Chapter 2

Data Preprocessing for Dynamic Social Network Analysis

Preeti Gupta
Ambedkar Institute of Advanced Communication Technologies & Research, India

Vishal Bhatnagar
Ambedkar Institute of Advanced Communication Technologies & Research, India

ABSTRACT

The social network analysis is of significant interest in various application domains due to its inherent richness. Social network analysis like any other data analysis is limited by the quality and quantity of data and for which data preprocessing plays the key role. Before the discovery of useful information or pattern from the social network data set, the original data set must be converted to a suitable format. In this chapter we present various phases of social network data preprocessing. In this context, the authors discuss various challenges in each phase. The goal of this chapter is to illustrate the importance of data preprocessing for social network analysis.

INTRODUCTION

Social network has become one of the most important communication media. Social network consists of a wide range of social media services which are used by the members for various purposes. The number of members registered with these social media services is huge and is further increasing. Social network analysis has been established as an important area of research due to its increased usage. Social network contains abundant information which is structurally complex, heterogeneous, high dimensional and incremental in nature.

Analyzing the social network data for information mining has the potential of revealing information of great value. Social network analysis provides a systematic method to identify, examine, visualize and support processes of knowledge sharing in social networks (Müller-Prothmann, 2006) such as expert finding and search ranking. Social network analysis plays an integral in
knowledge extraction from the rich data source. Social network analysis can help in determining various queries such as how the information flows, important members or links and network dynamics. Social network analysis may also help in revealing otherwise unobservable information. Aim of social network analysis is to discover and retrieve useful and interesting patterns from a large dataset. Social network data contains different kind of data such as profile data, log data, structural data, documents, audio and video which could be broadly classified as (Gupta & Bhatnagar, 2012)

- **Profile Data**: Profile data is the data member provides to social network provider for registering itself.
- **Posted Data**: The data which is explicitly posted by the members such as messages, comments and blog entries.
- **Derived Data**: The data which is derived or mined by correlating other information.

Social network data could be viewed and modeled as a graph where individuals or groups can be represented by nodes or vertices and relationship or flow between individuals represent links. The social network graph is highly dynamic, very large and sparse in nature. Social network analysis enables mapping of relationships between people, groups and other related information entities. Social network analysis thus helps in understanding and analyzing the social network structure and evolution. Social network analysis is a powerful approach to understand the subtleties of social network. Social network analysis has great utility value due to which it has been applied fields such as community detection, evolution in dynamic social networks, social influence analysis, link prediction, privacy in social networks, data mining and text mining (Aggarwal, 2011). Ferguson and Buckingham (2012) have broadly classified social analytics into network analytics, discourse analytics, content analytics, dispositions analytics and context analytics. Social network analysis provides a systematic method to identify, examine, visualize and support processes of knowledge sharing in social networks (Müller-Prothmann, 2006).

The social network is just not the social networking sites such as Facebook or Twitter; it is much more. A social network consists of service provider, user, social media services and third parties. Kaplan and Haenlein (2010) have broadly categorized social media into six parts namely collaborative projects, blogs, content communities, social networking sites, virtual game worlds, and virtual communities. Growth of social media services have made people to use these services for sharing and exchanging the information at an
Related Content

Discovery of Latent Patterns with Hierarchical Bayesian Mixed-Membership Models and the Issue of Model Choice
www.igi-global.com/chapter/discovery-latent-patterns-hierarchical-bayesian/7568?camid=4v1a

A Semi-Automatic Annotation Method of Effect Clue Words for Chinese Patents Based on Co-Training
Na Deng, Chunzhi Wang, Mingwu Zhang, Zhiwei Ye, Liang Xiao, Jingbai Tian, Desheng Li and Xu Chen (2018). International Journal of Data Warehousing and Mining (pp. 1-19).
www.igi-global.com/article/a-semi-automatic-annotation-method-of-effect-clue-words-for-chinese-patents-based-on-co-training/215003?camid=4v1a

Mathematical Statistical Examinations on Script Relics
Gábor Hosszú (2014). Data Mining and Analysis in the Engineering Field (pp. 142-158).
www.igi-global.com/chapter/mathematical-statistical-examinations-on-script-relics/109980?camid=4v1a