Paper:

# A Proposal of Visualization Method for Interpretable Fuzzy Model on Fusion Axes

**Kosuke Yamamoto, Tomohiro Yoshikawa, and Takeshi Furuhashi**

Department of Computational Science and Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
E-mail: yamamoto@cmplx.cse.nagoya-u.ac.jp, tom@fcs.coe.nagoya-u.ac.jp, furuhashi@cse.nagoya-u.ac.jp

[Received July 1, 2005; accepted September 13, 2005]

**Interpretability of fuzzy models has become one of the major topics in the field of fuzzy modeling. Visualization that makes input-output relationships interpretable is effective in extracting useful knowledge from unknown data. This paper presents visualization method that considers the visibility of fuzzy models. This method identifies clusters that have different statistical features, and projects the data to the "fusion axes", which are linear combinations of the multiple input variables, considering the distribution of each cluster in the projected space. This paper applies the proposed method to artificial data and also to collected data from the mobile robot, and shows that the proposed method can extract useful knowledge from the obtained visible and interpretable models.**

**Keywords:** interpretability, visualization, fusion axis, fuzzy modeling

## 1. Introduction

Interpretability of fuzzy model has become one of the major topics in the field of fuzzy modeling [1]. Many methods that construct interpretable fuzzy models from numerical data have been proposed [2–5]. So far, interpretability that has been discussed is related to shape and degree of overlapping of membership functions, the number of input variables, and/or the number of linguistic terms. These parameters are considered to be essential for the quality of linguistic expression of fuzzy models. For the interpretability, we have emphasized the importance of visualization of the model. A visible model can give human a clear image of input–output relationships, and it helps to interpret the data. This technique is often used in the field of data-mining [6]. However, there had been few methods that sought an interpretable fuzzy model from the viewpoint of visualization. We have proposed a fuzzy modeling in a visible space with "fusion axes" which are linear combinations of multiple input variables [7]. The purpose of this study is to extract simple structures in the nonlinear data to interpret their features through the visible model.

Projection of high-dimensional data can reduce its dimensionality and make it visible. Criteria for the projection are vital for the visible and interpretable model. In [7], we defined the criteria of projection as follows: (i) features of the system should be preserved, and (ii) each feature in the projected space should be separated as wide as possible from others. We employed Fuzzy C-Means (FCM) [8, 9] for finding the features of data, and Fuzzy Multiple Discriminant Analysis (FMDA) for identifying the fusion axes. The FMDA is able to identify the fusion axes that maximize the distance among clusters (features of data). That facilitates allocation of membership functions on the fusion axes. In [7], we showed that the proposed method was able to construct a visible and interpretable fuzzy model with a clarified structure of Box and Jenkins' gas furnaces data [10]. This method, however, does not always identify fusion axes that satisfy the criteria of interpretability, (i) and (ii). This is because the FCM can construct only clusters with uniform hyper sphere distributions, and the FMDA cannot consider the distribution of each cluster in the projected space, and it is not aimed at obtaining visible models.

This paper presents a new method to construct a visible and interpretable fuzzy model that avoids the above limitations. We employ two clustering methods for extracting different statistical features of clusters. One is a method with EM algorithm [11] and the other is kernel fuzzy clustering [12]. The hyper sphere distribution of each cluster can vary with EM Algorithm. The kernel fuzzy clustering is a method to construct clusters in a higher-dimensional feature space using a kernel function. This paper introduces new criteria of interpretability considering the correlation of the clusters to the fusion axes in a projected space as well as the distance among the clusters, and proposes a fuzzy modeling method which can construct an interpretable fuzzy model in a visible space formed by the identified fusion axes. The fusion axis is defined as a linear combination of input variables. If we can express the meanings of these fusion axes with appropriate words that reflect their coefficients and meanings of corresponding input variables, the model will have higher interpretability.

Some studies employed the linear combinations of input variables in the antecedent parts [13–15]. Kim et al. [13] proposed a fuzzy modeling method using Principal Component Analysis (PCA) and Fuzzy C-Regression

Models (FCRM) [14] which could identify partially linear structures from input–output data. However, the objective of [13] is to improve the accuracy of model, not to obtain an interpretable model. Yam et al. [15] proposed a method that simplify a fuzzy model using orthogonal transformation by Singular Value Decomposition (SVD). Although this method was effective for reducing complexity of model, the interpretation of fuzzy model was not discussed.

A lot of methods for dimensionality reduction discussed so far in the field of data-mining have been based on statistical linear transform or projection. The linear projection that leads to the maximum variance in the projected space is obtained through PCA. Independent component analysis (ICA) [16] is a statistical technique by which observed random data is linearly transformed into components that are maximally independent of each other. Since these methods were devised for feature extraction of data, they were not intended to identify the input–output relationships and to visualize the data structure considering the data distribution in the projected space.

Self Organizing Map (SOM) [17] is a nonlinear method for extracting hidden features of data in 2 dimensional plane. Runkler's fuzzy nonlinear projection [18] is a method to visualize arbitrarily shaped multiple nonlinear manifolds based on Sammon mapping. These nonlinear projection methods can discover topological structures contained in data, but interpretation of nonlinearly projected data is difficult. The visualization for enhancing the interpretability of fuzzy model is the unique feature of the proposed method in this paper.

This paper shows through experiments with artificial data that it is possible to construct adaptive clusters and identify appropriate fusion axes. This paper also applies the proposed method to collected data of a mobile robot as one of multi-inputs systems while it was passing an aisle, and shows that it is possible to construct a visible and interpretable fuzzy model from this data. We can interpret actions of the mobile robot easily with the identified fuzzy rules.

## 2. Proposed Method

The proposed method consists of the following 4 steps to construct a fuzzy model.

- Step 1: Clustering data with predetermined number of clusters.
- Step 2: Determining fusion axes for visualization of these clusters.
- Step 3: Determining the antecedent part of each rule.
- Step 4: Determining the consequent part of each rule.

**Table 1** shows a list of denotations to be used in this paper. A system with $D$ input – 1 output non-linear relationships is assumed.

**Table 1.** List of denotations.

| | |
|---|---|
| $i, i', i''$ | Indexes related to $i$-th cluster (rule) |
| $j, j', j''$ | Indexes related to $j$-th input variable $x_j$ or $X_j$ |
| $k, k', k''$ | Indexes related to $k$-th data |
| $l$ | Index related to $l$-th fusion axis |
| $t$ | Index related to times of training |
| $\boldsymbol{x}_k$ | Input vector of $k$-th data $\boldsymbol{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kD})^t$ |
| $y_k$ | Output of $k$-th data |
| $y^i$ | Output from $i$-th rule (singleton) |
| $\boldsymbol{X}_k$ | $\boldsymbol{X}_k = (x_{k1}, x_{k2}, \ldots, x_{kD}, y_k)^t$ |
| $\xi_l$ | $l$-th fusion axis |
| $\boldsymbol{e}_l$ | Weight vector of $l$-th fusion axis |
| $A_l^i$ | Membership Function on $l$-th fusion axis of $i$-th rule |
| $N_C$ | Number of clusters (rules) |
| $D$ | Dimension of input vectors |
| $N_D$ | Number of data |
| $\mu_k^i$ | Degree of belonging of $k$-th data to $i$-th cluster |
| $w^i$ | Antecedent fitness value of $i$-th rule |
| $\boldsymbol{m}^i$ | Center of $i$-th cluster |
| $\hat{y}_k$ | Inference value of $k$-th data |

The rule structure is given by

$$\text{if}(\xi_1 \text{ is } A_1^i \text{ and } \xi_2 \text{ is } A_2^i) \text{ then } y = y^i \quad . \; . \; . \; . \quad (1)$$

where $\xi_l (l = 1, 2)$ is the "fusion axis" expressed as

$$\xi_l = \boldsymbol{e}_l^t \boldsymbol{\xi} = e_{l1}x_1 + e_{l2}x_2 + \cdots + e_{lD}x_D. \quad . \; . \; . \; . \quad (2)$$

(1) is the $i$-th rule. In this paper, we call the 2 dimensional projection plane constructed by the fusion axes "fused plane".

The unique feature of this rule is that every rule uses the same membership functions defined on the fusion axes $\xi_1$ and $\xi_2$, and this is the point different from Kim et al.'s method that uses different principal component axes for each rule.

A fuzzy model is constructed on the fused plane formed by the two fusion axes. Since each fusion axis is defined as a linear equation as in (2), it is easy to know influential input variables from its coefficients. If we could express the meanings of these fusion axes with appropriate words reflecting the coefficients and meanings of each input variable, the model would have a higher interpretability. The fuzzy model in this paper uses singletons in the consequent parts for simplicity. Though these simple rules sacrifice the accuracy of the model, we use this type of rules for obtaining an interpretable model.

The proposed method identifies the data structure by clustering and then derives the fusion axes based on the obtained clusters. This method has two choices for each of the clustering and the derivation of fusion axes depending on the characters of data. These methodologies and the fuzzy modeling on the fused plane are described below.