Data Mining in Telecommunication

Mohsin Nadaf & Vidya Kadam

Department of IT, Trinity College of Engineering & Research, Pune, India
E-mail : mohsinanadaf@gmail.com

Abstract – **Telecommunication is one of the first industry to adopt Data mining technology. Generally these companies have many customers. So companies save this data and hence there is generation of tremendous amount of data. This data includes call-detail data, customer data, network data. To handle this large amount of data and to discover useful information from this data the automatic or semi-automatic method should be used as it simplifies the work. And the Data Mining fits for this criteria. Therefore use of Data Mining technology is must for telecommunication companies. This paper describes how data mining and its applications helps Telecommunication industry in various aspects.**

Keywords – Applications, Data, Data Mining, Telecommunication

I. INTRODUCTION

The Telecommunication industry generates large amount of data. These data includes call-detail data which describes calling behavior of that customer and that data traverse through the network that is nothing but network data. And the customers who are subscribed to companies describes the customer data. So these large amount of data was needed to be handled hence the concept of Human based Expert systems came into existence. These systems helped to detect the frauds and network issues. But later it came to know that this is very time consuming system as it involves much of knowledge from Human experts. Hence this method was not used later. At that time the Data mining was new. Data Mining is search for the relationships and global patterns that exist in large database but are hidden vast amounts of data. This relationship represents valuable knowledge about the database, and the objects in the database. So Telecommunication companies started to using this technology. They were the first to adopt Data mining technology. But telecommunication data pose very interesting issues for data mining. The first issue is telecommunication data may contain billions of records and amongst largest in world. The second one often raw data is not always suitable for data mining. And the last

issue is concerned with the real time performance of data mining applications such as fraud detection which requires learned protocol, rules to be applied in real-time. If we deal with these issues then the efficiency of data mining will definitely increase. But if we only consider its usefulness then we come to know that it is a necessary tool for the industry.

II. DATA MINING

Data Mining means extracting knowledge hidden in large volumes of data. It is a part of knowledge discovery process that offers new way to look at data. Simply, identifying potentially useful and understandable data. It is the advanced analysis step of the Knowledge Discovery in database. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indexes. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. In simple terms it uses machine learning, statistical and visualization techniques to discover and present the knowledge in a form that is easily comprehensible to humans. Typically in Telecommunication industry the Data mining process involves Customer segmentation, Profiling, Data Preparation and Clustering. These tasks are discussed in this paper.

III. NEED OF DATA MINING IN TELECOMMUNICATION

As Data is the base of Telecommunication so data mining is used to perform operations on data and obtain the desired results. Some of the reasons to use data mining are as follows:

- *To detect frauds*

Fraud is very serious problem for Telecommunication companies, resulting in billions of dollars of lost

- *To retain customers*

By applying data mining tools we can study and research on the customers' database which will help to know how to deal with our customers and how to satisfy them. In this way we can maintain good relations with the old customers as well as get new ones.

- *To know the Customers*

By studying the behavior of the customers we can better know them, what is their calling period, their likes, tastes, attitudes etc. This will help to approach them in terms of marketing activity.

- *Products and services which yield highest amount of profit*

As it is very large database, so it contains transactions made by customers. That transactions maybe of purchased product, services. The study of these transactions can give us statistics that which of my services and products are popular that yield highest amount of profit.

- *Factors that influence customers to call more at certain times*

By studying behavior of customers, we can come to know what are the factors that influence customers to call more at certain times.

IV. CUSTOMER SEGMENTATION AND PROFILING

Customer Segmentation is the process of dividing customers into the homogeneous groups according to their common attributes. These attributes include their likes, habits, actions. And Customer profiling means to describe the customers by their characteristics like age, gender, economic conditions, income, culture etc. This Customer segmentation and profiling helps the telecommunication companies to decide which marketing actions should be taken for respective segment and allocating the resources. Customer segmentation process requires the knowledge of an expert. First of all the segmentations are defined in advance then customers are divided over these segmentations. And then the profile for every customer is determined with the help of customer data. Now, To find the relation between customer profile and segments the SVM technique can be used. SVM is a Data Mining technique which stands for Support Vector Machine[2]. SVM estimates segment of a customer by his profile

information(age, gender, etc). So if the segment is estimated based on customer profile then we can easily determine usage behavior of that respective customer profile.

In today's competition world to compete with other Telecommunication companies, it is very much important to know about the your customers and their needs. And hence Customer Segmentation and Profiling is important. Overall the goal of doing all this is to predict the behavior of customer using all the information we have. The profiling is the process done after the segmentation.
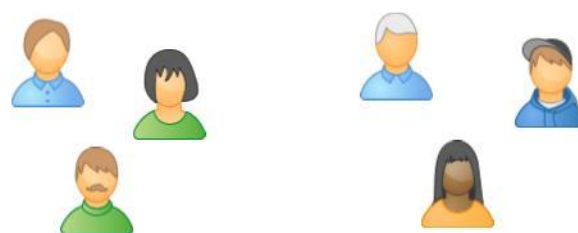
A. Customer Segmentation

Fig.4 :(a) Customer Segmentation

In simple words, it is act of partitioning the customer groups as per their similar characteristics. Customer segmentation helps us to know about the behavior of the customer like what is his usage, calling time, frequency of calling. This helps companies to approach and to have communication with customers so that their needs can be understood and marketing in this way can be done effectively. Hence more resources can be channeled effectively. But while doing the segmentation some difficulties arises like-

- Non Relevant Data

Although company have large amount of data but if it is not relevant to need then it is useless. Hence only the relevant data should be chosen.

- Not Updating Data

Here we should understand that segmentation is continuous process and it should be updated frequently whenever the new customer gets added.

- Over Segmentation

While defining segmentations care should taken that particular segment should not be too small that it will be difficult to treat separate segments.

B. Customer Profiling

Customer Profile is built on the basis of his personal information such as age, gender, location.

It is done so as to reach the customers and serve them better. Also as per the needs like to offer/build new product the relevant profiles are selected. The basic characteristics of personal information includes-

- Age and Gender-Age and gender of customers.

- Location- The geographic, national location of customer .

- Economic conditions- Income of a customer or his family, Payment mode.

- Attitude- The attitude of customer towards using the product

- Lifestyle- Lifestyle of customer as per he uses the service

- Knowledge- How marketing can be done for respective customers as per their knowledge

V. TYPES OF TELECOMMUNICATION DATA

Before performing the data mining process it is necessary to understand the data i.e. what type of data it is. So basically there are main three types of data viz. Call-Detail data, Network data and Customer data.

A. Call-Detail Data

Call-detail data is a data which contains details about when a call is placed, duration of call i.e. everything about the call. Everyday large numbers of calls are generated so when call takes place on the network then a detailed information about call is generated and is saved. That is nothing but call detail data. This information includes location, call duration, call time, calls originated/received, etc. That is it contains sufficient information to describe important characteristics of each call. The information about each call is saved in format of sequence of records in database. Then the summarization of all these records is done and actual data mining process is applied to extract useful knowledge as per our needs.

However the call description includes:

- average call duration
- % no-answer calls
- % calls to/from a different area code
- % of weekday calls (Monday – Friday)
- % of daytime calls (9am – 5pm)
- average # calls received per day
- average # calls originated per day
- # unique area codes called during P

From this, the information graphs, histograms and tables are drawn. And then these are analyzed and used for knowing current status and future predictions.
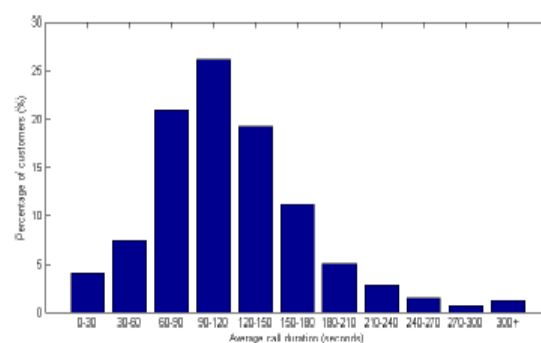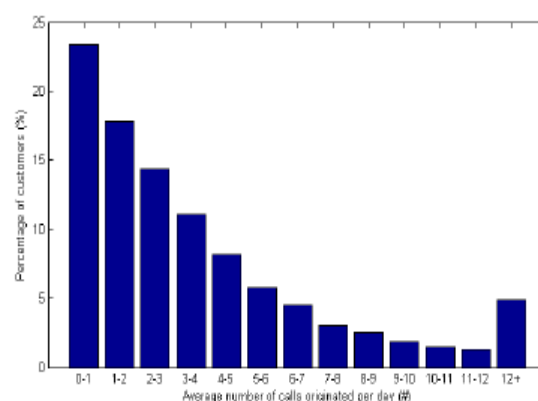

Fig. 5.(a) Call Duration

Fig. 5.(b) Originated Calls

For example, the above histograms shows the relationship between customers and calls. Fig. V.(a) shows call duration of customers i.e. 25% of customers call for about 90-120 seconds. Fig. V.(b) gives information about average number of calls generated by customers.

The call detail records are also used to distinguish between Business and Residential calls. For example, the calls made during the time period from 9-4/5 pm indicates it is business customer. And the calls which take place mostly at evening time and on weekdays indicates a residential customer.

B. Network Data

The Telecommunication network is very complex in nature. It consists of many elements and components. Sometimes an error may occur in these equipments while being in network. So they generate error and status messages. And these messages should be stored in database so as to support network management functions.

Fig. 5.(c) Network Architecture

Suppose if particular component is down for some time or it is overloaded and if status messages indicating theses problems are stored then proper action can be taken to resolve them by analysis. But every time manual analysis is difficult. Hence, Data mining technology is used to identify and resolve the network faults by extracting information from network data.

C. Customer Data

Age:	<25	25-40	40-55	>55
	21.2%	29.5%	27.9%	21.4%
Gender:	Male	Female		
	60.2%	39.8%		
Telephone type:	simple	basic	advanced	
	33.5%	38.7%	27.8%	
Type of subscription:	simple	advanced	expanded	
	34.9%	36.0%	29.1%	
Company size:	small	intermediate	big	
	31.5%	34.3%	34.2%	
Living area:	(big) city	small city/town		
	42.0%	58.0%		

Fig. 5.(d) Customer Data Table

The Telecommunication companies have huge amount of customers. So information about each of them needs to be maintained. Customer data includes details about their personal information like age, gender, income and telephone type, subscription type. Maintaining this customer information important so as to know profiles of customers. In the above chart certain parameters are given which indicates current status of them like there are total 29.5% customers of age 25-40 and also there are 21.2% of customers whose age is less than 25. The 38% customers use basic telephone type and 27.8% use advanced.

VI. DATA PREPARATION AND CLUSTERING

These are the tasks of Data Mining. They are discussed below:

A. Data Preparation

Before using actual data mining process the data should be prepared in the desired format. So data preparation involves many tasks. They are

- Discover useful data and repair inconsistency between these data formats
- Remove the spelling mistakes, and check for syntax errors
- Remove unnecessary data fields
- Normalize the tables
- Check blank data fields

B. Clustering

Clustering means grouping of similar things. So here we can see how clustering is done in Data mining.

Cluster Analysis:

A cluster means collection of objects which are similar between them and dissimilar to the objects belonging to other objects. The cluster analysis is done to organize objects into their respective groups as per the similarities between them[2].
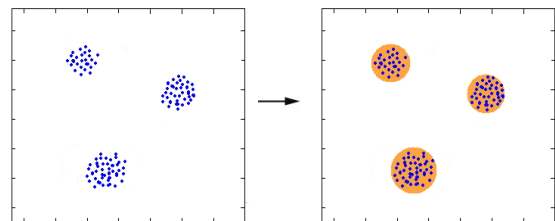

Fig. 6.(a) Clusters

In this case in first three clusters, similarities are found and then they are grouped accordingly.

VII. DATA MINING APPLICATIONS

A. Fraud Detection

Fig. 7.(a) Fraud Detection

Fraud is a dangerous problem for the Telecom industry. It results in huge loss to them. The common method to identify fraud is to build customer's profile of calling behavior which contains his previous calling pattern. Then his recent activity is compared against this behavior. So if any suspicious activity is seen then that fraud is caught. Thus, the data mining application depends on deviation detection. The call-detail records are summarized to obtain calling behavior. If the call detail summaries are updated in real-time then the fraud can be detected soon after it occurs.

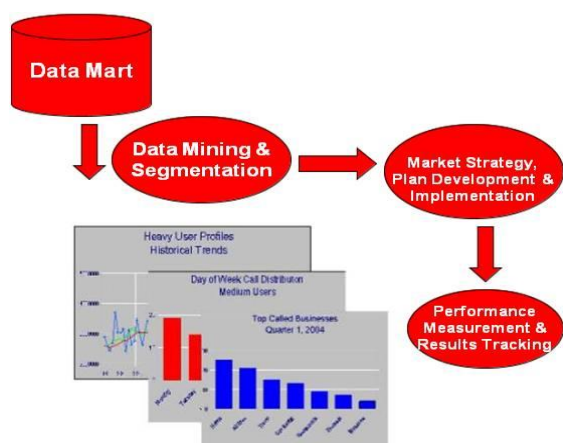*B. Marketing and Customer Profiling*

Fig. 7(b) Marketing Strategy

Marketing is an important thing for the Telecommunication companies. The companies maintains huge amount of data about their customers. This information includes his personal information like age, gender, lifestyle, knowledge etc which might be stored in Data mart. Then the Data mining tools are applied and segmentations is done on this large amount of data. And that is used to build customers' profile and then it is helps to build marketing strategies, planning for the future decisions, performance measurement, result tracking, etc. Hence companies come to know who are my customers ?, what is their gender?, what are their needs?, and how I can approach and offer new services to them? Therefore Customer profiling is an important aspect.

C. Network Fault Isolation

As mentioned earlier Telecommunication networks are of extremely complex configuration and they consist of many interconnected components. These network components generate status as well as alarm messages every time. So in order to identify the network faults the alarms should be analyzed automatically. So the data mining helps to automatically analyze and detect the faults so that they can be resolved.

VIII. CONCLUSION

The telecom industry is the early adopter of data mining technology. It reduced much of human based analysis. It has very helpful applications. It helps to detect frauds, to know the customers and serve them better as per their needs. And apart from customer side, to come to know how we can yield more profit. So overall it is very much essential in fact must for Telecommunication companies.

IX. REFERENCES

[1] Data mining in Telecommunication by Gray M. Weiss, Fordham University

[2] Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior, S.M.H Jansen, July 17, 2007

[3] IJSETT -Applications of Data Mining by Simmi Bagga and Dr. G.N. Singh

[4] Data Mining in Telecommunication Industry by Gray M. Weiss

[5] Industry Applications of Data Mining, Sept 6, 1999

[6] From Data mining to Knowledge Discovery in Databases by Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996)

❖ ❖ ❖