

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/43980503>

Improvement of the Cluster Searching Algorithm in Sugeno and Yasukawa's Qualitative Modeling Approach

Conference Paper in Lecture Notes in Computer Science · October 2001

DOI: 10.1007/3-540-45493-4_54 · Source: OAI

CITATIONS

22

READS

75

5 authors, including:



Kok Wai Wong

Murdoch University

253 PUBLICATIONS 3,196 CITATIONS

[SEE PROFILE](#)



Laszlo T. Koczy

Budapest University of Technology and Economics

454 PUBLICATIONS 5,453 CITATIONS

[SEE PROFILE](#)



Tom Gedeon

Australian National University

430 PUBLICATIONS 5,943 CITATIONS

[SEE PROFILE](#)



Domonkos Tikk

Gravity Research & Development Zrt

191 PUBLICATIONS 4,331 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Hand written character classification [View project](#)



Mineral Prospectivity Mapping [View project](#)

Improvement of the Cluster Searching Algorithm in Sugeno and Yasukawa's Qualitative Modeling Approach

K.W. Wong¹, L.T. Kóczy^{1,2}, T.D. Gedeon¹, A. Chong¹, D. Tikk²

¹School of Information Technology
Murdoch University
South St, Murdoch
Western Australia 6150
{ k.wong | t.gedeon | cchong } @murdoch.edu.au

²Department of Telecom. & Telematics
Budapest University of Technology and Economics
H-1117 Budapest, Pázmány sétány 1/d
Hungary
{ koczy | tikk } @ttt.bme.hu

Abstract. Fuzzy modeling has become very popular because of its main feature being the ability to assign meaningful linguistic labels to the fuzzy sets in the rule base. This paper examines Sugeno and Yasukawa's qualitative modeling approach, and addresses one of the remarks in the original paper. We propose a cluster search algorithm that can be used to provide a better projection of the output space to the input space. This algorithm can efficiently identify two or more fuzzy clusters in the input space that have the same output fuzzy cluster.

1 Introduction

Fuzzy modeling has become very popular because of the main feature of its ability to assign meaningful linguistic labels to the fuzzy sets [1] in the rule base [2,3]. Sugeno and Yasukawa's qualitative modeling (SY) method [4] has gained much attention in the fuzzy research field mainly due to its advantage of building fuzzy rule bases automatically from sample input-out data. The fuzzy rule bases extracted by the SY method are sparse fuzzy rule bases, i.e., where there are "gaps" among the rules, so that these regions can be interpolated from the remaining areas and rules [5,6,7]. In our approach, we intend to extend this method – with the necessary alterations and additional steps. The usual fuzzy controller identification methods generate dense fuzzy rule bases, so that the rule premises form a fuzzy partition of the input space. In a dense fuzzy rule base, the number of rules is very high, as it depends on the number of inputs k and the number of partitions per variable T in an exponential way. Assuming all the partitions are consistent in all premises and consequents, the total number of rules is $|R| = O(T^k)$. In order to avoid this exponential number of rules, the SY method puts emphasis on the rule consequents, i.e., the output space, and first finds a partition in Y . The determination of premises in the input space X is done by splitting appropriately the inverse images of the output clusters. Using this approach,

the partitioning of the input space is done in a secondary manner, thus the number of fuzzy rules does not increase exponentially with the number of inputs.

One of the important remarks made in the paper by Sugeno and Yasukawa [4], is the condition under which more than one fuzzy cluster can be found in the input space, so that they correspond to the same output cluster. It is made clear that special care has to be taken to form two or more convex input clusters. However, the details of this particularly important step are not given in any detail in the paper. With regards to this remark, it has to be stated that two conditions have to be considered when proposing an algorithm that could efficiently handle the problem. First, the algorithm has to be able to identify the occurrence of several (at least two) rules in one fuzzy output cluster, i.e., the presence of more than one corresponding fuzzy cluster in the input space. Second, the algorithm should be able to decide whether these input fuzzy clusters identified can be merged or any of them can be discarded. This paper shows a detailed analysis of these two conditions and proposes a cluster search algorithm that can address this problem efficiently.

2 The Sugeno and Yasukawa's Qualitative Modeling Method

In a given data set, with k inputs, the given input-output data pairs for n patterns are:

$$(x_1^i, x_2^i, x_3^i, \dots, x_k^i; y^i) \\ \text{where } i = 1, 2, 3, \dots, n$$

The SY method [4] is able to perform two main steps, *identification* and *qualitative modeling* in obtaining a fuzzy model as:

$$R^i : \text{if } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i \dots \text{ and } x_n \text{ is } A_n^i \text{ then } y \text{ is } B^i \quad (1)$$

In the identification step, it can be subdivided into *structure identification I and II*, and *parameter identification*. In structure identification I, the main purpose is to find appropriate input candidates and input variables in building the model. Structure identification II is concerned with the input-output relations by concentrating on the number of rules and partitions of the input space. The parameter identification step is basically used to tune the parameters in the membership functions of the fuzzy sets. Finally, linguistic labels can be assigned to the finalized fuzzy sets in the rule base. In this paper, we are mainly concerned with the structure identification II stage.

In order to extract a fuzzy rule base, the SY method uses two distinctive characteristics. First, it partitions the consequents of the rule and then finds a relationship concerning the premises. It is important that it does not use an ordinary fuzzy partition of the input space as shown in Figure 1. In order to satisfy the first characteristic, it makes use of the fuzzy means method (FCM) [8] to search for the fuzzy clusters in the output space using all available data. When determining the best clusters in the output space of the following selecting criterion [4] is used:

$$S(c) = \sum_{k=1}^n \sum_{i=1}^c (\mathbf{m}_{ik})^m (\|x_k - v_i\|^2 - \|v_i - \bar{x}\|^2) \quad (2)$$

where

- n : number of data to be clustered;
- c : number of clusters, $c \geq 2$;
- x_k : k th data;
- \bar{x} : average of data;
- v_i : vector expressing the center of the i th cluster;
- $\|\bullet\|$: norm;
- \mathbf{m}_{ik} : grade of the k th data belonging to the i th cluster
- m : adjustable weight

The best cluster is determined when the selecting criterion $S(c)$ reaches a minimum when the number of clusters, c , increases. The final results of the clustering have the relation between the grade of membership belonging to a fuzzy cluster B for every output y :

$$y^i \text{ in } B^j (1 \leq j \leq c):$$

$$(x^i, y^i), B^1(y^i), B^2(y^i), \dots, B^c(y^i) \quad (3)$$

From the output fuzzy cluster B , a fuzzy cluster A is induced in the input space as shown in Figure 2. If there are two input dimensions x_1 and x_2 , the fuzzy cluster A can then be projected to the axes x_1 and x_2 as shown in Figure 3.

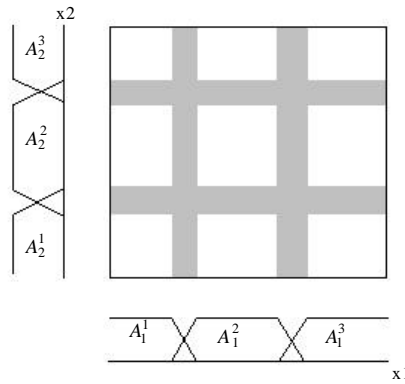


Fig. 1. Ordinary fuzzy partition of the input space

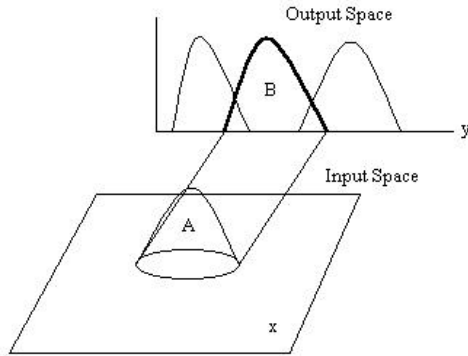


Fig. 2. Fuzzy cluster *A* from output cluster *B*

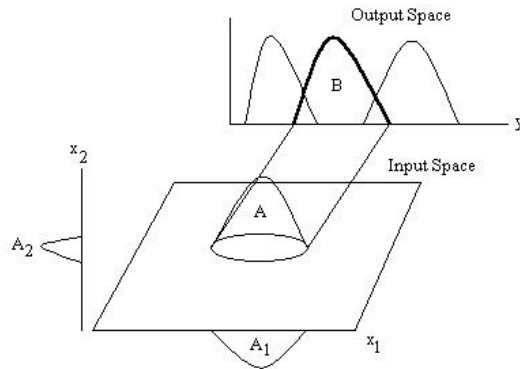


Fig. 3. Fuzzy cluster *A* for two input dimensions

From these cluster relations, this can be formulated as follows:

$$A_1(x_1^i) = A_2(x_2^i) = B(y^i) \quad (4)$$

From this output cluster, a fuzzy rule is generated as:

if x_1 is A_1 and x_2 is A_2 , then y is B .

3 The Problem of Clustering Input Space

In the original approach, the output cluster is typically used to obtain the corresponding (single) input cluster as well, by projecting the output cluster onto the input space. However, in cases where there is more than one cluster in the input space corresponding to the same output cluster, e.g. A^i to A^j corresponding to the same

output cluster B , a further step has to be done in order to separate them. The SY method suggests very generally that in the case of two (or more) input fuzzy clusters, two (or more) convex fuzzy clusters should be “formed carefully” in the input space as shown in Figure 4 and Figure 5. However, the details of the method are not shown in the paper as to how this can be done “carefully”. The first problem that we address here is how to separate two or more input clusters and corresponding antecedents, especially in multi-dimensional cases.

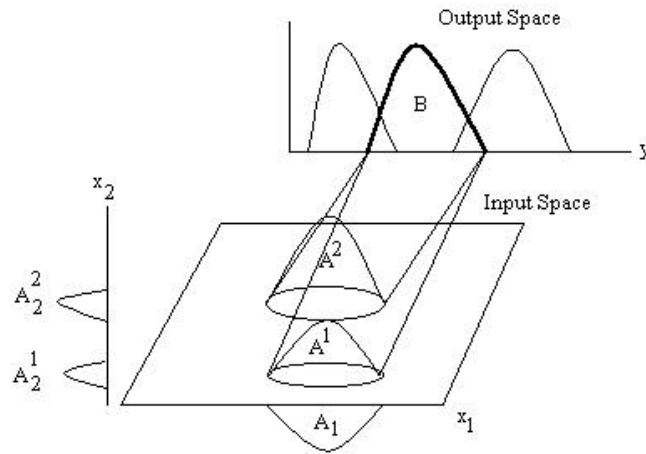


Fig. 4. Two fuzzy clusters for one output cluster.

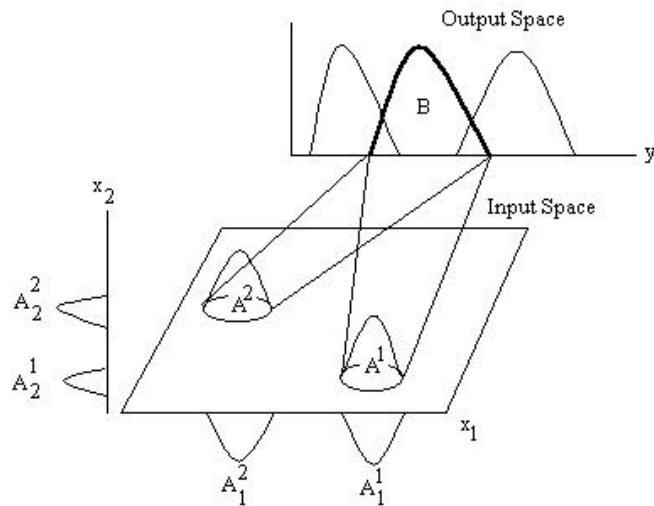


Fig. 5. Another example of two fuzzy clusters for one output cluster.

Besides identifying the fuzzy clusters that may correspond in the input space to a specific output fuzzy cluster, some other factors need to be considered. They can be classified in two main areas as follows:

(1) Merging of Input Clusters

After identifying the possible clusters in the input space, the next question is whether all the identified clusters are necessary to produce a reasonable model. Depending on the distribution of the sample data as well as the nature of the problem, the seemingly separate two or more input clusters may appear just because there is a slight decrease between two peaks, indicating a lower frequency of data there, due to the uneven distribution of the available data. In this case, separating the clusters may not really improve the model, but may increase the number of fuzzy rules constructed.

When sampling the data, noise may accidentally be included in the given input-output sample. Therefore it is not possible to define an exact model that describes the relationship between X and Y when noise exists in the input-output pair. However, a probabilistic relationship governed by a joint probability law $P(\mathbf{n})$ can be used to describe the relative frequency of occurrence of input-output pairs (x^i, y^j) for n training patterns. The joint probability law $P(\mathbf{n})$ can be further separated into an environmental probability law $P(\mathbf{m})$ and a conditional probability law $P(\mathbf{g})$. For notation expression, the probability law is expressed as:

$$P(\mathbf{n}) = P(\mathbf{m})P(\mathbf{g}) \quad (5)$$

The environmental probability law $P(\mathbf{m})$ describes the occurrence of the input X . The conditional probability law $P(\mathbf{g})$ describes the occurrence of the Y based on the given input X . An input-output pair (X, Y) is considered to be noise if X does not follow the environmental probability law $P(\mathbf{m})$, or the output Y based on the given X does not follow the conditional probability law $P(\mathbf{g})$. In this case, even when the input clusters have been identified, it may be a noise motivated clustering result. In this case, using all the clusters may not really improve the model, but may increase the number of fuzzy rules constructed and thus the computational complexity.

(2) Computational Efficiency

The main advantage of the SY method as compared to traditional methods is that it works on the consequents rather than the premises. This is very computationally efficient. Besides, after the fuzzy model; which is a sparse fuzzy rule base; has been built, the model does not require intensive computation to generate an answer, as the number of fuzzy rules is small for most cases. Care must be taken when constructing an algorithm to identify the input fuzzy clusters within the output cluster. The algorithm has to be computationally efficient and at the same time does not increase the number of fuzzy rules generated tremendously.

4 Identification of the Clustering in the Input Space

Cluster the output as per the SY method using fuzzy means. In all our analysis here, we are assuming that the distribution in each cluster can be approximated by a normal distribution.

For cluster C_i :

(1) Random selection

In each cluster, the population (U_i) of that cluster consists of data with membership not smaller than 0.2:

$$U_i = \{x_j \mid m_j \geq 0.2, \text{ where } j = 1, \dots, N\} \quad (6)$$

where N is the number of data in cluster i .

Those data that have membership grades smaller than 0.2 are left out, and are considered to be insignificant in contributing to the features of the cluster. When dealing with very large number of data, which is common in most real world applications, if we are to project all the output points back to the input space, this may be very time consuming. The total projections, p required are:

$$p = \sum_{i=1}^c (N_i) \quad (7)$$

and $p > n$

where N_i is the number of data in cluster i

We propose here to project a subset of N by using sampling distribution theory, being s . Regardless the distribution of the data in the cluster, s is always smaller than N [9].

When performing random selection, it is always necessary to ensure that the sample contains the major information in the population, U_i . The standard of measure is by using sampling distribution theory [10,11]. The basic idea is to ensure that the selected data can approximate the population in the cluster (U_i), by investigating the mean (\bar{u}_i), variance ($\mathbf{s}_{\bar{u}}^2$) and standard deviation ($\mathbf{s}_{\bar{u}}$) of the population.

For random sampling without replacement in a finite population, the variance ($\mathbf{s}_{s_i}^2$) and standard deviation (\mathbf{s}_{s_i}) of the sample are as follows (with the assumption that $s > 0.05N$):

$$\mathbf{s}_{s_i}^2 = \frac{\mathbf{s}_{\bar{u}}^2}{s} \cdot \frac{N-s}{N-1} \quad (8)$$

$$s_{s_i} = \frac{s_u}{\sqrt{s}} \cdot \sqrt{\frac{N-s}{N-1}} \quad (9)$$

where $\frac{N-s}{N-1}$ is the finite population correction factor.

However, according to the Central Limit Theorem, the finite population correction factor can be ignored if and only if $s \leq 0.05N$.

If the sample (S_i) we have selected is a good representation of the population (U_i) in the cluster, then the sampling error should be small. Sampling error (E) is the measure of the difference between the sample average and the population average:

$$E = \bar{u}_i - \bar{s}_i \quad (10)$$

In order to obtain a better random sample for use in the projection, three random data sets are obtained and their sampling errors are compared. The best sampling error of the three are then selected for use in the next step. However, the next issue we have to deal with is the size of s . In this paper, we make use of the mean as a guideline. The size of sample, s , required for accuracy of s_u , A , for a confidence level, l are:

$$s = \frac{s_u^2}{\frac{A^2}{Z^2} + \frac{s_u^2}{N}} \quad (11)$$

where Z is the z-value corresponding to the desired confidence level, l

As we are using 5% accuracy and a confidence level of 95%, s can be simplified to:

$$s = \frac{s_u^2 N}{0.00065 N + s_u^2} \quad (12)$$

As the denominator is a factor of 0.00065, s will be very small compared to N if N is very large, thus the computational complexity of projecting p will be reduced.

(2) Partitioning of input

After the random sample s has been selected, partitioning of the input variables into intervals needs to be carried out in order to identify the clusters in the input space. In this paper, we make use of the *partial completeness measure* K to account for the information lost by partitioning [12]. The decisions on how to partition an input variable and the number of partitions required depends on how this partial completeness measure reflects the information lost. Information lost while partitioning mainly occurs in the next stage. The main reason for information loss is

due to combining the adjacent intervals in searching for the normal distributed clusters. Besides, the weighting function in discarding any rules as noise is also implied by the partitions, this will in turn cause information to be lost. Let Ru be the set of rules obtained by considering all ranges over the variables, then Ru' be the set of rules obtained by considering all ranges from the partitioning of the variables. When Ru is transformed into Ru' , information is lost. The partial completeness measure is basically used to measure how far or close is Ru' is to Ru .

The partial completeness measure used to identify the number of intervals has direct effect from the human's decision on the confidence level and support level for each rule. It is known that in order to guarantee close rules will be generated, the minimum confidence must be set to $1/K$ times the desired level [12]. It is also shown that equi-depth partitioning minimizes the partial completeness level for any specified number of intervals.

After the analysis, the number of partitions I required can be calculated by the formula:

$$\text{No. of Partitions } (I) = \frac{2k}{\min \text{ sup}(K-1)} \quad (13)$$

where k is the number of input variables,
 $\min \text{ sup}$ is the minimum support (as a fraction) specified
 by the user,
 and K is the partial completeness level.

It is important to have the partitions as small as possible in order to obtain a constructed rule base Ru' that is as close to Ru as possible, but this will effectively increase the number of partitions. Care needs to be taken while selecting the appropriate partial completeness level, which is normally chosen to be between 1.5 and 5 [12]. If the number of partitions for an input variable is large, the support for a partition may be low. If the support for a partition is too low and just happens to be away from the searched distribution as shown in Figure 6, it will be treated as an outlier and be discarded while forming the cluster. In order to solve this problem, we introduce an algorithm for combining adjacent partitions while searching for input clusters. The details of the combination algorithm will be discussed in the next section.

(3) Identifying input clusters

In this stage, the s projections are made and the counter $Count_{Bin_j}$ for each bin Bin_j are also set when it is hit by the projection. Bins refer to the intervals of the partitions. Figure 7 shows the bins and counters for a one-dimensional input space and the projections from an output cluster.

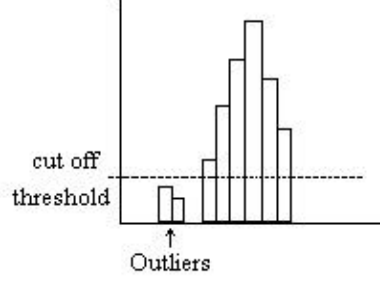


Fig. 6. Partitions with support less than cut off threshold are treated as outlier.

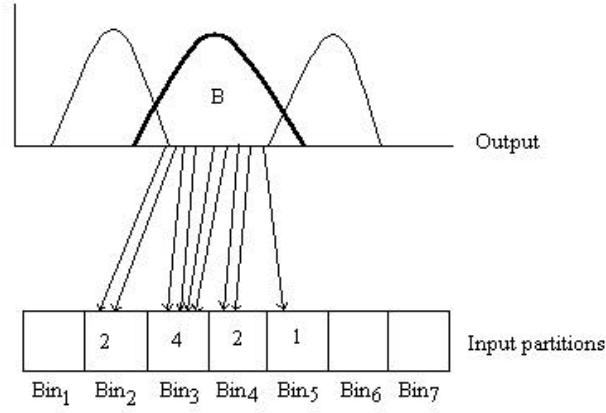


Fig. 7. Example of the projections and counters in the bins.

Beside the counters, a relation is formed by all the input variables and the output. Therefore, a total of s relations are formed. The relations are constructed as:

$$\{[x_1^i(Bin_{1j}), x_2^i(Bin_{2j}), x_3^i(Bin_{3j}), \dots, x_k^i(Bin_{kj}); C_i]\} \quad (14)$$

where $i = 1, 2, 3, \dots, s$

and $j = 1, 2, 3, \dots, I$

Based on the counter value in each bin, they are recombined in order to search for fuzzy sets in the input space. The combination algorithm is as follow:

A. Identifying the centre of distribution

By moving from left to right along the bins, the counter values are compared. If $\max(\text{Count}_{Bin_{j-1}}, \text{Count}_{Bin_{j+1}}) < \text{Count}_{Bin_j}$, then Bin_j is the centre of the

distribution. Figure 8 shows an example of the centres of the distribution with the circled bins with highest hits, i.e. centres of the distribution.

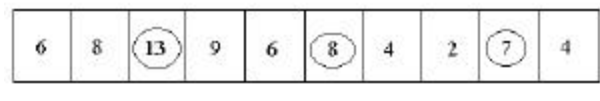


Fig. 8. Illustration of centre identification.

B. Distance measure.

Find the distance between the left and the centre of the distribution:

$$dist_L = Count_{Bin_j} - Count_{Bin_{j-1}} \quad (15)$$

If $dist_L$ is small (below a threshold) the two bins are combined into one.

The distance between the right and the centre of the distribution:

$$dist_R = Count_{Bin_{j+1}} - Count_{Bin_j} \quad (16)$$

Similar to the left hand side, if $dist_R$ is below the threshold, the two bins are combined. Figure 9 shows how they are combined. The above two steps are repeated until the two neighbouring distance measures are above the threshold.

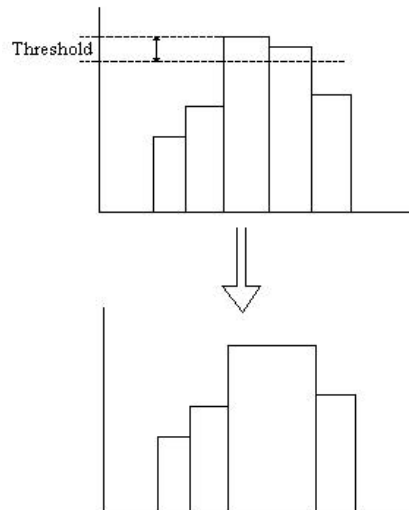


Fig. 9. Combination of bins.

With these normal distributions found in the input space, input clusters are constructed. Trapezoidal approximations based on Ruspini partitions [13] are then used to convert the input clusters to trapezoidal fuzzy membership functions. Basically, the condition below is met when performing the trapezoidal approximations:

$$\sum_{i=1}^T A_i(x) = 1 \quad (17)$$

Trapezoidal approximations based on Ruspini partitions are illustrated in Figure 10.

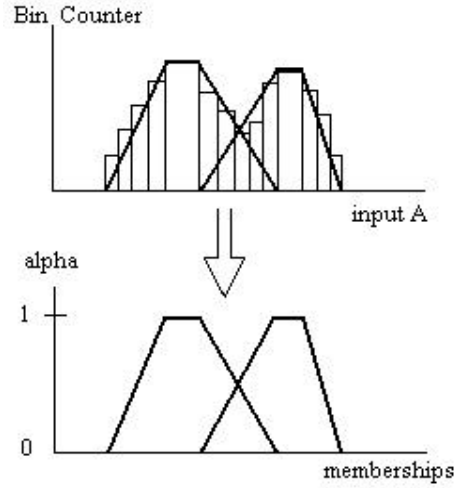


Fig. 10. Approximating trapezoidal fuzzy membership functions.

The bins information in the relation is then substituted with their respective fuzzy membership information. Any repetition relations are then discarded. Relations with small weighting functions or support level are treated as noise and are also discarded.

(4) Constructing fuzzy rules

For this output cluster with the relation information of the input space, fuzzy rules can be constructed. If any fuzzy memberships are found to be next to each other, they are merged into one fuzzy membership as shown in Figure 11. A prediction index (PI) [4] is then calculated based on the predicted output and observed output. The prediction index (PI) [4] is:

$$PI = \sum_{i=1}^n \left[\frac{(y^i - o^i)^2}{n} \right] \quad (18)$$

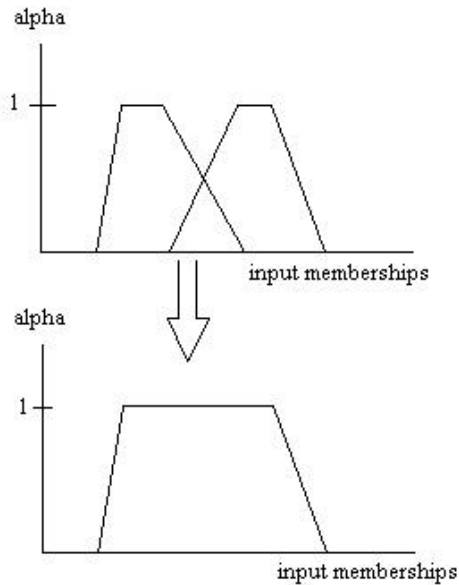


Fig. 11. Merging of fuzzy membership functions.

If it does not affect the PI, they will remain merged. This merging of fuzzy rules will help to reduce the number of fuzzy rules extracted.

5 Conclusion

This paper has examined the Sugeno and Yasukawa's qualitative modeling (SY) method [4]. The popularity of the SY method is mainly due to its rule extraction algorithm. SY method finds the partition in the output space and then projects back to the input space in searching for the input partitions. This results in a sparse fuzzy rule base. This method also avoids the possibility of exponential growth of the fuzzy rule base. One of the important remarks made in the paper [4] is the condition under which more than one fuzzy cluster can be found in the input space that corresponds to the same output cluster. This paper has examined this issue and formulated an algorithm to address it. There are two conditions that are taken into consideration while formulating the approach. First, the algorithm has to be able to identify the occurrence of several rules in one fuzzy output cluster, i.e., the presence of more than one corresponding fuzzy cluster in the input space. Second, the algorithm should be able to decide whether these input fuzzy clusters identified can be merged or any of them can be discarded. The main objective of our approach was to formulate an algorithm that preserves the advantages of the original SY method, which are computational efficiency and small number of fuzzy rules produced.

6 References

- [1] Zadeh, L.A. (1968) "Fuzzy Algorithm", *Information and Control*, vol. 12, pp. 94-102.
- [2] Sugeno, M., and Takagi, T. (1983) "A New Approach to Design of Fuzzy Controller," *Advances in Fuzzy Sets, Possibility Theory and Applications*, pp. 325-334.
- [3] Nguyen, H.T., and Sugeno, M. (Eds) (1998) *Fuzzy Systems: Modeling and Control*, The Handbook of Fuzzy Sets Series, Kluwer Academic Publishers.
- [4] Sugeno, M., and Yasukawa, T. (1993) "A Fuzzy Logic Based Approach to Qualitative Modeling," *IEEE Transactions on Fuzzy System*, vol. 1 no. 1, pp. 7-31.
- [5] Kóczy, L.T. and Hirota, K., (1993) "Approximate reasoning by linear rule interpolation and general approximation," *Int. J. Approx. Reason*, Vol. 9, pp.197-225.
- [6] Gedeon, T.D. and Kóczy, L.T., (1996) "Conservation of fuzziness in rule interpolation," *Intelligent Technologies, International Symposium on New Trends in Control of Large Scale Systems*, vol. 1, Herlany, pp. 13-19.
- [7] Tikk, D., and Baranyi, P., (2000) "Comprehensive analysis of a new fuzzy rule interpolation method," *IEEE Trans. on Fuzzy Systems*, vol.8 no. 3, pp. 281-296.
- [8] Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press.
- [9] Wadsworth, G.P., and Bryan, J.G., (1974) *Applications of Probability and Random Variables*, Second Edition, McGraw-Hill.
- [10] Cochran, W.G. (1977) *Sampling Techniques*, Wiley.
- [11] Anderson, T.W. (1996) *The new statistical analysis of data*, Springer
- [12] Srikant, R., and Agrawal, R., (1996) "Mining Quantitative Association Rules in Large Relational Tables," *Proceedings of ACM SIGMOD Conference on Management of Data*, Montreal, Canada, pp. 1-12.
- [13] Ruspini, E.H., (1969) "A new approach to clustering," *Information and Control*, vol. 15, pp. 22-32.