The Improvement of Response Modeling: Combining Rule-Induction and Case-Based Reasoning

F. Coenen, G. Swinnen, K. Vanhoof and G. Wets

Limburg University Centre, Department of Applied Economics,
B-3590 Diepenbeek, Belgium
{filip.coenen;gilbert.swinnen;koen.vanhoof;geert.wets}@luc.ac.be

Abstract. Direct mail is a typical example for response modeling to be used. In order to decide which people will receive the mailing, the potential customers are divided into two groups or classes (buyers and non-buyers) and a response model is created. Since the improvement of response modeling is the purpose of this paper, we suggest a combined approach of rule-induction and case-based reasoning. The initial classification of buyers and non-buyers is done by means of the C5-algorithm. To improve the ranking of the classified cases, we introduce in this research *rule-predicted typicality*. The combination of these two approaches is tested on synergy by elaborating a direct mail example.

1 Introduction

One of the most typical examples where response modeling comes into play is direct mail. This marketing application goes further than just sending product information to randomly chosen people, as mass marketing does. A key characteristic of direct mail is that a specific market or geographic location is targeted, while selecting receptors by age, buying habits, interests, income, etc. In order to decide which people will receive the mailing, the potential customers are divided into two groups or classes: buyers and non-buyers. This division, which is based upon the above-mentioned socio-demographic and/or economic information of the potential customers, is called response modeling and can be realized by means of artificial intelligence. A learning algorithm is then applied to predict the class of unseen cases or records, i.e. possible customers. As known from literature, the accuracy of the prediction never reaches 100%, as there are always cases attributed to the wrong class. When applied to the direct mail example again, this means that, at a given mailing depth, there are always people receiving mail concerning products that appear uninteresting to them while buyers are left out of the mailing. As a consequence, costs are made that can be avoided, e.g. by creating a better response model. One way to come to a better response model would be by choosing a better classifier [8]. In this paper, however, we suggest an other approach; i.e. the combination of multiple classification methods.

The remainder of this paper is organized as follows. In the next section, a theoretical background of the performed approach will be discussed and a following

J.M. Zytkow and J. Rauch (Eds.): PKDD'99, LNAI 1704, pp. 301-308, 1999.

section deals with the empirical evaluation of the suggested approach. To illustrate this, a direct mail example is further elaborated. The last section will be reserved for conclusions and topics for future research.

2 Suggested Approach

2.1 Classifiers

C5. One of the possible classifiers that can be used in the response modeling of a data set is the C5-algorithm, the more recent version of C4.5 [10]. The reason that we preferred this algorithm is based upon previous research Van den Poel and Wets [14]. They used the same data set as we did to provide a comparison between a number of classification techniques. They selected techniques in the field of statistical, machine learning and neural network applications, and compared them by means of the overall accuracy on the data set. We preferred to use the C5-algorithm to do the initial classification, since this algorithm attained the highest accuracy on the test set (see also section 3.2).

The goal of response modeling is to rank the cases by probability of response. Since each case is classified with a certain confidence by C5, the most trivial way to rank the cases would be by the confidence figure of the applied rule. This means that when the assigned class label is the non-responding class, the complement of the confidence should be taken before sorting the whole data set on this confidence. However, as mentioned before, we propose in this paper an other method to improve response modeling, as will be explained.

Case-Based Reasoning. Case-based reasoning methods are based on similarity and try to use the total information of a given unknown case. In our research, we used typicality as similarity measure. To determine the typicality of each case in the context of this research, the following approach was used.

a) Firstly, for each case i a distance measure dist(i, j) is determined as follows; the attribute values of i are compared with the attribute values of a case j <> i. If the values of the considered cases differ, dist(i, j) is increased by one (independent of the size of the difference).

b) After determining dist(i, j) according to the above-mentioned method, the class value of the cases i and j is compared. If i belongs to the same class as j, a measure intra(i) is increased by (1 - (dist(i, j) / number of attributes)). If, on the other hand, both cases belong to a different class, a measure inter(i) is increased by (1 – (dist(i, j) / number of attributes)). The above calculations are made for all the cases j <> i. This is the point where the global character of our approach comes into play, since all other cases j<>i are taken into account in the calculation of the typicality of just one case i.

c) In a next step the measure intra(i) is divided by the number of cases that belong to the same class as i, and inter(i) is divided by the number of cases that belong to the other class.

d) Finally, the typicality of case i is determined by dividing intra(i) by inter(i). For each case the typicality was calculated, allowing these cases to be ranked by this measure.

The above steps lead to the following definition:

$$\text{Typicality}\ (i) = \frac{\text{intra}(i)/p}{\text{inter(i)}/n}. \quad \textbf{(1)}$$

with p the number of cases that belong to the same class as case i, and n the number of cases that belong to the other class. The cases with typicality higher than 1 are considered as typical cases for the class they belong to. Used as a classifier, this method looks at the similarity between the considered case and the different classes and assigns the label of the most similar class to the case. In response modeling however, it is sufficient to look at the similarity between the considered case and the responding class, in order to use this similarity as a ranking criterion.

A Combined Approach. As it is known from previous research, the accuracy of such a model almost never reaches 100% for real world cases. Also in our direct mail example, the classification of buyers and non-buyers was not completely correct; some non-buyers were classified in the class of the buyers, and vice-versa. Since the accuracy of our model attained 76.32 %, a percentage of 23.68% of the cases were misclassified. The fact that errors are made implies that there is room left for improvement if we choose not to write to all persons in the data set as is often the case in direct mailing. This will further be explained in section 3.3

In order to upgrade the response model and have more control over these mislabeled instances, we decided to rank the classified cases. Empirical results taught us that C5 is a better classifier than typicality on the one hand, and also better than other considered classifiers on the other hand This is why we opted for this algorithm to do the initial classification. By applying C5, a case obtains a response probability from just one rule, i.e. the rule with the highest confidence that meets the case. The other rules or cases are not taken into account. The classification by C5 can thus be considered as a local approach; only a part of the information carried by the case and the rule-base is used. In contrast to this, a case-based reasoning method displays a global character; a case obtains a response probability by looking at the total data set. Empirical results (see section 3.2) showed us that typicality outperforms confidence in ranking the cases. That is why we opted for this method to improve the ranking of the classified cases. By combining the strengths of both methods, i.e. C5 as the best classifier, and typicality as the best ranker, we could investigate the effects of the combination between a global and a local approach. The new response modeling method that is suggested in this paper can then be described as follows. An unknown case obtains the class label from the C5 classifier and obtains as response measure the typicality for the given class label. The latter has as consequence that the calculation of the assigned typicality is based on the predicted class label of the case. This typicality will further be called *rule-predicted typicality*. Thus, the cases are firstly ranked by class label and secondly by rule-predicted typicality.

2.2 Evaluation

To compare the ranking by typicality on the one hand with the ranking by confidence and the original situation on the other hand, we selected the Coefficient of Concordance (CoC) [6] and the cumulative response rate as objective measures and graphs as a visualization tool. The CoC takes into account the ranking of the cases, and gives a percentage as outcome. The higher the percentage, the better the sorting. The main reason for choosing this measure is that it looks at the distribution of the cases in the predicted class as a whole. Therefore, the distribution is calibrated on a 10-class rating scale. This means that the distribution is split up into 10 intervals, each with a score higher than the previous interval. The CoC is defined as follows:

$$\text{CoC} = \frac{1}{n_g n_b}\left(\sum_{i=\min score}^{\max score} nb_i ng'_i + 0.5 \sum_{i=\min score}^{\max score} nb_i ng_i \right). \quad \textbf{(2)}$$

with nb_i respectively ng_i the number of bad, respectively well classified cases with a score equal to i, ng'_i the number of well classified cases with score better than i. With a given mailing depth, we know how many cases will be mailed, and the different methods can be evaluated with the cumulative response rate. Further, the graphs can help us by discovering in which range a certain method is superior.

3 Empirical Validation

3.1 The Data Set

The data set that was used for empirical validation was collected from an anonymous mail-order company and consists of 6800 records or cases, each record described by 15 attributes. These records are equally divided between the classes 0 (non-buyers) and 1 (buyers). The information that was available concerns transactional data, as well as socio-demographic information o the customers. All variables were categorized after careful consideration with the mail-order company. They provided the data to us at the level of the individual customer. The specific model that we have built is based on all available data, and predicts whether a person is a possible buyer or not. The outcome is a binary response variable (0/1) representing buying or not buying. Before the induction of the C5 classifier, a training set was composed by randomly selecting approximately 2/3 of the cases from the original data set. The remaining part was used for purposes of testing.

3.2 Results

Evaluation of the Classification. As mentioned in the section concerning the suggested approach, the C5-algorithm was used to classify the cases in a first step. By

applying C5, an accuracy of 76,32% was obtained on the test set, which consisted of 2052 cases (1018 buyers and 1034 non-buyers). Since the accuracy didn't reach 100%, and the cases were randomly divided into a training and a test set, there are a number of incorrectly classified cases randomly divided among the predicted ones. In order to implement our approach, we separated the cases that were predicted to belong to class 0 (1120 cases) from the cases that were predicted to belong to class 1 (932 cases). This means that our model considered 1120 out of 2052 persons as non-buyers, and 932 persons as buyers. An overview of the situation in the test set after classifying by C5 is shown in table 3.

Table 3. Confusion Matrix of the Test Set

	Real 0	Real 1	Total
Predicted 0	834	286	1120
Predicted 1	200	732	932
Total	1034	1018	2052

In order to deduce a better response model, we improved the ranking of the cases by sorting them by typicality within the predicted class, under the assumption that the cases that were misclassified would have a lower typicality. This means that after sorting by typicality, the misclassified cases would appear lower in the rank than the correctly classified ones.

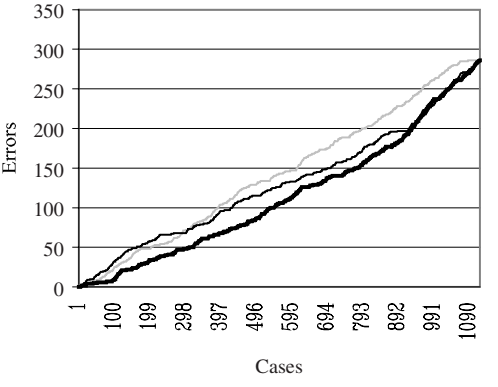
Evaluation of the Ranking. To compare the outcome of our experiments with the initially unsorted situation on the one hand and the sorting by confidence and rule-predicted typicality on the other hand, we used the Coefficient of Concordance. As mentioned in section 2.2, the distribution has to be ranked on a 10-class rating scale to be evaluated. To evaluate the ranking by rule-predicted typicality in the context of this research, we decided to use the rule-predicted typicality of the cases as score. This means that if the highest rule-predicted typicality of a case in the set attains 1.5, and the lowest rule-predicted typicality equals 0.5, the cases with rule-predicted typicality between 0.5 and 0.6 will be considered as belonging to the same group, and thus have the same score. This implies that the score in the definition (see section 3.2) is replaced by rule-predicted typicality. The same method is used to evaluate the sorting by confidence. The exact results of these calculations can be found in table 4.

Table 4. The Coefficient of Concordance

	Predicted Class	
	0	1
Sorted by Confidence	55,2%	65,9%
Sorted by Rule-Predicted Typicality	62,9%	65,0%

Table 4 shows us that the ranking of the test cases that were predicted to belong to class 0, as well as the test cases that were predicted to belong to class 1, becomes better after sorting by rule-predicted typicality or by confidence. In both cases the coefficient of concordance is higher than 50%, i.e. the percentage that can be

expected by a random division of the misclassified cases among the correctly classified cases. If the sorting by rule-predicted typicality is compared with the sorting by confidence, a difference between the predicted class 0 and the predicted class 1 can be noticed. For the predicted class 0, the rule-predicted typicality produces a better result since the coefficient of concordance equals 62.9% against 55.2% after sorting by confidence. This observation is further illustrated by figure 1; the rule-predicted typicality curve is less steep over a larger distance than the confidence curve. The X-axis shows the number of cases in the predicted class 0, whereas the Y-axis shows the number of misclassifications as they appear gradually among the considered cases



.

Fig. 1. The appearance of the errors among the cases that are predicted to belong to class 0. The *gray colored graph* represents the occurrence of errors among the cases that were predicted to belong to class 0 for the unsorted situation. The *black* and the *bold black graph* describe the same after sorting by confidence, respectively rule-predicted typicality.

3.3 Application on the Direct Mail Example

In normal circumstances, a mail-order company will try to cut off between 10% and 40% of its unattractive part of the mailing list. This means that between 60% and 90 % of all the persons in the data set will receive the mailing. Often, a mailing depth of 75% is used [14]. The reason for this is that the profit generated by converting a non-buyer into a buyer is considered higher than the cost of sending a letter to a person that is not interested in the products that are subject of the mail. In our further calculations, we will also consider a mailing depth of 75% of the test set (0,75 * 2052 = 1539 persons). To reach these people, we will direct a letter to all the persons that are considered as buyers by our system, i.e. 932 persons, of which 732 are classified right and thus are buyers in reality. 1539 – 932 = 607 persons from the predicted class 0 will complete this number so that a total amount of 1539 persons are reached. Applied to this direct mail example, the sorting by confidence and rule-predicted typicality produced the following results.

Sorting by Confidence. The predicted non-buyers were sorted by an increasing confidence. As the non-buyers with low confidence are more likely to be misclassified than the ones that were predicted to be non-buyers with a high confidence, the 607 non-buyers with the lowest confidence are included in the mailing list. Among these 607 persons there were 169 buyers. This means that by mailing 1539 persons, we would reach 732 + 169 = 901 buyers out of the 1018 buyers (88,5%) that are present in the test set.

Sorting by Rule-predicted Typicality. Analogously on the sorting by confidence, we sorted the cases that were predicted to belong to class 0 by increasing rule-predicted typicality and included the 607 persons with the lowest rule-predicted typicality in the mailing list. Among these 607 persons there were 194 buyers, so that we would reach 732 + 194 = 926 buyers out of 1018 (91%) by mailing 1539 persons.

Unsorted Situation. To illustrate the improvement that is made by sorting the cases of the predicted class, we finally give an overview of the situation as it would be without any sorting. Among the 607 persons of the predicted class 0, there would be approximately 0,26 * 607 = 158 buyers since 286 out of 1120 (+/- 26%) cases were misclassified and the errors are randomly divided in the predicted class. This means that we would reach 732 + 158 = 890 buyers out of 1018, or 87.4%. An overview of the results can be found in table 5.

Table 5. The number of reached buyers

Unsorted	Sorted by Confidence	Sorted by Rule-Predicted Typicality
87.40%	88.50%	91.00%

The fact that the improvement after ranking by confidence is limited to 1.10% shows us that the sorting of the classified cases is a difficult topic. Our approach proved to be a useful one, since it outperforms sorting by confidence by an improvement more than twice as high (2.50%) as the existing improvement of 1.10%.

4 Conclusions

This article describes a method for improving response modeling by using a combined approach of rule-induction and case-based reasoning. The proposed approach consists of classifying the cases by means of the C5-algorithm in a first step, and ranking the classified cases by a typicality measure in a second step. In this way, we could test the combination of the use of local and global information on synergy.

Based on empirical results we decided that the C5-algorithm was the best classifier to do the initial classification. This algorithm provides the local aspect of our approach, since it classifies each case by just one rule, i.e. the rule with the highest confidence that meets the case. The other rules or cases are not taken into account. In contrast to this, a case-based reasoning approach displays a global character, since a case obtains a response probability by looking at the total data set. Empirical results showed us that sorting by typicality was the best method to improve the ranking of the

classified cases. To do so, we introduced the concept rule-predicted typicality, as the calculation of the typicality of a test case is based on the predicted class value of the considered case. Finally, the application of our approach on a direct mail example has shown this method to be a promising one. It proves to yield an improvement of 2.50% over the improvement of 1.10% that is generated by the ranking of the classified cases by the existing confidence figures. This implies that we were able to reach 91% of the buyers in our test set, under the consideration of a mailing depth of 75%. Although it is only about a small improvement in absolute terms, yet the total success of a direct mail can depend on this. Since we were not able to test this approach on more than one data set so far, opportunity for future work lies within this topic.

References

1. Aijun, A., Cercone, N.: Multimodal Reasoning with Rule Induction and Case-Based Reasoning, in Multimodal Reasoning, AAAI Press (1998)
2. Bayer, J.: Automated Response Modeling System for Targeted Marketing (1998)
3. Brodley, C.E. and Friedl, M.A.: Identifying and eliminating mislabeled training instances, in Proceedings of Thirteenth Nat. Conference on Artificial Intelligence, AAAI Press (1996)
4. Domingos, P.: Knowledge Discovery vie Multiple Methods, IDA, Elsevier Science (1997)
5. Domingos, P.: Multimodal Inductive Reasoning: Combining Rule-Based and Case-Based Learning, in Multimodal Reasoning, AAAI Press (1998)
6. Goonatilake, S., Treleaven, P.: Intelligent Systems for Finance and Business, Wiley (1995) 42 – 45
7. Holte, R., Acker, L.E., Porter, B.W.: Concept learning and the problem of small disjuncts, in Proceedings of the Eleventh Int. Join Conference on AI, Morgan Kaufmann (1989) 813-818
8. Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms, Thirteenth National Conference on Artificial Intelligence (1996)
9. Ling, C.X., Li, C.: Data Mining for Direct Marketing: Problems and Solutions, in Proceedings of the Fourth Int. Conference on Knowledge Discovery and Datamining, AAAI Press (1998) 73-79
10. Quinlan, J.R.: C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, 1993.
11. Sabater, J., Arcos, J.L. and Lopez de Mantaras, R.: Using Rules to Support Case-Based Reasoning for Harmonizing Nelodies, in Multimodal Reasoning, AAAI Press (1998)
12. Surma, J. and Vanhoof, K.: Integrating Rules and Cases for the Classification Task, Case-Based Reasoning, Research and Development, First International Case-Based Reasoning Conference, - ICCBR'95, Springer Verlag (1995) 325-334
13. Surma, J., Vanhoof, K. and Limere, A.: Integrating Rules and Cases for Data Mining in Financial Databases, in Proceedings of the Nineth Int. Conference on AI Applications – EXPERSYS'97, IIIT-International (1997)
14. Van den Poel, D., Wets, G.: Data Mining for Database Marketing: a mail-order company application, in Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, RSFD '96 (1996) 383- 389
15. Zhang, J.: Selecting Typical Instances in Instance-Based Learning, in Proceedings of the Ninth Int. Conference on Machine Learning, Morgan Kaufmann (1992) 470-479