ABSTRACT

With the success of social media, social network analysis has become a very hot research topic and attracted much attention in the last decade. Most studies focus on analyzing the whole network from the perspective of topology or contents. However, there is still no systematic model proposed for multi-dimensional analysis on big social media data. Furthermore, little work has been done on identifying emerging new popular phrases and analyzing them multi-dimensionally. In this paper, the authors first propose an interactive systematic framework. In order to detect the emerging new popular phrases effectively and efficiently, they present an N-Pat Tree model and give some filtering mechanisms. They also propose an algorithm to find and analyze new popular phrases multi-dimensionally. The experiments on one-year Tencent-Microblogs data have demonstrated the effectiveness of their work and shown many meaningful results.

Keywords: Multidimensional Analysis, Popular Phrase, Social Media, Social Network Analysis

DOI: 10.4018/IJDWM.2015070105
1. INTRODUCTION

Social media, such as Flickr, Facebook, YouTube, MySpace, Twitter etc., has emerged as one of the most popular platform for people to communicate with each other. People engaging in those online social activities can post their ideas, browse others' posts, write comments and forward interesting posts, and follow anyone that they are interested in. Thus, analyzing those data has many potential applications. Therefore, it has attracted tremendous attention of researchers.

Recent studies are paying more attention to users' characteristics (Krishnamurthy, Gill, & Arlitt, 2008), statistical analysis (Java, Song, Finin & Tseng, 2007; Du, Wu, Pei, Wang & Xu, 2007), community detection (Zhao, Feng, Wang, Huang, Williams & Fan, 2012), propagation models (Galuba, Aberer, Chakraborty, Despotovic & Kellerer, 2010) and predictive methods (Zhao & Rosson, 2009). Data mining and learning techniques, such as clustering (Kwok, Smith, Lozano & Taniar, 2002), association rule mining (Taniar, Rahayu, Lee & Daly, 2008; Daly & Taniar, 2004) and model learning (Tan & Taniar, 2007), can be applied in the area of social network analysis. Java et al. have conducted preliminary analysis on microblogging behaviors by studying the topological and geographical properties of Twitter network (Java, Song, Finin & Tseng, 2007). Krishnamurthy, Gill, & Arlitt (2008) have analyzed users’ characteristics through following-followed relationships. Zhao & Rosson (2009) have qualitatively investigated the motivation of using Twitter. Galuba, Aberer, Chakraborty, Despotovic & Kellerer (2010) have proposed a propagation model, which has been used to predict which users tweet about which URLs based on their historical activities. Jansen et al. have studied the word-of-mouth branding in Twitter (Jansen, Zhang, Sobel & Choudury, 2009). Aral, Muchnik & Sundararajan (2009) have distinguished the effects of homophily from influence as motivator for propagation. As to the study of influence within Twitter, Cha, Haddadi, Benevenuto & Gummadi (2010) performed a comparison of three different measures of influence-degrees, retweets, and users’ mentions. In a word, researchers have done a lot of work on social media data mining and analysis.

However, the previous studies have ignored the multi-dimensional analysis on the big social media. Furthermore, social media is generating a variety of new popular phrases every hour. It is a challenging problem to detect those new popular phrases efficiently and analyze them multi-dimensionally.

In this paper, we propose an interactive and multi-dimensional analyzing framework for big social media data. Then we discover and analyze the new popular phrases from different dimensions based on the proposed framework. We propose a N-Pat Tree model, an efficient data structure, to discover phrases from social texts. We also present some filtering mechanisms to get new popular phrases. Finally, we conduct some experiments on one-year Chinese Microblogs data from more than 0.4 million users’ Microblogs. We detect a large number of popular phrases from those data and then analyze them multi-dimensionally. The results have demonstrated the effectiveness of our work, and shown us many meaningful results from different dimensions.

The paper is organized as follows. In section 2, we give a multi-dimensional and interactive analyzing framework. In section 3, we present a new data structure called N-Pat Tree. Section 4 proposes some filtering mechanisms to find new popular phrases. Section 5 presents the evaluation of the approach of detecting popular phrases. The discovery and multi-dimensional analysis on new popular phrases of Tencent Microblogs are shown in section 6. Section 7 concludes the main work of this paper.

2. THE FRAMEWORK TO ANALYZE SOCIAL MEDIA DATA MULTI-DIMENSIONALLY

It is an undeniable fact that social media is generating large volume data every day. Those
Elasticity in Cloud Databases and Their Query Processing
[www.igi-global.com/article/elasticity-cloud-databases-their-query/78284?camid=4v1a](www.igi-global.com/article/elasticity-cloud-databases-their-query/78284?camid=4v1a)

Dimensionality Reduction with Unsupervised Feature Selection and Applying Non-Euclidean Norms for Classification Accuracy
[www.igi-global.com/chapter/dimensionality-reduction-unsupervised-feature-selection/61170?camid=4v1a](www.igi-global.com/chapter/dimensionality-reduction-unsupervised-feature-selection/61170?camid=4v1a)

I-BAT: A Data-Intensive Solution Based on the Internet of Things to Predict Energy Behaviors in Microgrids
Antonio J. Jara, Luc Dufour, Gianluca Rizzo, Marcin Piotr Pawlowski, Dominique Genoud, Alexandre Cotting, Yann Bocchi and Francois Chabbey (2016). *International Journal of Data Warehousing and Mining* (pp. 39-61).
[www.igi-global.com/article/i-bat/146852?camid=4v1a](www.igi-global.com/article/i-bat/146852?camid=4v1a)