A Comparative Study of Fuzzy Target Selection Methods in Direct Marketing

João M. Sousa\(^1\)  Uzay Kaymak\(^2\)  Sara Madeira\(^1\)

\(^1\) Technical University of Lisbon, Instituto Superior Técnico Dept. Mech. Eng./GCAR–IDMEC
1049-001 Lisbon, Portugal, Email: jmsousa@ist.utl.pt, saramadeira@mail.telepac.pt

\(^2\) Erasmus University of Rotterdam, School of Economics, Dept. of Computer Science
P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, Email: u.kaymak@ieee.org

Abstract - Target selection in direct marketing is an important data mining problem for which fuzzy modeling can be used. This paper compares several fuzzy modeling techniques applied to target selection based on recency, frequency and monetary value measures. The comparison uses cross validation applied to mailing campaigns of a charity organization.

I. INTRODUCTION

Direct marketing has become an important application field for data mining. In direct marketing, companies or organizations try to establish and maintain a direct relationship with their customers in order to target them individually for specific product offers or for fund raising. Large databases of customer 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 the customer’s personal characteristics like profession, age and purchase history. Apart from commercial firms and companies, charity organizations also apply direct marketing for fund raising. Charity organizations do not have customers in the regular sense of the word, but they must be able to trace people who are more likely to donate money in order to optimize their fund raising results. However, there is a major difference between direct mailing in a commercial environment and direct mailing for the benefit of a charity organization: the response rate to a commercial direct marketing campaign seldom exceeds 5%, whereas a charity campaign among a group of known supporters often triggers a much higher response. Modeling of charity campaigns/donations has been considered in [1].

Target selection is an important data mining problem in direct marketing. The main task in target selection is to determine the potential customers for a product from a client data base by identifying profiles of customers that are known to have shown interest in a product in the past. Many techniques have been applied to select the targets in commercial applications, such as decision tree methods like CHAID or CART [2], statistical regression [3], neural computing [4] and fuzzy clustering [5].

Recently, fuzzy clustering has also been applied to target selection using the so-called recency, frequency and monetary value (RFM) variables [6]. It is claimed in [6] that it is often sufficient to use fuzzy c-means clustering for target selection instead of other fuzzy algorithms. This paper tests this assertion explicitly by comparing different fuzzy clustering algorithms applied to target selection. A collection of new and classical fuzzy clustering techniques are applied to a target selection problem. Each methodology derives multiple models using parts randomly selected from the total data set. The different fuzzy modeling strategies are then compared using measures obtained from cross validation, namely the mean and the standard deviation of the ensemble of models obtained.

The paper starts by presenting briefly direct marketing in Sec. II, where the concept of target selection models and the important step of feature selection are introduced. Section III describes the various fuzzy modeling techniques used in this paper, and their application to target selection. The cross validation technique used to compare the different target selection models is described in Sec. IV. The performance of different fuzzy models are compared in Sec. V, where the cross validation methodology is applied to fuzzy target selection models obtained from data from a charity organization. Finally, the paper presents the conclusions in Sec. VI.

II. TARGET SELECTION

Direct marketing is characterized by the use of existing and new marketing channels to develop a direct one-to-one relationship with the customers. Usually, one sifts through large amounts of customer data in order to identify patterns and regularity that can help develop the relationship. Target selection is an important data mining problem in direct marketing. The main task in target selection is to determine the potential customers for a product from a client data base by identifying profiles of customers that are known to have shown interest in
a product in the past.

The methods for target selection can be divided into two main groups:

1. segmentation methods,
2. scoring methods.

Both the segmentation approach and the scoring approach are possible within the framework of fuzzy clustering, as illustrated in [6] by using data from the actual target selection campaigns of a large charity organization.

The segmentation approach divides the customer database into customer segments with similar properties. Segmentation based target selection models thus divide the customers into several groups depending on similarity in relevant features. An estimate of the response percentage can be made for each group given the training data available. The customers within the groups that have a high response percentage are then selected for targeted offers, i.e. they are sent a product offer by mail or otherwise.

The scoring approach assigns a separate score to each individual customer, and is interesting for tailoring the marketing campaign to individual customers. In the scoring based models, therefore, a score is assigned to each individual customer, where the score is indicative of the likelihood of response of the customer. The customers are then ordered according to their likelihood of response based on the predictions of the target selection model. Only the customers who are likely to respond (e.g. their score is above a threshold value) are sent the product offer. One way of evaluating target selection models is by means of the so-called hit probability charts, which show what percentage of the mailed customers will respond, given a certain mailing size. Hence, the point (20%,70%) on the hit probability chart indicates that within the 20% selected, 70% will be responders and 30% will not be. An example of a hit probability chart is depicted in Fig. 1. In a random model, the hit probability is constant at the average level of responders within the data set, as indicated by the horizontal line in the figure. The ideal target selection model produces a hit probability chart that starts at the 100% level, and which decreases to the average level as customers with smaller likelihood of response are added to the selected group. Hit probability charts are used in the following for evaluating the target selection models.

Just like in any modeling, an important step of building target selection models is selecting the features that will be used as the explanatory variables in the model. In this paper, we consider features based on 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?). The advantage of the so-called RFM features is that the customers’ behavior can be captured by using a relatively small number of features, which improves the transparency of the target selection models that are developed. It is often assumed in marketing literature that the RFM-features are appropriate for capturing the specifics of the customer’s purchase behavior [7].

III. FUZZY TARGET SELECTION MODELS

Fuzzy models have gained in popularity in various fields such as control engineering, decision making and data mining. 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 used extensively 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.

Let \( \{x_1, \ldots, x_N\} \) be a set of \( N \) data objects where \( x_k \in \mathbb{R}^n \). The set of data objects can then be represented as a \( N \times n \) data matrix \( X \). The fuzzy clustering algorithms determine a fuzzy partition of \( X \) into \( C \) clusters by computing a \( N \times C \) partition matrix \( U \) and the \( C \)-tuple of corresponding cluster prototypes \( V = \{v_1, \ldots, v_C\} \). Often, the cluster prototypes are points in the cluster space, i.e. \( v_i \in \mathbb{R}^n \), but they can also be closed volumes in the clustering space as in the case of extended fuzzy c-means algorithm [8]. The elements \( u_{ki} \in [0,1] \) of \( U \) represent the membership of data object \( x_k \) in cluster \( i \).

Many clustering algorithms are available for solving for \( U \) and \( V \) iteratively. Fuzzy c-means and the Gustafson–Kessel clustering algorithms (or variations thereof) are the most popular. 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. Let \( \mathbf{r} \) be the vector of response indication, where \( r_k \in \{0, 1\}, k = 1, \ldots, N \). The average response rate \( \rho_k \) for each cluster can be calculated as the weighted average of response indication according to

\[
\rho_k = \frac{\sum_{k=1}^{N} u_{i,k} r_k}{\sum_{k=1}^{N} u_{i,k}} \tag{1}
\]

A score for each customer can then be computed as

\[
s_k = \frac{\sum_{i=1}^{C} u_{i,k} \beta_k}{\sum_{i=1}^{C} u_{i,k}} \tag{2}
\]

The customers are ordered according to \( s_k \) in descending order, and the hit probability charts can be produced. The target selection model consists of the computed cluster centers and the corresponding \( \rho_k \).

This paper compares several fuzzy modeling methods applied to the target selection problem in direct marketing. The following methods are applied to the target selection problem.

1. Fuzzy clustering in the product-space of the RFM variables by using the fuzzy c-means (FCM) algorithm [9]. Fuzzy c-means algorithm searches for spherically distributed clusters within the data, and have been applied extensively in the literature. The fuzzy clusters obtained from clustering methods are defined on the product space of variables.

2. Fuzzy clustering in the product-space of the RFM variables by using the Gustafson-Kessel (GK) algorithm [10]. This algorithm uses an adaptive distance measure in order to adapt the shape of the clusters to the distribution of the data points, thereby leading to ellipsoidal clusters. The details of FCM and GK algorithms are well-known, and they are not studied further in this paper. This paper intends to study the advantages or drawbacks of this method compared to the fuzzy c-means for target selection.

3. Fuzzy clustering by using the Extended FCM algorithm as described in [8]. This clustering technique uses FCM clustering with volume prototypes and similarity driven cluster merging. This method can be made to determine the number of clusters automatically.

4. Takagi-Sugeno (TS) fuzzy models obtained from the projections of the fuzzy clusters obtained by using fuzzy c-means on to the individual features [11], [12]. The fuzzy clusters defined on the product space of variables need not be interpretable yet, since humans find it difficult to visualize groupings in high dimensional spaces. Therefore, the model must be transformed into one where the groupings are made on each feature separately. This is the more common approach in fuzzy modeling. Since a single score \( \rho_k \) can be associated with each grouping, a Takagi–Sugeno model with constant consequents is natural for target selection models (although not necessary strictly). Transparent Takagi–Sugeno models can be obtained by projecting each fuzzy cluster on to individual variables and by approximating the projections by membership functions. Afterwards, the rule consequents can be estimated from (1), where now the degree of firing \( \beta_{ik} \) of each rule replaces the values of membership. Similarly, the scoring per individual is obtained from (2) by replacing membership values with degrees of rule firing. The fuzzy clusters are derived using the FCM algorithm.

5. Takagi-Sugeno fuzzy models obtained as in the previous item, but using the fuzzy GK algorithm instead of FCM [13]. One disadvantage of cluster projection is that the projected clusters almost always overlap to a large degree. For interpretability, therefore, a simplification step must follow cluster projection, where similar (largely overlapping) membership functions are replaced by common representative membership functions. This simplification step is not considered in this paper. The reader is referred to [12] for a similarity based simplification of rule bases obtained from fuzzy clustering.

6. Optimization of Takagi-Sugeno models using Genetic Algorithms (GA). An initial TS fuzzy model is derived using the FCM algorithm. The model obtained is further optimized by using the GA algorithm proposed in [14].

IV. COMPARISON OF FUZZY MODELS USING CROSS VALIDATION MEASURES

In order to compare the performance of different fuzzy modeling approaches, several target selection models must be identified. From the several possibilities to compare and validate the different models, statistical cross validation measures are one of the most utilized. This paper uses the mean and the standard deviation of a certain number \( M \) of models to compare the different fuzzy identification techniques.

Let the data set \( \mathbf{X} \) be divided in a certain number of equal parts \( M \). Then, \( M \) models are identified and they are cross validated using these \( M \) partial data sets. For these models we can derive hit probability charts. Then, the mean and the standard deviation of the hit probability charts can be computed. The comparison of these curves gives an idea of the differences between the models. The cross validation algorithm used in this paper can be summarized as follows.

1. Let the data set \( \mathbf{X} \) be randomly divided into \( M \) equal parts, \( \mathbf{X}_1 \) to \( \mathbf{X}_M \).
2. For \( k = 1 \) up to \( M \), train a model using \( \mathbf{X}_k \). Compute a hit probability chart \( h\mathbf{P}_k \) using \( \mathbf{X}_k \).
3. Assume that all $h_{pc_k}$ have equal probability of occurrence (i.e. $\frac{100}{M}$ %), compute the mean of $h_{pc}$ as

$$\text{mean}(h_{pc}) = \frac{\sum_{i=1}^{M} h_{pc_i}}{M} \tag{3}$$

4. Calculate the standard deviation of $h_{pc}$ as

$$\text{std}(h_{pc}) = \sqrt{\frac{\sum_{i=1}^{M} (h_{pc_i} - \text{mean}(h_{pc}))^2}{M - 1}} \tag{4}$$

The means of $h_{pc}$ indicate the average accuracy that we can expect from training a model by using the whole data set $X$. The standard deviation of $h_{pc}$ gives an error margin around the models obtained by training on $X$. The different models (methods) are compared using the mean and standard deviation of the hit probability charts. These measures indicate how significant the differences are amongst the models considered.

V. APPLICATION: CHARITY DONATIONS

The fuzzy clustering methods described in Sec. III using the cross validation presented in Sec. IV are applied to target selection modeling from the database of a Dutch charity organization. A charity organization does not have customers in the usual sense of the word. However, it must be able to trace people who are more likely to donate money (supporters) in order to optimize their fund raising results. The targeted supporters are then contacted by mail preferentially in relation to other individuals in the database. A training data set of about 4000 supporters has been collected for modeling purposes. This data set is prepared for identification as described in Sec. IV. Three RFM features have been used for characterizing the donation history of the supporters, following the suggestion in [6].

1. Number of weeks since last response (TIMELR).
2. Number of months as a supporter (TIMECL).
3. Fraction of mailings responded (FRQRES).

After normalizing the data, target selection models are derived. A target selection model based on logistic regression (a widely–used method for obtaining target selection models) is also derived. The performance of the logistic regression model has been used in the following as a benchmark, with which the performance of the various fuzzy models has been compared.

An important parameter to select in the fuzzy clustering models is the number of clusters into which the data should be partitioned. The number of clusters should be large in order to capture important details, but on the other hand, it should be small in order to obtain sufficient generalization from the model. Therefore, this number is carefully chosen for each fuzzy model, and it is a trade-off between obtaining sufficiently predictive models and having a sufficient number of data points with large membership in each cluster.

The charity donation data is divided into $M = 10$ randomly chosen data sets. Thus, 10 models are identified following the algorithm described in Sec. IV, by using logistic regression and the six different fuzzy clustering techniques proposed in Sec. III. The comparison of hit probability charts is adopted in this paper, as these types of charts revealed to be more informative than the gain charts that are sometimes used.

Figure 2 shows the mean of the 10 hit probability charts obtained using logistic regression, fuzzy clustering with fuzzy c-means and fuzzy clustering with the Gustafson-Kessel algorithm. The fuzzy modeling techniques are the ones described in Sec. III in items 1 and 2. Both the FCM and the GK algorithms divide the data into 40 clusters, which is a trade-off between obtaining sufficiently predictive models and having a sufficient number of data points with large membership in each cluster. It is clear that the GK algorithm performs worse than the other algorithms. The FCM algorithm performs worse when 0 up to 10% of the clients are addressed, but it is the best from 10% up to 40%, which is actually the region of interest, i.e. the region that is usually selected for mailing the clients in direct marketing [6]. Therefore, the FCM algorithm is preferable to both logistic regression and the GK algorithm.

A comparison between the mean of the 10 hit probability charts obtained using logistic regression and the extended FCM algorithm, described in Sec. III in item 3, is presented in Fig. 3. The extended FCM algorithm has an average of 10 clusters for each variable. The results are comparable except for the first decile, where the logistic regression presents best results again. Therefore, the extended FCM algorithm has less clusters than FCM alone, but also slightly worse results.

Next, the mean of the 10 models obtained using Takagi–Sugeno fuzzy models are presented in Fig. 4. These ap-
Fig. 3. Comparison of the mean values of target selection models using logistic regression and the extended FCM algorithm.

Fig. 4. Comparison of the mean values of target selection models using logistic regression and Takagi–Sugeno models identified using FCM and GK algorithms.

Fig. 5. Comparison of the mean values of target selection models using logistic regression, FCM and TS models optimized by GA algorithms.

The last algorithm to be tested is the derivation of TS fuzzy models optimized by using a genetic algorithm as described in Sec. III, item 6, and introduced in [14]. As FCM performed slightly better than GK, we use this clustering method to identify the fuzzy models. The model is initialized with 20 clusters, which have proven to be enough before. The genetic algorithm runs for 100 generations for each of the 10 target selection models identified. The mean of the target selection models identified using logistic regression, FCM and the GA algorithm are depicted in Fig. 5. This figure shows clearly that the TS fuzzy model performs better than the others methods presented. However, the mean of the hit probability charts is very close to the TS models using FCM. As the GA algorithm has a much larger computational effort, it seems to be preferable to use FCM clustering to identify TS fuzzy models.

In order to compare the different modeling techniques it is not possible to use only the mean of the obtained hit probability charts. In fact, a measure such as the variation or the standard deviation of the models is useful to know how much they vary. Figure 6 presents the standard deviations for the model chosen as benchmark and the fuzzy models that revealed to perform better, i.e. logistic regression, fuzzy c-means, TS models using FCM, and TS models optimized using GA. No significant changes are noticed in the regions of interest (from 10 up to 50% of clients addressed). The standard deviations range from 0.025 up to 0.05, which means that the models have variations of 0.06 up to 0.25%. Between the methods themselves, Fig. 6 shows that logistic regression has in general the smallest standard deviation, followed closely by the FCM method. The TS model using FCM has a slightly smaller standard deviation than the one obtained using genetic algorithms.
VI. CONCLUSIONS

Target selection of customers in direct marketing by using various fuzzy clustering methods are considered in this paper. These target selection models are generated by analyzing client data obtained from similar previous marketing campaigns. This paper applied target selection in charity organizations. Logistic regression has been chosen as benchmark, and it is compared to fuzzy modeling techniques, namely, fuzzy c-means in the product space of variables, the Gustafson–Kessel algorithm, extended fuzzy c-means, identification of Takagi–Sugeno fuzzy models derived using both FCM and GK clustering techniques, and finally, Takagi–Sugeno models optimized by a genetic algorithm. The comparison is performed by using cross validation whereby the mean and the standard deviation of a certain number of models derived from similar data are calculated. The best method in terms of the mean value of hit probability charts is the identification of Takagi–Sugeno fuzzy models using the FCM algorithm. Note however that other algorithms obtained better results at specific regions. The logistic regression method leads to a more stable model, but its average performance is not as good. The fuzzy models are slightly less stable, but they have the advantage of superior numeric accuracy. Moreover, it is known that linguistic rules of the fuzzy models contribute to the explanatory properties of the target selection model [6].

As future work, algorithms that are parameter dependent, such as extended FCM and the optimization using GA are going to be more deeply explored. Moreover, the best model obtained in this paper is going to be compared to neural network modeling techniques. Finally, a fuzzy algorithm more suitable to deal with target selection problems using RFM variables is under development. This algorithm must achieve better results than the best obtained in this paper, i.e. more accurate or as accurate but simpler, in order to be worthy.

ACKNOWLEDGMENTS

This work is partially supported by the project “On Soft Computing in Modeling and Control”, POCTI/34058/EME/2000, FCT, Ministério da Ciência e Tecnologia, Portugal.

References