Chapter 2
Role of Data Mining and Knowledge Discovery in Managing Telecommunication Systems

Ibrahiem Mahmoud Mohamed El Emary
King Abdulaziz University, Saudi Arabia

ABSTRACT

This chapter is interested in discussing how to use data mining techniques to assist in achieving an acceptable level of quality of service of telecommunication systems. The quality of service is defined as the metrics which are predicated by using the data mining techniques, decision tree, association rules and neural networks. Routing algorithms can use this metric for optimal path selection which in turn will affect positively on the system performance. Also, in this chapter management axis using data mining techniques were handled, i.e., check the status of the telecommunication networks, role of data mining in obtaining optimal configuration, how to use data mining technique to assure high level of security for the telecommunication. The popularity of data mining in the telecommunications industry can be viewed as an extension of the use of expert systems in the telecommunications industry. These systems were developed to address the complexity associated with maintaining a huge network infrastructure and the need to maximize network reliability while minimizing labor costs (Liebowitz, J. 1988). The problem with these expert systems is that they are expensive to develop because it is both difficult and time consuming to elicit the requisite domain knowledge from experts.

INTRODUCTION

Normally collecting and storing data is outpaced the ability to analyze, summarize and extract knowledge from the continuous stream of input data. In database, knowledge discovery represents the complex process of identifying valid, novel, potentially useful and ultimately understandable patterns of data. Data mining which is a particular step in the KDD process consists of particular algorithms that under acceptable computational efficiency limitations produce a particular enumeration of patterns (models) over the data. Data mining technology has emerged as a means of

DOI: 10.4018/978-1-61520-805-0.ch002

identifying patterns and trends from large quantities of data. The Data Mining technology normally adopts data integration method to generate data warehouse, on which to gather all data into a central site, and then run an algorithm against that data to extract the useful module prediction and knowledge evaluation.

The major Data Mining (DM) tasks are: classification which means assigning each record of a database to one of a predefined set of classes, clustering that deals with finding groups of records that are close according to some user defined metrics or association rules that determines implication rules for a subset of record attributes. In research community, Data Mining have many challenges covering the following:

- In various application domains, the mined data is produced with high rate or come in stream, so in those cases, knowledge has to be mined fast and efficiently in order to be updated and useful. Accordingly, the input data is changed rapidly.
- In a lot of organizations, security is of major concern. Here, there may be willing to release data mining results but not the source data itself.
- It is urgently to partition and distribute the data for parallel processing to achieve an acceptable space and time performance. Based on this fact, data mining is very computationally intensive process involving very large data sets.
- Data mining deal with huge amounts of data located at different sites where the amount of data can easily exceed the terabyte limit;

According to the above challenges, a basic approach for data mining which is to move all of the data to a central data repository and then to analyze them with a single data mining system, even though it guarantees accurate results of data analysis, it might be infeasible in many cases. A an alternative approach to the above is to use high level learning with in-place strategies in which all the data can be locally analyzed, and the local results at their local sites are combined at the central site to get the final result which mean building a global data model. This approach is less expensive but may produce incorrect and ambiguous global results.

Telecommunication sector represents one of the tremendous growths among the technology during the last century. With the debut of intelligent networks (INs), service providers and specialized ISVs gained a hand in developing new services. The basic function of a service management system is to manage a service network for contracted customers. This system can efficiently support proactive troubleshooting in case of a network error, analyzing errors on the service level and correlating the network level error messages with the service topology purchased by the specific client. If any matching result is found, the operators receive an error message on the service involved. A service is defined as a network with a P2P or VPN topology running on any network technology based on customer needs and technical conditions.

In order to perform the tasks above, the service management system must be in contact with the network management systems managing the network technologies. The service management system receives the network-level alarms from these network management systems. The provider receives the network topology needed to interpret the specific alarms from the technical inventory management system using a special interface as well, although they often took a long time to deliver to customers. One of the major requirements for IT systems in telecommunications environments is achieving an acceptable level of reliability and scalability. In addition, there is a strong need for integration due to the above-average number and significance of heterogeneous IT support systems in modern day telecommunications companies. Accordingly, the IT industry attempts to meet those requirements by deploying innovative soft-