Integrated Clustering Modeling with Backpropagation Neural Network for Efficient Customer Relationship Management

Tijen Ertay[1*] and Bora Çekyay[2]

[1] Istanbul Technical University Faculty of Management, Management Engineering Department, Macka, 34367, Istanbul, Turkey
[2] Istanbul Technical University Faculty of Management, Industrial Engineering Department, Macka, 34367, Istanbul, Turkey

1 Introduction

The rapid progress in digital data acquisition and storage technology has lead to the fast growing tremendous and amount of data stored in databases, data warehouses, or other kinds of data repositories such as the World Wide Web [40].

The advent of the network world induced by the rapid development of the Internet and the accompanying adoption of the Web has promoted the changes to create greater business opportunities and to reach customers more easily. This situation is required the capability to both generate and collect data and this capability has been expanded enormously and provides us with huge amounts of data. Although valuable information may be hiding behind the huge amounts of data, it makes difficult for human to extract them without powerful tools. *Data Mining (DM)* emerged to extract knowledge from huge volumes of data with the help of computing device during the late 1980s. *DM* has become a research area with increasing importance with the amount of data greatly increasing [5, 7, 9, 27].

Due to its interdisciplinary nature, data mining has received contributions from many disciplines such as databases, machine learning, statistics, information retrieval, data visualization, and parallel and distributed computing etc. The field of *DM* is a new discipline in engineering and computer science to address these new opportunities and challenges. Industrial Engineering (IE), with the diverse areas it encompasses, presents unique opportunities for the

[1*]Corresponding Author: ertay@itu.edu.tr

Tijen Ertay and Bora Çekyay: *Integrated Clustering Modeling with Backpropagation Neural Network for Efficient Customer Relationship Management*, Studies in Computational Intelligence (SCI) **5**, 355–373 (2005)
www.springerlink.com

application of *DM* and for the development of new concepts and techniques in this field. Specific techniques used in *DM* applications include market basket analysis, memory based reasoning, cluster detection, link analysis, decision trees, neural networks and genetic algorithms.

Reference [3] defines data mining as the exploration and analysis, by automatic or semiautomatic means of large quantities of data in order to discover meaningful patterns and rules. Reference [12] defines data mining as the process of discovering interesting knowledge from large amounts of data stored either in data bases, data warehouses, or other information repositories. Reference [13] considers data mining as the analysis of observational data set to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. DM business applications can be found in a diverse group of business including for example, banks [32], healthcare [20], insurance [15].

Taxonomy of DM tasks can be broadly divided into five categories: dependency analysis, class identification, concept description, and data visualization. The taxonomy reflects the emerging role of data visualization as a separate DM task, even as it is used to support other data mining tasks. Different DM tasks are grouped into categories depending on the type of knowledge extracted by the tasks. The identification of patterns in large data set is the first step to gaining useful marketing insights and making critical marketing decisions. The DM tasks generate an assortment of customer and market knowledge, which form the core of knowledge management process. Knowledge management is prerequisite for e-business and its increasing customer centric focus. DM is usually used to answer two main types of application questions: to generate predictions on basis of available data; to describe behavior captured in the data [8].

To answer the first type of question, classification can be used as one of the most popular approaches. Classification and clustering are prime targets of most empirical research of the real world and of DM. These aim to group entities or objects into classes so that there is maximum intra-class proximity between members, and maximum inter-class proximity between members and maximum inter-class distinction among groups. Commonly, a clustering model provides a representation scheme in the form of a data structure, an index to calculate similarity and a grouping technique. There are many different methods, which may be used data predict the appropriate class for situations. Among the most popular one are logistic regression, case-base reasoning, neural networks, decision trees, and rule induction [17].

Especially, neural networks have been applied to almost all types of data mining applications including forecasting [29], credit approval problems [19], target marketing [25], cost comparison of assembly systems [30, 31], development of maintenance policies [2].

In general, development of good neural network applications can be very time consuming and requires the building, training and testing of many different network structures to arrive at an efficient model. Neural networks are characterized by learning capability, the ability to improve performance over time. A closely related feature is that of generalization, relating to the recognition of new objects, which are similar but not identical to previous ones. A typical neural network consists of a number of simple processing elements called neurons. Each neuron is connected to other neurons by means of directed information being used by the net to solve a problem. Neural networks are usually modeled into one input layer, one or several hidden layers and one output layer [34].

The most popular neural network method for practical applications is bacpropagation algorithm. In spite of its simple structure, the presence of one or more hidden units, together with a non-linear activation function, gives it the ability to solve many complex problems. A backpropagation neural network (BP) is the most common multi-layer network estimated to be used in percent 80 of all applications.

The research presented in this study emanates from the apparent use of neural network modeling for customer segmentation and the apparent primary advantages of self-organizing mapping (SOM) neural networks over cluster analysis. SOM network has been considered as one of the most popular unsupervised competitive neural network learning models, for clustering and visualization in a number of real world problems [18].

The primary objectives of the research are to consider the use of SOM NN for segmenting customers and to analyze the predictive ability of BP NN for classifying customers from follow-up surveys by using the output provided by SOM NN. Data collected by the *Group Lens Research Project* at the University of Minnesota (http://www.cs.umn.edu/research/GroupLeans/data) are used in this research. The rest of this study is organized as follows. Section 2 provides an overview on customer relationship management. Section 3 includes the backgrounds and the methodology to be used in this study. Section 4 provides the results of modeling process based on the used methodology. Section 5 is related to the conclusions of this study.

2 Customer Relationship Management: An Overview

Among business practitioners and marketing scientists today, there is an increasing interest in customer relationship management (CRM). The focus of CRM is to forge closer and deeper relationships with customers, "being willing and able to change your behavior toward an individual customer based on what the customer tells you and what else you know about the customer" [28].

In this manner, CRM is an enterprise wide business strategy designed to optimize profitability, revenue and customer satisfaction by organizing the enterprise around customer segments, fostering customer-satisfying behaviors and linking processes from customers through suppliers [24]. It requires a clear focus on the service attributes that represent value to the customer and create loyalty [14].

CRM deals with the mechanics of building relationships, including data capture and analysis and business process simplification. Proponents of CRM argue that may benefits can be derived from implementing CRM; including sales functionality by developing customer profiles and history; customer service support through warranty management, tracking and problem resolution; cross selling and upselling higher – margin products or services to targeted customer segments; and attracting additional customers by offering personalized service such as direct mail outs. Another key benefit claimed by the CRM industry is that customers are segmented and communication programs are developed to retain the most profitable customers. With CRM, emphasis is placed on selling more products and services through data mining to determine the types of customers that would be most likely to buy a particular product. This is achieved by developing sophisticated predictive models that assess a segment's propensity to purchase product based on the purchasing behavior of individuals with similar demographic other profiles. Segmentation is product focused rather than customer focused and does not consider any element of the emotional connection with the customer. The relationship is driven entirely by historical behaviors. CRM does have the capability to allow companies to better understand customer purchasing behavior, or at least that portion that is captured in the system, and to determine the type of communications that should be undertaken with the customer [1].

CRM is essentially a two-stage concept. The task of the first stage is to master the basics of building customer focus. This means moving from a product orientation to a customer orientation and defining market strategy from outside-in and not from inside-out. The focus should be on customer needs rather than product features. The second stage includes company's development of customer orientation by integrating CRM across the entire customer experience chain, by leveraging technology to achieve real-time customer management, and by constantly innovating their value proposition to customer. The goal of CRM is to identify a customer, understand and predict the customer-buying pattern, identify an appropriate offer, and deliver it in a personalized format directly to the customer. CRM means that companies manage relationships with individual customers with the aid of customer databases and interactive and mass customization technologies. The adoption of CRM has been enhanced by developments in information and communication technology (e.g., Database Technology, E-commerce and the Internet) [28, 35].

The first step in CRM is based on information processing about company's customers through internal customer data or the purchased data from outside sources. An enterprise data warehouse is a critical component of a successful CRM. By using an efficient enterprise data warehouses, companies can invest in the customers that are potentially valuable for the company, but also minimize their investments in non-valuable customers. Hence, customers are categorized by the enterprise in terms of their different values and served with different relationship strengthening practices. In other words, on the turnover of each customer or customer profitability can be used as segmentation variables to distinguish between valuable and nonvaluable customers. Besides, considering customers' demographic information, it makes possible the segmentation according to the customer values. This indicates what direction customers' preferences to move. In this situation, it can be said that the *customer value* (CV) provides a good framework for applying data mining to CRM. On the "input" side of data mining, the customer value tells what is likely to be interesting. In general, marketers say there are three ways to increase a customer value – increase their use (or purchases) of products they already have – sell them more or higher margin products – keep the customers for a longer period of time. However, to obtain information on the potential value of a customer, analysts need on the customer's purchasing behavior at their own company, as well as at other companies in the market. Usually, companies only have data on customers' purchasing behavior at their own company in their customer information file. Hence, models are needed to predict the potential value of a customer based on the purchasing behavior and any available socio-demographic data. Behavioral clustering and segmentation help drive strategic marketing initiatives, while sub-segments based on demographic lifestyle or value characteristics could also be determined and used for tactical marketing.

The process of establishing a traditional static CRM begins with the identification of customers of a retailer and the construction of a customer purchase data mart from customer history of purchases within a certain time period. Once the customer purchase data mart has been built on the enterprise intranet, the summarized data such as the time of purchase, quantity, frequency, and product's rating are extracted from the customer information of the data mart for mining customer purchase behavior patterns. According to the calculated CV, with the aid of the SOM, the customers can be divided into a set of segments of similar customers with similar CV. The last component of a CRM is campaign execution and tracking. Implementation of decisions made as a result of data mining is based on campaign execution and tracking. Campaign execution is based on recommendation about company's products for the define segments. Tracking is also based on getting users' opinions about the promotions made for the company's products. These are the processes and systems that allow the user to develop and deliver targeted messages in a test-and-learn environment.

3 Backgrounds and Methodology

The first part of this section will be related to a discussion on customer value and a segmentation method for CRM that uses customer value. Next, it will be considered the methodology that combines SOM being one of the most popular unsupervised competitive neural network learning models for clustering and backpropagation neural networks (BP) for predictive classifying the customers from follow-up surveys by using the output provided by a SOM.

3.1 Customer Value Concept

Customer Value (CV) has turned out to be a very important concept in marketing strategy and research in spite of the fact that the growing body of knowledge about the construct is fragmented. Costumer value measurement is a strategic marketing tool to clarify a company's proposition to its customers, thus creating a differential superior offering compared with the competition. The tool assesses a company's performance in comparison with its main competitors as perceived by former, present, and potential customers. Reference [39] has proposed that customer value was a customer's perceived preference for and evaluation of those product attributes, attribute performances, and consequences arising from use that facilities achieving the customer's goals and purposes in use situations. There are some areas of consensus about customer value. First, customer value is inherent in or linked to the use of a product. Second, customer value is something perceived by customers rather than determined by sellers. Third, these perceptions include a trade-off between what the customer receives, and what he/she gives up in order to acquire and use a product. Despite the increasing attention being focused on customer value, extant definitions of the construct are somewhat ambiguous, because they typically rely on other terms such as utility, worth, benefits and quality which are too often not well defined [26]. In general, definitions take a rather narrow perspective, because value frequently is measured as attribute-based desires or preferences, which are influencing purchase. Besides, researchers have assessed extensively the construct of value in marketing. Reference [10] proposed that value can be approached in three different concepts: values, desired values and value judgments. Value is defined as centrally held enduring core beliefs, desired end states, or higher order goals of the individual customer or customer organization that guide behavior. Desired customer value, to the contrary, are the customers' perceptions of what they want to have happen in a specific kind of use situation, with the help of a product or service offering, to accomplish a desired purpose or goal [38]. Value judgment is "the customer's assessment of the value that has been created for them by a supplier given the trade-offs between all relevant benefits and sacrifices in a specific-use situation". For this situation, value judgment is the most important for a definition of customer-perceived value.

The most important success factor for a film is the ability to deliver better customer value than the competition. Good customer value can be achieved only when product quality, service quality, and value-based prices are in harmony and exceed customer expectations. Maximizing customer value must flow from a firm's culture, beliefs, values, management style, reward systems, and structure. As a concept, customer value is fairly simple. But since it is

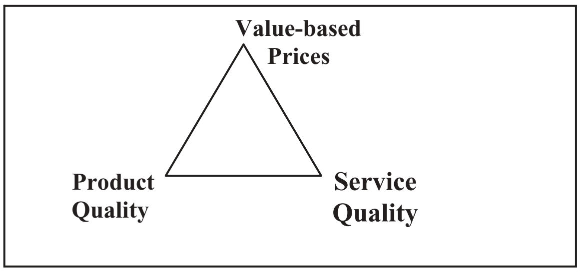

Fig. 1. Customer Value Triad

defined by the customer, it is inherently ambiguous. The customer value triad as seen in Fig. 1 consists of only three things: product quality, service quality and value based prices. For this reason, customer value is created when customer expectations in any one of the three areas, the firms have not delivered good customer value. Only when all three are in harmony will customer value be maximized. Therefore a firm must implement an organized system to capture customer expectations of product quality, service quality and reasonable price. The first subject is that customer expectations must be formed relative to competitor alternatives.

A firm can steadily improve product and service quality, and reduce price and still convey less value than competitors. The competitors may simply be improving faster. The second subject is that customer expectations are dynamic. Therefore, determining customer expectations of product quality, service quality and price relative to the competition should be part of a continuous, ongoing program. The third subject is that product and service quality must extend throughout the channel. Upstream, suppliers must adhere to higher and higher quality standards. Downstream, all of the channel intermediaries must be coordinated to deliver maximum value to the end user. Maximizing customer value would achieved by achieving superior product and service quality at a price competitive with the competition ([23] pp. 15–24). Besides, the expected customer value is ultimately determined by the comparison of expected benefits to expected sacrifice. Decreasing the sacrifice while holding benefits constant also enhances customer value. Increasing the benefits faster than the sacrifice enhances customer value too ([23] p. 117).

3.2 Self-Organizing Neural Network for Customer Segmentation

In CRM, managers develop specific strategies for different segments of their customer base. Segmentation proposes that groups of customers with similar needs and purchasing behaviors are likely to demonstrate a more homogenous response to marketing programs that target specific consumer groups. A critical issue to successful market segmentation is the selection of segmentation (clusters) variables. Segmentation variables can be expanded to general variables and product specific variables [36]. General variables have been used in many researches because these ones are intuitive and easy to operate [11]. Besides, [22] proposed artificial neural network clustering method incorporating both clusters and segment discriminate analysis to estimate the relationship between customer demographics. Clustering is prime target of most empirical research of the real world and of data mining. That aims to group entities or objects into classes so that there is maximum intra-class proximity between members and maximum inter-class distinction among groups. Customer segmentation based on general variables is more intuitive and easier to conduct than product specific variables. However, the assumption that customers with similar demographics will exhibit similar purchasing behavior is doubtful. Each customer pursues personalized products and services even within groups with similar demographics in order to present uniqueness and identity. This makes customer's purchase patterns difficult to determine using only general variables. Self-Organizing Map (SOM) can be used for market segmentation. The versatile properties of SOM make it a valuable tool in data mining. SOM is designed to do unsupervised clustering of data; i.e. given training patterns that contain inputs only, the SOM assigns output units that represent clusters to inputs. The SOM not only assigns cluster centers to units, but also it tends to group similar centers on units close together, thus giving on impression of the relationship of the clusters and the topological behavior of the network. Clustering customers via a SOM is divided into two phases. On phase is to train a SOM using the customer value patterns. The other is to map input customer value patterns to output customer segments. When an input pattern is imposed on the network, an output node is selected from among all the output nodes as having the smallest Euclidean distance between the presented input pattern vector and its weight vector. This output unit is declared the winner in the competition among all the neurons in the output layer. Only the winning neuron generates an output signal from the output layer. All the other neurons have a zero output signal. As learning involves adjustment of weight vectors, only the neurons within the neighborhood of the winning neuron are allowed to learn with this particular input pattern. The neighborhood means the physical proximity of the neurons to the winning neuron. Learning for the nodes within the neighborhood is carried out by adjusting their weights closer to the input vector. The size of the neighborhood is initially chosen to be large enough to include all units in the output layer. As learning proceeds, the size of the neighborhood is progressively reduced to a pre-defined limit. Thus during these stages, fewer neurons have their weights adjusted closer to the input vector. Data variables representing continuous data (e.g. rating of costumer) are scaled to assume a value between zero and one, while dichotomy variables (e.g. male/female) are recoded to assume values of zero or one. In this manner each data is considered by the network as binary-valued input. In order to create the training and test sets, the data set is randomly, nearly sub-divided so that percent 70 of the data is allocated to the training set, percent 20 to the test set, percent 10 to the validation set. This portion is based on the principle that the size of the validation set must procure a balance between obtaining a sufficient sample size to evaluate both the training and test tests. Reference [16] In this study, The MATLAP 6.0 software package will be used for training the selected SOM neural network model.

3.3 Backpropagation Neural Network for Predictive Clustering

Backpropagation NNs have been extremely popular for their unique learning capability [37] and have been shown to perform well in different applications in our previous research such as medical application [33] and game playing [6] A typical backpropagation neural network consists of a three layer structure: input layer nodes, output layer nodes, hidden layer nodes. In this second stage of methodology, in order to specify an output variable, which is required for supervised learning, each respondent belonging to a specific segment is mapped back to the original data set and represented the dependent variable for the training of the BP neural network. Backpropagation networks are fully connected, layered, feed-forward models. Activitations flow from the input layer through the hidden layer, then to the output layer. A backpropagation network typically starts out with a random set of weights. The network adjusts its weights each time it sees an input-output pair. Each pair is processed at two stages, a forward pass and backward pass. The forward pass involves presenting a sample input to the network and letting activitations flow until they reach the output layer. During the backward pass, the network's actual output is compared with the target output and error estimates are computed for the output units. The weights connected to the output units are adjusted to reduce the errors (a gradient descent method). The error estimates of the output units are then used to derive error estimates for the units in the hidden layer. Finally, errors are propagated back to the connections stemming from the input units. The backpropagation network updates its weights incrementally until the network stabilizes. The algorithm details can be shown in [4] and [37]. In this study, we followed the standard neural network architecture because it provides comparable results to the optimal architecture and works well as a benchmark for comparison. In this study, a BP NN model is trained using the demographic variables as inputs and twelve variables each representing a segment as outputs. The primary aim of this stage is to develop a trained model, which learns the pattern in the data and has the ability to generalize the pattern for predictive purposes. The accuracy of the trained model is validated by the validation patterns and tested on a portion of the data set to determine its ability to generalize. The size of the training, validation and test sets used for the SOM neural network modeling is also applied to the BP NN model. A validation set is used for visual comparative purposes during training. The accuracy of the model is verified using a test set of data not included for the purposes of training or validation. The selected sigmoid transfer function requires that each data point be scaled between 0 and 1. For this reason, each of the input variables is scaled to primarily enable the network to learn the relevant patterns. The sigmoid function scales the data by channel to match the range of the first hidden transfer function and the desired data to match the range of the output transfer function. A transfer function is used to prevent outputs from reaching very large values, which can “impair” neural networks. The BP NN training is monitored by using two stopping rules. The second rule is related to the acceptable error rate as measures of overall accuracy and applied whenever the output error term indicated signs of over-training.

4 Finding of the Modeling Procedure

4.1 Data Preparation

In this study, the data was collected through the Movie Lens web site (movielens.umn.edu) during the seven month period from September 19th, 1997 through April 22nd, 1998. The historical data consists of 100000 ratings from 943 users (customer) on 1682 movies with every user having at least 20 ratings and simple demographic information for the users such as age, gender, occupation, and zip code is included. The purpose of data preparation is to integrate, select and transform the data from one or more data bases into the data required for the proposed methodology. For this reason, it will be sampled a reference set that has enough rating information to discover similar user patterns. First, the data from web environment have been transferred to ACCESS program. Rating values that the user has given to the films, the demographic information related to the user, the contents of the films have been transferred to each separate file. Later, how many the users to go to the determined film kinds, the rating values given to these films and how many the films to be gone are extracted by using the interrogation commend. In order to accelerate the interrogation, it has been prepared as table sheet the information to be used for interrogation. A data preparation example is shown in Fig. 2. This information will be used for calculation of the customer value in the following section. The above extracted information has been combined with the extracted demographic information for data mining. Sexual ones from demographic information are coded as “1” for male and “0” for female. Occupation information can be also coded as follows.

$$X_i^j = \begin{cases} 1 & \text{if person } i \text{ has occupation } j, \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, 2, \ldots, n \; j \in Q$$

n : The number of selected person

Q : The set of all occupations

Nine genre types and 308 people gone and voted to these film types are selected. The selected people are considered each occupation, age, and sexual groups.

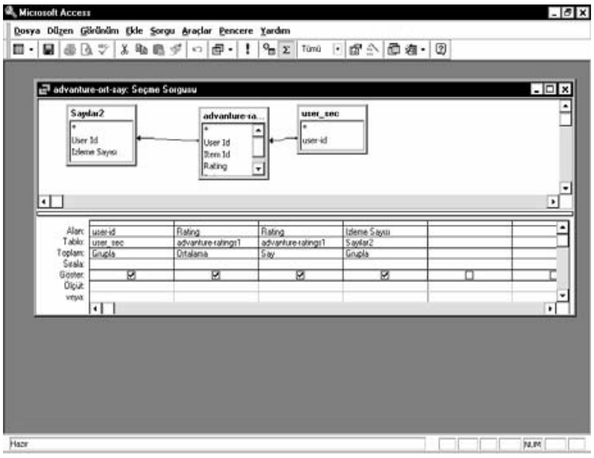

Fig. 2. A Data Preparation Example

4.2 Evaluation Metric: Calculating Customer Value

Value is the customer's overall assessment of the utility of a product based on perceptions of what is received and what is given. In other words, value in business markets is the perceived worth in monetary units of the set of economic, technical, service and social benefits received by customer firm in exchange for the price paid for a product, taking into consideration the available suppliers' offerings and prices. Buyers' perceptions of value represent a trade off between the quality and benefits they perceive in the product relative to the sacrifice they perceive by paying the price [39].

Most customer value importance scaling techniques used in practice today, such as rating scales, ranking scales and derived importance through regression and conjoint analysis can handle only a relatively few value dimensions at one time. In this study, we evaluate customer value with two viewpoints. Equation (1) evaluates the customer value having been developed for this study.

$$
\begin{aligned}
CV_i^j &= \text{Preference rate of } j. \text{ person for } i. \text{ film type} * \text{rate of the} \\
&\quad\ \text{being met expectation} \\
CV_i^j &= \frac{\text{The number of films of } i. \text{ type being seen by } j. \text{ person}}{\text{The number of films seen by } j. \text{ person}} \\
&\quad \cdot \frac{\text{Expectation being met}}{\text{Total expectation}} \\
CV_i^j &= \frac{\left\|\alpha_i^j\right\|}{\sum\limits_{k \in G} \left\|\alpha_i^k\right\|} \cdot \frac{\sum\limits_{k \in \alpha_i^j} r_i^k}{\mu \cdot \left\|\alpha_i^j\right\|}, \quad i = 1,2,\ldots,308, \quad j \in G \qquad (1)
\end{aligned}
$$

r_i^j: rating value that i. person gives j. film

α_i^j: the set of films of j. film type having been seen by i. person.
$\left\|\alpha_i^j\right\|$: the number of members in α_i^j

$$G = \left\{ \begin{array}{l} \text{Adventure, Children's, Comedy, Crime, Horror,} \\ \text{Romance, Sci-Fic, Thriller, War} \end{array} \right\}$$

μ : Maximum rating value (5 for this study).

4.3 Results of the SOM Neural Network Model

First, it has been clustered the customers according to the customer value, the demographic variables, the occupations, the sexual information. Preliminary the SOM studies, it appears that twelve possibly segments could be distinguished among the customers that visit the internet web page related to data collected by the Group Lens Research Project at the University of Minnesota. It is considered the ones to be in the different states from the point of view of the customer value of the segments in the determination of the cluster numbers. The following parameters and values are used for training the selected SOM neural network model. Ordering phase learning rate 0.9; ordering phase steps 1000; tuning phase learning rate 0.02; topology function adopts *"hextop"* and distance function uses *"mandist."* Topology and distance functions are selected to the best being seen distinction for segments. Table 1 indicates the centroid values of the segments according to the customer values for genre. Also Table 2 indicates the centroid values of the segments according to the demographic variables.

Table 1. The centroid values to the customer values of the segments

	Genre (Customer Value)								
Cluster No	Adventure (I)	Children's (II)	Comedy (III)	Crime (IV)	Horror (V)	Romance (VI)	Sci-Fi (VII)	Thriller (VIII)	War (IX)
1	0.25	0.18	0.37	0.21	0.07	0.39	0.19	0.24	0.20
2	0.24	0.13	0.25	0.23	0.05	0.28	0.17	0.28	0.30
3	0.27	0.14	0.31	0.20	0.09	0.32	0.22	0.22	0.33
4	0.30	0.14	0.29	0.27	0.13	0.26	0.26	0.28	0.31
5	0.26	0.15	0.35	0.19	0.05	0.46	0.18	0.23	0.32
6	0.29	0.10	0.30	0.19	0.10	0.30	0.25	0.23	0.34
7	0.26	0.08	0.29	0.23	0.06	0.33	0.25	0.28	0.29
8	0.21	0.12	0.32	0.22	0.10	0.23	0.23	0.24	0.19
9	0.22	0.17	0.30	0.25	0.12	0.30	0.22	0.31	0.20
10	0.25	0.18	0.29	0.22	0.08	0.33	0.25	0.24	0.26
11	0.34	0.13	0.27	0.24	0.08	0.25	0.28	0.25	0.32
12	0.14	0.09	0.27	0.27	0.11	0.19	0.29	0.30	0.20

Table 2. The centroid values to the demographic values of the segments

Cluster	Demographic Centroid Values																					
No	Age	Sex	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	30.14	0.00	0.48	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.00	0.00
2	43.00	0.27	0.03	0.00	0.00	0.02	0.00	0.00	0.00	0.81	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.05	0.01	0.01	0.03	0.00	0.00
3	46.31	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.27	0.33	0.00	0.00	0.00	0.00	0.34	0.04	0.00	0.00	0.00	0.00
4	35.29	1.00	0.33	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	36.57	0.05	0.03	0.01	0.00	0.82	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.02	0.02	0.02	0.01	0.00	0.00	0.01	0.02	0.00	0.01
6	46.71	0.93	0.01	0.01	0.01	0.80	0.00	0.00	0.00	0.02	0.00	0.01	0.01	0.01	0.00	0.02	0.00	0.00	0.01	0.05	0.00	0.04	0.00
7	35.71	1.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.36	0.00	0.36	0.00
8	26.69	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.01	0.47	0.01	0.00
9	34.71	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.06	0.00	0.11	0.04	0.19	0.00	0.08	0.00	0.11	0.02	0.06	0.06	0.00	0.02	0.19
10	31.81	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	34.23	0.99	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	30.32	1.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.23	0.00	0.00	0.00	0.29

1: administrator, 2: artist, 3: doctor, 4: educator, 5.engineer, 6: entertainment, 7: executive, 8: healthcare, 9: homemaker, 10: lawyer, 11: librarian, 12: marketing, 13: none, 14: other, 15: programmer, 16: retired, 17: salesman, 18: scientist, 19: student, 20: technician, 21: writer

Segment 1: In general, this cluster includes the women, which are of the occupations of the students and administrator and the average 30 years old. The order of the film types (genre) from point of view of the customer value has been determined as VI-III-I-VIII.

Segment 2: In general, this cluster includes the women, which are of the occupation of the healthcare and the average 43 years old. The order of the film types (genre) from the point of view of the customer value has been determined as IX-VI-VIII-III.

Segment 3: In general, this cluster includes the men, which are of the occupations of librarian, lawyer, and retired and the average 46 years old. The order of the film types (genre) from the point of view of the customer value has been determined as IX-VI-III-I.

Segment 4: In general, this cluster includes the men, which are of the occupations of administrator, entertainment and executive and the average 35 years old. The order of the film types (genre) from the point of view of the customer value has been determined as IX-I-III-VIII.

Segment 5: In general, this cluster includes the women, which are of the occupation of educator and the average 36 years old. The order of the film types (genre) from the point of the customer value has been determined as VI-III-IX-I.

Segment 6: In general, this cluster includes the men, which are of the occupation of educator and the average 36 years old. The order the film types (genre) from the point of view of the customer value has been determined as VI-III-IX-I.

Segment 7: In general, this cluster includes the men, which are of the occupation of doctor, scientist, technician, and the average 35 years old. The order of the film types (genre) from the point of view of the customer value has been determined as VI-IX-III-VIII.

Segment 8: In general, this cluster includes the men, which are of the occupation of programmer, student, and the average 26 years old. The order of the film types (genre) from the point of view of the customer value has been determined as III-VIII-VI-VII.

Segment 9: In general, this cluster includes the women, which are of the occupation of artist, marketing and other and the average 32 years old. The order of the film types (genre) from the point of view of the customer value has been determined as VIII-VI-III-IV.

Segment 10: In general, this cluster includes the women, which are of the occupation of artist, marketing and other and the average 32 years old. The order of the film types (genre) from the point of view of the customer value has been determined as VI-III-IX-I.

Segment 11: In general, this cluster includes the men, which are of the occupation of artist, marketing and other and the average 34 years old. The order of the film types (genre) from the point of view of the customer value has been determined as I-IX-VII-III.

Segment 12: In general, this cluster includes the men, which are of the occupation of engineer, salesman and writer and the average 30 years old. The order of the film types (genre) from the point of view of the customer value has been determined as VIII-VII-III-IV.

4.4 Results of Backpropagation Neural Network Model

It has been designed a BP neural network model to predict the segments determined in SOM model according to the demographic values of the new customers. This model required for the development of BP neural network considers the variables that are also used for the SOM neural network modeling process. In order to determine output variables, which is required for supervised learning, each respondent belonging to a specific segment is mapped back to the original data set and represented the output variable for the training of the BP neural network. A BP neural network model is trained using the 23 variables as inputs and 12 variables each representing a segment as outputs. This model is shown in Fig. 3.

The aim of this stage is to develop a trained model, which learns the pattern in the data and gets the ability to generalize the pattern for predictive purposes. The accuracy of the trained model is validated by unseen data and tested on a portion of the data set to determine its ability to generalize. A validation set is used for comparative purposes during training. The accuracy of the model is verified using the test data not used for training or validation. In this study, the data set belonging to 260 people was used for training.

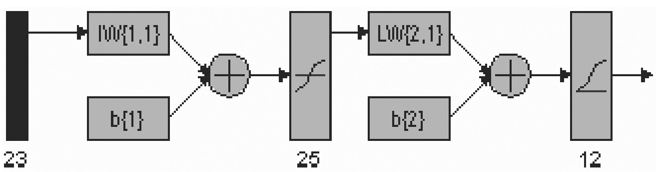

Fig. 3. BP Neural Network Model

Also the data set related to 48 people was used for testing. The one intermediate (hidden) layer with 25 neurons used for the model specification is a function of the non-linearity of the data. The tansig function has been used as the activity one for the hidden layer because of indicating the better generalization for training. The logsig function has been also used as the activity one for the output layer because of being 0–1 intervals values. The neuron number in the hidden layer has been expanded one by one and continued as to the acceptable error level. The accuracy of BP neural network model is determined by how well the training set can learn the pattern in data when compared to the test data set, which comprises unseen data. Figure 4 indicates the error graph of the training and validation sets for BP NN.

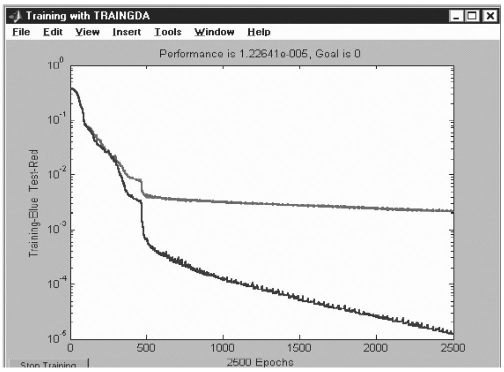

Fig. 4. The error graph of BP-NN model

In Fig. 4, the training set and validations set converge as the error between the trained data and the validation data starts to decrease. The BP NN training is monitored by using stopping rule if the total number of iterations reached 1000. The overall accuracy for BP NN model could also be considered using the unseen 161 people's data. BP NN has been simulated by these people's demographic information and the segments in which the people are located have been surmised. According to these predictions, the film's types having the highest customer value have been recommended to the person and the accuracy of these recommendations related to the first three preferences has been determined as shown in Table 3.

Table 3. The prediction accuracy of segments according to BP-NN model

Film Preference of Cluster	Film Preference of Person		
	1	1&2	1&2&3
1	%16.15	%32.30	%43.50
1&2	%39.75	%58.40	%67.70
1&2&3	%75.77	%85.70	%89.44
1&2&3&4	%89.44	%97.52	%99.38

When the first four film's types with the highest customer value are recommended to the related person for each segment, at least one of the first three film's type on the person's preferences could be predicted accurately. For this reason, it can be said that the more accurate segmentation requires more detailed data and more definite recommendations are based on more accurate predictions.

5 Conclusion

The requirement for the detailed knowledge on film industry and the preference of the customer and the need to cope with the limitations in analyzing non-linear relationship are the aim of this research. Knowledge discovery is a complicated process of extracting useful information from raw data. It includes many steps such as data warehousing; target data section, data cleaning, preprocessing, and transformation. In a majority of the real cases, the knowledge discovery process is iterative and interactive in nature. At the core of the knowledge discovery process, there are the data mining methods for extracting patterns from data. These methods can have different goals and may be applied successively to achieve the desired result in the knowledge discovery process. For example, SOM neural network models based on artificial intelligence technology can be developed to create clusters based on combinations of natural characteristics within a set of customer data (e.g. information in this study). In addition, BP-NN models could be also used to predict the segmentation of new customers as part of an already existing segment. This study demonstrates the usefulness of Artificial Neural Networks as means of grouping the respondents reached according to the survey in web page and predicting the segmentation of new respondents. Profiling the clusters and tracking the dynamic behavioral changes of segments can be considered in the future research. Besides, additional inputs required for the more segmentation and calibration of BP-NN model should be obtained through a process of data enrichment. The knowledge generated by the predictive ability of the BP-NN model could induce greater effectiveness in film industry and financial efficiency. The application of the BP-NN could generate further knowledge, which may be used to order the composition of the film recommendation with the changing behavior of customers. Using the BP-NN model to generate knowledge related to high customer value segments could increase revenue for the local economy and assist with the development of film industry's management strategies. A competitive advantage could be gained by researching the differences in behavior of customer segments that go and do not go to a film. The matching of cluster profiles of high value customers that go to a film to similar profiles of customers that do not go to the film by using the BP-NN model could procure a target market opportunity.

References

1. Barnes, J.G., 2000. Secrets of Customer Relationship Management: It's All About How You Make Them Fell, Mc Graw-Hill. 358
2. Bellandi, G., Dulmin, R., Mininno, V., 1998. Failure rate neural analysis in the transport sector. International Journal of Operations & Production Management, 18(8), 778–793. 356
3. Berry, M.J.A., Linoff, G., 1997. Data mining techniques for marketing, sales and customer support, New York, Wiley. 356

4. Bishop, C.M., 1995. Neural Networks for Pattern Recognition, Oxford University Press, New York. 363
5. Changchien, S.W., Lu, T., 2001. Mining association rules procedure to support on-line recommendation by customer and products fragmentation. Expert Systems with Applications, 20, 325–335. 355
6. Chen, H., Buntin, P., She L., Sutjahjo, S., Sommer, C., Neely, D., 1994. Expert prediction, symbolic. IEEE Expert, 9(6), 21–27. 363
7. Chiang, D., Chow, L.R., Wang, Y., 2000. Mining time series data by a fuzzy linguistic summary system. Fuzzy Sets and Systems, 112, 419–432. 355
8. Edelstein, H., 1997. Mining for gold. Information Week, 4/21/1997. 356
9. Fayyads, U.M., Ratetsky-Shapiro, G., Smyth, P., 1996. The KDD processes for extracting useful knowledge from volumes of data. Communications on the ACM, 39(11), 27–34. 355
10. Flint, D.J., Woodruff, R.B., Gardial, S.F., 1997. Customer Value Change in Industrial Marketing Relationships-A Call for New Strategies and Research. Industrial Marketing Management, 26(2), 163–175. 360
11. Hammond, K., Ehrenberg, A.S.C., Goodhandt, G.J., 1996. Market segmentation for competitive brands. European Journal of Marketing, 30(12), 39–49. 362
12. Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. Academic Press, New York. 356
13. Hand, D., Mannila, H., Smyth, P., 2001. Principles of Data Mining. London, England: MIT Press. 356
14. Handen, L., 2000. Putting CRM to work: The rise of the relationship. In S.A. Brown (Ed.), Customer relationship management: A Strategic imperative in the world of e-Business, 7–18, Toronto, Wiley. 358
15. Iannotte, D., 2000. Up-front analysis helps control losses. National Underwriter, 104(32), 23. 356
16. Kaastra, I., Body, M., 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing, 10, 215–236. 363
17. Kennedy, R.L., Lee, Y., Van Roy, B., Reed C.D., Lippman, R.P., 1997. Solving data mining problems through pattern recognition, New York, Wiley. 356
18. Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J., 1996. SOM-PAK: The Self-organizing Map Program Package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150. 357
19. Lee, T.H., Jung, S., 1999. Forecasting credit worthiness: Logistic vs. artificial neural net. The Journal of Business Forecasting, 18(4), 28–30. 356
20. Milley, A., 2000. Healthcare and data mining. Health Management Technology, 21(8), 44–45. 356
21. Moody, J., Utans, J., 1995. Architecture selection strategies for neural networks application to corporate bound rating, in: A. Refenes (Ed.), Neural Networks in the Capital Markets, Wiley, Chichester, 277–300.
22. Natter, M., 1999. Conditional market segmentation by neural networks: a Monte Carlo Study. Journal of Retailing and Consumer Services, 6(4), 237-2-48. 362
23. Naumann, E., 1994. Creating Customer Value: The Path to sustainable competitive advantage. Thomson Executive Press, Cincinnati, Ohio, International Thomson Publishing. 361
24. Nicolett, M., Andren, E., Gilbert, M., 2000. Challenges of aggregating and managing catalog content. Gartner Group report, April 12; (Online) available: http://gartner4.gertnerweb.com:80/gg/purchase/0/00/877/63/doc/00087763/. 357

25. O'Brien, T.V., 1994. Neural nets for direct marketers. Marketing Research, 6(1), 47–49. 356
26. Parasuraman, A., 1997. Reflections on Gaining Competitive Advantage Through Customer Value. Academy of Marketing Science Journal, 25(2), 154–161. 360
27. Park, S.C., Piramuthu, S., Shaw, M.J., 2001. Dynamic rule refinement in knowledge-based data mining systems. Decision Support Systems, 31, 205–222. 355
28. Peppers, D., Rogers, M., 1999. The One to One Manager: Real-World Lessons in Customer Relationship Management. Doubleday, New York. 357, 358
29. Poli, I., Jones, R.D., 1994. A Neural netmodel for prediction. Journal of the American Statistical Association, 89(425) 117–121. 356
30. Schalkoff, R.J., 1997. Artificial Neural Networks. McGraw-Hill, New York. 356
31. Shtup, A., Versano, R., 1999. Estimating the cost of steel pipe bending, a comparison between neural networks and regression analysis. International Journal of Production Economics, 62(3), 201–207. 356
32. Tillett, L.S., 2000. Banks mine customer data. Internetweek 830, 45–48. 356
33. Tolle, K.M., Chen, H., Chow, H., 2000. Estimating drug/plasma concentration levels by applying neural networks to pharmacokinetic data sets, Decision Support Systems, Special Issue on Decision Support for Health Care in a New Information Age, 30(2), 139–152. 363
34. Turban, E., Trippi, R.R., 1996. Neural Network Fundamentals for financial analysts In R.R. Trippi & E. Turban (Eds.), Neural Networks in Finance and Investment, 3–24, New York, McGraw-Hill. 357
35. Verhoef, P.C., Donkers, B., 2001. Predicting customer potential value an application in the insurance industry. Decision Support Systems, 32, 189–199. 358
36. Wedel, S., Kamura, W., 1997. Market segmentation: Conceptual and Methodological foundations. Boston, Kluwer. 361
37. Widrow, B., Rumelhart, D.E., Lehr, M.A., 1994. Neural Networks: Applications in industry, business and science, communications of the ACM, 37, 93–105. 363
38. Woodruff, R.B., Gardial, S.F., 1996. Know Your Customer: New Approaches to Customer Value and Satisfaction. Blackwell Publishers, Cambridge, MA. 360
39. Woodruff, R.B., 1997. Customer Value: The Next Source for Competitive Advantage. Journal of the Academy of Marketing Science, 25(2), 139–153. 360, 365
40. Zhou, Q., 2003. Adaptive Knowledge Discovery Techniques for Data Mining, Ph.D. Thesis, University of Otago, New Zealand. 355