Construction of Fuzzy Signature from Data: An Example of SARS Pre-clinical Diagnosis System

Kok Wai Wong[1], Tamás Gedeon[2], and László Kóczy[3, 4]

[1]School of Computer Engineering
Nanyang Technological University
Nanyang Avenue,
Singapore 639798
Email: askwwong@ntu.edu.sg

[2]Department of Computer Science
The Australian National University
Acton, ACT 0200, Australia
Email: tom.gedeon@anu.edu.au

[3]Department of Telecommunication & Telematics
Budapest University of Technology and Economics
Budapest, Hungary
E-mail: koczy@ttt.bme.hu

[4]Institute of Information Technology and Electrical Engineering
Széchenyi István University
Győr, Hungary
E-mail: koczy@sze.hu

Abstract - **There are many areas where objects with very complex and sometimes interdependent features are to be classified; similarities and dissimilarities are to be evaluated. This makes a complex decision model difficult to construct effectively. Fuzzy signatures are introduced to handle complex structured data and interdependent feature problems. Fuzzy signatures can also used in cases where data is missing. This paper presents the concept of a fuzzy signature and how its flexibility can be used to quickly construct a medical pre-clinical diagnosis system. A Severe Acute Respiratory Syndrome (SARS) pre-clinical diagnosis system using fuzzy signatures is constructed as an example to show many advantages of the fuzzy signature. With the use of this fuzzy signature structure, complex decision models in the medical field should be able to be constructed more effectively.**

I. Introduction

The main motivation of soft computing research has been on dealing with very complex and often analytically unknown systems. Most of the soft computing techniques do this in the sense of identifying approximate models, controlling or generating decision support. Fuzzy modelling has become a very popular field in soft computing research because of its ability to assign meaningful linguistic labels to the fuzzy sets [1] in the rule base [2,3]. This approach still has not successfully tackled problems with reasonably high numbers of input variables because of the high computational cost involved, nor have they solved the problem of dealing with problems with complex and interdependent features or where data is missing.

A serious problem is caused by the high computational time and space complexity of rule bases describing systems with multiple inputs with proper accuracy. The complexity allows little general systems application (or real time control application) of classical fuzzy algorithms, where the inputs exceed about 6 to 10. These traditional fuzzy systems deal with very simple structured data, where the number of inputs is well defined, and values for each input occur for most or all data items. This further reduces their general applicability.

When dealing with high dimensional of data, a new branch of computer science and applications known as data mining has gained much attention over the last few years. The known situation is used to build the model, and then it is applied to another situation where it is not known. The objective of the data mining algorithm is to automate the detection of relevant patterns in a large database. This signature is then widely applied, mainly for data organisation, retrieval and data mining [4]. A fuzzy signature is mainly an extension of this basic concept to include fuzzy sets theory. Problems like those in the economy and medical fields normally have objects with very complex and sometimes interdependent features that need to be classified and evaluated. The fuzzy signature is thus introduced to model complex structured data [5,6,7].

Fuzzy signatures that structure data into vectors of fuzzy values, each of which can be a further vector, are introduced to handle complex structured data. This will widen the application of fuzzy theory to many areas where objects are complex and sometimes interdependent features are to be classified and similarities / dissimilarities evaluated. Often, human experts can and must make decisions based on comparisons of cases with different numbers of data components, with even some components missing. Fuzzy signatures are created with this objective in mind. This tree structure is a generalisation of fuzzy sets and vector valued fuzzy sets in a way modelling the human approach to complex problems.

One of the advantages of using fuzzy signature for complex structured decision modelling is the underlying fuzzy signature can be extracted directly from data. The constructed fuzzy signature can then be modified if necessary without changing much of the decision nature of the fuzzy signature. The objective of this paper is to show how fuzzy

0-7803-8353-2/04/$20.00 © 2004 IEEE

signatures can be constructed, and a Severe Acute Respiratory Syndrome (SARS) Pre-clinical Diagnosis System is used as an application example of fuzzy signatures.

II. Fuzzy Signature

The original definition of fuzzy sets was $A: X \to [0,1]$, and was soon extended to *L-fuzzy sets* by Goguen [8],

$$A_S : X \to [a_i]_{i=1}^{k}, a_i = \begin{cases} [0,1] \\ [a_{ij}]_{j=1}^{k_i} \end{cases}, a_{ij} = \begin{cases} [0,1] \\ [a_{ijl}]_{l=1}^{k_{ij}} \end{cases},$$

$A_L : X \to L$, L being an arbitrary algebraic lattice. A practical special case, *Vector Valued Fuzzy Sets* was introduced by Kóczy [9], where $A_{V,k} : X \to [0,1]^k$, and the range of membership values was the lattice of k-dimensional vectors with components in the unit interval. A further generalisation of this concept is the introduction of fuzzy signatures and signature sets, where each vector component is possibly another nested vector as shown in Figure 1 and on the right.

Each signature corresponds to a nested vector structure or, equivalently, to a tree graph. The internal structure of the signature indicates the semantic and logical connection of state variables, corresponding to the leaves of the signature graph. The fuzzy signatures can be described as a generalised vectorial fuzzy set with possible recursive vectorial components. It can be denoted as:

$$x = \begin{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \\ \begin{bmatrix} x_{21} \\ \begin{bmatrix} x_{221} \\ x_{222} \\ x_{223} \end{bmatrix} \\ x_{23} \end{bmatrix} \\ \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \end{bmatrix}^T$$

$$A : X \to S^{(n)} \qquad (1)$$

where $n \geq 1$ and

$$S^{(n)} = \prod_{i=1}^{n} S_i \qquad (2)$$

$$S_i = \begin{cases} [0,1] \\ S^{(m)} \end{cases} \qquad (3)$$

and $\prod$ describes Cartesian product.

Fuzzy signatures can be considered as special multi-dimensional fuzzy data. Some of the dimensions are inter-related in the sense that they form sub-groups of variables, which jointly determine some feature on a higher level. Here $[x_{11}\ x_{12}]$ form a sub-group that corresponds to a higher level compound variable of x_1. $[x_{221}\ x_{222}\ x_{223}]$ will then combine together to form x_{22} and $[x_{21} [x_{221}\ x_{222}\ x_{223}] x_{23}]$ is equivalent on a higher level with $[x_{21}\ x_{22}\ x_{23}] = x_2$. Finally, the fuzzy signature structure will become $x = [x_1\ x_2\ x_3]$ in the example.

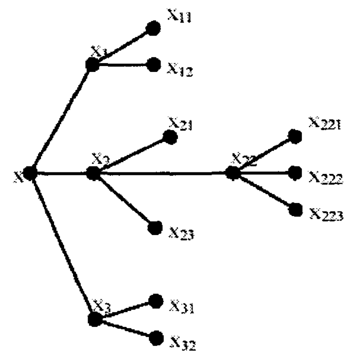

Fig. 1. A Fuzzy Signature

The relationship between higher and lower level is governed by a set of fuzzy aggregations. The results of the parent signature at each level are computed from their branches with appropriate aggregation of their child signatures. Let a_1 be the aggregation associating x_{11} and x_{12} used to derive x_1, thus $x_1 = x_{11}\ a_1\ x_{12}$. By referring to Figure 1, the aggregations for the whole signature structure would be a_1, a_2, a_{22}, and a_3. The aggregations a_1, a_2, a_{22}, and a_3 are not necessarily identical or different. The simplest case for a_{22} might be the *min* operation, the most well known t-norm. Let all aggregations be *min* except a_{22} be the averaging aggregation.

III. Constructing Fuzzy Signature

Let S_{S_0} denote the set of all fuzzy signatures whose structure graphs are sub-trees of the structural ("stretching") tree of a given signature S_0. Then the signature sets introduced on S_{S_0} are defined by

$$A_{S_0} : X \to S_{S_0}.$$

In this case, the prototype structure S_0 describes the "maximal" signature type that can be assumed by any element of X in the sense that any structural graph obtained by a set of repeated omissions of leaves from the original tree of S_0 might be the tree stretching the signature of some A_{S_0}. There are two ways to determine the sub-trees of the fuzzy signature structure, S_0. One way is predetermined by a human expert in the field. Alternatively, the structure of the fuzzy signature can be determined by finding the separability from the data [7,10]. However, in this paper, as we are dealing with a medical diagnosis system, so only the first method will be used.

In the following example, we will show how this can be done. Let us think about some patients, whose daily symptom signatures are based on doctors' assessments according to the following scheme:

$$A_S = \begin{bmatrix} \text{fever} \begin{bmatrix} \text{8 a.m.} \\ \text{12 p.m.} \\ \text{4 p.m.} \\ \text{8 p.m.} \end{bmatrix} \\ \text{blood pressure} \begin{bmatrix} \text{systolic} \\ \text{diastolic} \end{bmatrix} \\ \text{nausea} \\ \text{abdominal pain} \end{bmatrix}.$$

The medical practitioners know that for certain symptoms, they need to check the patient for possible fever, blood pressure, and conditions of nausea or abdominal pain. However, the medical practitioners know that it is fairly important to monitor the possible fever 4 times in a day, and so on. Therefore, the fuzzy signature could be shown as above.

Let us take a few examples with linguistic values and numerical signatures:

$$A_1 = \begin{bmatrix} \begin{bmatrix} \text{none} \\ \text{none} \\ \text{slight} \\ \text{slight} \end{bmatrix} \\ \begin{bmatrix} \text{normal} \\ \varnothing \end{bmatrix} \\ \text{slight} \\ \text{slight} \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.2 \end{bmatrix} \\ \begin{bmatrix} 0.5 \\ \varnothing \end{bmatrix} \\ 0.25 \\ 0.25 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} \begin{bmatrix} \varnothing \\ \varnothing \\ \text{moderate} \\ \text{moderate} \end{bmatrix} \\ \begin{bmatrix} \text{slightly high} \\ \text{rather high} \end{bmatrix} \\ \text{slight} \\ \text{none} \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} \varnothing \\ \varnothing \\ 0.4 \\ 0.4 \end{bmatrix} \\ \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \\ 0.25 \\ 0.0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} \begin{bmatrix} \text{rather high} \\ \text{high} \\ \text{rather high} \\ \text{rather high} \end{bmatrix} \\ \begin{bmatrix} \text{rather high} \\ \text{very high} \end{bmatrix} \\ \text{none} \\ \varnothing \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} 0.8 \\ 0.6 \\ 0.8 \\ 0.8 \end{bmatrix} \\ \begin{bmatrix} 0.8 \\ 1.0 \end{bmatrix} \\ 0.0 \\ \varnothing \end{bmatrix}$$

Of course, normally the blood pressure values would initially rather be expressed by the physician as e.g. 75 / 120, which could then be converted to the linguistic values as appropriate for the patient, taking into account contextual information such as the higher normal blood pressure of infants and children and so on. As for most techniques, there is a significant role for the use of background knowledge of domain experts in data preprocessing.

Note that the structures above are different, which is the point as much real world data is like this. For patient 2, we have only 2 measurements for fever. The structure of the fuzzy signature contains some information by the association of vector components. The use of aggregation operators allows us to compare components regardless of the different numbers of sub-components. Such aggregation operators would in general be designed for each vectorial component with the assistance of a domain expert. In this case, let us assume that the time of day of fever is less significant, and that the maximum value is most important. (In this assumption, the spacing of measurements must therefore ensure a reasonable coverage of the data source.) The three signatures will be reduced to the following forms finally:

$$A_{1f} = \begin{bmatrix} 0.2 \\ \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \\ 0.25 \\ 0.25 \end{bmatrix}, A_{2f} = \begin{bmatrix} 0.4 \\ \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \\ 0.25 \\ 0.0 \end{bmatrix}, A_{3f} = \begin{bmatrix} 0.8 \\ \begin{bmatrix} 0.8 \\ 1.0 \end{bmatrix} \\ 0.0 \\ \varnothing \end{bmatrix}$$

The "fever component" can be verbally rewritten as "slight", "moderate" and "rather high", respectively. The signatures above still contain sufficient information about the "worst case fever" of each patient, while the detailed knowledge of the daily tendency of the fever is lost. This hierarchically structured access to the information is a key benefit of fuzzy signatures.

We could continue this process further completely, and determine an overall "abnormal condition" measure $A_{1o} = [0.25], A_{2o} = [0.4], A_{3o} = [1.0]$.

This example shows the process of converting patients' data into individual fuzzy signatures. After which, by using some fuzzy operation and aggregation of their child signature, the fuzzy signature can give an indication for medical diagnosis about the "abnormal condition" measure for each patient. However, in some cases, where there is new disease that has just been detected like the outbreak of Severe Acute Respiratory Syndrome (SARS) at the beginning of year 2003, a fuzzy signature could be used. Fuzzy signatures in this case, can be constructed by case-to-case basis, and the combined fuzzy signature using some fuzzy aggregation [5] across different fuzzy signatures can be used as knowledge discovery for the disease. The combined fuzzy signature can also be used as a medical pre-clinical diagnosis system. The advantages of using fuzzy signatures for this purpose are they allow modelling of vague information, and in some cases symptoms for each new patient may be different. The other main advantage of using fuzzy signatures for this purpose is that any new information related to the disease can be added easily without the need of any prior structure design. The next section will give an outline of how fuzzy signatures can be constructed from patients' data for a SARS pre-clinical diagnosis system.

IV. EXAMPLE OF SARS PRE-CLINICAL DIAGNOSIS SYSTEM

Severe acute respiratory syndrome (SARS) became an emerging viral infectious disease, since February 2003 when it started in Asia and spread to other countries in North America, South America and Europe. It is a viral respiratory disease that is caused by a kind of coronavirus [11]. To date, the symptoms of SARS are clear and a fuzzy signature for medical diagnosis can be easily formulated by using the method described in section 3 with the vast amount of known symptoms and information. However, when the first few cases occurred in early 2003, it was difficult to have a generalised structural tree for the fuzzy signature, as some information about patients were either redundant for pre-clinical diagnosis, or more information should be included in the pre-clinical diagnosis. This section will present examples of how a medical pre-clinical diagnosis system for a new disease can be constructed.

When the SARS outbreak occurred early 2003, the three countries that were among the first to present reports are Canada, Hong Kong and Singapore [12,13,14]. Fuzzy signatures for each patient can be constructed based on information taken from the reports. After which, three sets of fuzzy signatures each corresponding to each country will be constructed by combining the fuzzy signatures of all patients in that country as: $A_{S(CAN)}$, $A_{S(HK)}$, and $A_{S(SIN)}$

To construct a fuzzy signature from the available data for a patient, and assuming that we have no prior knowledge of the disease, it is safer to include as much information as possible. We can remove them later when we realise that they are not necessary. One of the advantages of fuzzy signatures is in allowing this flexibility, and it can be done easily without a need to re-design the structure of the fuzzy signature.

We will show the construction of a fuzzy signature for patient 1 in Hong Kong here. The information available that may be useful for pre-clinical diagnosis is: Patient 1 is a male patient age 64, without any smoking history. He has symptoms of cough, dyspnea, pleurisy, malaise, myalgia, rigor, fever (over 38°C for 24 hour), and headache [13]. In most cases here, the fuzzy signature will have a fuzzy measure of 0.9 indicating the symptom is present, and it will have a fuzzy measure of 0.1 to indicate a symptom is absent. However, if more information is available, more fuzzy sets can be constructed to better model the nature of the symptoms. For example, fever can be divided into febrile (temperature from 38.4°C to 40°C), low-grade fever (temperature around 37.9°C), and hypothermia (temperature between 35.5°C to 36.5°C). The fuzzy signature for this patient can be arranged in the following structure:

$$A_{S(HK1)} = \begin{bmatrix} Fever \begin{bmatrix} 1st hour \\ 6th hour \\ 12th hour \\ 18th hour \\ 24th hour \end{bmatrix} \\ Sex \\ Age \\ Smoking \\ Cough \\ Dyspnea \\ Pleurisy \\ Malaise \\ Myalgia \\ Rigor \\ Headache \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \end{bmatrix} \\ 1 \\ 0.6 \\ 0.1 \\ 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \\ 0.9 \end{bmatrix}$$

After the fuzzy signature for this first patient has been constructed, fuzzy signatures for the rest of the 9 patients with information available in [13] can also be constructed. For these 10 patients, not all the symptoms presented in $A_{s(HK1)}$ are common to all of them. There are some patients with extra symptoms like sputum production, sore throat, and rhinorrhea, which are absent for patient 1. The use of fuzzy signatures allow complex and incomplete information to be modelled easily. After collecting the information on the 10 patients in Hong Kong, it is realised that the signature features like sex, age, and smoking history may not be too significant in determining the SARS infection. They can then be removed from the fuzzy signature. The final signature structure that best describes the cluster in Hong Kong based on the first 10 patients can be as shown in Figure 2. Fuzzy signatures for the Canada SARS cluster [12] and Singapore cluster [14] can also be constructed similarly. After which, the three main fuzzy signatures $A_{S(CAN)}$, $A_{S(HK)}$, and $A_{S(SIN)}$ can be combined to form a general SARS pre-clinical diagnosis system. The flexibility of fuzzy signature and its non a-priori feature allows any new information to be added easily. For example, it is realised that SARS can be transmitted through close contact with SARS patient, another feature that indicates whether the patient has visited SARS affected countries or was in contact with any SARS patient could be included. Later in April 2003, knowledge about results on different types of thermometry and whether the patient is an adult or child is also available. It is known that this information plays a part in the accuracy of determining fever. For example, an adult consider to have a fever if the following thermometer readings are taken:

- 37.3 and above (oral temperature)
- 37.0 and above (armpit temperature)
- 37.7 and above (ear temperature)

These can be added easily by introducing more levels in the structure of the fuzzy signature under the "Fever" branch as shown in Figure 3.

$$A_{S(HK)} = \begin{bmatrix} Fever \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \\ Cough \\ Sputum \\ Dyspnea \\ Pleurisy \\ Sorethroat \\ Rhinorrhea \\ Malaise \\ Myalgia \\ Rigor \\ Headache \end{bmatrix}$$

Fig. 2. Fuzzy Signature Structure for Hong Kong SARS Cluster

V. Conclusion

We have described a technique for dealing with problems consisting of complex and interdependent features or where data is missing. This was done by the notion of fuzzy signatures, which extends the concept of vectorial fuzzy sets. We have demonstrated with an example the benefit of the hierarchical structuring of data. The hierarchical structuring allows the further use of domain experts, as the information can be abstracted to higher levels analogous to patterns meaningful to human experts. A Severe Acute Respiratory Syndrome (SARS) pre-clinical diagnosis systems using fuzzy signatures is constructed as an example to show the flexibility of the fuzzy signature. The advantages of using fuzzy signatures for knowledge management in this case are its ability to deal with cases without a-priori information, to handle complex structure data, to include evolving information easily, and to handle missing information.

Acknowledgement

Research supported by the Australian Research Council, National Scientific Research Fund OTKA T034233 and T034212, a Main Research Direction Grant 2002 by Széchenyi István University, and the National Research and Development Project Grant NKFP-2/0015/2002.

$$A_{S(HK)} = \begin{bmatrix} Fever \begin{bmatrix} Adult \begin{bmatrix} Oral \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \\ Armpit \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \\ Ear \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \end{bmatrix} \\ Child \begin{bmatrix} Oral \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \\ Armpit \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \\ Ear \begin{bmatrix} 1sthour \\ 6thhour \\ 12thhour \\ 18thhour \\ 24thhour \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ Cough \\ Sputum \\ Dyspnea \\ Pleurisy \\ Sorethroat \\ Rhinorrhea \\ Malaise \\ Myalgia \\ Rigor \\ Headache \end{bmatrix}$$

Fig. 3. Expanded Fuzzy Signature

References

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

[2] M. Sugeno, and T. Takagi, "A New Approach to Design of Fuzzy Controller," *Advances in Fuzzy Sets, Possibility Theory and Applications*, 1983, pp. 325-334.

[3] H.T. Nguyen, and M. Sugeno, (Eds) *Fuzzy Systems: Modeling and Control*, The Handbook of Fuzzy Sets Series, Kluwer Academic Publishers, 1998.

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

[5] L.T. Kóczy, T. Vámos, and G. Biró, "Fuzzy Signatures," *Proceedings of EUROFUSE-SIC'99*, 1999, pp. 25-28.

[6] T.D. Gedeon, L.T. Kóczy, K.W. Wong, and P. Liu, "Effective Fuzzy Systems for Complex Structured Data," *Proceedings of IASTED International Conference Control and Applications (CA 2001)*, 2001, pp. 184-187.

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

[8] J.A. Goguen, "L-fuzzy Sets," *J. Math. Anal. Appl.*, vol. 18, 1967, pp. 145-174.

[9] Kóczy, L. T., "Vectorial I-fuzzy Sets," in: M. M. Gupta & E. Sanchez (eds.): Approximate Reasoning in Decision Analysis, North Holland, Amsterdam, 1982, pp. 151-156.

[10] Chong, A., Gedeon, T.D. and Kóczy, L. T., "Projection Based Method for Sparse Fuzzy System Generation," in *Proceedings of 2nd WSEAS Int. Conf. on Scientific Computation and Soft Computing*, 2002, pp. 321-325

[11] Centre for Disease Control and Prevention, Fact Sheet: Basic Information about SARS, January 13, 2004.

[12] Poutanen, S.M., et al., "Identification of Severe Acute Respiratory Syndrome in Canada," *The New England Journal of Medicine*, 2003, 11pages.

[13] Tsang, K.W. et al., "A Cluster of Cases of Severe Acute Respiratory Syndrome in Hong Kong," *The New England Journal of Medicine*, 2003, 9 pages.

[14] Hsu, L.T. el al., "Severe Acute Respiratory Syndrome (SARS) in Singapore Clinical Features of Index Patient and Initial Contacts," *Emerging Infectious Diseases*, CDC, Vol. 9, No. 6, 2003.