

Chapter 10

Efficient Metaheuristic Approaches for Exploration of Online Social Networks

Zorica Stanimirović
University of Belgrade, Serbia

Stefan Mišković
University of Belgrade, Serbia

ABSTRACT

This study presents a novel approach in analyzing big data from social networks based on optimization techniques for efficient exploration of information flow within a network. Three mathematical models are proposed, which use similar assumptions on a social network and different objective functions reflecting different search goals. Since social networks usually involve a large number of users, solving the proposed models to optimality is out of reach for exact methods due to memory or time limits. Therefore, three metaheuristic methods are designed to solve problems of large-scaled dimensions: a robust Evolutionary Algorithm and two hybrid methods that represent a combination of Evolutionary Algorithm with Local Search and Tabu Search methods, respectively. The results of computational experiments indicate that the proposed metaheuristic methods are efficient in detecting trends and linking behavior within a social network, which is important for providing a support to decision-making activities in a limited amount of time.

1. INTRODUCTION

The term *big data* refers to vast amounts of information that originates from variety of sources, such as transactional records, log files, social media, sensors, third parties, Web applications, etc. *Big data* should be distinguished

from *large amount of data*, since big data is not just about giant data volumes. It's also about an extraordinary diversity of data types, delivered at various speeds and frequencies. There are two important issues concerning big data, which occupy attention of researchers in past several years.

DOI: 10.4018/978-1-4666-4699-5.ch010

- **Big Data Storage:** Big data is characteristically generated in large volumes per individual data set and in high frequency, meaning that information is collected at frequent time intervals. Additionally, big data is usually not nicely packaged in a spreadsheet or even a multidimensional database and often involves unstructured, qualitative information as well. Data warehouses are popular technologies for managing large volumes of data. However, they mostly rely on a relational format for storing data, which works fine for structured data, but not so successful for unstructured data. For example, relational databases are good for handling discrete packets of information, but they are less able to handle content such as video or sound files or emails, which do not necessarily conform to a rigid structure.
- **Big Data Analytics:** In contrast to traditional data, big data varies in terms of volume, frequency, variety and value. It is difficult to analyze big data with traditional data analytics tools that were not designed having these massive data sets in mind. With data volumes growing at an alarming rate, the importance of big data analytics has never been greater. For business enterprises, it is important to have a real time or near-real time information delivery that will allow analysts to quickly spot trends and avoid business problems. With the ability to comprehensively analyze large volumes of disparate and complex data, data analytics can help senior and board-level executives in better understanding and managing business opportunities.

In this study, we focus our attention to big data analytics techniques. The big data represents great challenges for analytics in general (Cuzzocrea et al., 2011). Developing adequate big data analysis techniques may help to improve decision making

process and minimize risks by unearthing valuable insights that would otherwise remain hidden. In some perspective, an automated decision making software can be provided by using big data analytics to automatically fine-tune inventories in response to real-time sales.

In the literature, clustering techniques are mainly used for big data analysis. Clustering is a process of organizing data into groups according to certain property or similarity. It is used for discovering natural groups or underlying structure of a given data set in many fields, for example in text mining (Dhillon et al., 2002), social network analysis (Sharma & Gupta, 2010), bioinformatics (Enright & Ouzounis, 2000), market research (Vakharia & Mahajan, 2000), etc. Parallel clustering technique showed to be an effective way to clustering big data (Jain et al., 1999). Recently, Chen et al. (2011) investigated parallel spectral clustering in distributed computing systems, and showed that a parallel clustering algorithm can effectively handle massive data problems.

Lin and Cohen (2010) introduced power iteration clustering (PIC), which replaces the eigendecomposition of the similarity matrix required by spectral clustering by a small number of matrix-vector multiplications, which leads to a great reduction in the computational complexity. They have demonstrated that the PIC algorithm outperformed several spectral clustering methods in terms of clustering accuracy. Yan et al. (2013) further expand PIC's data scalability by implementing a parallel power iteration clustering strategy (p -PIC). They propose a strategy to fit the data and its associated similarity matrix into memory, which makes the p -PIC algorithm appropriate for applications in big data analysis. Experimental results demonstrate that the parallel p -PIC implementation scales well, regarding the data size, and minimizes computation and communication costs.

Chiba and Sato-Ilic (2012) consider analysis of web survey data based on similarity of fuzzy clusters. The authors propose a method that cap-

46 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/efficient-metaheuristic-approaches-for-exploration-of-online-social-networks/85458

Related Content

Towards Autonomic Workload Management in DBMSs

Baoning Niu, Patrick Martinand Wendy Powley (2009). *Journal of Database Management* (pp. 1-17).
www.irma-international.org/article/towards-autonomic-workload-management-dbmss/4122

E-R Approach to Distributed Heterogeneous Database Systems for Integrated Manufacturing

Hemant Jainand Mohammed I. Bu-Hulaiga (1992). *Journal of Database Administration* (pp. 21-29).
www.irma-international.org/article/approach-distributed-heterogeneous-database-systems/51107

Data Warehousing and OLAP

Jose Hernandez-Orallo (2005). *Encyclopedia of Database Technologies and Applications* (pp. 127-133).
www.irma-international.org/chapter/data-warehousing-olap/11134

Information Search Patterns in E-Commerce Product Comparison Services

Fiona Fui-Hoon Nah, Weiyin Hong, Liqiang Chenand Hong-Hee Lee (2012). *Cross-Disciplinary Models and Applications of Database Management: Advancing Approaches* (pp. 131-145).
www.irma-international.org/chapter/information-search-patterns-commerce-product/63665

Methodology of Schema Integration for New Database Applications: A Practitioner's Approach

Joseph Fong, Kamalakar Karlapalem, Qing Liand Irene Kwan (1999). *Journal of Database Management* (pp. 2-18).
www.irma-international.org/article/methodology-schema-integration-new-database/51209