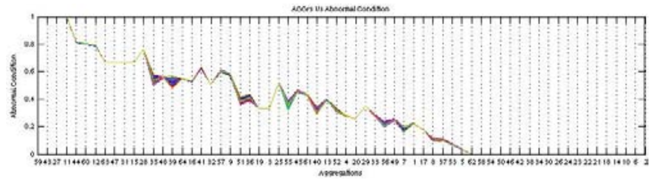

Fig. 12.3. Pneumonia patient data with weighted Wong's method.

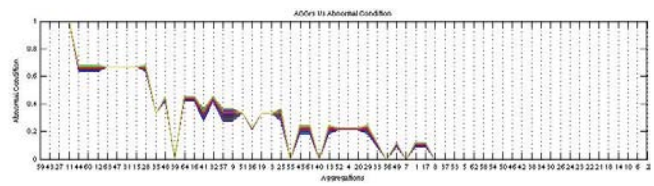

Fig. 12.4. High Blood pressure patient data with weighted Wong's method.

The overall experiment with all methods shows that no aggregation method will work properly for the new SARS fuzzy signature structure. Therefore, only the weighted Wong's method has been shown as an example for discussion of this experiment.

According to the results of experiment two, it can be stated that there is a relationship between the fuzzy signature structure and the aggregations for better performance of the overall fuzzy signature. On the other hand, Wong in [4] discussed that when data are missing or removed from the inputs an optimal fuzzy signature, which is derived from original fuzzy signature, can be used to find accurate results. Our experiment 2 thus opens a new research direction, to find the limits and constraints of deriving an optimal fuzzy signature from the original fuzzy signature.

IV. The applicability of the fuzzy signature method for signatures other than SARS signature.

Finally, a different fuzzy signature has been used with the same set of aggregations to evaluate the applicability of the aggregations for a different signature. In addition to the main scope, our experiment has been organized to discover the performance of the new weighted aggregation method for different fuzzy signatures. Therefore, the third experiment has been conducted for both weighted and non-weighted Wong's method.

The High Salary selection signature in [6] has been used as the example for the experiment. Figure 13 shows the High Salary Fuzzy Signature extracted from [6], where receiving a high salary of a person is given by contacts ("*Poor*", "*Normal*", and "*Quality*"), age ("*Young*", "*Middle*", and "*Old*"), and experience

("*Little*", "*Some*", and "*Good*"). Also, figure 13 shows how weights are applied to the High Salary fuzzy signature when calculated using weighted aggregation method.

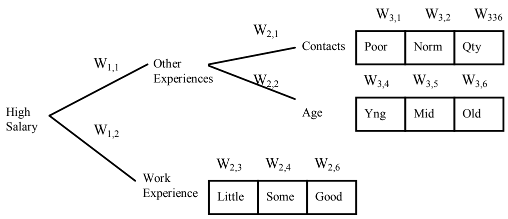

Fig. 13 High Salary Fuzzy Signature

Figure 14 below shows the result of the experiment 3 using Wong's method. All the graphs in figure 14

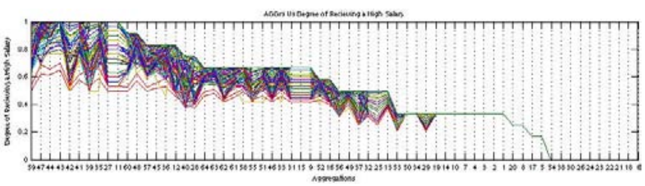

Fig. 14.1. Membership to High Salary is [0, 0.3] data with Wong's method.

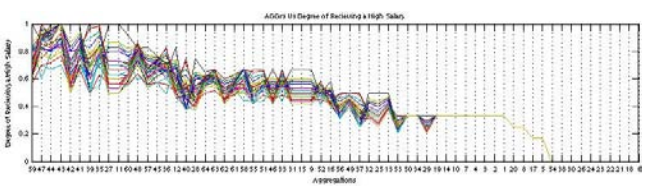

Fig. 14.2. Membership to High Salary is (0.3, 0.6] data with Wong's method.

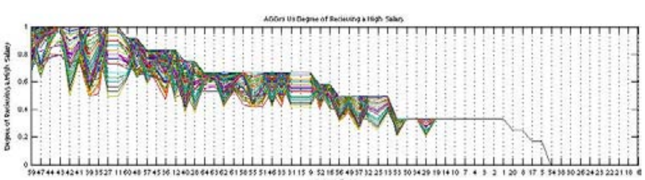

Fig. 14.3. Membership to High Salary is (0.6, 1] data with Wong's method.

show different combinations of aggregations against the degree of membership to receiving a high Salary. The test data contains 200 records. Graphs in figure 14.1, 14.2, and 14.3 are categorized according to the results of the hierarchical fuzzy system data [6] with membership to have high salary [0, 0.3], (0.3, 0.6], and (0.6, 1] respectively. The contents of these graphs show the results after application to the High Salary fuzzy signature. All graphs in figure 14 have been sorted into descending order of the maximum results of the aggregations. The only difference between figures 14 and 15 is that figure 15 shows the results of experiment 3 using the weighted Wong's method.

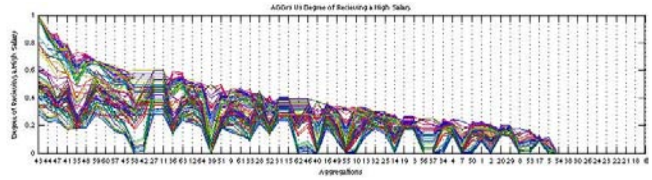

Fig. 15.1.Membership to High Salary is [0, 0.3] data with weighted Wong's method.

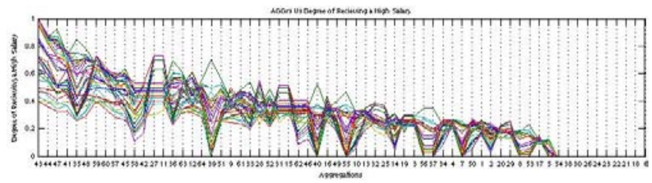

Fig. 15.2. Membership to High Salary is (0.3, 0.6] data with weighted Wong's method.

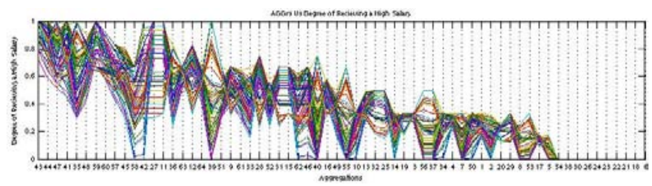

Fig. 15.3. Membership to High Salary is (0.6, 1] data with weighted Wong's method.

Figure 14 shows that there is no aggregation method will work well for Wong's method. The major drawback of Wong's method is there is no option to fine tune the results. On the other hand, the weighted aggregation method can be used to fine tune the final results. During the experiment the weights have been fine tuned to get optimal results. Thus, figure 15 shows the optimal results of the weighted aggregation method. Also, it has been observed that the best results of the weighted aggregation method depend on effective weightings. Therefore, our third experiment also suggests a new research direction, to find effective weighting methods for aggregation of fuzzy signatures. Finally, experiment 3 shows that the fuzzy signature method will work for different types of real world data effectively.

V. Conclusion

The special benefits of fuzzy signatures for decision making and classification over conventional rule based fuzzy systems has been discussed. Moreover, differences between hierarchically structured fuzzy signatures and sparse hierarchical fuzzy systems have been mentioned. There experiments were carried out to find the applicability of different aggregation methods, the relationship between the fuzzy signature structure and the aggregation methods, and applicability of fuzzy signatures for different real world problems. The first experiment concludes that some aggregations, e.g. [max, max, min] perform excellently. Also, it can be concluded that our new aggregation MA will mostly contributes to good aggregation combinations. The results of the second experiment pointed out the need for constraints when finding an optimal fuzzy signature from the original fuzzy signature. The third experiment shows the ability of fuzzy signatures to handle different real world problems effectively. In addition to the three experiments a new aggregation method for fuzzy signatures has been proposed. All three experiments conclude that the importance and accuracy of the proposed weighted aggregation method is high.

References

[1] L.A. Zadeh, "Fuzzy Algorithm", Information and Control, vol. 12, 1968, pp. 94-102.

[2] T. Vámos, L.T. Kóczy, G. Biró, "Fuzzy Signatures in Data Mining," Proceedings of the joint 9th IFSA World Congress, 2001, pp. 2842-2846.

[3] L.T. Kóczy, T. Vámos, G. Biró, "Fuzzy Signatures," Proceedings of EUROFUSE-SIC '99, 1999, pp. 210-217.

[4] K.W. Wong, T.D. Gedeon, L.T. Kóczy, "Construction of Fuzzy Signature from data," Proceedings of IEEE International Conference on Fuzzy Systems, 2004, vol. 3, pp 1649-1654.

[5] K.W. Wong, A. Chong, T.D. Gedeon, L.T. Kóczy, T. Vámos, " Hierarchical Fuzzy Signature Structure for Complex Structured Data," Proceedings of International Symposium on Computational Intelligence and Intelligent Informatics 2003 (*ISCIII'03*), 2003, Nabeul, Tunisia, pp 105-109.

[6] T.D. Gedeon, K.W. Wong, D. Tikk, "Constructing Hierarchical Fuzzy Rule Bases for Classification," Proceedings of the 10th IEEE International Conference on Fuzzy Systems (*FUZZ-IEEE 2001*), 2001, pp. 1388–1391.