Multi-Fuzzy-Objective Graph Pattern Matching with Big Graph Data

Lei Li, Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Hefei, China & School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China
https://orcid.org/0000-0002-5374-7293

Fang Zhang, Zhongxing Telecommunication Equipment Corporation, Nanjing, China
Guanfeng Liu, Macquarie University, Sydney Australia
https://orcid.org/0000-0001-8980-4950

ABSTRACT

Big graph data is different from traditional data and they usually contain complex relationships and multiple attributes. With the help of graph pattern matching, a pattern graph can be designed, satisfying special personal requirements and locate the subgraphs which match the required pattern. Then, how to locate a graph pattern with better attribute values in the big graph effectively and efficiently becomes a key problem to analyze and deal with big graph data, especially for a specific domain. This article introduces fuzziness into graph pattern matching. Then, a genetic algorithm, specifically an NSGA-II algorithm, and a particle swarm optimization algorithm are adopted for multi-fuzzy-objective optimization. Experimental results show that the proposed approaches outperform the existing approaches effectively.

KEYWORDS

Big Graph Data, Fuzzy, Graph Pattern Matching, Multi-Objective

INTRODUCTION

Since the concept of Big Data has been proposed by the paper Big data: How do your data grow? published in Sep 2008 at Nature (Lynch, 2008), Big data have become the focus of both academics and industry. Recently, the rising and popularity of new network applications, such as composite services, social networks, biological networks and internet of things, on the one hand, increase the velocity and scale of the data generation (Li et al., 2016); on the other hand, these lead to a dramatic promotion of the complexity of big data processing and analyzing. The intricate relationship between data makes the processing complexity of big data involving network structures and data connections (named big graph data) become exponential. In addition to the exponential increase of data scale, this leads to a double exponential increase of complexity of big graph data processing and analyzing. However, there is a significant leap between this situation and the current slow increase of big graph data storage and computation. Hence, it is a common basic problem for big graph data computation that how to effectively and efficiently process and analyze big graph data. Sophisticated analysis and intelligent understanding of big graph data have become an important bottleneck for big graph data cognition and value mining.
Big graph, particularly composite service or social network attracts more and more attention from scholars, where one of the most important research areas is graph pattern matching, especially with big graph data. With the help of graph pattern matching, we can design a pattern graph satisfying special personal requirements, and locate the subgraphs which match the required pattern, which can finish the required task with the matched subgraph, such as: social status detection (Cheng et al., 2008), expert finding (Cheng et al., 2013; Tong et al., 2007), travel planning (Tong et al., 2007), research group selection, and etc. Big graph data are different from traditional data with a simple relation, and they usually contain complex relationship. For example, in composite service, vertices represent services, such as selling a product online (i.e., the traditional online service), or a functional component implemented by using Web service technologies, edges represent the relation of services, and together they can function as a whole; in social networks, vertices represent participants, and edges represent social relations which can be friendship or working relationship in real, co-authorship of a paper, or superior-subordinate relationship etc. In addition, vertices or edges can have some attribute information, such as the vertex of a social contextual graph (Henzinger et al., 2002) can have the information about social roles, and its edge can have attribute information about social trust and social intimacy. Generally, this rich metadata information of big graph data enriches the description of the data in a wide range of contextual information, such as source, time, occurrence probability, and other related metadata about the data (Li et al., 2015), which is significant for the subsequent data processing. Taking big graph data related to medical issues as an example, due to difference of medical equipment and software, and personal experience of doctors and other reasons, the data collected by certain medical institutions may be taken as untrustworthiness by others. These data with different attributes can produce intricate correlations and then converge and form a big graph. Then, how to locate a graph pattern with certain attributes in the big graph effectively and efficiently becomes a key problem to analyze and deal with big graph data, especially for a specific domain (Pienta et al., 2014).

The major contribution of the paper can be summarized as follows:

1. In traditional multi-constraint graph pattern matching, there may be a subgraph with a certain attribute value slightly lower than the requested value and other attributes significantly better than the requested ones. This should be a good subgraph which satisfies the requested graph pattern in many applications, but it has been ruled out by traditional graph pattern algorithms. In this paper, we introduce fuzzy numbers to include the above cases.

2. As the increment of graph scale and network complexity, given a graph pattern, there may be thousands of matched subgraphs. In the literature, as for multi-attributes, we usually focus on multi-constraint graph pattern problems, which have just one objective and several constraints. However, in real applications, we usually require several attributes to be optimized, which leads to a multi-objective optimization problem. In addition to fuzziness, in this paper, a genetic algorithm, specifically an NSGA-II algorithm, and a particle swarm optimization algorithm are adopted for multi-fuzzy-objective optimization, with which the requested optimized subgraph can be selected. Experimental results show that our proposed approaches outperform the existing approaches effectively.

This paper is organized as follows. Firstly, we describe the research progress about big knowledge engineering and graph pattern matching. Secondly, it introduces main ideas and steps of our proposed multi-fuzzy-objective graph pattern matching algorithm. In the following section, the experimental results of the proposed algorithms are calculated on real networks, showing that the proposed approach can not only greatly improve the matching accuracy but also can reduce the computation time. Finally, it concludes our work in this paper.
Matching Attributes across Overlapping Heterogeneous Data Sources Using Mutual Information
Huimin Zhao (2010). *Journal of Database Management* (pp. 91-110).
www.igi-global.com/article/matching-attributes-across-overlapping-heterogeneous/47421?camid=4v1a

On Efficient Evaluation of XML Queries
www.igi-global.com/chapter/efficient-evaluation-xml-queries/52959?camid=4v1a