

Network Security and Social Network Analysis in Data Mining

I.Kalaiselvi^{#1}, R. Anitha^{*2}

¹Master of Computer Applications, S.A. Engineering college, Chennai-77.
kalai17121993@gmail.com

²Asst. Prof., Department of Computer Applications, S.A. Engineering College, Chennai-77.
anitha@saec.ac.in

Abstract— Data mining can be a controlling tool for extracting useful information from tons of data. The growing popularity and development of data mining technologies is a serious threat to the security of individual's sensitive information. Data mining is a collection of stores and organizes data for use in areas such as medicine, finance, intelligence, law enforcement, security, logistics, education, and process control. Key to avoiding the pitfalls is a basic understanding of what data mining is a network security and planning a data mining project. A social network is defined as a social structure of individuals, who are related based on a common relation of interest. The basic motivation is the demand to exploit knowledge from copious amount of data collections, pertaining to social behavior of user in online environments.

Keywords— Data mining; Sensitive Information; Environments.

1. Introduction

Data mining has involved more and more attention in recent year, probably because of the popularity of the “big data”. Some presentations are used such as business intelligence, web search, scientific discovery, ordinal libraries, etc. Real world data are dirty data; sound handling is a defining representative for data mining research and applications. Data mining in social network analyses a brief overview of research directions.

In individual, we identify four different types of users involved in data mining network security applications, namely, data provider, data collector, data miner, and decision maker. Recent developments in dynamism market deregulation and provision of justifiable energy have contributed to increased interest in this area. Electricity meters record energy ingestion and power quality at a preset interval, usually an hour or less. Readings from meters in a specific geographical area are transmitted in real-time to a central location where it can be analysed to produce immediate outcomes that support decision-making. The availability of accurate and updated information on energy consumption presents the potential to transform demand forecasting and energy conservation from passive historical data-based activities to active real-time data driven operations. Research in energy consumption

analysis is traditionally conducted in three distinct areas based on the time scale of investigation; short, medium and long term. A short-term forecast attempts predict several hours or days in advance, medium- term predictions focus on weeks to months and long-term forecasting looks at years.

1.1 Information Security in Big Data

The major worry of a data provider is whether he can control the understanding of the data he provides to others. The provider should be able to make his very private data, specifically the data comprising information that he does not want anyone else to know, unapproachable to the data collector. The provider has to provide some data to the data collector, he requests to hide his sensitive statistics as much as possible and get enough returns for the possible loss in privacy.

1.2 Data Mining Of Social Network Analysis

A social network is a defined as a social structure of individuals, who are associated (directly or indirectly to each other) based on a common relative of interest, e.g. friendship, trust, etc. Social network analysis is the learning of social networks to recognize their structure and behavior. Social network analysis has gained prominence due to its use in changed requests - from product marketing (e.g. viral marketing) to search engines and organizational crescendos (e.g. management). Freshly there has been a rapid increase in interest regarding social network analysis in the data mining network security community. The basic incentive is the demand to exploit knowledge from copious amounts of data collected, relating to social conduct of users in working environments. A prime example of this are the research efforts devoted towards the Enron email dataset.

1.3 Signal-Integrity Optimization

A black-box model of high-speed MIMO interconnect networks in N-port S-parameters can be considered as a large data set, especially when the port number is large. The port to- port S-parameters has fully characterized the MIMO system under arbitrary input patterns. The overall

system responses under multiple inputs, which are concealed within the S-parameters, must be extracted by special means. For a passive linear time-invariable MIMO system, weighting method can be used. For convenience, the M+N ports are grouped into M input ports (stimulated ports) and N output ports (response ports), as shown in Fig. 1(b). According to the linear time-invariable system theory, the M+N-port S-parameter of the MIMO network is given. If there are multiple inputs with variable magnitude and phase stimulating simultaneously, then the final response of the MIMO system.

1.4 Mining with Noise Knowledge

Data discretization errors - Data discretization is a general procedure of discretizing the field of a nonstop variable into a limited number of intervals [46], [47]. Because this process uses a certain number of distinct values to estimate infinite continuous values, the difference between the discrete value and the actual value of the uninterrupted variable thus leads to a possible error.

Information transformation errors - Information transformation, particularly wireless networking, often raises a certain amount of errors in communicated data. For error organize purpose, the mathematics errors of the signal broadcast channel should be explored in expansion and can be utilized to estimate the error rate in the distorted info.

1.5 Data Mining Framework

The dimensional model is able to organize dimensional data on granularity. In addition to the time dimension which is a natural hierarchy, geography and meter dimensions pose a further interesting hierarchy. Information regarding the highest energy consuming week, month, quarter, and meter can be extracted from the granular structure. The presence of trends across time at different levels of granularity can also be identified using interim summarization. The availability of accurate and updated information on energy consumption presents the potential to transform demand forecasting and energy conservation from passive historical data-based activities to active real-time data driven operations [1]. Research in energy consumption analysis is traditionally conducted in three distinct areas based on the time scale of investigation; short, medium and long term.



Fig.1: Data Mining Process

The data miner applies mining processes to the data provided by data collector, and he wishes to abstract useful information from data in a privacy- preserving manner. Current models and algorithms proposed for PPDM mainly focus on how to hide that searching information from certain mining operations. Though, as demonstrates in Fig. 1, the entire KDD procedure engrosses multi-phase maneuvers.

2. Methodology

2.1 The KDD Process

- Data preprocessing: Tundamental processes comprise data collection, data onslaught, and data incorporation
- Data transformation: The goal is to transform data into forms appropriate for the mining task, find useful features to represent the data. Feature selection and feature transformation are basic operations.
- Data mining: This is an essential process is intelligent methods are employed to extract data patterns (association rules, clusters, classification rules, etc.)
- Pattern appraisal and presentation: Basic operations include identifying and presentation patterns which represent knowledge, and giving the mined knowledge in an easy-to-understand fashion.

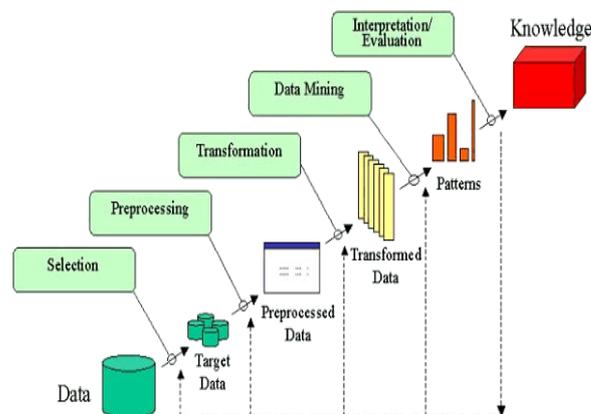


Fig.2: KDD Process

- Data integration: Integration of various databases, data cubes, or files.
- Data cleaning: Data cleaning tasks is used to fill in missing values, identify outliers and smooth out noisy data, correct unreliable data and smooth the noisy data.
- Noisy Data: It is random error or variance in a measured variable. Noisy data comes from the process of data collection, data entry, data transmission, In the words, incorrect attribute values may due to,
 - Injured data compilation appliances.
 - Data access troubles.
 - Data communication problems.

- Machinery negative aspect.
- Contradiction in naming decision.

2.2 Social Network study

Social network study has expanded significance due to its employ in different apps from creation marketing to hunt engines and prearranged dynamics. The necessary inspiration is the command to commotion acquaintance from plentiful amounts of data composed, be appropriate to social recital or clients in online backgrounds.

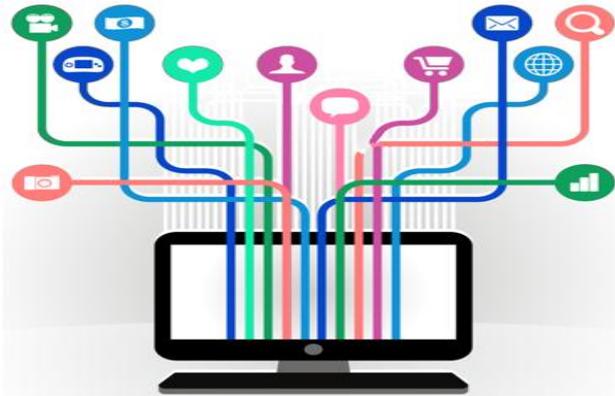


Fig. 3: Social Network Analysis

Some applications used in social network analysis are,

- Message
- Search
- Contacts
- Google
- Camera

3. Existing System

The most fundamentals challenge for big data applications is to search the large volumes data and remove useful information or knowledge for future action. In many

situations the knowledge removal process has to be very efficient and close a real time because storing all observed data is nearly infeasible. Data analysis and prediction platform to achieve fast reply and real time classification for such big data.

4. Conclusion

In order to apply data mining, a large amount of quality data is required. The aim of data mining is acquiring rules and equations which can be used predict future. To be successful on such a work is dependent on working with database authorities and data mining specialists. They need to work together. Work may take longer, you need time and patience. we differentiate four different user roles that are commonly involved in data mining applications, It can provide data provider, data collector, data miner and decision maker.

5. Future Enhancement

- Data speed is very high.
- Third parties should not be attack the information.
- Efficient.
- Avoid redundancy.

References

- [1] R. Agrawal and R. Srikant, "The Privacy -Protective data mining," in Proc. ACM SIGMOD, 2000, pp. 439-450.
- [2] J. Han, M. Kamber, and J. Pei, "Data Mining : Concepts and Techniques, San Mateo, CA, Morgan Kaufmann", Vol.3, 2006, pp.611-623.
- [3] E.Rasmusen, "Games and information: An Introduction to game Theory", Vol. 2, Cambridge, MA, USA: Blackwell,1994, pp.451-457.
- [4] R.Loganathan and N.Vivekananda Moorthy, "GPS-Based Smart System for Enhancing Driving Directions for Finding Fastest Route using Driver's Intelligence", International Journal of Linguistics and Computational Applications, Volume 2, Issue 1, January - March 2015, pp.18-24.
- [5] Jyothi Mandala and Suneetha Merugula, "Review of Privacy Preserving Data Mining Techniques", Engineering and Scientific International Journal, Volume 2, Issue 1, January - March 2015, pp.1-3.