Behavior-Based Telecom Tariff Service Design with Neural Network Approach

Yongbin Zhang Ronghua Liang Yeli Li

Information & Mechanical Engineering Department
Beijing Institute of Graphic Communication
Beijing, China

zhangyongbin@bigc.edu.cn liangronghua@bigc.edu.cn liyl@bigc.edu.cn

***Abstract*—This paper presents a behavior-based telecommunication tariff service design model. Unlike traditional service design models, which includes demographic attributes to design telecommunication tariff services, this model only uses history data about telecommunication customer behavior, such as voice service usage, text service usage, and multimedia service usage to design new tariff services. An unsupervised algorithm, the self-Organizing Map neural network algorithm, is adopted to train the model. The model segments telecommunication customers into different groups only according to their behavior, service usage. The characteristics of customer service usage can be identified with the results. With the generated behavior-based groups and demographic attributes of customers as training data set, decision tree algorithm is adopted to reveal the characteristics of potential customers for the new designed tariff services.**

Keywords-behavior-based; tariff services design; neural network; Self-Organizing Map (SOM); decision trees

I. INTRODUCTION

Telecommunication has entered a new age of development with advanced technology and increased competition with established players [1]. With the increasing competition, service providers are looking for new strategies to attract new customers, hold on the existing ones, and increase the revenue [2]. Some researchers have paid attention to enhance customer satisfaction with total quality management [3]. Other researchers have discussed how to design and provide a flexible, scalable and open network infrastructure so that new services can be delivered in a faster manner and at lower costs [1, 4, 5, 6].

The above research has contributed to enhancing competition of service providers. However, there is little research on how to design new services or provide flexible tariff packages that are suitable for customers. In this paper, efforts are put to address the issue of telecommunication tariff service design according to customer behavior only. All data are from user profile repository from which user service usage history and basic user information such as gender and birthday can be accessed [10]. Unsupervised neural network algorithms are adapted to segment customers into different groups according to the usage of different services. With the results, it is easy to identify which services are used together by different customer segments. After that, customer basic information is analyzed with decision tree algorithms to find out the demographic characteristics of different segments.

II. PROCESSES OF TELECOMMUNICATION TARIFF SERVICE DESIGN

A telecommunication tariff service is a service package provided by telecommunication service provider, which puts some of their services together with an open contract between a telecommunication service provider and the public, filed with a regulating body such as a Public Utilities Commission. Such tariffs outline the terms and conditions of providing telecommunication service to the public including rates, fees, and charges [7].

At present, the 'five-step model' is one of the mostly used processes to design new services [8]. The first step of the model is to define market, the second step is to define customers' need, the third step is to advertise and test the service, the forth step is to introduce the new service to the market, and the fifth step is to monitor and analyze the response.

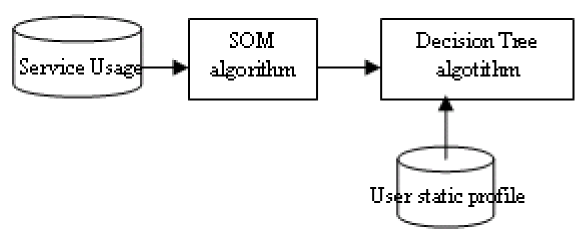

Figure1 . Architecture of the behavior-based service design model

This model plays an important role in success of a newly developed service for telecommunication service providers. However, it is difficult to predict which service will be widely accepted and generate profits. Some researchers have developed different models to design new services. Ahn and Skudlark have introduced a new modified process for a new telecommunication service, which includes factors such as technological innovations, political influences, and changes of management environment [9].

However, it is not easy to quantify these additional factors. To address these problems, a behavior-based model is presented, in which the first step is to analyze the service usage of telecommunication customers and to design new tariff services according to the analysis, and the second step is to decide potential customers who will need the new tariff

This work is partially supported by Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality Grant # PXM2010_014223_095557 to Y.L. Li.

978-1-4244-6255-1/11/$26.00 ©2011 Crown

services. The other steps are the same as provided in the "five-step model".

The architecture of the behavior-based service design model is shown as Fig.1. The model uses service usage as customers' behavior and segments customers' behavior into different groups with SOM algorithm, and then the results of the SOM model work as target attribute with customer static profile, such as gender, age and location as input attributes for decision tree algorithm to reveal the customer characteristics of different service usage.

III. ATTRIBUTES OF THE MODEL

User profile repositories have been designed by many researchers to facility the development of new telecommunication services [10]. With user profile repository, it is easy to access the information on service usage history and basic user information, such as gander, birthday and name. Unlike other model, data on services usage are used only to analyze customer behavior in this model, which is named behavior-driven model. The basic attributes used in this model include voice service, data services, and multimedia service. These attributes are related with the service usage behavior of current customers. So assumed A is the basic attribute set, then set A is :

$$A=\{voice, data, multimedia\} \quad (1)$$

where *voice* represents the volume of voice service used by a telecommunication customer, *data* is the quantity of data service usage and *multimedia* is the quantity of multimedia service usage.

Because the value for each attribute depends on different time, the value varies with different periods. It is more reasonable to specify a period for different service usage. So the attributes set A can be transformed into the following formula:

$$A= \left\{\frac{\sum voice}{p}, \frac{\sum data}{p}, \frac{\sum multimedia}{p}\right\} \quad (2)$$

with (2), it is easy to analyze the usage of the three different kinds of services used by current customers.

At the same time, it is still difficult to distinct the service usage in different time of a day. It is necessary to calculate volume of different services in different period of a day. In this model, day time and night time are used to distinguish the differences, it is shown as following:

$$voice = voice_{day} + voice_{night} \quad (3)$$

Different periods of a day can be adopted to analyze the behavior of service users. If a shorter interval is used, it will be easier to identify users' behavior varying with time. However, more data will be generated and more time will be needed to execute the model.

At present telecommunication service charge rates vary in different kinds of voice services such as local city call, long-distance domestic call and international call. Different kinds of voice service should also be considered. So each voice element in set A can be divided into local call, domestic call and international call. For example, three attributes can be derived from $voice_{day}$. They are shown as following:

$$voice_{day} = voice_{day,local} + voice_{day,long} + voice_{day,international} \quad (4)$$

For data service, at present, the fee for international data and domestic data are different. So for each data service attribute, two attributes can be derived. One is for international data service, and the other is for domestic data service. Two attributes can be developed from $data$. They are $data_{local}$ and $data_{international}$.

The same rule is applied to multimedia service. That is $multimedia_{local}$ and $multimedia_{international}$. There are six attributes derived from the voice service, four attributes on data service and four attributes on multimedia data services. So 14 attributes used to build the service analysis model are shown in table 1.

TABLE I. ATTRIBUTES OF THE MODEL

Attribute	meaning
$voice_{day,local}$	Average local voice call in day time
$voice_{day,long}$	Average long-distance voice call in day time
$voice_{day,international}$	Average international voice call in day time
$voice_{night,local}$	Average local voice call at night
$voice_{night,long}$	Average long-distance voice call at night
$voice_{night,international}$	Average international voice call at night
$data_{day,domestic}$	Average domestic plain data service volume in day time
$data_{day,international}$	Average international plain data service volume in day time
$data_{night,domestic}$	Average domestic plain data service volume at night
$data_{night,international}$	Average domestic plain data service volume at night
$multimedia_{day,domestic}$	Average domestic multimedia service volume in day time
$multimedia_{day,international}$	Average international multimedia service volume in day time
$multimedia_{night,domestic}$	Average domestic multimedia service volume at night
$multimedia_{night,international}$	Average international multimedia service volume at night

All of the attributes shown in table 1 are related to services that used by telecommunication customers. The algorithms in following section will be adopted to train the model and segment the customers.

Besides set A, which is set of customer behavior attributes, another attribute set is needed to identify characteristics of each segment. These attributes are also from user profile. The attribute 'age' is an important factor and the living area is another one. Different kinds of occupation can also affect the services usage. So the attributes set used in this research is:

$$B=\{gender, age, location, occupation\} \quad (5)$$

The attributes in set B are used to identify the characteristics of different services users.

IV. Algorithms of the Model

To cluster customer into different segments according to their behavior of service usage, unsupervised algorithms such as K-means, nearest neighbor clustering and self-organizing map (SOM) means can be adopted.

The k-means is the simplest and most commonly used algorithm employing a squared error criterion [11]. However, with k-means algorithm, the choices of initial centroid for each cluster have great effects on the final results. The Nearest neighbors clustering algorithm serves as the basis of clustering procedures. The Self-organizing map (SOM) algorithm is an artificial neural network technology. The SOM algorithm uses a two-dimensional map of the multidimensional data set. Kohonen has successfully applied the SOM for vector quantization [14].

In this research, SOM algorithm is adopted to cluster the services usage patterns of telecommunication users into different segments according to their behavior of service usage. The attributes in set A discussed above are used as inputs of the model, so the model have 14 nodes as inputs. The output nodes of the model represent the number of customer segment. There is no precise way to determine the number of the output nodes. Various output nodes can be tried in practice in order to get the proper output nodes. However, the principle is that fewer output nodes will cause the segment vague. For the extremity condition, if the model has only one output, then only two groups can be determined and it can not distinct the differences among different customers. In this research, ten output nodes are used in the mode. Each input node is connected with every output node. There are no connections among output nodes.

The SOM algorithm can be described as follows:

1. Randomly choose an input vector x from the input dataset.
2. Determine the "winning" output node i according to

$$|w_i - x| \leq |w_k - x|, \forall k \in [1,2...,N] \qquad (6)$$

Where w_i is the weight vector connecting the inputs to output node i.

3. Given the winning node i, the weight update is

$$w_k = w_k + \lambda h(i,k)(x - w_k) \qquad (7)$$

Where $h(i,k)$ is called the neighborhood function that has value 1 when $i = k$ and falls off with the distance between units i and k in the output array. Units close to the winner as well as the winner, have their weights updated appreciably. Weights associated with far away output nodes do not change significantly.

4. Go to step 1 until the system is stable.

The system is said to be stable if no input data changes its category (output node) any more after a finite number of learning iterations. When the system is stable, only one of the output nodes is active for each input observer. It is easy to decide which category the input observer belongs to according to the active output node.

After each customer in the training set has been assigned to a unique group, the decision tree algorithms are used to generate a set of rules while the group identification works as the target attribute.

Decision trees are popular tools for classification and prediction [15]. The attractiveness of decision trees is that, in contrast to neural networks, decision trees represent rules. Rules can readily be expressed so that humans can understand them. In this research, decision tree algorithm is used to generate the output as a binary tree-like structure, which gives fairly easy interpretation to the marketing people.

A decision tree is a class discriminator that recursively partitions the training set until each partition consists entirely or dominantly of examples from one class. Each non-leaf node of the tree contains a split point that is a test on one or more attributes and determines how the data is partitioned. The tree is built by recursively partitioning the data. Partitioning continues until each partition is either 'pure' (all members belong to the same class) or sufficiently small (a parameter set by the user). The initial lists created from the training set are associated with the root of the decision tree. As the tree is grown and nodes are split to create new children, the attribute lists for each node are partitioned and associated with the children.

A decision tree classifier is built in two phases: growth phase and a prune phase. After the initial tree has been built (the 'growth phase'), a sub-tree is built with the least estimated error rate (the 'prune phase'). The process of pruning the initial tree consists of removing small, deep nodes of the tree resulting from 'noise' contained in the training data, thus reducing the risk of 'overfitting', and resulting in a more accurate classification of unknown data. While the decision tree is being built, the goal at each node is to determine the split attribute and the split point that best divides the training records belonging to that leaf. The value of a split point depends on how well it separates the classes. Several splitting indices have been proposed in the past to evaluate the quality of the split.

V .Results and Discussion

In this research, data were prepared according to the attribute sets A and B. Because the total number of customers is too large and it is also not necessary to deal with the whole population, a sample with 10000 customers was chosen randomly. In this research, the value of the parameter p is set to 30. It means different service volume is the average service volume in 30 days during the latest active service usage.

After all data are prepared, they work as input information of the SOM model and the model is trained as SOM algorithm process. When the SOM model is stable, each input belongs to a single segment, which means that every customer's service usage behavior pattern in the sample set has unique group identification. Customer who has the similar behavior will be in the same group.

With the group identification as the target attribute and attributes in set A as input attributes, decision tree algorithm C5.0 is used to train the same customers' behavior patterns trained by SOM algorithm. The results of the decision tree algorithm are rules which describe the characteristics of services usage for each group.

With the generated rules set, telecommunication tariff service designers can design new telecommunication tariff services according to the behavior patterns.

However, with behavior data only, it is difficult to understand who are using or would likely to use these services. The next step is to identify the characteristics of potential users. With the generated group identification, and the attributes in set B, which is $\{gender, age, location, occuption\}$, the new attribute set C will be shown as following:

C= $\{groupId, gender, age, location, occuption\}$ (8)

With set C, the decision tree algorithm C5.0 is used to generate the rules about customers' characteristics of new services. The attribute 'groupId' works as the target attribute for the decision tree algorithm. With the output of decision tree algorithm, which is a set of rules and describes the characteristics of services users for each group, potential customers of each designed telecommunication tariff service can be determined.

In practice, service designers can modify the attributes in set A according to telecommunication services provided by different service providers. For example, attributes such as MOU (outbound call usage) and MIU(inbound call message) can be added into set A. Attributes in set B can also be modified according to market demand. Different telecommunication service providers may choose different attributes to the model.

VI. CONCLUSION

This research provides a new method for telecommunication tariff services design. With this model, new telecommunication tariff services can be designed according to customer behavior patterns that represent different telecommunication services usage. At the same time, the characteristics of potential customer can be identified with customer profile attributes. It provides an easy way to decide the target market for the new designed tariff services.

REFERENCES

[1] R. Shankar, "Benchmarking telecommunication service in India An application of data envelopment analysis," Benchmarking: An International Journal Vol. 15 No. 5, pp. 584-598, 2008

[2] N. Kryvinska, C. Strauss, L. Auer, and P. Zinterhof, "Conceptual Framework for Services Creation/Development Environment in Telecom," Domain.Proceedings of iiWAS2008., Linz, Austria. pp.324-331,2008

[3] S. Mehra, and S. Ranganathan, " Implementing total quality management with a focus on enhancing customer satisfaction," International Journal of Quality & Reliability Management Vol. 25 No. 9,pp. 913-927, 2008

[4] NGOSS (New Generation Operations Systems and Soft-ware) Overview, Technical Program, TeleManagement Forum, http://www.tmforum.org/browse.aspx?catID?1912.

[5] R. Christian, and H. E. Hanrahan, " Converging towards Service Centric Networks: Requirements for a Service Delivery Platform Framework," Southern African Telecommunications Networks and Applications Conference (SATNAC) Proceedings, vol. 1, pp. 295-300, September, 2005

[6] V. Pandey, D. Ghosal, and B. Mukherjee, " Exploiting User Profiles to Support Differentiated Services in Next-Generation Wireless Networks, " IEEE Network, Volume 18, Issue 5, pp. 40 – 48, 2004.

[7] H.L. Jee, ".Enterprise Customer's Needs for SLA on Telecom Services in Korea," PICMET 2007 Proceedings. USA:portland Oregon. pp.180-184,2007

[8] G. L. Urban, and R. H. John, Design and marketing of new products 2nd ed.Englewood Cliffs, NJ: Prentice-Hall, Inc,2008

[9] J. H. Ahn,, and A. Skudlark, "Managing risk in a new telecommunications service development process through a scenario planning approach," Journal of Information Technology, vol. 17, pp. 103–118,2002.

[10] H. Wang, X. m. Shi, Y. Li , H. Chang, W.l. Chen, and J. Tang. "User Profile Management for Personalized Telecom Service,".The 9th International Conference for Young Computer Scientists,pp.1087-1092

[11] J. MCQUEEN, "Some methods for classification and analysis of multivariate observations," In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp.281–297,1967

[12]E. DIDAY, "The dynamic cluster method in non-hierarchical clustering," J. Comput. Inf.Sci. 2,pp. 61–88,1973

[13]M. J. SYMON, "Clustering criterion and multi-variate normal mixture, " Biometrics 77, pp.35–43,1977

[14]T. KOHONEN, Self-Organization and Associative Memory, 3rd ed. Springer information sciences series. Springer-Verlag, New York, NY,1989

[15]J.R. QUINLAN, "Induction of Decision Trees". Machine Learning vol.1, pp.81-106, 1986