Finding strongly connected components of simple digraphs based on generalized rough sets theory

Taihua Xu\textsuperscript{a,b}, Guoyin Wang\textsuperscript{a,b,*}

\textsuperscript{a}School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 610031, PR China
\textsuperscript{b}Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, PR China

\textbf{Article history:}
Received 2 April 2017
Revised 31 January 2018
Accepted 25 February 2018
Available online 27 February 2018

\textbf{Keywords:}
Rough sets theory
Graph theory
k-step R-related set
Strongly connected components

\textbf{Abstract}
Rough sets theory is not good at discovering knowledge from digraphs which is a kind of relational data. In order to solve this problem, we introduce binary relations derived from simple digraphs and propose a new concept of k-step R-related set in the framework of generalized rough sets theory. In addition, we first investigate the relationships between generalized rough sets theory and graph theory on the basis of mutual representation between binary relations and digraphs. The relationships established in this work make it possible to use generalized rough sets theory to find strongly connected components of simple directed graphs, which previously can be solved only by graph algorithms. An algorithm is correspondingly developed based on the above works, especially k-step R-related set. A series of experiments are carried out to test the proposed algorithm. The results show that our algorithm provides comparable performance to the classical Tarjan algorithm. In addition, the proposed algorithm can be implemented in parallel. And its parallel performance is comparable to existing state-of-the-art parallel algorithms.

\textcopyright 2018 Elsevier B.V. All rights reserved.

1. Introduction

Rough sets theory (RST) proposed by Pawlak [1] is a useful mathematical tool to handle uncertain and ambiguous information. RST has a wide range of applications in the fields of data mining [2,3], pattern recognition [4], decision making analysis [5–7], artificial intelligence [8–10], knowledge discovery [11], machine learning [12,13], intelligent data analyzing [14], etc.

Relational data, describing the relationships between objects, plays an increasingly important role in knowledge discovery in recent years [15–17]. It can be regarded as contacts, ties and connections, which relate one object to another and so cannot be reduced to the properties of the individual object themselves [18]. As one of the most widely studied instances of relational data, graph is also an important object of knowledge discovery. A graph may be either directed or undirected according to whether its edges (relationships) between vertices (objects) are directed or undirected. In general, graph theory is widely used to deal with relational data.

Strongly connected components (SCCs) is the problem of decomposing a digraph into a set of disjoint maximal connected components. It is one of crucial problems of graph theory. SCCs is also involved in many other research fields, including ecologival subsystems [19], transitive closure [20], suspicious transaction recognition [21], small world web [22], program analysis [23,24], etc.

Many serial/parallel algorithms based on graph theory have been proposed to find the SCCs of digraphs. Among several serial algorithms [25–30], the most well-established serial algorithm was Tarjan algorithm [26] based on depth-first search (DFS). It can perform in linear time $O(n+m)$, where $n$ is the number of vertices and $m$ the number of edges. However, the application of serial algorithms is limited nowadays because the computation of DFS is known to be P-complete [31]. Then many parallel algorithms were proposed to speed up the computation of SCCs. Early studies were concentrated on NC algorithms [32–36]. Current studies included recursive algorithms [37–42], GPU-based algorithms [43–46] and parallel randomized DFS algorithms [47–53]. The milestone of recursive algorithms was Forward-Backward (FB) algorithm [37] devised by Fleischer et al. It achieves parallelism by partitioning the given digraphs into three disjoint subgraphs, which can be processed independently in a recursive manner. FB algorithm has quadratic time complexity $O(n(n+m))$. Following FB algorithm, Coloring algorithm [38], FB-Trim algorithm [39], Hong’s algorithm [40] and OBFB algorithm [42] were proposed. All of them run in quadratic time. Barnat et al. [44] firstly implemented the FB-Trim, Coloring and OBFB algorithms on GPUs. His pioneering works on the computation of SCCs on GPU drew much attentions [43,45,46]. As a novel approach, parallel randomized DFS algo-
Algorithms show that DFS-based algorithms can be parallelized more directly without sacrificing time complexity \([47–52]\). All of them requires \(O(p(n + m))\) time, where \(p\) is the number of processors. In 2016, Lowe \([53]\) proposed a concurrent DFS algorithm based on Tarjan algorithm. Lowe’s algorithm requires \(O(n^2)\) time.

Although graph theory can deal with relational data, it is viewed as a theory of solving deterministic problem \([54]\). In practical applications, non-deterministic factors appeared in relational data produce uncertainty information \([55]\). For handling uncertain relational data, graph theory is not an effective tool. In \([55]\), Gao and Gao proposed the concept of uncertain graph. For the uncertain digraph in Fig. 1, whether two vertices of it are joined by a directed edge can not be determined. Similarly, whether the uncertain digraph is strongly connected can not be determined via conventional approaches based on graph theory. In contrast to graph theory, RST can approximate an uncertainty problem by means of two deterministic concepts of lower and upper approximation. RST may provide a new way to deal with the SCC problems in digraphs.

We try to solve this problem by taking the two following considerations.

Firstly, as core concepts of generalized RST and graph theory, binary relations and digraphs can be represented mutually. This fact motivates us to believe that there are some relationships between generalized RST and graph theory. In this work, we first investigate the relationships between generalized RST and graph theory. It is found that some basic concepts in the generalized RST and graph theory are equivalent.

Secondly, the relationships between generalized RST and graph theory investigated in this paper and the ability of RST to handle uncertainty information motivate us to find a way that RST can handle relational data no matter uncertain or certain it is. The investigated relationships will provide a reasonable theory basis for handling relational data (especially digraphs) based on generalized RST approach. The reason for choosing generalized RST is that the binary relations derived from simple digraphs are not definitely reflexive, symmetric or transitive. Therefor the primary objective of this research is to improve generalized RST so that it can extract SCCs knowledge from simple digraphs. Consequently, an algorithm based on generalized RST named RSCC is proposed to compute SCCs of simple digraphs. The parallel version of RSCC, PRSCC, is also proposed in this paper because that parallel computing \([56,57]\) and incremental learning \([58,59]\) are useful to improve the efficiency of discovering knowledge in datasets. The proposed algorithms can be applied in fields which require to compute SCCs, including ecological subsystems \([19]\), transitive closure \([20]\), suspicious transaction recognition \([21]\), small world web \([22]\), program analysis \([23,24]\), etc.

![Fig. 1. An uncertain digraph.](image)

About relationships between generalized RST and graph theory, extensive studies have been done. Chiaseletti et al. \([60–62]\) interpreted adjacency matrix of a simple undirected graph as an information table so that RST can handle simple undirected graph. Additionally, the works of Wang et al. are notable \([63–65]\). In \([63]\), they established the relationships between transversal matroids and covering-based RST. And they used graph and matrix approaches to study RST through matroids \([64]\). The works of Chiaseletti et al. and Wang et al. have made outstanding contributions to combining generalized RST with graph theory. However, the studies of the relationships between generalized RST and graph theory from the perspective of mutual representation between binary relations and digraphs are rare \([57]\). Thus, in this work, we will reveal the relationships of some fundamental concepts between graph theory and generalized RST on the basis of mutual representation between binary relations and digraphs.

For handling relational data, classical RST is incapable. Many researchers have paid attentions to extend the application scope of RST to relational data \([15,57,66–75]\). Fan \([15]\) extended the application scope of rough sets analysis from table-style information systems to relational information structure. In \([57]\), Chen et al. introduced a novel concept of \(k\)-step upper approximation, and designed an algorithm based on the concept to find the connected components of simple undirected graphs, a kind of relational data. Although the remarkable work of Chen et al. makes RST can extract knowledge from simple undirected graphs, it still fails to deal with digraphs. And the fact that undirected graphs can be easily converted into digraphs but not vice versa leads one to pay more attention on digraphs than undirected graphs.

This paper is organized as follows. In Section 2, some basic concepts related to generalized RST and graph theory are reviewed. In Section 3, binary relations derived from simple digraphs are introduced and related propositions are investigated. Meanwhile, a new concept named \(k\)-step \(R\)-related set is proposed. And the relationships between generalized RST and graph theory are first explored. In Section 4, the strongly connection of digraphs and SCCs are analyzed by means of generalized RST. In Section 5, according to the above theoretical work, we propose a new algorithm to find the SCCs of simple digraphs. And its implementation in parallel is also provided. Subsequently experiments details and related discussions are also given. Finally, in Section 6 a conclusion is given.

2. Preliminaries

In this section, we briefly review some basic concepts of generalized RST \([76]\) and graph theory \([77]\).

2.1. Generalized rough sets theory

Binary relations play a crucial role in RST. For instance, Pawlak’s RST \([1]\) is based on equivalence relation and dominance-based RST \([78]\) on dominance relation, and so on.

Let \(U\) be a non-empty finite set, \(P(U)\) the power set of \(U\), \(X^c\) the complement of \(X\) in \(U\), and \(U \times U\) the product set of \(U\) and \(U\). Any subset \(R\) of \(U \times U\) is called a binary relation on \(U\). For any \((x, y)\) \(\in U \times U\), if \((x, y) \in R\), we say \(x\) has relation \(R\) with \(y\), and denote it as \(xRy\).

Let \(R\) be a binary relation on \(U\), it is:

- (a) Serial, if for all \(x \in U\), there exists \(y \in U\) such that \(xRy\);
- (b) Reflexive, if for any \(x \in U\), \(xRx\);
- (c) Irreflexive, if for any \(x \in U\), \((x, x) \notin R\);
- (d) Symmetric, if for any \(x, y \in U\), \(xRy\) implies \(yRx\);
- (e) Transitive, if for any \(x, y, z \in U\), \(xRy\) and \(yRz\) imply \(xRz\).

For instance, equivalence relation is reflexive, symmetric and transitive. Dominance relation is reflexive and transitive. In \([79]\), Yao used the pair \((U, R)\) to denote a generalized approximation.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات