

Application of Data Mining in Telecommunication Industry

Manasvi Gawande, Manjiri Pise

E-Mail : manaswigawande96@gmail.com

pisemanjiri@gmail.com

Abstract: - Data Mining is a logical procedure intended to investigate data (normally a lot of data - commonly business or market related - otherwise called "enormous data") looking for predictable examples as well as methodical connections amongst factors, and after that to approve the discoveries by applying the recognized examples to new subsets of data. The telecommunications industry inside the division of data and correspondence technology is comprised of all Telecommunications/telephone companies and web access suppliers and assumes the urgent part in the development of versatile interchanges and the data society. Customary telephone calls keep on being the industry's greatest income generator, yet because of advances in arrange technology, Telecom today is less about voice and progressively about content (informing, email) and pictures (e.g. video gushing). Fast web access for PC based data applications, for example, broadband data administrations and intelligent stimulation, is unavoidable. Digital Subscriber Line (DSL) is the primary broadband telecom technology. The quickest development originates from (esteem included) administrations conveyed over portable systems.

Keywords:- Data Mining, Telecommunication Industry, Data, Knowledge, Churn.

I. Introduction

Data mining is one of advanced type of analytical tools at this time available; these tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that increase their performance automatically during experience, such as neural networks or decision trees) but data analysis system that does not deal with large amounts of data[2].

Data mining consists of collecting and managing data and includes analysis and prediction; but this process has limitations to its capability. One limitation is it does not tell the user the value or significance of the patterns and the relationships that were discovered let it to the user. A second limitation is that does not identify a causal relationship that was identified connections between behaviours and/or variables. Because this process to be successful requires skilled technical and logical specialists who can structure the analysis and explain the output that is produced [3].

In the telecommunications industry, huge amounts of data are produced and stored, including: Call data describing the calls that traverse the telecommunication networks, Network data concerning the state of hardware and software components and Customer related data. Within such tremendous amounts of business critical data, valuable knowledge may be hidden. Indeed, as seen above, it has previously been demonstrated how data mining can be used to discover new patterns and correlations within such sets of data. Applications include telecommunication fraud detection, improving market efficiency, and fault detection and localization. At the same time, a continuous work on customer retention and churn prevention becomes a

necessity, because the competition has similar acquisition issues. Retention of the existing users is important since it is 5 up to 7 times cheaper to retain a consumer than to acquire a new one[4].

Two basic groups of churners are voluntary and involuntary churners. Involuntary churners are the customers that Telecommunication Company decides to eliminate from the subscribers list. This class includes customers that are churned for fraud (customers who cheat), non-payment (customers with credit problem), and under-utilization (customers who don't use the phone)[5].

Telecommunications companies reserves data on telephone calls that cross their networks as a Call detail records, which contain descriptive information per phone call. Because of detailed call records maintained for months, billions of detailed calls records available and useful to be mined. Calls details data is useful for applications marketing and fraud detection[6]. Telecommunication companies also maintain large information about Customer, such as invoices information, in addition to information obtained from external parties, Information like credit record. A number of distinct approaches to extract the data has been identified, namely, classification, and association rule learning, clustering, and multi-dimensional scaling (data visualization)[7].

II. Informative Data

Data mining is a process that takes data as input and outputs knowledge. One of the earliest and most cited definitions of the data mining process, which highlights some of its distinctive characteristics, is provided by Fayyad, Piatetsky-Shapiro and Smyth (1996), who define it as "the nontrivial

process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.” Note that because the process must be non-trivial, simple computations and statistical measures are not considered data mining. Thus predicting which salesperson will make the most future sales by calculating who made the most sales in the previous year would not be considered data mining. The connection between “patterns in data” and “knowledge” will be discussed shortly. Although not stated explicitly in this definition, it is understood that the process must be at least partially automated, relying heavily on specialized computer algorithms (i.e., data mining algorithms) that search for patterns in the data.

It is imperative to point out that there is some vagueness about the expression "data mining", which is in expansive part deliberate. This term originally alluded to the algorithmic step in the data mining process, which initially was known as the Knowledge Discovery in Databases (KDD) process. Be that as it may, after some time this distinction has been dropped and data mining, depending on the specific circumstance, may allude to the whole procedure or simply the algorithmic step. This whole procedure, as originally imagined by Fayyad, Piatetsky-Shapiro and Smyth (1996), is appeared in Figure 1. We talk about the whole procedure, in any case, as is normal with most messages regarding the matter, we concentrate the vast majority of our consideration on the algorithmic data mining step.

The first three steps in Figure 1 involve preparing the data for mining. The relevant data must be selected from a potentially large and diverse set of data, any necessary preprocessing must then be performed, and finally the data must be transformed into a representation suitable for the data mining algorithm that is applied in the data mining step. As an example, the preprocessing step might involve computing the day of week from a date field, assuming that the domain experts thought that having the day of week information would be useful. An example of data transformation is provided by Cortes and Pregibon (1998). If each data record describes one phone call but the goal is to predict whether a phone number belongs to a business or residential customer based on its calling patterns, then all records associated with each phone number must be aggregated, which will entail creating attributes corresponding to the average number of calls per day, average call duration, etc.

While data planning does not get much consideration in the exploration group or the data mining group by and large, it is basic to the achievement of any data mining venture in light of the fact that without top notch data it is regularly

difficult to gain much from the data. Besides, albeit most research on data mining relates to the data mining algorithms, it is normally recognized that the decision of a particular data mining algorithms is for the most part less imperative than completing a great job in data arrangement. By and by it is regular for the data preparations steps to take additional time and exertion than the genuine data mining step. Along these lines, anybody undertaking a data mining venture ought to guarantee that adequate time and exertion is apportioned to the data arrangement steps. For those intrigued by this point, there is a book (Pyle 1999) that spotlights only on data planning for data mining.

The fourth step in the data mining process is the data mining step. This step involves applying specialized computer algorithms to identify patterns in the data. Many of the most common data mining algorithms, including decision tree algorithms and neural network algorithms . The patterns that are generated may take various forms (e.g., decision tree algorithms generate decision trees). At least for predictive tasks, which are probably the most common type of data mining task, these patterns collectively can be viewed as a model. For example, if a decision tree algorithm is used to predict who will respond to a direct marketing offer, we can say that the decision tree models how a consumer will respond to a direct mail offer. Finally, the results of data mining cannot simply be accepted, but must be carefully evaluated and interpreted. As a simple example, in the case of the direct-mail example just described, we could evaluate the decision tree based on its accuracy—the percentage of its predictions that are correct. However, many other evaluation or performance metrics are possible and for this specific example return on investment might actually be a better metric.

The data mining process is an iterative process, although this is not explicitly reflected in Figure 1. After the initial run of the process is complete, the user will evaluate the results and decide whether further work is necessary or if the results are adequate. Normally, the initial results are either not acceptable or there is an expectation that further improvements are possible, so the process is repeated after some adjustments are made. These adjustments can be made at any stage of the process. For example, additional data records may be acquired, additional fields (i.e., variables) may be generated from existing information or obtained (via purchase or measurement), manual cleaning of the data may be performed, or new data mining algorithms may be selected. At some point the results may become acceptable and the mined knowledge then will be communicated and may be acted upon. However, even once the mined knowledge is acted upon the data mining process may not be complete and have to be repeated, since the data distribution

may change over time, new data may become available ,or new evaluation criteria may be introduced[1].

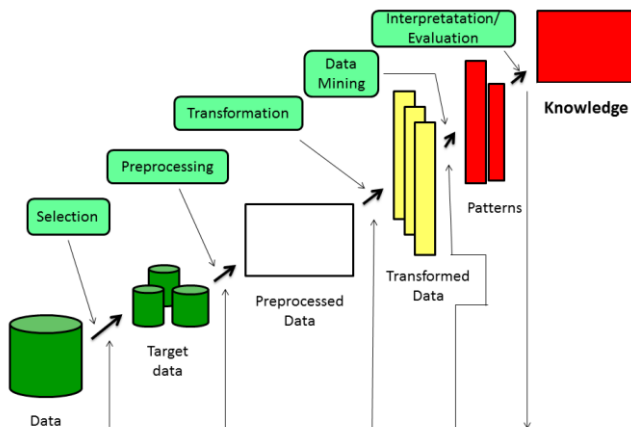


Figure 1

III. Application

Here is the list of areas where data mining is widely used –

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

Telecommunication Industry:

Today the telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, internet messenger, images, e-mail, web data transmission, etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business.

Data mining in telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list of examples for which data mining improves telecommunication services –

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.

- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis

Telecommunication market everywhere throughout the world is confronting a serious loss of income because of savage rivalry and loss of potential customers. To keep the competitive advantages and secure whatever number customers as could reasonably be expected, most administrators contribute a gigantic measure of income to extend their business in the absolute starting point [7]. Along these lines, it has turned out to be key for the administrators to obtain the sum contributed and to pick up no less than a base benefit inside a brief timeframe. Since it is especially testing and repetitive issue to keep the customers in place for a long span because of the opposition engaged with this business field. To make due in the market, telecom administrators generally offer an assortment of maintenance arrangements to draw in new customers. This is the real reason for the subscribers abandoning one system and moving to another which suits their necessities. As per telecom market, the procedure of subscribers (either paid ahead of time or post-paid) changing from one specialist organization is called „customer churn“ [8]. On the off chance that churning keeps on occurring for any telecom industry, it would prompt the immense loss of income to the organization. In this circumstance, the main solution for defeat such business dangers and to hold in the market, administrators are compelled to search for elective methods for utilizing data mining strategies and factual instruments to recognize the reason ahead of time and to take prompt endeavors accordingly. This is conceivable if the previous history of the customers is investigated efficiently. Luckily, telecom businesses create and keep up an expansive volume of data. They incorporate Billing data, Call detail Data and Network data. This voluminous sum data guarantees the extension for the utilization of data mining strategies in telecommunication database.

As plenty of information is hidden in the data generated by the telecom industries, there is a lot of scope for the researchers to analyze the data in different perspectives and to help the operators to improve their business in various ways. The most common areas of research in telecom databases are broadly classified into 3 types, i) Telecom Fraud Detection ii) Telecom Churn Prediction iii) Network Fault Identification and Isolation. Moreover, not all the data items of the telecom database are used by all the techniques. Only the relevant data items which really contribute to the specific analysis must be considered for any study. This study focuses on churn prediction, the importance of feature extraction (attribute selection) and the use of data mining techniques in churn prediction in telecomm data.

IV. Churn Prediction

In a business environment, the term, customer steady loss basically alludes to the customers abandoning one business administration to another. Customer churn or endorser churn is likewise like whittling down, which is the process of customers changing starting with one specialist co-op then onto the next secretly. From a machine learning viewpoint, churn prediction is a managed (i.e. named) issue characterized as takes after: Given a predefined gauge skyline, the objective is to foresee the future churners over that skyline, given the information related with every endorser in the system [9]. The churn prediction issue spoke to here includes 3 phases, namely, i) training phase, ii) test phase, iii) prediction phase. The contribution for this issue incorporates the information on past requires every versatile endorser, together with all individual and business data that is kept up by the specialist organization. Likewise, for the training phase, names are given as a rundown of churners. After the model is prepared with most noteworthy exactness, the model must have the capacity to foresee the rundown of churners from the genuine dataset which does exclude any churn name. In the point of view of knowledge discovery process, this issue is sorted as prescient mining or prescient demonstrating.

Churn Prediction is a phenomenon which is used to identify the possible churners in advance before they leave the network. This helps the CRM department to prevent subscribers who are likely to churn in future by taking the required retention policies to attract the likely churners and to retain them. Thereby, the potential loss of the company could be avoided. This study utilizes data mining techniques to identify the churners.

References:

- [1] Data Mining, Gary M. Weiss
- [2] H. Newton, Newton's Telecom Dictionary, CMP Books, <http://www.cmpbooks.com>
- [3] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth (1996), Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the 2nd ACM international conference on knowledge discovery and data mining (KDD), Portland, OR, pp 82-88
- [4] Sheth, Jagdish, and Rajendra Sisodia. The 4 A's of Marketing: Creating Value for Customer, Company and Society. Routledge, 2012.
- [5] Kraljević, Goran, and Sven Gotovac. "Modeling data mining applications for prediction of prepaid churn in telecommunication services." AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije 51.3 (2010): 275-283.
- [6] Privacy Preserving Data Mining in Telecommunication Services) Granmo, O., and VLADIMIR A. Oleshchuk.
- [7] Yu-Teng Chang, "Applying Data Mining To Telecom Churn Management", IJRIC, 2009 67-77.
- [8] Rongbin Zhang, Ronghua, Yanying Zhrng, Michael Berry. "Behavior -Based Telecommunication Churn Prediction with Neural Network Approach", page 307-310, ISCCS.2011.
- [9] Yossi Ritchev, Elad Yom-Tov, Noam Slonim, "Predicting Customer Churn in Mobile Networks through Analysis of Social Groups". SIAM.