Usage of Data Mining Techniques on Marketing Research Data

PAVEL TURCINEK, JIRI STASTNY, ARNOST MOTYCKA
Department of Informatics
Mendel University in Brno
Zemedelská 1, 61300 Brno
CZECH REPUBLIC
pavel.turcinek@mendelu.cz, jiri.stastny@mendelu.cz, mot@mendelu.cz

Abstract: This contribution contains problems of marketing research data classification by means of data mining algorithms. Three basic methods are described, classification with the aid of Multi-layer Perceptron neural network with Back-propagation algorithm, classification with the aid of Bayesian Networks and classification with the aid of Decision Tree. Finally, applicability of these algorithms is compared. These algorithms are applied over the data from a survey about consumer behavior in the food market in the Czech Republic.

Key-Words: Data mining, Classification, Multi-Layer Perceptron neural network, Bayesian Networks, Decision Tree, Marketing Research

1 Introduction
Marketing research is a process of collecting and using information for marketing decision making [1] and plays an essential role in marketing management. Tools for supporting individual phases of marketing research, especially collection and analysis of information can be successfully facilitated by increased use of databases and data mining techniques [2]. As a part of a Marketing Information System [3] such tools provide decision makers with a continuous flow of information relevant to their area of responsibility [1].

In the area of marketing research are commonly used traditional statistical methods. Our goal is to try modern approaches of artificial intelligence tools on data from the marketing research which deals with consumer behavior in the food market in the Czech Republic.

The issue of consumer behavior falls into the field of marketing. Into issue of consumer behavior fall categories of recognition and understanding of how consumers think, feel, evaluate, choose among different alternatives, how consumers are influenced by their surroundings, how they act during the decision-making and purchasing, how is their behavior limited by their knowledge or ability to process information, what motivates them and how they differ in their decision-making in different ways depending on the importance or product interest [4].

For the purposes of marketing research can be used secondary data from national or international sources such as the Czech Statistical Office or Eurostat, or be process primary data obtained, for example, from surveys [5]. The vast majority of data that provide statistical offices today exists in electronic form. Even data from the surveys, it is usually recorded electronically or they are converted to electronic form.

Perception of information and communication technology is gradually transformed from something rather unique, bringing a competitive advantage in the market, to the necessity of conditioning the existence or not existence of business between the competitive business organizations [6].

Nowadays it is hard to imagine the processing greater volume of data without the use of information technology and machine processing. The aim of this paper is to test the classification algorithms over a set of data from the questionnaire survey of marketing research. The issue of data mining in terms of customer behavior prediction also deals [7]. Using neural networks for classification of customer behavior problems in [8].

2 Problem Formulation
This article will address the possibilities of using artificial intelligence tools for classification [9, 10, 11, 12] over data on consumer behavior in the food market and assess their applicability to this type of data. Article will be based on primary data that was taken out of the research Department of Marketing and Trade at the Faculty of Business and Economics.
of Mendel University in Brno. These data were processed only by means of statistical methods. More details about this research are given in [13].

2.1 The use of layered Multi-Layer Perceptron neural network
Multi-Layer Perceptron neural network (MLP) is an acyclic forward network. Neurons can be divided into disjunctive layers so that the output of each neuron of one layer is connected to the inputs of each neuron layer following. There are no links between non-neighboring layers of neurons or between neurons in the same layer. Each neuron has as many inputs as there are neurons in the lower layer. The input layer serves only to distribute input values to the first hidden layer. A network with one hidden layer and one output layer is known as a two-layer network, a network with two hidden layers of a three-layer, etc. [3, 8]. As a learning algorithm for MLP neural network is the most widely used back-propagation algorithm.

Back-propagation algorithm is an iterative method where the network gets from an initial non-learned state to the full learned one [3, 8, 9].

Back-propagation algorithm is based on minimization of neural network energy given with the formula (SSE) [3, 8]:

$$E = \frac{1}{2} \sum_{i=1}^{n} (y_i - d_i)^2$$  \hspace{1cm} (1)

where $n$ means the number of network outputs, $y_i$ means the $i$-th output and $d_i$ means the $i$-th output required.

2.2 The use Bayesian Networks
Bayesian classifiers are statistical classifiers. They can predict class membership probability that a given tuple belongs to a particular class. Bayesian classification is based on Bayes theorem described in [16].

Learning Bayesian networks from data is known for long time. This is a form of unsupervised learning, in the sense that the learner does not distinguish the class variable from the attribute variables in the data. The objective is to induce a network (or a set of networks) that “best describes” the probability distribution over the training data. This optimization process is implemented in practice by using heuristic search techniques to find the best candidate over the space of possible networks. The search process relies on a scoring function that assesses the merits of each candidate network. [17]

2.3 The use decision tree
Decision trees are a way to represent rules underlying data with hierarchical, sequential structures that recursively partition the data. A decision tree can be used for data exploration in one or more of the following ways:

- **Description:** To reduce a volume of data by transforming it into a more compact form, which preserves the essential characteristics and provides an accurate summary.
- **Classification:** Discovering whether the data contains well-separated classes of objects, such that the classes can be interpreted meaningfully in the context of a substantive theory.
- **Generalization:** Uncovering a mapping from independent to dependent variables that is useful for predicting the value of the dependent variable in the future [18].

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label. Basic algorithm is described in [16].

3 Problem Solution
From members of the Department of Marketing and Trade we received a set of 5809 responses to a questionnaire on research of consumer behavior in the food market. This research was conducted in the Czech Republic. It included all regions, age groups, types of education and occupation of respondents. The entire data set consists of thirty-one questions related to consumer behavior, where the respondent determines the influence of factors (composition of the product, low price, recommendations, etc.) on his/her shopping behavior on a scale from 1 to 10, where 10 specifies the highest importance in decision-making to purchase a specific product. Furthermore, the questionnaire contains eight questions that characterize the respondent (age, gender, etc.). This data is described in detail in [13].

Employees of Department of Marketing and Trade, based on of traditional statistical methods, broke down respondents into eight groups. This classification was created based on three factors:

- I prefer stores that are closest to my place of residence (or employment), I don’t commute.
- I rather shop less often and bigger purchases.
- Before shopping I first look for inspiration in offering in leaflets.
The results of this classification has been provided together with a complete set of input data. Our goal was to divide the respondents into given classes based on three factors. The thirty-one factors, as shown in the calculation of 2, can be assembled 4495 triplet combinations.

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (2)
\]

\[
\binom{31}{3} = \frac{31!}{3!(31-3)!} = 4495
\]

This number was reduced so that each factor occurred in a trio with another factor just once. We created an algorithm which generated such triplets of factors, which limited the total number of tested triplets to 155.

For classification of data was used freely available software WEKA version 3.7 [19]. This program offers many classification algorithms. We used three of them. They were a function BayesNET, MultilayerPerceptron REPTree, which are specific implementations of above mentioned algorithms.

BayesNET use various search algorithms and quality measures.

Base class for a Bayes Network classifier provides data structures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2 and B.

MultilayerPerceptron is a Classifier that uses backpropagation to classify instances.

This network can be built by hand, created by an algorithm or both. The network can also be monitored and modified during training time. The nodes in this network are all sigmoid (except for when the class is numeric in which case the the output nodes become unthresholded linear units).

REPTree is fast decision tree learner. Builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting). Only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces (i.e. as in C4.5)[19].

By using these algorithms were created models designed to classify into groups proposed by employees of Department of Marketing and Trade based on various triplets of factors.

AS the first three factors, we used those that were used in the earlier mentioned classification. In this case, all three models were designed that all of them were able to classify without any mistake. Figure 1 shows the decision tree, which exactly corresponds to the mentioned classification.

![Decision tree generated based on the original factors by Weka](image)

In creating models for classification based on a combination of other factors were focused on percentage of correctly classified instances and the time needed to build the model. Overview of the results can be found in Table 1. This table is based on 154 combinations of input factors. I does not include in the original triplet.
Table 1: Summary results achieved by individual methods

<table>
<thead>
<tr>
<th>classification method</th>
<th>average classification success [%]</th>
<th>average time to build the model[s]</th>
<th>number of cases where the method was the best</th>
<th>the best case</th>
</tr>
</thead>
<tbody>
<tr>
<td>BayesNET</td>
<td>34.2344</td>
<td>0.0212</td>
<td>109</td>
<td>57.8033</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>33.8044</td>
<td>10.2395</td>
<td>40</td>
<td>56.9517</td>
</tr>
<tr>
<td>REPTree</td>
<td>33.0077</td>
<td>0.0357</td>
<td>5</td>
<td>57.3166</td>
</tr>
</tbody>
</table>

The table shows that the success in classifying into classes based on the three input factors (second column) was very low. The best results were accomplished with use of combination of factors, which included one of the factors used in the first classification. Even in these cases, success did not achieve the rate of 60%. Only in two cases, all methods have classified with a success rate of more than 50%. In all cases, methods classified with comparable success. The biggest difference in the success of correct classification was 3.5370%.

It is also possible to read from the table that the most successful method for creating classification models was BayesNET, but the differences in success rate were insignificant. BayesNET method was also the fastest in creating the model. Method REPTree didn’t need much more time. However method MultilayerPerceptron needed to create a model much more time.

The table 2 shows the influence of the inclusion of a specific factor into triplet on the success of classifications of the model. Factors 0, 1 and 2 represent those factors that were used in the original combination. It is interesting that the use of factor 0 in input triplet brought significantly lower average success classification than the remaining two original factors. Factor 0 is a factor in this view even worse than factor 9.

Table 2: Average accuracy of classification based on the input factor

<table>
<thead>
<tr>
<th>factor</th>
<th>average classification success [%]</th>
<th>average classification success [%]</th>
<th>average classification success [%]</th>
<th>average classification success [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46.505</td>
<td>33.056</td>
<td>31.795</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>45.810</td>
<td>32.905</td>
<td>31.789</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>38.028</td>
<td>32.033</td>
<td>31.780</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>37.097</td>
<td>32.019</td>
<td>31.731</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>36.129</td>
<td>31.886</td>
<td>31.709</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>36.126</td>
<td>31.826</td>
<td>31.689</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>34.089</td>
<td>31.824</td>
<td>31.634</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33.903</td>
<td>31.821</td>
<td>31.594</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>33.892</td>
<td>31.805</td>
<td>31.566</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>33.626</td>
<td>31.802</td>
<td>31.443</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>33.128</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to table 2 another classification was made based on factors 1, 2 and 9. The results are shown in Table 3. Figure 2 shows the decision tree for this case.

Table 3: results for combination of factors 1, 2 and 9

<table>
<thead>
<tr>
<th>classification method</th>
<th>average classification success [%]</th>
<th>average time to build the model [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BayesNET</td>
<td>74,9871</td>
<td>0,24</td>
</tr>
<tr>
<td>Multilayer-Perceptron</td>
<td>74,9182</td>
<td>10,90</td>
</tr>
<tr>
<td>REPTree</td>
<td>74,4879</td>
<td>0,04</td>
</tr>
</tbody>
</table>
Fig. 2: Decision tree for the input factors 1, 2 and 9 generated by Weka

4 Conclusion
The results of the classification of all methods were comparable. All created models showed that the original classification is highly dependent on three factors mentioned above. With use of two original factors of success of classification improves. The decision tree in Figure 2 shows that the model is first decided based on factors 2 and 1.

It is obvious that chosen classification of consumer behavior does not affect all factors, but only three selected. Selecting three other factors will not create a model good enough to classification into the correct class.

From achieved results it is obvious that the method MultilayerPerceptron needed much more time to create the model, while providing comparable results. As the best method seems BayesNET.

Our research did not explore all possible combinations of triplets of factors. However, it may be said that without using the original input factors is not possible to create a usable model for classification in the proposed classes.

Currently, our research focuses on testing different topologies of neural networks MLP and also other types of neural networks. Partial achieved results are encouraging and provide a prerequisite for using these methods for solving problems.

Acknowledgments:
This work has been supported by the research design of Mendel University in Brno MSM 6215648904/03.

References:


