INTELLIGENT ANALYSIS OF MARKETING DATA

Łukasz Paśko, Galina Setlak

***Abstract**: The main goal of this paper is to present and evaluate the possibility of using the methods and tools of Artificial Intelligence and Data Mining to analyze marketing data needed to support decision-making in the process of market segmentation. This paper describes the application of Kohonen's Neural Networks and Classification Trees (including tools such as CART-Classification and Regression Tree, Chi-squared Automatic Interaction Detector (CHAID) and Boosted Tree) to solving problems of classification and grouping of data. The main part presents the results of market segmentation that can be used by the company producing household products. Finally conclusions and further research plans have been described.*

***Keywords**: data analysis, artificial intelligence, data mining, classification, clustering, Kohonen's neural networks.*

***ACM Classification Keywords**: I.2.m Miscellaneous : I.2.6 [Artificial Intelligence]: Learning – Connectionism and neural nets; I.5.1: Models – Neural nets; I.5.3: Clustering – Algorithms.*

Introduction

In order to effectively run a modern enterprise operating on international markets, in conditions of intensive and constantly growing competition, there is a need for knowledge that is often hidden in massive data sets. Every organization conducts its business with the use of huge amount of information. This information constitutes the sources of knowledge regarding firm's processes and functions, and is essential for making informed decisions.

Today the words of Philipp Kotler, well-known specialist in the field of marketing become more and more important. As he says: "If the data cannot be transformed into information, which will be the basis of knowledge, and knowledge - the source of wisdom, we lose far more than we get." [Kotler Ph., 1984]. These words show the importance and the need to develop modern tools and systems for the processing and analysis of all these huge streams of information in order to obtain the knowledge necessary to make optimal decisions in the process of firm's management.

The main responsibility of marketing analysis is to provide managers with very important market knowledge which concerns the potential markets, specificity of demand and the trends related to pricing as well as advertising effectiveness. It gives a manager an opportunity to learn how the characteristics of the market, distribution channels and strategies to stimulate sales look like.

Traditionally, marketing data analysis uses statistical methods such as regression and correlation analysis or discriminant and factor analysis. As noted by Stanimir "The effectiveness of marketing activities may be multiplied by interpreting customer behavior using modern IT solutions." [Stanimir, 2006]. This is why nowadays marketing research is supported by advanced tools, methods and techniques of artificial intelligence.

In recent years, the field of research known as "data mining" has been developing very intensively. Data mining is primarily focused on automating the analysis of large data sets stored in databases and/or data warehouses. Data mining is about finding hidden patterns or relationships in large data sets. The knowledge regarding these discovered patterns may be used as a basis for decision making processes. Data mining is also known as an intelligent analysis of the data, because the paradigm is based on the use of artificial intelligence methods and techniques (neural networks, genetic algorithms) [Hand David et all, 2005], [Larose , 2006], [Cios K., 1998].

In this paper the evaluation of the possibility of using the methods and tools of Artificial Intelligence and Data Mining to analyze marketing data, needed to support decision-making in the process of market segmentation has been presented. The paper describes the study of the application of Kohonen's Neural Networks and Classification Trees (including CART-Classification and Regression Tree, Chi-squared Automatic Interaction Detector (CHAID) and Boosted Tree) to solving problems of classification and grouping of marketing data. The main part presents the results of market segmentation that can be used by the company producing household products. Finally conclusions and further research plans have been described.

Theoretical foundations

Classification and grouping belong to basic tasks of the data mining. Problems of classification are very common (which products should be offered on which markets, which customers are credible and which are not, recognition of signals and images etc.) and are being solved in the everyday business activity, in the process of managing organization, in the medical or technical diagnostics, and in data analyses as well as measuring experiments.

Especially significant are the developments of tools and systems that are able to effectively support the problem solving in situations of high complexity, large number of problem's parameters or time constraints imposed on the solution. Problems of classification and grouping of marketing data belong to the above mentioned categories and usually require sophisticated computational techniques and methods [Migut, 2010], [Mynarski, 2010] , [StatSoft, 2010].

Classification is about assigning the object to one of the model classes based on selected distinctive features. The process may be called classification when the categories to which elements of input set will be assigned are determined in advance. The term classification is also defined as an automatic determination of objects affiliation to specified class on the basis of their images [[Adamczak, 2001], [Hand David et all, 2005], [Stąpor, 2011], [Szczuka, 2000]]. Classification is the most frequently solved problem both in technology and economy. Formally, classification is defined as the process of mapping data into the set of predefined classes.

$$f_c : R^p \supset X \rightarrow C, \tag{1}$$

where $C = \{C_1, C_2, ..., C_n\}$ is the finite set of classes, whereas the set $X \subset R^p$ is the attribute space, and the decision about the classification result is based on these attributes. Classification mapping f_c divides space X into n decision areas, grouping the attributes patterns that belong to one category [Zieliński, 2000]. Input data is the set of examples, observations and samples which are the list of the descriptive features values. Output data constitutes a model (classifier). The main aim of data classification is to build formal classifier. The classification process consists of two stages: building a model and testing unknown values.

Grouping also known as clustering is a primal in relation to classification because is responsible for defining classes. The process may be called grouping when the number of groups and the ranges of distinctive features values are not known before the division of the input set. Grouping is about matching similar and separating different objects [Hand David et al., 2005], [Li S., 2000], [Żurada et al., 1996]. It also means the division of set of elements into subsets on the basis of distinctive features detected during the process of division.

Neural networks are widely used as classifiers [Jang et al., 1997], [Moon et al., 1998], [Takagi et al., 2000], [Setlak G., 2004]. Classification and clustering problems have been addressed in many research works and disciplines such as statistics, machine learning as well as databases. The basic algorithms of the classification methods are presented in [Nauck et al., 1997], [Stąpor, 2005], [Żurada et al., 1996].

The applications of the clustering procedure can be divided into following categories [Hand David et all, 2005]:

- hierarchical form trees in which the leaves represent particular objects, and the nodes represent their groups. The higher level concentrations include the lower level concentrations. In terms of hierarchical methods, depending on the technique of creating hierarchy classes (agglomerative methods and divisive methods);
- graph-theoretic clustering,
- fuzzy clustering,
- methods based on evolutionary methods,
- methods based on artificial neural networks.

The most often used classification methods, also applied to marketing data analysis, are the following: decision trees (aka classification trees) [Li S., 2000], [Mynarski, 2010], methods of naive Bayes classification [Migut, 2010], [Stanimir et al., 2006], memory classification (e.g. nearest neighbor method or neural networks [Li S., 2000], [Szczuka, 2000]). Nowadays for solving problems of classification and grouping scientists are using also artificial neural networks [Hand David et al., 2005], [Larose, 2006], [Setlak, 2000], [Stąpor, 2011] and algorithms of fuzzy grouping (*aka* fuzzy clustering algorithms) [Czogala et al., 2000], [Jang, 1997]. What is more, the solutions that are combination of several artificial intelligence tools (neural networks, fuzzy logic, genetic algorithms) are applied, what enables to create more efficient hybrid neuro-fuzzy classifiers in which genetic algorithms are used for a teaching process [Nauck, 1997], [Rutkowska, 2000], [Setlak, 2001], [Setlak, 2008], [Lotfi Zadeh, Rutkowska, 2000].

In this research, artificial neural networks and decision tree algorithms have been used as a basic tool for creating classifiers. Neural Networks have been selected because they possess such characteristics as approximation abilities, interference immunity and adaptability. However, it is worth to remember that it may be good idea to use neural networks along with other AI tools such as decision trees, expert systems, and fuzzy sets as well as fuzzy logic. Such combination enables to design more efficient support systems, free from neural networks deficiencies.

Experiment

An experiment includes several tasks of vacuum cleaners market segmentation. Market segmentation is about dividing market into smaller parts called segments. Segments are created according to segmentation criteria. These criteria may be related to customers'

preferences, method of purchase, gender, age or other characteristics selected by marketer.

After segmentation is done, company has to choose the segment that is best suited for specific product. The primary goal of segmentation is customer preferences analysis. The secondary is the product positioning, which is the process of giving the product specific features that will differentiate it from similar products offered by competitors [Migut, 2010], [Stanimir et al., 2006].

Before the segmentation is performed the criteria of segmentation have to be determined. Next important step is to identify, if it is possible, optimal number of segments (groups or classes). In specialist literature two types of segmentation are distinguished: descriptive segmentation and predictive segmentation.

Descriptive segmentation is usually used in situations where:

- there is lack of criteria used for groups selection,
- all the variables (attributes describing objects) are independent variables,
- there is a lack of information allowing the use of methods of learning with the teacher, and one can use only the methods of so-called not-directed data mining (learning without a teacher) [Migut, 2010].

Descriptive market segmentation may be conducted with the use of such methods and tools as agglomeration method, the method of K-means and one of the modern methods – Kohonen's Neural Networks.

Table 1. Characteristics of the products (vacuum cleaners)

The attribute	Description	The data type
ENGINE_W	Engine power, W	Numerical
PRICE	Price	Numerical
FILTR_SYS	Advanced filtration system	{yes, no}
AUTOFUNC	Automatic power control	{yes, no}
AUTOCORD	Auto cord rewinder	{yes, no}
SPD_CTRL	Electronic adjustment of suction strength	{yes, no}
NOISSYS	Noise suppression system	{yes, no}
WASH	Wet cleaning option	{yes, no}
VIEW	Style and design	{yes, no}
FEATURE	Additional features	{yes, no}
BRAND	Brand	{yes, no}
SERVICE	After sales service	{low, medium, high}

Predictive Segmentation is used in situations where:

- it is possible to determine the criterion of segmentation,
- the variables are independent,
- directed data mining methods based on learning with the teacher may be used.

The data sets used in the study have been developed as a result of marketing research (surveys) conducted for company producing vacuum cleaners in years 2003-2005. The data includes vacuum cleaners characteristics. For these products, according to research goals, the tasks of classification and grouping have been carried out. Table 1 contains the input parameters (products' characteristics) for the classification and grouping processes.

An output parameter, in a prepared data set, is one of the market segments that is selected in the process of classification. It is marked in the training set with CLASS label.

Descriptive segmentation with Kohonen's Neural Networks

At first, the process of data clustering has been done in Statistica Neural Networks environment with the use of Kohonen's Neural Network.

Kohonen's Neural Network is also called the self-organizing maps (SOM) due to the way of learning – it uses unsupervised learning approach. Neural Networks developed by Teuvo Kohonen in 1982 [Kohonen,1989], [Kohonen,1990] constitute the special category of Artificial Neural Networks. These are unidirectional networks, made up of two layers. In the first layer (input layer) are only neurons corresponding to the input signals for transmission from the data source to the network. Each neuron of the input layer is connected to all output neurons (so called full network). Output layer acts as both computing and presenting the results. Neurons form a topological map, thanks to which, data clusters found in the data set may be observed.

The basic method of self-organizing networks learning, including Kohonen's network, is a competitive learning method [Kohonen,1989]. The method is based on competitive learning in which only one output neuron in the group is active at a given moment. All output units compete with one another, so this rule is sometimes called the "winner takes all". Purpose of this type of network is grouping or classification of input patterns. It is done in accordance with the principle that similar input signals trigger the same output units of the neural network. Groups are defined based on the correlation of the input data. Kohonen's network is able to process complex input signals and thanks to this it may be used to test a set of products that may have both quantitative and qualitative attributes. It was decided that during the learning process, the CLASS attribute is not taken into account. It will be treated as an attribute, which the groups found by the network will be

compared with. Disabling CLASS variable in the process of learning is necessary to ensure that the comparative analysis of clusters found by the network and market segments gathered in CLASS attribute is reliable. Besides CLASS attribute, all other variables have been selected for analysis.

The data that will be used by the neural network has to be properly prepared in advance. All the records in the data set should contain all values of attributes. In addition, all the values should be stored in the same structure for all records. The data set has met these assumptions. Therefore, the initial stage of data processing (called preprocessing) is limited to encoding the attributes' values. In the case of continuous attributes (ENGINE_W and PRICE) their maximum and minimum values in the data set are known. In addition, it is assumed that the network learned will be used to group objects which attributes' values do not exceed the scope defined in the training set. For encoding continuous attributes min-max normalization method was used. It allows to normalize the value of the variable that will belong to <0,1> interval. It is done according to the formula (2):

$$X^* = \frac{X - \min(X)}{range(X)} \tag{2}$$

where:

X* - Scaled continuous value attribute of the record,

X - The attribute value of the record before scaling,

min(x) - Minimum value of the attribute in the data set,

range(x)=max(x)-min(x) - The range of attribute values in the entire data set.

In the case of SERVICE attribute, neural network uses three input neurons corresponding to one of three possible attribute states: "service = low", " service = medium ", "service = high". The possible tags' values are true (1) or false (0). All other attributes are encoded using two values: 0 and 1. In the next step, the data set used for the analysis has been divided into three subsets: for learning, for validation and for testing. Training set is responsible for the proper modification of neuronal weights during learning process. A validation set controls learning error value, what allows the selection of the best network training algorithm, and stops the process if the symptoms of network over-training have been identified.

The validation set is a basis for an independent test to check the correctness of the network operation and will be used at the end of the learning process. For the validation and test sets, 39 cases have randomly been selected. The remaining 116 records constitute training set. Table 2 shows the number of cases of each segment divided into training, validation and testing sets.

Table 2. Number of cases in the collection used in Statistica Neural Networks

	CLASS attribute value (market segments)				Sum
Set	**m1**	**m2**	**m3**	**m4**	
Training	25	47	27	17	116
Validation	8	11	10	10	39
Testing	13	14	5	7	39
Sum	46	72	42	34	194

The assumption related to data division into training, validation, and testing sets was to keep similar proportion of training cases to total number of cases for every value of CLASS attribute. Thanks to drawing process that has been used for data set division, this assumption has been met. The result is that the training set is as representative as the entire data set. Such prepared data set was analyzed with the use of Kohonen's network. The structure of the network is shown in Figure 1. It includes one input neuron for each quantitative and binary attributes and three neurons for SERVICE attribute (one neuron for each possible value of this attribute). In the output layer there are 98 neurons arranged according to 14 × 7 configuration.

The size of the output layer has been determined experimentally. The minimum number of neurons needed to create a topological map that reflects the market segments is four (each neuron corresponds to a different segment). However, after learning the network with 2×2 output layer and analysis of its behavior it turned out that the error it makes is not acceptable. Every neuron was the winner for the cases related to different market segments. Therefore it was not possible to tag every neuron with market segment label. Extending the size of topological map enabled the identification and tagging of clusters recognized by the network.

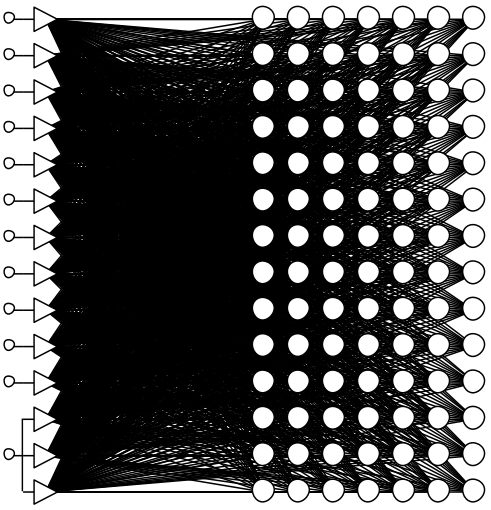

Figure 1. Architecture of Kohonen's network developed for the research

Teaching the Kohonen's network in Statistica Neural Networks environment was conducted with the Kohonen's algorithm. Learning consists of three stages. In the first one (called preliminary learning) network "is reading" the data set and initially arranges clusters on the topological map. The second stage is called tutoring or fine tuning, and provides a better fit of the output layer neurons to the training set by appropriate adjustment of the position of individual neurons. For both phases the values of three learning parameters have been experimentally determined (Table 3).

Table 3. The parameter values Kohonen network learning

	Number of epochs (periods)	**Learning rate**		**Neighborhood**	
		Initial value	**The final value**	**Initial value**	**The final value**
Initial learning	200	0.50	0.10	7	6
Tutoring	500	0.10	0.01	6	0

Number of epochs is the number of iterations of the learning algorithm. During a single iteration the network examines the training set and modifies the weights. It was assumed that in the first two stages of learning there is a need for 200 periods. In the second step, for more precise clusters identification, this number was increased to 500. Adoption of

smaller number of epochs resulted in deterioration in cluster recognition done by network. However, the greater number of iterations resulted in only a slight reduction of the error.

Learning rate is responsible for a speed of learning. The higher this value is, the stronger the network weights are modified after every iteration of the learning algorithm. It was determined that during the initial learning this factor had the greatest value. Over the iterations, the learning rate was decreasing linearly. During the process of network training the value of learning rate adopted was five times smaller than during the first stage of learning. This allows the Kohonen's network to learn more slowly but more accurately.

Neighborhood size determines the number of neurons within the winning neuron, which are involved in the adaptation process. After selecting the winner from the training set for a particular case, the weights of the winner and all neurons within the neighborhood have been modified. It causes that these neurons have a greater chance of winning the competition in the future, when similar case occurs in the training set. In the first stage of learning greater value of neighborhood has been adopted in order to enable the creation of clusters on the topological map. In the second stage of learning the neighborhood size was decreased to 0. As a result only weights of winners were modified, what has made the clusters boundaries more clear.

After learning process the Kohonen's network has been run for all cases from the data set. A large number of wins of the neuron indicates the existence of the center of cluster in the given place of the topological map. All the neurons of the output layer, which won at least once, have been named. For labeling the neurons the CLASS attribute values were used (label "mn" denotes winning neuron for the cases in the segment tagged "n"). The results of labeling process are shown in Figure 2.

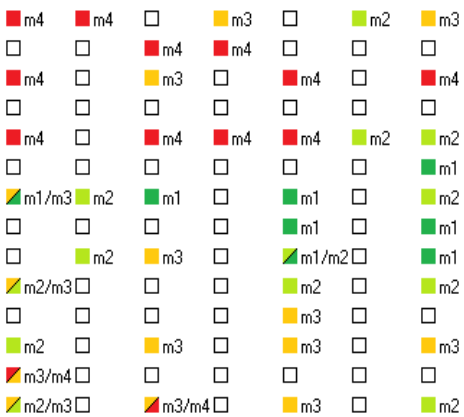

Figure 2. The topological map of Kohonen's network with segments' labels next to winning neurons

On the topological map presented in figure 2 it is possible to observe the regions, which include neurons that won for the cases from one specific market segment. However, in some areas of the map it is quite difficult to clearly determine the boundaries between clusters, because some neurons that are next to each other and have labels of different segments. What is more, some neurons won for cases belonging to two different segments of the market. Therefore, in the next step the cases that belong to hardly recognizable areas of the map have been analyzed.

During the comparative analysis of maps regions it turned out that there are groups of products with similar characteristics, yet belonging to different segments of the market. This is the reason why the map has been labeled again. The new labels reflect the boundaries of clusters identified by the Kohonen's network and now take into account the results of the comparative analysis in the areas that have initially been difficult to interpret. A topological map after re-labeling is shown in Figure 3.

In order to distinguish neurons after re-labeling, they have been marked with "cn" labels, where "n" is a number of recognized market segment. Then, based on new topological map for each case from data set new market segment has been assigned. Information about which product belongs to which market segment is stored in the data set as the CLUSTER attribute's value. This attribute is the second dependent variable (apart from CLASS variable). All analyzes have been conducted with the use of decision trees for both dependent variables. It enabled to assess what was the impact of Kohonen's network on decision trees built.

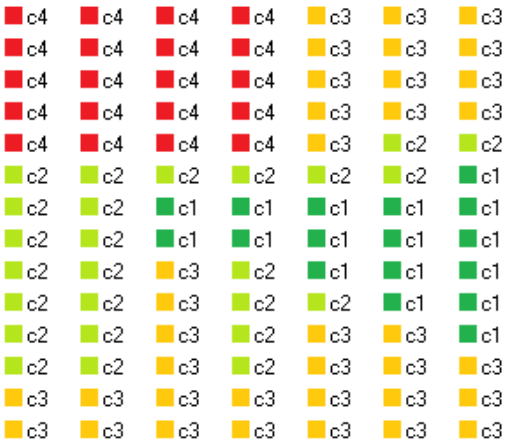

Figure 3. Kohonen's network topological map after re-labeling neurons

Classification and Regression Trees

Statistica Data Miner contains a wide range of techniques for classification as well as construction and implementation of appropriate models. For classification task one can use such techniques as: Classification Trees Models, General Classification and Regression Trees (GTrees), general CHAID models, cluster analysis, cluster analysis using generalized EM and k-means (with a cross test). In the study presented in the paper, the first models have been developed with Classification and Regression Trees (CART).

The main idea behind the classification and regression trees is to present a set of data in the form of a decision tree. Algorithms that allow to generate decision trees belong to the teacher based learning algorithms category (also known as supervised learning). In these algorithms the training set with the qualitative goal variable that divides cases from the training set into classes has to be provided. The algorithm learns what values of attributes correspond to each class of goal variable. Generated tree can classify not only cases from the training set, but it should also have the ability to generalize the knowledge gained. This allows to classify new cases, that are not present in the training data. Thus it can be a part of a decision support system.

The decision tree is a directed graph consisting of decision nodes and leaves, interconnected by branches. Each decision node corresponds to an attribute of the data set. The branches coming out from the node symbolize the possible values of a given attribute. Each branch may lead to the next decision node or a leaf node. Decision nodes divide the data set into subsets. The leaves contain a solution of the classification problem, which assigns a specific class to the subsets of cases. In the decision nodes there are variables that are able to divide the data set according to the goal variable.

Classification and Regression Tree (CART) is an algorithm that can generate a strictly binary tree, where each node can divide the data set into two subsets only. Trees of this kind are rarely used in solving marketing problems. Nevertheless it was decided to compare the predictive abilities of the CART trees created for CLASS and CLUSTER dependent variables. For every variable the distinct trees with the similar level of complexity (with regard to number of decision nodes and leaves) have been selected for analysis.

Program generated 8 CART decision trees with different degree of pruning for both CLASS and CLUSTER dependent variables. The pruning of decision tree protects it from overfitting, the situation in which tree depends too much on irrelevant features of the training instances, with the result that it performs well on the training data but relatively

poorly on unseen instances. CART algorithm is able to properly trim the nodes and branches of the tree. The resulted tree makes bigger classification error but it is less complicated. Lower tree complexity results in the ability to knowledge generalization, what allows for better classification of unknown cases during learning. In order to select the best pruned tree V-fold cross-validation was used. It is designed especially for small data sets, because it does not require a separate set of test cases isolated from the training set. It draws the V samples from the training set, which will serve as test sets. Then they are tested with varying degrees on trees with different degree of pruned branches. The result is the cost of cross-validation calculated on the basis of the results of classification of each of the trees. As best pruned tree is selected the one that is the least complex and at the same time its cost of cross-validation is the smallest one. The program has selected the models for CLASS and CLUSTER variables with regard to the selection criteria described above. Table 4 presents the results.

Histograms placed in the leaves nodes represent the distribution of the dependent variable for training cases assigned to the leaf. In case of tree for the dependent variable CLASS, the highest classification uncertainty may cause the leaves with numbers 18, 26, 30, 31. Their histograms show that with such complexity of the decision tree, several cases are classified into other segments of the market than they have nominally been assigned. It was decided to analyze these training cases. They have been compared with the cases put by the Kohonen's network in the areas of the topological map that were difficult to describe. It turned out that several records assigned to four of these leaves also belong to these regions of a topological map of the network. This example shows that both the Kohonen's network and the CART algorithm have found a group of products that are difficult to unambiguously classify to one of the four segments described in CLASS variable.

Histograms placed on the tree for the CLUSTER dependent variable show that cases have been classified with higher level of certainty. None of them indicates that to one leaf similar number of cases belonging to different classes has been assigned. Although the CART algorithm used the same predictors during the creating of both two trees, variables that were included in the decision nodes are different. PRICE is the attribute that differentiates CLASS variable the most. For the CLUSTER variable there is SPD_CTRL attribute located at the root of the tree. This also reflects the different approach to the classification of products for both trees.

Comparing the performance of CART trees presented in Table 4 it can be concluded that a better ability to predict has a tree created for CLUSTER variable. It is backed by cross-validation cost which is twice lower than the cost of CLASS variable. In addition, the model for the CLUSTER variable is simpler. The CART algorithm found more easily

the interdependences between attributes when analyzing a set of data for the CLUSTER dependent variable. In the case of CLASS variable, the algorithm needed more attributes to map the data set, and despite of this the tree makes larger predictive error.

Table 4. Summary of selected parameters CART trees

Dependent variable	Number of leaves	Number of nodes making	The cost of cross-validation
CLASS	12	11	0,288660
CLUSTER	8	7	0,113402

Two trees have been generated with the use of the same algorithm, where each of them describes the same data sets. Only difference is in the dependent variable values. Despite the fact that the dependent variable of the second CLUSTER was created on the basis of CLASS dependent variable, these two trees vary considerably with regard to nodes arrangement as well as the number of erroneous predictions.

Chi-Squared Automatic Interaction Detector

CHi-squared Automatic Interaction Detector (CHAID) is one of the oldest algorithms to create decision trees [Kass, 1980]. CHAID tree can be used for classification and regression tasks. CHAID algorithm is very efficient for large data sets. It creates non-binary trees, that is those in which from decision node can go more than two branches. This is the primary feature that enables to distinguish CHAID trees from CART trees.

To select the variables that are included in decision nodes, the algorithm uses the chi-square test. Important parameters that should be defined in the construction of tree are stop conditions. The modification of these parameters affects the degree of pruning the tree. These parameters include such values as: the minimum number of training cases in the node, which is subject to division, the probability of splitting and merging categories and the maximum number of tree nodes. It is assumed that the minimum number of cases for the divided node will be 19. Other parameters have been determined experimentally. The aim was to generate the tree which has the size similar to the size of CART tree previously created for CLASS dependent variable.

It was determined that the probabilities of splitting and merging categories are the same and equal 0.01. These values allowed to generate trees for the CLASS and CLUSTER variables.

When analyzing the CHAID tree for the CLASS variable it can be seen that the histograms placed in the leaves of the tree reveal the relationship, which has already occurred in the

CART tree for this variable. Some leaves allocate cases belonging to different market segments to the same class. This trend has been preserved despite the fact that both the trees are different with regard to nodes arrangement and the variables selection. The root of CHAID tree is the VIEW attribute. CHAID algorithm used to build the tree such variables as BRAND and SPD_CTRL. These variables were not present on the CART tree. Comparing CART and CHAID trees for the CLUSTER variable few differences may be observed. There are the sizes of the trees, the ways the attributes have been selected in the nodes and the presence of leaves that connect cases belonging to different classes. These leaves cause greater prediction uncertainty, already recorded in the CLASS variable. Analyzing the number of cases attributed to the wrong class, it can be said that the CART tree better adapted to the training data. A common feature of two trees is their binary nature, because the CHAID algorithm has not created decision nodes that divide the data set into more than two parts. In case of CHAID trees V-fold cross validation provides the value which is known as the risk of estimation. This measure substitutes the cost of cross validation which is used for CART trees validation. The basic parameters of both trees are presented in table 5.

Table 5. Summary of selected parameters of CHAID trees

Dependent variable	Number of leaves	Number of nodes making	Risk of estimation for cross-validation
CLASS	14	10	0,243523
CLUSTER	11	12	0,144330

Boosted Trees

Boosted Trees are one of the newest methods of data mining. They can be used for quantitative and qualitative variables for regression as well as classification. In this method, the data model is created using the basic decision tree. Although each tree when considered separately gives a large classification error, together they form a model that has an excellent ability to predict. All trees are usually binary trees, and therefore each decision node divides the data set into two subsets. Algorithm is trying to determine the best distribution of data (to create a single binary tree) and calculates the residues for each division, the deviations observation from the mean values. Then another tree is built, which fits the calculated residuals and makes another division of the data set. With every new tree added the error related to a whole sequence of trees is reduced.

Boosted Tree algorithm is able to recognize even very sophisticated relationships among variables and thanks to this is the data fit is perfect. As with other methods, it is adverse

phenomenon, which reduces the prediction correctness of the model. Boosted Tree algorithm solves this problem by creating for each iteration two sets: training and testing. The cases belonging to both sets are randomly selected from the data set. Training cases are used to generate another tree that is added to the trees sequence and will be used for the prediction of residuals calculated for preceding trees in the sequence. Creating trees for subsequent samplings results in the decrease of prediction error and is called stochastic gradient reinforcement. Test cases are not used for creating the tree. They are applied to the process of model verification in a given iteration.

When the model was created two parameters used by the Boosted Tree algorithm were modified: learning factor (responsible for the speed and accuracy of learning) and the size of the test set (the proportion of the number of test cases to the number of all cases in the data set). Several models have been created which differ with regard to the number of trees and the prediction error. In order to allow comparison of models for the CLUSTER and CLASS variables and adopted the same learning parameters were assumed for both models. Experimentally determined value of the learning factor is 0.1 and the size of the test set is 0.3.

While generating both models graphs were created. Figure 4 presents the graph for the CLASS variable and figure 5 - for the CLUSTER variable. In both cases, it can be observed that every new tree added to the model reduces the training set error. However, there is a point in which the test set error is not decreasing any more. When a certain number of trees is exceeded there is an increase in the value of the test sample error, what indicates the occurrence of overfitting. The optimal number of trees in the model is the value at which the test set error is minimal.

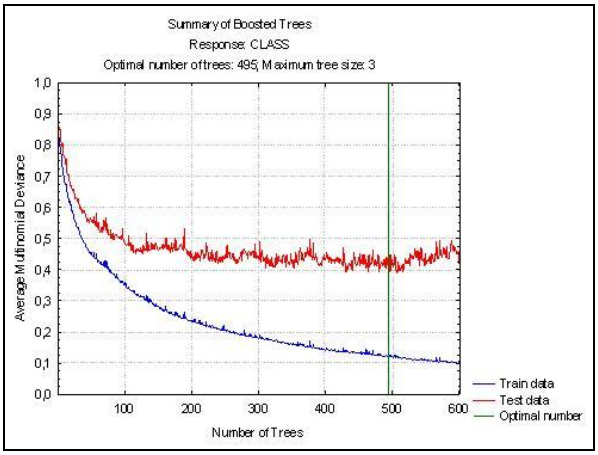

Figure 4. Prediction errors for learning and testing sets in relation to the number of trees in the model for the CLASS dependent variable

It is hard to compare Boosted Tree models with models developed with CART and CHAID algorithms, because their architectures are different. However, the analysis of sequences of trees created generated by Boosted Tree algorithm reveals similarities with the CART models. The first tree in the trees sequence generated for the CLASS dependent variable contains PRICE attribute in the decision node, and the first tree generated for the CLUSTER variable contains SPD_CTRL the attribute. These same attributes are in the CART tree roots for both dependent variables. This means that in case of the analyzed data set both algorithms, Boosted Tree algorithms as well as CART algorithm select variables a similar manner. The parameters of both models generated by Boosted Tree algorithm are summarized in Table 6.

Table 6. Summary of selected parameters of Boosted Trees algorithm

Dependent variable	Learning Ratio	The size of the training set	Number of trees	Risk estimation for cross-validation
CLASS	0,1	0,3	495	0,160714
CLUSTER	0,1	0,3	540	0,035714

As can be seen, the same learning parameters adopted, gave different number of trees and different fit to the data. Risk estimation indicates that the sequence of elementary decision trees for the CLUSTER variable will make the smaller prediction error than the corresponding model for the CLASS variable. This may be confirmed by the graphs created during the construction of models. The application of CLUSTER attribute as a dependent variable resulted in data set that is more "convenient" to learn. The relationships between the variables in this set have been pre-ordered by the Kohonen's network. In case of CART and CHAID decision trees as well as Boosted Tree algorithm early identification of clusters on topological map allowed to easier classification of uncertain records from the data set and therefore reduction of prediction error.

Conclusions and Further Research

For performing analyzes of decision trees, data set describing household products was used. Each case in the data set regards one of the four segments of the market, described with the CLASS dependent variable. Before starting to build decision trees models a grouping of data set with the use of Kohonen's network has been carried out. The aim of the grouping was to identify data clusters corresponding to the market segments. After clusters have been found, all characteristics of products placed by

the network in the areas of topological map that were difficult to identify have been analyzed.

As a result of the analysis the market segments have been assigned again to the products. Information about to which market segment each product belongs has been stored in an additional dependent variable called CLUSTER, which has been included in the analyzed data set. All analyzes were performed for market segments initially defined in the CLASS variable and for market segments described by CLUSTER variable and ordered by Kohonen's network.

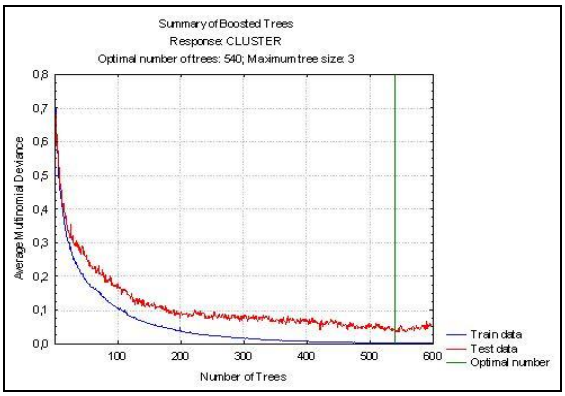

Figure 5. Chart of prediction error learning and test sets relative number of trees in the model for the dependent variable CLUSTER

In conducted analyses, in order to build decision trees, CART, CHAID and Boosted Tree algorithms have been used. These algorithms were used to build predictive models for two dependent variables. Every model developed was assessed with Statistica Data Miner metrics. These metrics include cross-validation cost and risk estimation. The results of the analysis and comparisons of the developed models have been presented in the previous sections. It can be said that all kinds of trees have made the largest predictive error for classifications done for CLASS dependent variable. In case of CLUSTER variable, prediction errors for the test sets (or the V-fold cross-validation) were smaller. This could be due to the fact that the Kohonen's network initially recognized and ordered the interdependences between the attributes that affect the membership of market segments.

The smallest error was made by Boosted Tree algorithm, what confirms that it is the most advanced approach from all the models analyzed. Boosted Tree algorithm adapts very well to the training data, and controls the symptoms of overfitting.

One of the advantages of the Statistica Data Miner environment is that it is possible to quick develop the models with the use of automatic code generator (C / C + +, Statistica

Visual Basic or PMML). After generating the code, all programs may run concurrently in order to determine the value of the dependent variable. This is done thanks to technique called models aggregation. In the case of qualitative dependent variables the voting is used. During the voting process the separate prediction is made by each of the models. As a final result only one class is selected that has been recommended by majority of models. The following tables (table 7 and table 8) show the classification matrix of aggregated model (voting based prediction) as well as decision trees.

Table 7. Classification matrix for models developed for CLASS dependent variable

Observed class	Model	Predicted class			
		m1	m2	m3	m4
m1	CART Tree	28	3	3	
	CHAID Tree	28	5	1	
	Boosted Trees	31	3		
	Voted prediction	28	2	4	
m2	CART Tree	6	32	4	
	CHAID Tree	6	27	9	
	Boosted Trees	1	37	4	
	Voted prediction	4	35	3	
m3	CART Tree		12	58	2
	CHAID Tree	4	7	60	1
	Boosted Trees		3	68	1
	Voted prediction		6	64	2
m4	CART Tree			1	45
	CHAID Tree			1	45
	Boosted Trees				46
	Voted prediction			1	45

After comparing values from both tables it is possible to notice that the aggregated model is trying to average of the prediction errors of individual decision trees. Predictions of this model are worse than the best classifier results (Boosted Tree), but it gives better results than CART and CHAID trees. For several classes the results for aggregated model are the same as those for CART tree. If you want to use an intermediate model and avoid generating aggregated models for data set, CART algorithm should be used. What is more, the aggregated model demonstrates the same tendency as decision trees acting separately do. This model better classifies the cases according to CLUSTER dependent variable. This confirms the positive impact of the Kohonen's network on analyzed data set.

Table 8. Classification Matrix for models developed for the CLUSTER dependent variable

Observed class	Model	Predicted class			
		c1	c2	c3	c4
c1	CART Tree	34		2	
	CHAID Tree	36			
	Boosted Trees	36			
	Voted prediction	36			
c2	CART Tree		37	6	
	CHAID Tree	3	37	3	
	Boosted Trees		43	1	
	Voted prediction		37	6	
c3	CART Tree		3	77	1
	CHAID Tree	3	3	74	1
	Boosted Trees		1	79	1
	Voted prediction		3	77	1
c4	CART Tree		3	1	30
	CHAID Tree			4	30
	Boosted Trees				34
	Voted prediction			1	33

The developed models of decision trees can also be used in combination with neural networks. Neural networks and decision trees are two types of predictive models that differ with regard to many features. The combination of such discrete models is possible thanks to a technique called stacking. It involves the use of at least two models to classify cases from the same data set. Then the results of the classification are passed to the next model, which tries to merge the results and provide a final solution to the problem of prediction.

In the case of developed models it is possible to make separate predictions with CART, CHAID, and Boosted Tree algorithms. The results generated by each of them can be transferred to the neural network input. The network will learn how to combine the results of individual trees so that the final predictive model gave the best possible results. This approach is called "meta-learning", as the collective model learns on the basis of what other models learned before.

Another way of combining decision trees and neural networks may be to use tree-building algorithms to determine what attributes of the data set are the most significant. These attributes can divide the data set in the best possible way and therefore they are placed in the nodes of decision tree. The higher given attribute is located in the tree (closer to the root of the tree), the better the data set is divided. The most significant

variables can be used as inputs to the neural network. After learning process based on such "truncated" data set, network can easily recognize all interdependences between variables. It will result in a better ability to predict new cases, which are not present in the training data set.

Bibliography

[Kotler Ph., 1984] Kotler Ph.: Marketing Essentials, Prentice-Hall International, Englewood Cliffs, 1984, 733 pp.

[Adamczak, 2001] R. Adamczak. Zastosowanie sieci neuronowych do klasyfikacji danych doświadczalnych, Praca doktorska, Uniwersytet M. Kopernika, Toruń, 2001.

[Hand David et al., 2005] Hand David, Mannila Heikki, Smyth Padhraic: Eksploracja danych, WNT, Warszawa, 2005

[Cios et all, 1998] Cios K., Pedrycz W., Świniarski R. Data mining methods for knowledge discovery, Kluwer, Norwell MA, 1998

[Czogala et al., 2000] E.Czogala, Lęski J.: Fuzzy and Neuro-Fuzzy Intelligent Systems, Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York, 2000

[Jang et al., 1997] Jang S.R., Sun C.T., Mizutani E.: Neurofuzzy and Soft Computing, Prentice-Hall, Upper Saddle River 1997, p. 245.

[Kass et al., 1980] Kass, Gordon V.; An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics, Vol. 29, No. 2 (1980), pp. 119–127

[Kohonen, 1990] Kohonen T.: Self-organizing Maps, Proc. IEEE, 1990, 78, NR.9, pp. 1464-1480.

[Kohonen,1989] Kohonen T.: Self-organization and associative memory, Berlin, Springer Verlag,1989

[Larose, 2006] D. T. Larose, Odkrywanie wiedzy z danych, Wyd. Nauk. PWN, Warszawa, 2006

[Li S., 2000] Li S.: The Development of a Hybrid Intelligent System for Developing Marketing Strategy, Decision Support Systems, 2000, Vol 27, N4

[Migut, 2010] Migut G.: Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku, StatSoft, Materiay Seminarium StatSoft "Zastosowanie nowoczesnej analizy danych w marketingu i badaniach rynku", Kraków, 2010.

[Moon et al., 1998] Moon Y.B., Divers C.K., and H.-J. Kim: AEWS: An Integrated Knowledge-based System with Neural Network for Reliability Prediction // Computers in Industry, 1998, Vol.35, N2, pp.312-344.

[Mynarski, 2010] Mynarski S.: Metody ilościowe i jakościowe badań rynkowych i marketingowych, StatSoft, Kraków, 2010

[Nauck et al.,, 1997] D.Nauck, F. Klawonn, R.Kruse: Foundations of Neuro-Fuzzy Systems, J.Wiley&Sons, Chichester, 1997.

[Rutkowska, 2000] Rutkowska D.: Implication-based neuro-fuzzy architectures.- Applied mathematics and computer science, V.10, N4, 2000, Technical Unieversity Press, Zielona Gora, 675-701.

[Setlak, 2001] Setlak G.: „Fuzzy Neural Networks in Intelligent Manufacturing Systems", // Proceedings of the International Workshop on Intelligent Data Acquisition and Advanced Computing Systems, Copyright Clearence Center IEEE, Piscataway (USA), Ternopil (Ukraine), Foros, 2001, pp.203-206. and http://ieeexplore.ieee.org/servlet/opac?punumber=7498

[Setlak, 2008] Setlak G.: The Fuzzy-Neuro Classifier for Decision Support //International Journal Information Theories and Applications, Pub. of the Institute of Information Theories and Applications FOI ITHEA, Sofia, 2008, Vol. 15, N.1, PP.22-28.

[Setlak, 2004] Setlak G.: Intelligent Decision Support System, // LOGOS, Kiev, 2004, (in Rus.), pp. 250.
[Setlak, 2000] Setlak G.: Neural networks in intelligent decision support systems for management // Journal of Automation and Information Sciences, Kiev, N1, 2000r., pp. 112-119.
[Stąpor, 2011] Stąpor K.: Metody klasyfikacji obiektów w wizji komputerowej, Wydawn. Naukowe PWN, Warszawa, 2011
[Stąpor, 2005] K. Stąpor. Automatyczna klasyfikacja obiektów, Wyd. Exit, Warszawa, 2005.
[Stanimir et al., 2006] Red.: A. Stanimir: Analiza danych marketingowych. Problemy, metody, przykłady, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2006
[StatSoft, 2010] Praktyczna analiza danych w marketingu i badaniach rynku, materiały z konferencji StatSoft Polska, wrzesień 2010
[Szczuka, 2000] Szczuka M.: Metody symboliczne i sieci neuronowe w konstrukcji klasyfikatorów, autoreferat rozprawy dokt., Uniwersytet Warszawski, Warszawa, 2000
[Lotfi Zadeh, Rutkowska, 2000] Ed. By Lotfi A.Zadeh and D. Rutkowska: International Journal of Applied Mathematics and Computer Science, Special Issue: Neuro-fuzzy and soft Computing.– Zielona Gora : Technical University Press, Vol. 10, N4, 2000.
[Zieliński, 2000] J. Zieliński. Inteligentne systemy w zarządzaniu – teoria i praktyka, PWN, Warszawa, 2000.
[Żurada, 1996] J.Żurada, M.Barski, W.Jędruch: Sztuczne sieci neuronowe, Warszawa: PWN, 1996, 375 pp.

Authors' Information

***Łukasz Paśko**, **M.Sc., Eng** – Rzeszow University of Technology, Department of Computer Science, The Faculty of Mechanical Engineering and Aeronautics, Powstancow Warszawy ave. 8, 35-959 Rzeszow, Poland;*
e-mail: lukasz.pasko48@gmail.com

Major Fields of Scientific Research: artificial intelligence, decision support systems, data mining.

Galina Setlak- **D.Sc, Ph.D., Associate Professor**, *Rzeszow University of Technology,* Department of Computer Science, *Powstancow Warszawy ave. 8, 35-959 Rzeszow, Poland,*
e-mail: gsetlak@prz.edu.pl

Major Fields of Scientific Research: decision-making in intelligent manufacturing systems, knowledge and process modeling, artificial Intelligence, neural networks, fuzzy logic, evolutionary computing, soft computing.