



Topic knowledge map and knowledge structure constructions with genetic algorithm, information retrieval, and multi-dimension scaling method



Deng-Yiv Chiu*, Ya-Chen Pan

Department of Information Management, ChungHua University, 707, Sec. 2, Wu-Fu Rd., Hsin-Chu 300, Taiwan, ROC

ARTICLE INFO

Article history:

Received 30 March 2013

Received in revised form 15 March 2014

Accepted 16 March 2014

Available online 26 March 2014

Keywords:

Knowledge structure

Topic knowledge map

Information retrieval

Genetic algorithm

Independent chi-square

Multi-dimension scaling

ABSTRACT

This work presents a novel automated approach to construct topic knowledge maps with knowledge structures, followed by its application to an internationally renowned journal. Knowledge structures are diagrams showing the important components of knowledge in study. Knowledge maps identify the locations of objects and illustrate the relationship among objects. In our study, the important components derived from knowledge structures are used as objects to be spotted in a topic knowledge map. The purpose of our knowledge structures is to find out the major topics serving as subjects of article collections as well as related methods employed in the published papers. The purpose of topic knowledge maps is to transform high-dimensional objects (topic, paper, and cited frequency) into a 2-dimensional space to help understand complicated relatedness among high-dimensional objects, such as the related degree between an article and a topic.

First, we adopt independent chi-square test to examine the independence of topics and apply genetic algorithm to choose topics selection with best fitness value to construct knowledge structures.

Additionally, high-dimensional relationships among objects are transformed into a 2-dimensional space using the multi-dimension scaling method. The optimal transformation coordinate matrix is also determined by using a genetic algorithm to preserve the original relations among objects and construct appropriate topic knowledge maps.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Individuals can search for and retrieve an enormous amount of information and knowledge in a certain domain from search engines or subscribed journal databases. However, the amount of information retrieved is enormous, making it difficult and time-consuming for users to acquire knowledge. Therefore, knowledge management focuses on assisting users to obtain information from enormous journal databases and retrieve implicit knowledge in the databases.

Readers and editors of academic journals are especially concerned with understanding the relationship among knowledge attributes. Reliable information retrieval technologies can enable users to retrieve necessary information efficiently and identify journal-related topics. Moreover, journal editors can also understand trends of research topics and distribution of published

articles. In knowledge management, a hierarchical knowledge structure is an effective means of presenting knowledge. Knowledge structure depicts the important components of knowledge in study. Such a structure can help individuals to thoroughly understand knowledge in a certain domain and reach a decision efficiently. For instance, a knowledge structure has been applied to manage legal knowledge [26]. By considering the importance weight of legal terms and pair-wise correlation of legal terms, all legal terms are clustered based the contents of the documents to construct a hierarchical knowledge structure for legal knowledge.

In addition to a knowledge structure, knowledge map is an effective means of presenting knowledge. Knowledge maps identify the location of objects and present the relationship among objects. In this work, the important components derived from knowledge structures are used as objects to be located in the knowledge map. Devising a knowledge structure allows us to analyze knowledge in a specific domain and, then, present the analysis result on a knowledge map to help understand the knowledge. For instance, some studies have analyzed patent document-related topics of the National Science Council (NSC) of the Republic of

* Corresponding author. Tel.: +1 3518 6531; fax: +1 3518 6543.

E-mail addresses: chiuden@chu.edu.tw (D.-Y. Chiu), d09403019@cc.chu.edu.tw (Y.-C. Pan).

China, Taiwan, to facilitate a presentation of the knowledge of patent documents in a knowledge map [28].

A knowledge map has been viewed from various perspectives in literature emphasizing knowledge representation. From the perspective of knowledge representation, the cognitive map (CM) is a proper means by which the perceptions of humans can be retrieved [20,27,31]. For instance, some studies built a visualized cognitive knowledge map integration system, called *VisCog*, to facilitate knowledge management on peer-to-peer networks [17]. *VisCog* can retain an individual peer's knowledge structure while articulating with those of other peer to build its cognitive knowledge map. Some research employs graph theory and authoritative sources identification techniques with visualization tools to find critical research topics from citation graph. They create the citation graph from data retrieved from the CiteSeer database. In the experiments, they identify important research topics from the citation graph with promising results. They successfully identify the established research topics and new research topics with the proposed method [2]. Some research proposes a novel tool to enhance the inadequacies of existing co-citation visualization tools. They demonstrate the visualized effects of the tool with a case study of a larger data set to explore the cross-field studies among different computer science fields. The tool displays co-citation graphs with latent visual cues and allows direct manipulation of the visualized graphs. The tool also allows the exploration of the relationships between articles in the graphs [3].

How to present an efficient map based on a large amount of data has been extensively studied in knowledge management literature. Based on information retrieval methods, some studies have developed Knowledge Based News Miner (KBNMiner) to acquire important news information from the Internet, along with cognitive maps used to effectively present new information [12]. However, this method requires manually setting related parameters before cognitive maps can be presented. Some other researches focus on the automatic generation of a hierarchical knowledge map NewsMap, based on online Chinese news, particularly the finance and health sections [19]. The hierarchical knowledge map can be used as a tool for browsing business intelligence and medical knowledge hidden in news articles. Some research uses domain analysis and journal mining to obtain an overview of an emerging research field. The approach extends the traditional domain analysis method by adding two additional aspects: objectives and applications. They demonstrate a case study of applying it to explore the knowledge management field [18].

Some research tries to automatically build a domain knowledge map for e-learning using text mining techniques. TF/IDF algorithm is used to extract keywords with a set of documents about a specific topic. They use ranking pairs of keywords according to the number of appearances in a sentence and the number of words in a sentence. The number of relations required to identify the important ideas in the text is analyzed in the experiments [15]. Some research uses a language-neutral method to handle the linguistics difficulties in the text mining. The Self-Organizing Maps (SOM) is employed to generate two maps. They include word cluster map and document cluster map. The maps present the relationships among words and documents. They incorporate these two maps and effectively find the relevant documents according to the keywords specified in the query. The conceptually associated web documents are found by specific keywords and relevant words found by the word cluster map [6]. Some research uses clustering techniques and Apriori association rule to classify textual data. The Apriori association rule is employed to obtain multiple key term phrasal knowledge sequences. The approach can be used to analyze any free formatted textual data. They demonstrate it on an industrial dataset consisting of Post Project Reviews (PPRs). That is collected from the construction industry [22]. To facilitate the

construction of knowledge maps, some studies implemented an incremental two-stage clustering method as a mechanism that could maintain the knowledge maps with the increment of documents [16]. Although incorporating a text-clustering mechanism, this method cannot present the hierarchical relations among clusters. Additionally, each cluster of documents cannot be named automatically. Moreover, some studies devise an artificial neural network model, called Growing Hierarchical Self-Organizing Map (GHSOM). GHSOM comprises a multi-player hierarchical framework, where each layer contains multiple independent SOMs. GHSOM is used to construct knowledge maps of legal documents in the securities and futures market. Furthermore, a topic selection module is designed to enhance the readability of topic names [26]. A topic is a subject representing a collection of related articles. Twenty labels are produced for each cluster by the LabelSOM algorithm. The topic name for each cluster is selected from the top two highest-frequency labels based on the statistics of all article titles mapped to the cluster.

In the literature, some research uses genetic algorithms to select text features for text clustering. They apply term variance to measure groups of terms and apply genetic based algorithm to find the most valuable groups of terms. These terms then would be utilized to generate the final feature vector for the clustering process [25]. Some research focuses on the extraction of drug utilization knowledge. They use a two-stage SOM technique to process concepts in laboratory test and diagnostic data, respectively. Later, rough set theory is applied to extract the drug use knowledge [5]. Some research proposes a data clustering method using bacterial foraging optimization. They use a group of bacteria forage to converge to certain positions as final cluster centers by minimizing the fitness function. The algorithm could be used to handle data sets with various cluster sizes, densities, and multiple dimensions [29]. Some research uses vector space model to present the published papers and feature and adopts chi-square test to examine the independence of topics. Then they applied genetic algorithm to facilitate automatic topic selections used to construct the knowledge structures for the journal in study [4]. Some researches apply various approaches to process and/or visualize research trend and knowledge management [30,8].

In this study, we propose a method to first construct knowledge structures to explore the major topics as well as the related methods of journal articles and help identify the research tendencies among these published papers. The major topics and related methods are employed as object plotted in a 2-dimensional topic knowledge maps to mainly explore the relationship among selected topics, papers, and cited frequency to effectively help understand the complicated relationships among high-dimensional objects in a 2-dimensional space.

2. Related methods

Here, we introduce some of methods used in the study, including term weighting, vector space model, similarity computation, multi-dimension scaling, and genetic algorithm.

2.1. Term weighting, vector space model and similarity computation

The term weighting approach is adopted to identify important terms in documents, along with a vector space model to represent documents, as well as a similarity computation formula used to estimate the relations among objects.

Term-weighting is to find out the importance of each term in the document. The *tf-idf* approach is often used for term weighting that considers the term frequency (*tf*) and inverse document frequency (*idf*). It assumes that terms with higher frequency in a

Download English Version:

<https://daneshyari.com/en/article/405104>

Download Persian Version:

<https://daneshyari.com/article/405104>

[Daneshyari.com](https://daneshyari.com)