Rule Classification Visualization of High-Dimensional Data

Frank Rehm
German Aerospace Center
Braunschweig
frank.rehm@dlr.de

Frank Klawonn
University of Applied Sciences
Braunschweig/Wolfenbttel
f.klawonn@fh-wolfenbuettel.de

Rudolf Kruse
University of Magdeburg
kruse@iws.cs.uni-magdeburg.de

Abstract

This paper presents an approach to visualize high-dimensional fuzzy classification rules and the corresponding classified data set in the plane. This enables the observer to check visually to which degree a feature vector is classified by a certain rule. Also misclassified feature vectors can be well spotted and conflicting or error-prone rules can be identified.

Keywords: Fuzzy Rules, Visualization, Multi-Dimensional Scaling.

1 Introduction

Fuzzy rules are a powerful instrument to model classification problems. The strength of fuzzy rules is their good interpretability and their facile extraction from data or generation by hand. Nevertheless, if the data set is high dimensional in the feature space and the underlying data structure is rather complicated, a resulting rule system can be quite complex.

Visualizing high-dimensional data in the plane, i.e. on a computer monitor, is connected with dimension reduction which can be achieved by several techniques. With multidimensional scaling and related methods one tries to find a low-dimensional representation of the data while preserving distances or even angles between vectors as an objective function for such transformations [2, 4, 5]. We propose in this paper a method that maps high-dimensional rules, represented by their centre vectors, by means of Sammon's mapping. Classified feature vectors will be mapped while preserving membership degrees to the two rules that yield the highest response to the feature vector.

The paper is organized as follows: In section 2 we recall multidimensional scaling and Sammon's mapping as a common representative of it. In section 3 we describe the proposed method. In section 4 we illustrate the visualization technique on some benchmark data sets. Finally we conclude with section 5.

2 Multidimensional Scaling

Multidimensional scaling (MDS) is a method that estimates the coordinates of a set of objects $Y = \{y_1, \dots, y_n\}$ in a feature space of specified (low) dimensionality that come from data $X = \{x_1, \dots, x_n\} \subset \Re^p$ trying to preserve the distances between pairs of objects. Different ways of computing distances and various functions relating the distances to the actual data are commonly used. These distances are usually stored in a distance matrix

$$D^x = \left(d^x_{ij}\right),\ d^x_{ij} = \|x_i - x_j\|,\ i,j = 1,\dots,n.$$

The estimation of the coordinates will be carried out under the constraint, that the error between the distance matrix D^x of the data set and the distance matrix $D^y = \left(d^y_{ij}\right),\ d^y_{ij} = \|y_i - y_j\|,\ i,j = 1,\dots,n$ of the corresponding transformed data set will be minimised.

Thus, different error measures to be minimised were proposed, i.e. the absolute error, the relative error or a combination of both. A commonly used error measure, the so-called *Sammon's mapping*

$$E = \frac{1}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}^{x}} \sum_{i=1}^{n}\sum_{j=i+1}^{n} \frac{\left(d_{ij}^{y} - d_{ij}^{x}\right)^2}{d_{ij}^{x}}$$

describes the absolute and the relative quadratic error. To determine the transformed data set Y by means of minimising error E a gradient descent method can be used. By means of this iterative method, the parameters y_k to be optimised, will be updated during each step proportional to the gradient of the error function E. Calculating the gradient of the error function leads to

$$\frac{\partial E}{\partial y_k} = \frac{2}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}^{x}} \sum_{j \neq k} \frac{d_{kj}^{y} - d_{kj}^{x}}{d_{kj}^{x}} \frac{y_k - y_j}{d_{kj}^{y}}.$$

After random initialization for each projected feature vector y_k a gradient descent is carried out and the distances d_{ij}^{y} as well as the gradients $\frac{\partial d_{ij}^{y}}{\partial y_k}$ will be recalculated again. The algorithm terminates when E becomes smaller than a certain threshold.

3 Rule Classification Visualization

For our approach to visualize high-dimensional rules, we follow the terminology regarding fuzzy rules, according to the one defined in [1]. A trapezoidal membership function of a fuzzy rule is defined by four parameters $< a_i, b_i, c_i, d_i >$ (see figure 1). The rule's *core region* for attribute i is defined by parameter b_i and c_i. It describes the region of the membership function that is supported by training examples during the rule learning phase. The rule's *support region* for attribute i is defined by parameter a_i and d_i. The support region might be constrained as the figure shows, but also open to $\pm\infty$ depending on the training algorithm. In addition, a centre vector of each rule can be determined by means of the core region's centre for each attribute of the rule.

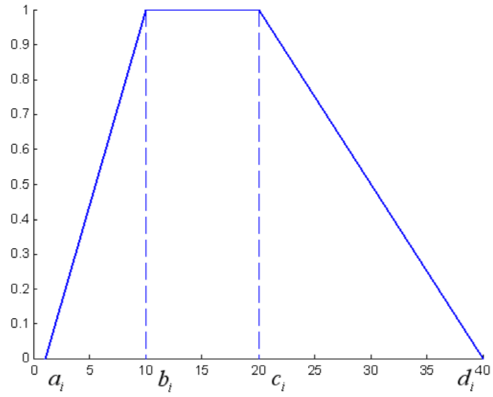

Figure 1: A trapezoidal membership function

Further, we define neighbourhood of rule centre vectors according to overlap regarding the core regions of the rule system. Neighbourhood N_{ef} of rule r_e and r_f can either be 1 if all core regions of both rules overlap, or 0 if not. N_{ef} will be used to enforce to preserve the neighbourhood relationship of the rules in the Sammon's mapping. As described above, Sammon's mapping uses a distance matrix D for the transformation minimizing differences of distances between rule centre vectors in the original space and distances in the target space. Thus, we use the normalized rule centre vectors to determine the required distance matrix and enlarge the distance of non-neighbouring rules by 1. Of course, no guarantee can be given, that all neighbouring rules will be placed appropriately, but considering core based neighbourhood might improve the chance to obtain feasible transformations.

Once the rule centre vectors are mapped in the plane, we propose to place the data set's feature vectors according to their membership degree to the two rules that yield the highest response. Thus, we place the feature vectors proportional to both rule centre vectors.

A visualization like this reveals some interesting aspects of the rule system. Similar rules and even neighbouring rules can be visualized by their distance and a drawn link respectively. Classified feature vectors symbolize by their colour and their proportional distance to the respective rule centre vector which rule fires to which degree. Misclassi-

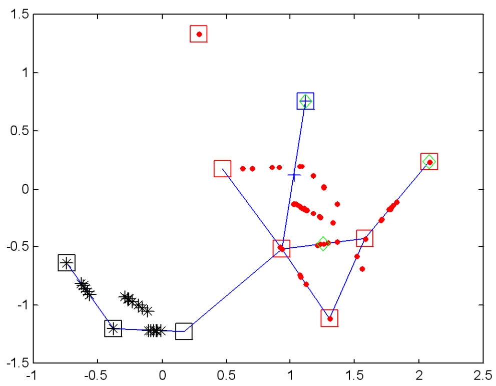

Figure 2: Visualization of the rule classifier on the Wine data set

fied feature vectors can be detected when using appropriate symbols or colours for them. In the next section we will demonstrate the proposed technique on some benchmark examples.

4 Results

Figure 2 shows some results on the well known wine data set. The wine data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

We applied the fuzzy rule learning algorithm as described in [1] and obtained 10 rules which classify the data set correctly. Additionally, we added some data with incorrect labels by hand to be able to visualize some misclassified data. Rule centre vectors are visualized by squares □. Connections between rule centre vectors indicate their neighbourhood regarding the core region. Rules of the same class are visualized by the same colour. We use different symbols ($\star, +, \bullet, \diamond$) to be able to differentiate feature vectors that lie upon each other if necessary.

The figure reveals some interesting facts. In consequence of placing vectors in the plane depending on their membership degree to the two rules that yield the highest response, classified feature vectors will be placed on an imaginary line that connects two rule centre vectors. Note, feature vectors may not only be represented by neighbouring rules corresponding to the core based neighbourhood definition whose neighbourhood is visualized by lines in the figure. As the figure reveals, for some neighbouring rules the data set contains no data that lies in the core regions of those rules. Two of ten rules represent data that lies not in all core regions of these rules. If two rules yield similar membership degrees to a feature vector, it will be placed in the middle between these rule centre vectors. Of cause, the classification that will be made in such cases is not that confiding since the decision comes randomly if no further information is available.

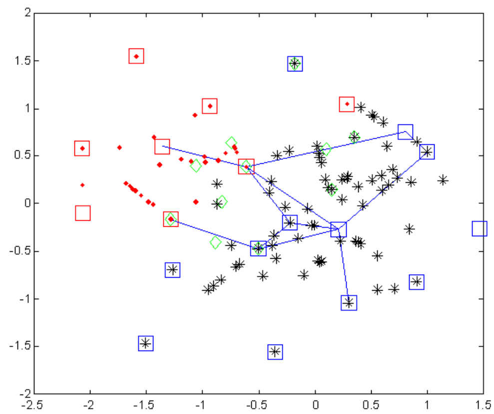

Figure 3: Visualization of the rule classifier on the Wisconsin breast cancer data set

The manually inserted data was partly misclassified which is not surprising since we labelled them intentionally incorrect. One vector was correctly classified but lies exactly in the middle between to rule centre vectors, which indicates, that classification was made at random since the feature vector is covered by the core regions of two neighbouring rules. This vector is visualized by the blue star in the figure. The misclassified feature vectors are visualized by diamonds. One of these is represented by two red rules. Again, we labelled the vector intentionally wrong to show this effect in the figure. The other feature vectors are clearly misclassified by a red or a blue rule.

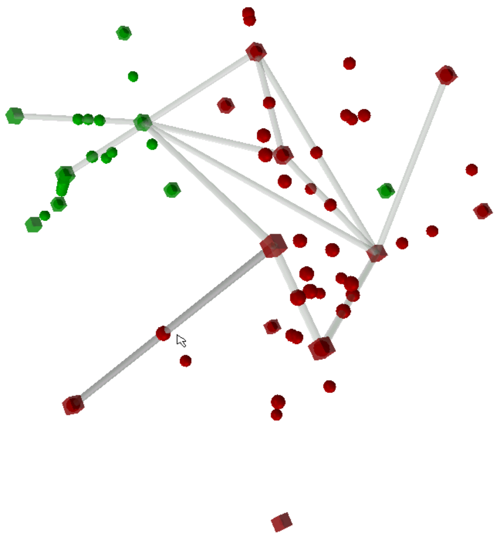

Figure 4: 3D-Visualization of the rule classifier on the Wisconsin breast cancer data set

The second example is the Wisconsin breast cancer data set [3]. Each patient in the database had a fine needle aspirate taken from her breast. Resultant, nine attributes where determined and analysed to discriminate benign from malignant breast lumps.

Figure 3 shows the visualization of the learned rule system and the corresponding classified data set. For this example we divided the data set into a training data set and a test data set by choosing randomly 50% of the data for each of both data sets. We used the training data set to learn the fuzzy rule classifier. The test data set was applied on the learned classifier which yields to the figure above using the proposed visualization technique.

The figure shows clearly that rule centre vectors which represent the same class are mapped in the same region in the plane. There are two neighbouring rules that represent different classes. These rules misclassify some feature vectors. Some rules respond only with small membership degrees to few feature vectors and do not yield high response to any other feature vector. This fact is shown in the figure by rule centre vectors that have no adjacent feature vectors. Thus, the figure reveals that the rule system can be pruned here.

Figure 4 shows a 3-dimensional visualization of the rule classifier on the training data of the Wisconsin breast cancer data set. Feature vectors of different classes are visualized by small spheres of different colours. Rule centres are visualized by cubes. Transparency helps to identify feature vectors which are positioned exactly on the same coordinate as rule centres. Light grey connections between rule centres indicate rule neighbourhood. Three-dimensional visualization is mainly efficient when interaction (zooming, rotating, etc.) is provided. The figure results from a Java3D implementation that enables the user to interact. In the foreground of the figure a dark grey connection can be found. In the actual implementation, feature vectors can be clicked and the two rules that yield the highest response to the feature vector will be visualized by an dark gray connection. Clicking the same feature vector again causes the disappearance of the according connection. This feature helps the analyst to identify interesting rules and feature vectors as well.

5 Conclusions

In this paper we presented a technique to visualize rules of a fuzzy rule classifier and the classified data as well. The visualization is useful for small rule systems. A huge number of rules are hard to map in the plane while preserving the core region neighbourhood. We discussed many aspects that can be extracted from the visualization. Subject of future work will be to visualize overlap of the support regions.

Acknowledgements

The breast cancer database was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg.

References

[1] Berthold, M.R.: Mixed Fuzzy Rule Formation. International Journal of Approximate Reasoning (IJAR), 32, pp. 67-84, Elsevier, 2003.

[2] Borg, I., Groenen, P.: Modern Multidimensional Scaling : Theory and Applications. Springer, Berlin, 1997.

[3] Mangasarian, O.L., Wolberg, W.H.: Cancer diagnosis via linear programming. SIAM News, volume 23, number 5, 1990, pp. 17–18.

[4] Rehm, F., Klawonn, F., Kruse, R.: MDS_{polar}: A New Approach for Dimension Reduction to Visualize High Dimensional Data, IDA 2005, LNCS 3646, Springer-Verlag Berlin, pp. 316-327, 2005.

[5] Sammon, J.W.: A nonlinear mapping for data structure analysis. IEEE Transactions on Computers, 18, pp. 401-409, 1969.