A Type 2 Fuzzy System Modelling Algorithm

Kemal Kılıç
FENS, Sabancı Üniversitesi
İstanbul, Turkey
kkilic@sabanciuniv.edu

Özge Uncu
MIE, University of Toronto, ON, Canada
uncu@mie.utoronto.ca

İ. Burhan Türkşen
MIE, University of Toronto, ON, Canada
uncu@mie.utoronto.ca

Abstract

In this paper, a modified fuzzy system modelling algorithm that incorporates Type 2 fuzzy sets, which is based on interval-valued membership degrees rather than singleton membership degrees, is proposed. The proposed algorithm is evaluated in terms of predictive performance and determination of the significance degrees and compared with other algorithms in the literature, namely Stepwise Multiple Linear Regression (SMLR) and Sugeno-Yasukawa [4] based fuzzy system modelling algorithm, i.e. Turksen-Bazoon (T-B) [3]. A nonlinear function, which is introduced as a benchmarking data set by Sugeno-Yasukawa, is used for validating the models. The proposed algorithm outperformed the other alternatives both in terms of the root mean square error (RMSE) and in terms of the determination of the significance of the inputs. These results showed that the proposed fuzzy system modelling algorithm could effectively approximate nonlinear functions with simple fuzzy if-then rules without assuming a priori structure for the model.

Keywords: Fuzzy System Modelling, Type 2 Fuzzy Sets

1 Introduction

Model building is an essential step in decision making processes in order to explain, predict and control a system. Crucial information, which is necessary for building a realistic model, is usually hidden in historical data. Therefore, data analysis is an important part of model building. Informally, one can define data analysis as a search for structure in data. The data can be viewed as a collection of *ND* (number of data) objects, where each object is represented by means of *NV* (number of variables) attributes. However, unless the structure that is hidden in the system is identified, the data provides very little information. Hence, the objective of data analysis is to bring the hidden structure to the surface.

Fuzzy system modeling emerged as an alternative approach for data analysis and reasoning, and claims to model data sets that have a high level of uncertainty, effectively. One of the advantages of fuzzy system modeling is the fact that it reduces the complexity of the data by using information granularities and presents the data to the user in perceivable fuzzy rules. Therefore, it reduces the uncertainty based on an abundance of information. Furthermore, fuzzy system modeling allows the formation of fuzzy granularities that handle vagueness of the concepts. This is a more realistic approach in a world where there are all shades of grey between black and white.

In fuzzy system modeling, the nonlinear relations in the data are approximated by means of fuzzy if-then rules. In older approaches, the fuzzy if-then rules were determined a priori from other sources such as experts' knowledge. However this approach is highly subjective, i.e., the fuzzy if-then rules usually change from expert to expert, even the same expert may suggest different rules at different times. Therefore there is a growing research domain on modeling approaches for objective identification of the structure in the data in terms of fuzzy if-then rules.

Many different approaches have been developed to date. Generally speaking these algorithms can be classified into two broad approaches in terms of the structure of their rules. In Takagi-Sugeno [5] type modeling approaches, the consequents are functions of the input data, whereas in Sugeno-Yasukawa [4] type approaches the consequents are fuzzy sets. The latter one has the advantage of being more descriptive and easier to implement. A major problem with the former one is the fact that the determination of the parameters associated with this model is computationally costly and the obtained rules are hard to interpret. However the latter one suffers in terms of the predictive performance because of some structural misrepresentations with the existing algorithms. In this paper, a new fuzzy system modeling algorithm has been proposed, which is simple, theoretically clear and attempts to solve the problems associated with the existing algorithms to provide better predictive performance.

In the following section the proposed algorithm will be briefly explained. Later in Section 3, some discussion on Type 2 fuzziness is provided and the way it is incorporated to the proposed algorithm is demonstrated. Section 4 will be the part where the performance of the Type 2 algorithm is demonstrated with a benchmark data set. We will conclude the paper with our final remarks.

2 The proposed algorithm

The proposed fuzzy system modeling algorithm is different from the existing ones both in terms of structure identification and inference. It utilizes the basic premise of the current algorithms, that is to say, "clustering the output data means clustering the input data as well". This is the primary reason for the simplicity of the Sugeno-Yasukawa [4] type modeling algorithms. In Takagi-Sugeno [5] type algorithms, since the consequents of the rules are not fuzzy sets, it is not possible to utilize this idea. However the major drawback of the algorithms based on output clustering is the fact that while projecting the output fuzzy clusters onto input space in order to partition the input variables, they break the natural ties within the input variables and partition each input variable separately. This approach neglects possible correlations among the input variables. The proposed algorithm addresses this problem and solves it by partitioning the input space into n-dimensional clusters. This is achieved by first clustering the output, next projecting the clustered outputs onto n-dimensional input space. Hence a rule structure, which has n-dimensional antecedents, is proposed unlike the existing algorithms. This rule structure keeps the natural ties among the input features. Another advantage of the proposed approach is that it does not assume any pre-specified shape of membership functions such as triangular, trapezoidal, etc. This is important because in the existing algorithms fitting a curve or a line to the projected data points is usually a source of misrepresentation.

Therefore, in the structure identification phase, where the fuzzy if-then rules are determined from the historical data, the first step is to cluster the output variable and then to project these output clusters onto n-dimensional input space. This approach approximates the relations hidden in the data by means of the fuzzy rules. For example, "Those type of objects with such and such values of input variables are related with low values of output", where "such and such" and "low" are fuzzy labels.

The major problem with this rule structure is the determination of the degree of firing in the inference stage. The existing algorithms determine the degree of firing by separately determining the membership degree of each dimension of the data and by conjuncting these membership degrees [4]. However, since the proposed approach has an n-dimensional input rule structure, the degree of firing cannot be determined with the same methods. Therefore a new algorithm to determine the degree of firing has been proposed which is based on a k-Nearest Neighbour (k-NN) algorithm. Note that degree of firing is actually a figure that represents how well the features (input variables) of the test data fits with the rules domain. The proposed k-NN algorithm determines the similarity of the input features of the test data to the domains of the rules. This similarity is measured in terms of a distance metric; in our case a weighted Euclidean distance is proposed where the weights are the significance degrees of each input variable. Please note that this concept will be discussed later in this section.

Output clustering is often achieved by the well known Fuzzy C-Means algorithm proposed by Bezdek [1]**.** However there are some problems that are associated with the application of this algorithm to fuzzy rule base development. These problems are namely, the problems of harmonics, the problems with the boundary fuzzy sets, the problems associated with the classification and the more general problem of cluster validity, i.e. determination of number of clusters (c) and level of fuzziness (m). The main reason for these problems is the fact that, whenever we can order the labels of the fuzzy sets, which is usually the case in single dimensional clustering, we shouldn't allow a data point to be member of more than two consecutive clusters. Otherwise the classification of the intermediate values results in cases where the FCM assumption (for each individual, the sum of the membership degrees to each fuzzy sets is equal to one) is violated. Therefore the solution of the problem may be achieved with a perspective that limits the assignment of each point to only two consecutive fuzzy sets. It may be possible to assign linear membership functions to each output cluster. Note that in this research a Type 2 clustering algorithm is introduced, which assigns interval value membership degrees rather than singleton membership degrees and a system modeling approach is developed that utilizes these interval valued membership degrees. The Type 2 approach will be described in more detail later in Section 3. In terms of the more general cluster validity problem, numerous cluster validity indices are proposed in the literature, which often optimize pre-specified functions. However, the value that optimizes these functions is not necessarily that which optimizes the predictive performance. Therefore a supervised approach, based on modeling error minimization is offered as an alternative.

In data analysis one of the most important steps is the determination of the significant input variables. The proposed algorithm introduces a fuzzy learning based algorithm that determines the significance degrees of the input variables. The existing fuzzy algorithms neglect the fact that it is not always possible to classify an input variable as significant or not. Whereas, it is more suitable to refer to an input as "more significant" or "less significant". That is to say the proposed algorithm fuzzifies the concept of "significant inputs" to "significance degrees of the inputs". The significance degrees of the inputs are determined by a local heuristic algorithm. In this algorithm an initial solution (significance vector) is created first and this solution is updated at each iteration based on the improvement in training error. This algorithm is repeated a number of times with different initial solutions . Then, it is proposed to select the significance vector that minimizes the training error. One may start with random initial solutions or with prespecified initial solutions, such as equal weights or weights based on the correlation of the input variables with respect to the output, etc. In order to avoid myopic local entrapments a probabilistic neighbourhood selection methodology is proposed. That is to say, while updating the significance vector at each iteration rather than the best neighbourhood in terms of the improvement of the training error it is suggested to first determine a set of ‘promising’ significance vectors (again in terms of the improvement in the training error) and jump to the neighbour randomly among the ones in the ‘promising set’. Note that the performance of the significance determination algorithm is assessed in Section 4 where two dummy variables are introduced to the nonlinear system.

To sum up, the output is clustered and projected onto n-dimensional input space and fuzzy rules are constructed. These rules are fired by means of the k-NN algorithm where the degree of firing is calculated in terms of the similarity of the test data and the domain of the rule. The similarity measure is based on the weighted Euclidean distance metric where the weights are the significance of each input variable. These significance figures are obtained via a fuzzy learning algorithm described above briefly. Readers may find more details of the algorithm proposed by Kilic in [3].

3 Type 2 Fuzziness

The existing algorithms in the literature mostly deal with Type 1 uncertainty. That is to say, a point membership degree is assigned to the elements of the fuzzy sets. However, there is further uncertainty associated with the determination of this membership degree. Therefore, in some cases it may be beneficial to develop Type 2 modeling. In this paper an approach that is based on interval valued membership degrees proposed by Uncu *et al.* [6, 7] is incorporated with the above proposed algorithm.

In this approach, the uncertainty associated with the determination of the level of fuzziness (m) is utilized. Recall that m is a parameter in FCM that specifies the degree of overlapping among the fuzzy sets. As $m \rightarrow \infty$ the algorithm performs in a totally fuzzy environment and the membership degrees of all points to all clusters becomes ($1/c$). As $m \rightarrow 1$ the algorithm behaves as a hard clustering algorithm and provides crisp membership degrees, i.e., each point is a member of only one cluster with full membership degree. That is to say, a single membership degree would be obtained, for different selections of level of fuzziness. The rule of thumb is selection of $m=2$. While modeling a system or analyzing the patterns in the data on hand, one cannot be sure about the behaviour of the environment. Certain systems may be less fuzzy than others. Therefore an uncertainty is hidden in this decision. Different values of m will lead to different modeling. Based on the work of Uncu [6, 7] one can take advantage of such an uncertainty and build a Type 2 fuzzy set at this level. In the proposed approach rather than specifying $m=2$, we propose to generate an interval valued membership where the boundaries are obtained by selecting a lower and upper level of fuzziness. Note that the lower level of fuzziness must be low enough to capture the hard clustering tendency of FCM and upper level of fuzziness must imitate the behaviour of $m\rightarrow\infty$. In this study we suggest to select $m^L=1.5$ and $m^U=3.5$ which also covers the rule of thumb value of 2.

In Figure 1, a Type 2 fuzzy cluster with $m^L=1.5$ and $m^U=3.5$ is presented. As it can be observed the S-curve shaped membership functions are obtained. The fuzzy sets represented by the dashed lines are obtained with $m=1.5$, and the solid lines are with $m=3.5$. As it can be observed as m decreases the fuzzy sets obtained becomes more crisp.

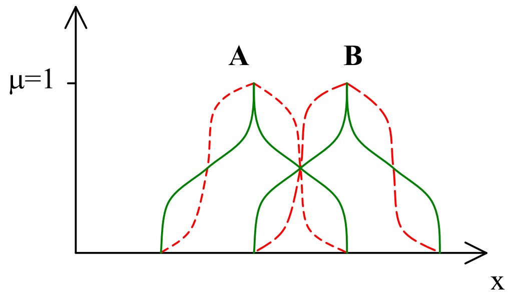

Figure 1: Type 2 Fuzzy Sets

4 Experimental Analysis

The performance of the proposed algorithm is tested with the nonlinear two-input single-output system that is introduced by Sugeno and Yasukawa [4] and has been used as benchmark data for the various modeling and data analysis applications. The nonlinear equation that relates the output, y, to the inputs, x_1 and x_2, is as follows,

$$y=(1+x_1^{-2}+x_2^{-1.5})^2 \qquad 1\leq x_1, x_2\leq 5$$

Ten experiments were constructed in order to conduct a statistical analysis. For each experiment ninety x_1, x_2 pairs were generated randomly and the corresponding y value calculated. In order to test the performance of the proposed significance determination algorithm, randomly generated dummy x_3 and x_4 input variables were added to the data set. Hence, a four-input single output data set was used for the experiments. The proposed fuzzy system modeling algorithm was compared with the Turksen-Bazoon [3] model (denoted as T-B in Table 2), which is based on Sugeno-Yasukawa [4] approach and Stepwise Multiple Linear Regression (denoted as SMLR in Table 2). Sixty data vectors were used to build the fuzzy rule base (the training set) and the remaining thirty data vectors were used as the test set.

The significance degrees determined by the proposed Type 2 algorithm from 10 experiments are tabulated in Table 1. On average, the significance degrees of the input variables obtained from 10 experiments are (0.48, 0.46, 0.03, 0.03) for x_1, x_2, x_3, and x_4, respectively. The overall results demonstrate that the proposed algorithm performed well in terms of significance degree determination for this experiment and correctly identified the dummy variables.

Table 1: Significance Degrees obtained for 10 Experiments

Exp. No	x_1	x_2	x_3	x_4
1	0.37	0.60	0.00	0.04
2	0.44	0.52	0.01	0.02
3	0.63	0.34	0.01	0.01
4	0.48	0.43	0.05	0.04
5	0.63	0.34	0.03	0.00
6	0.47	0.48	0.02	0.03
7	0.46	0.51	0.02	0.01
8	0.63	0.32	0.04	0.02
9	0.39	0.48	0.11	0.02
10	0.32	0.55	0.00	0.13
Average	0.48	0.46	0.03	0.03

The predictive performance of the proposed algorithm is evaluated in terms of the root mean square error and tabulated in Table 2. On the average the proposed algorithm performed better than the other two algorithms. It was statistically better than Turksen-Bazoon algorithm ($p \leq 0.02$) but not different than SMLR.

Table 2: RMSE results of the 10 experiments for the algorithms

Exp. No	T-B	SMLR	Type2
1	0.67	0.40	0.40
2	0.63	0.47	0.36
3	0.91	0.73	0.65
4	0.61	0.47	0.40
5	0.77	0.62	0.53
6	0.69	0.59	0.64
7	0.67	0.53	0.40
8	0.76	0.63	0.58
9	0.78	0.74	0.76
10	0.60	0.37	0.31
Average	0.71	0.55	0.51

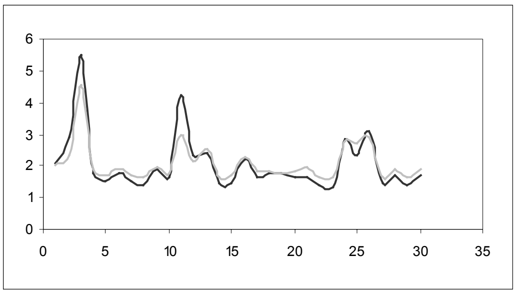

Figure 2: Prediction (gray line) vs. actual (black line) output graph for the test data

In Figure 2, for a sample experiment, actual versus prediction graph is presented for the proposed Type 2 algorithm. The depicted figures and tables demonstrate that the proposed algorithm is really a promising one.

5 Conclusion

Fuzzy system modelling has the potential to effectively approximate nonlinear and complex relations in terms of the fuzzy if-then rules. One of the challenges is the determination of the underlying fuzzy if-then rules that govern the system objectively. Many different approaches have been developed to date. In this paper we proposed a new approach, which may be classified as a Sugeno-Yasukawa type algorithm. However, the proposed algorithm is different from the ones that exist in the literature both in terms of the structure identification and the inference phases. It bears the advantage of simplicity which is common for Sugeno-Yasukawa type fuzzy system modelling algorithms. Furthermore, the proposed system modelling technique proved to have better predictive performance.

References

[1] J. C. Bezdek (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms,* Plenum Press.

[2] K. Kilic (2002). A Proposed Fuzzy System Modelling Algorithm with an Application in Pharmacokinetic Modelling. PhD Thesis, University of Toronto, 2002.

[3] B. A. Sproule, M. Bazoon, K.I. Shulman, I. B. Turksen and C. A. Naranjo (2000). Fuzzy Logic Pharmacokinetic Modelling: Application to Lithium Concentration Prediction. In *Clinical Pharmacology Therapy,* volume 62, pages 29-40

[4] M. Sugeno and T. A. Yasukawa (1993). A Fuzzy Logic Based Approach to Qualitative Modelling. In *IEEE Transactions on Fuzzy Systems,* volume 1, pages 7-31.

[5] T. Takagi and M. Sugeno (1985). Fuzzy Identification of Systems and Its Application to Modelling and Control. In *IEEE Transactions on Systems, Man and Cybernetics,* volume 15, pages 116-132.

[6] O. Uncu, I.B. Turksen and K. Kilic (2003). LOCALM-FSM: A New Fuzzy System Modelling Approach Using a Two-Step Fuzzy Inference Mechanism Based on Local Fuzziness Level. *International Fuzzy Systems Association World Congress, IFSA*, Istanbul, Turkey, 2003 (accepted)

[7] O. Uncu and I.B. Turksen (2003). A New Two-Step Fuzzy Inference Approach Based on Takagi-Sugeno Inference Using Discrete Type 2 Fuzzy Sets. *NAFIPS,* Chicago, USA, 2003 (accepted)