Fuzzy Clustering in Classification Using Weighted Features

Lourenço P.C. Bandeira[1], João M.C. Sousa[1], and Uzay Kaymak[2]

[1] Technical University of Lisbon, Instituto Superior Técnico
Dept. of Mechanical Engineering, GCAR/IDMEC
1049-001 Lisbon, Portugal,
jmsousa@ieee.org
[2] Erasmus University of Rotterdam, Faculty of Economics
P.O. Box 1738, 3000 DR Rotterdam, the Netherlands

Abstract. This paper proposes a fuzzy classification/regression method based on an extension of classical fuzzy clustering algorithms, by weighting the features during cluster estimation. By translating the importance of each feature using weights, the classifier can lead to better results. The proposed method is applied to target selection, where the goal is to maximize profit obtained from the clients. A real-world application shows the effectiveness of the proposed approach.

1 Introduction

Fuzzy modeling combines numerical accuracy with transparency in the form of linguistic rules [9]. One of the most important field of application of fuzzy models is classification, especially in data mining problems. A method that has been used extensively for obtaining fuzzy models is fuzzy clustering, which partition a data set into overlapping groups based on similarity measures. In classification, an object usually belongs to a certain class. However, in some applications it might be useful to give a degree to this classification, combining classification and regression of data. This paper proposes a classification/regression method where the data objects are sorted by real values. This method can be particularly useful in target selection.

Target selection is an important data mining problem from the world of direct marketing. Its goal is to determine the potential customers for a new product by identifying profiles of customers that are known to have shown interest in a product in the past. The key to target selection is maximizing the profits of selling the product, while minimizing the cost of the marketing campaign. In this type of problems, some features can be more relevant than others. Thus, weighting the different features by order of importance can lead to better classifiers. If there is some previous expert knowledge about this interaction between features, it should be looked upon. Therefore, this paper proposes to weight the features of the data based on expert knowledge. These weights are then added to fuzzy clustering techniques. Target selection models can be evaluated using gain charts (also called lift charts), which indicate the advantage obtained by using a derived model for target selection over random selection of targets. In this paper our objective is to maximize the profit that each customer can provide. We also propose a type of gain chart based on real outputs to describe the goal.

T. Bilgiç et al. (Eds.): IFSA 2003, LNAI 2715, pp. 560–567, 2003.

Summarizing, this paper proposes classification by using weighted fuzzy clustering based on the importance of each feature, and an adapted gain chart for classification/regression in target selection problems. The paper starts by presenting briefly fuzzy classification using fuzzy clustering in Section 2. The proposed weighting of features in fuzzy clustering is presented in Section 3. Target selection in direct marketing is presented afterwards in Section 4, where a new type of gain chart is proposed, and feature selection is also briefly discussed. An application is presented in Section 5, where the maximization of donations to a Dutch charity organization is considered. Finally, Section 6 presents the conclusions.

2 Fuzzy Classification

Fuzzy models have gained in popularity in various fields such as control engineering, decision making, classification and data mining [9]. One of the important advantages of fuzzy models is that they combine numerical accuracy with transparency in the form of linguistic rules. Hence, fuzzy models take an intermediate place between numerical and symbolic models. A method that has been extensively used for obtaining fuzzy models is fuzzy clustering. Fuzzy clustering algorithms are unsupervised techniques that partition a data set into overlapping groups based on similarity within the groups and dissimilarity amongst the groups. This paper uses fuzzy models for classifying objects, as described in the following.

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a set of N data objects where $\mathbf{x}_k \in \mathbb{R}^n$. The set of data objects can then be represented as a $N \times n$ data matrix $\mathbf{X}$, where n is the number of features (attributes) used to described the data. Note that each object is an instance represented by the vector $\mathbf{x}_k$, which is described by a set of features. The fuzzy clustering algorithm determines a fuzzy partition of $\mathbf{X}$ into C clusters by computing a $N \times C$ partition matrix $\mathbf{U}$ and the C-tuple of corresponding cluster prototypes $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_C\}$. Often, the cluster prototypes are points in the cluster space, i.e. $\mathbf{v}_i \in \mathbb{R}^n$, but they can also be closed volumes in the clustering space, as in the case of the extended fuzzy c-means algorithm presented in [6]. The elements $u_{ik} \in [0, 1]$ of $\mathbf{U}$ represent the membership of data object $\mathbf{x}_k$ in cluster i.

For the application used in this paper; charity donations (presented in Section 5), the objective is to maximize the number of responders, which is similar to target selection in a commercial environment, see Section 4. However, it is preferable to receive as much donations (positive responses to the direct marketing campaign) contacting as few target supporters as possible (minimizing the mailing costs). This approach has been considered recently [7]. In this paper we are not interested only in maximizing the number of responders, but also in maximizing the donation revenue. Therefore, there is not only the necessity of classification, but it is also required to estimate the donation of each of the identified supporters, by using a regression method. Thus, some new characteristics are added to the problem, as will be explained in Section 5.

3 Weighted Fuzzy Clustering

Many clustering algorithms are available for determining $\mathbf{U}$ and $\mathbf{V}$ iteratively. Fuzzy c-means and the Gustafson–Kessel (GK) clustering algorithms (or variations thereof) are the most popular [2, 4]. The fuzzy clustering algorithm introduced by Gustafson and Kessel introduces feature relevance through covariance matrices. However, this algorithm weight the inputs (features in classification problems) based on the distribution of data. In the algorithm proposed in this paper, the weights are based on the importance of the features directly. In our case, the weights are based on expert knowledge to determine the relevance of the features. A similar approach was proposed recently to handle missing data by fuzzy clustering methods [10], where an index for each data point is considered, in order to deal with missing data. The index has the value 0 when data is missing, and the value 1 when the data is available. This index can be seen as a weight, determining the importance of a certain data object. In our approach, each feature as a weight in the interval $[0, 1]$, where 1 stands for the most important features, and 0 correspond to features that are not relevant at all.

As the curse of dimensionality demands for less features, it is useful to reduce the features as much as possible. However, sometimes an important feature can be disregarded, which leads to poor classification results. Therefore, this paper explores the possibility of weighting the features in the clustering algorithm through expert knowledge, in order to reduce classification errors. Consider a vector of weights $\mathbf{w}$ with n elements, one for each feature:

$$\mathbf{w} = [w_1, w_2, \ldots, w_n], \tag{1}$$

where $w_i \in]0, 1]$, and $i = 1, \ldots, n$. Using this extension, the cluster centers are determined as follows:

$$v_{ij} = \frac{\sum_{k=1}^{N} u_{ik}^m w_j x_{kj}}{\sum_{k=1}^{N} u_{ik}^m w_j}, \quad j = 1, \ldots, n \tag{2}$$

where $i = 1, \ldots, C$, and m is an exponent that determines the fuzziness of the resulting clusters. The weights $\mathbf{w}$ are also used in the computation of the distance. Consider that an $n \times n$ diagonal matrix $\mathbf{W}$ is given by:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}. \tag{3}$$

The distance necessary to apply the algorithm is calculated by:

$$d^2(\mathbf{x}_k, \mathbf{v}_i) = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{W} (\mathbf{x}_k - \mathbf{v}_i) \tag{4}$$

Note that (4) is a modified version of the Euclidian distance. This paper applies fuzzy classification to target selection in direct marketing, in order to choose the best costumers, as explained in the following.

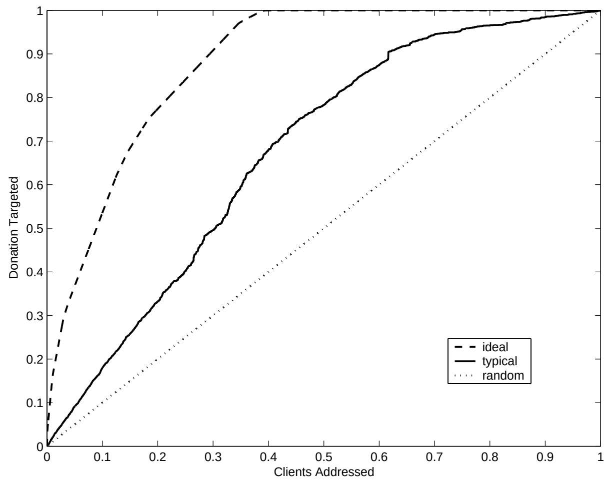

Fig. 1. Example of a gain chart: ideal model (- -), random selection (..), and typical model (–).

4 Target Selection in Direct Marketing

Target selection is an important data mining problem in direct marketing. The main task is to determine the potential customers for a product from a client database, by identifying profiles of customers that are known to have showed interest in a product in the past. Direct marketing is known as the use of existing and new marketing channels to encourage a direct relationship with the customers. Large databases of customers and market data are maintained for this purpose. The customers or clients to be targeted in a specific campaign are selected from the database, given different types of information such as demographic information and information on customer's personal characteristics like profession, age and purchase history.

4.1 Gain Charts

Target selection models can be evaluated in various ways. Often, gain charts (also called lift charts) are used to assess the gain to be expected from the utilization of a target selection model [8]. A gain chart indicates the advantage obtained by using a derived model for target selection over random selection of targets. When a classification problem intends to maximize a revenue, instead of the number of targeted responders, the gain chart presents a different aspect. Note that instead of considering the response percentage as in [8], the vertical axis must now represent the percentage gain over the total amount donated. Thus, this chart presents more explicitly the outcome to be expected from a certain marketing campaign. An example of this gain chart is shown in Fig. 1, where the horizontal axis indicates the percentage of customers that should be mailed to obtain a certain percentage of the total amount. For example, the point (20%, 30%)

indicates that 30% of the total donations can be expected to be captured by the target selection model, when 20% of the customers are selected. A random selection of customers corresponds only to 20% of the total amount raised by the same number of clients contacted. The ideal model (also shown in Fig. 1) produces a gain chart that rises as steeply as possible in this data set to the 100% level. This last chart is obtained by ordering the clients according to the amount of money donated, in descending order.

4.2 Feature Selection

An important step of building target selection models is selecting the features that will be used as the explanatory variables in the model. Internal databases typically contain customer-specific information about their purchase history and personal preferences. The purchase history can often be translated into measures of recency (e.g. how recent is the last purchase?), frequency (e.g. how often does a customer buy a product?) and monetary value (e.g. how much money does the customer spend per order?). It is often assumed in marketing literature that the RFM-variables are appropriate for capturing the specifics of the customer's purchase behavior [1]. From a modeling point-of-view, the RFM features have the advantage that the purchase behavior can be summarized by using a relatively small number of variables, which can be 10 or less for a given data set. Even though the number of available RFM features is typically small, one must still select the most relevant ones for a particular problem, since it is possible that certain variables do not have any explanatory power at all.

When the RFM variables are used, the clustering space can consist of the product-space of RFM features, since the dimensionality of the RFM feature space is often small enough. Fuzzy clustering divides the data into groups with similar properties on the RFM features considered. The clustering results must now be related to the known response behavior of the customers. An application of feature selection in determining the target models using fuzzy algorithms is presented in the next section.

5 Application: Charity Donations

The algorithm proposed in this paper is applied to target selection of a large data base from a Dutch charity organization. Such an organization does not have clients in the usual sense of the word. However, in order to optimize their fund raising results, it must be able to find the supporters who will probably donate more money, to optimize their fund raising results. These targeted supporters are then contacted by mail preferentially in relation to other individuals in the database. A training data set of about 8000 supporters has been collected for modeling purposes. Seven RFM features have been used for characterizing the donation history of the supporters:

1. Number of weeks since last response (TIMELR).
2. Number of months as a supporter (TIMECL).
3. Fraction of mailing responded (FRQRES).
4. Medium time of response (MEDTOR).
5. Average donation amount (AVGDON).

6. Amount of last donation (LSTDON).
7. Average donation per year (ANNDON).

Therefore, each instance $\mathbf{x}_k \in \mathbf{X}$ can be represented by the expression:

$$\mathbf{x}_k = [\text{TIMELR, TIMECL, FRQRES, MEDTOR,AVGDON, LSTDON, ANNDON}] \quad (5)$$

or by any subset of n attributes considered relevant to the problem. In any case, $\mathbf{x}_k$ is a vector with n elements, specifying the values of the n chosen attributes. The concept or function to be learned (target concept), denoted by c, is in this case the donation amount of the campaign used to build the target selection models:

$$c = \text{DONAMT} : \mathbf{X} \rightarrow \{0, \mathbb{R}^+\} \quad (6)$$

Note that c is both a classification and a regression value. A zero value indicates that the contacted person is not a supporter, and a value different than 0 indicates the expected donation amount. The training examples are described by the ordered pair $< \mathbf{x}_k, c(\mathbf{x}_k) >$. After the target selection model has been constructed, a score s_k is attributed to each supporter. This score corresponds to the predicted value $c(\mathbf{x}_k)$, which is the model prediction of the amount of money that the supporter will donate in the next fund raising campaign. The best supporters, in the sense of donating more money, will be the targets. Usually, the region of interest is composed by 10% up to 50% of the clients, i.e. the region that is selected for mailing purposes in direct marketing [5].

5.1 Comparing the Target Selection Models

In order to compare the performance of the obtained models, the sum squared error is used. Considering that a real number is being maximized, this is one of the most utilized performance measures. This error is computed as the difference between the ideal and the obtained gain charts, for all data points. This measure S can be computed as follows:

$$S = \sum_{k=1}^{N} (gi_k - gc_k)^2 \quad (7)$$

where gi_k are the values obtained for the ideal gain chart, and gc_k are the values obtained for the gain chart computed using the fuzzy model. The model features are selected and weighted, based on the analysis of the sum squared error S. Note that all models have been derived based on the same pseudo-random seed, in order to be comparable. After normalizing the data, a model using the fuzzy c-means has been obtained using all seven RFM variables with $w_i = 1$, for $i = 1, \ldots, n$. The automatic weighting using GK fuzzy clustering has also been derived. Based on expert knowledge and some trial-and-error experiments, the best weight vector obtained for the seven features in (5) is the following:

$$w = [0.10, 0.01, 0.01, 0.01, 0.725, 0.225, 0.05]. \quad (8)$$

The gain chart for the three tested methods is presented in Fig. 2. Note that the proposed weighted method clearly outperforms the other methods.

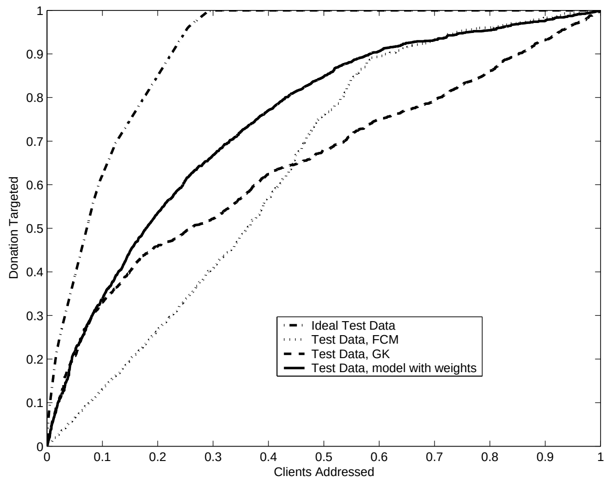

Fig. 2. Gain charts obtained using the seven features: ideal model (-.), c-means (..), weighted c-means(–) and GK model (- -).

By performing some experimentation, analyzing both gain charts and their respective sum squared error S, it was found out that four features presented the most accurate results. So, all the possible combination of target selection models using four features have been tested. The model with the smallest error S, is the one with the following features: TIMELR, AVGDON, LSTDON, ANNDON. Then, based on these four features, one feature was added or excluded, and as so, all models with three and five features have been analyzed, and none had a better result.

Table 1. Comparison of errors using different number of features.

Model	# features	Weights	Train Error	Test Error
c–means	7	no	473.3	490.3
GK	7	yes	166.2	343.9
c–means	7	yes	113.6	155.1
c–means	4	no	128.7	168.8
GK	4	yes	184.9	536.9
c–means	4	yes	113.4	155.8

The sum squared errors obtained in test data are shown in Table 1. Note that the tests using four features are only possible due to the application on expert weights to determine the most important features, as proposed in this paper. Using the seven features, the weighted approach is clearly better than the other approaches. When the four most relevant features are used, the test error S is still clearly better when using the weighted fuzzy clustering algorithm.

6 Conclusions

This paper proposed the weighting of features in fuzzy clustering based on expert's knowledge in classification/regression problems. The method is applied to the maximization of profit in a target selection problem. The application to a charity donations problem shows clearly the advantage of the proposed method.

Future research will deal with the choice of the most relevant features in an automatic way. One possibility is to use logistic regression [3] by testing statistically the parameters obtained for the features. Another possibility is the use of genetic algorithms to optimize the weights.

Acknowledgements

This research is partially supported by the "Programa de Financiamento Plurianual de Unidades de I&D (POCTI), do Quadro Comunitário de Apoio III", FCT, Ministério do Ensino Superior, da Ciência e Tecnologia, Portugal.

References

[1] C. L. Bauer. A direct mail customer purchase model. *Journal of Direct Marketing*, 2:16–24, 1988.

[2] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function*. Plenum Press, New York, 1981.

[3] J. S. Cramer. *The LOGIT Model: an introduction for economists*. Edward Arnold, Kent, 1991.

[4] D. E. Gustafson and W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings IEEE CDC*, pages 761–766, San Diego, USA, 1979.

[5] U. Kaymak. Fuzzy target selection using RFM variables. In *Proceedings of joint 9th IFSA World Congress and 20th NAFIPS Int. Conference*, pages 1038–1043, Vancouver, Canada, July 2001.

[6] U. Kaymak and M. Setnes. Fuzzy clustering with volume prototypes and adaptive cluster merging. *IEEE Transactions on Fuzzy Systems*, 10(6):705–712, December 2002.

[7] S. Madeira and J.M. Sousa. Comparison of target selection methods in direct marketing. In *Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems, Eunite'2002*, pages 333–338, Albufeira, Portugal, September 2002.

[8] J. M. Sousa, U. Kaymak, and S. Madeira. A comparative study of fuzzy target selection methods in direct marketing. In *Proceedings of 2002 IEEE World Congress on Computational Intelligence, WCCI'2002, FUZZ-IEEE'2002*, pages 1–6, Paper 1251, Hawaii, USA, May 2002.

[9] J.M. Sousa and U. Kaymak. *Fuzzy Decision Making in Modeling and Control*. World Scientific, Singapore and UK, December 2002.

[10] H. Timm, C. Doring, and R. Kruse. Fuzzy cluster analysis of partially missing data. In *Proceedings of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems, Eunite'2002*, pages 426–431, Albufeira, Portugal, September 2002.