Information Processing and Management 000 (2016) 1–13

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target

Weidong Zhao*, Ran Wu, Haitao Liu

Shanghai Key Laboratory of Data Science, School of Software, Fudan University, 200433 Shanghai, China

ARTICLE INFO

Article history:
Received 27 November 2015
Revised 16 March 2016
Accepted 13 April 2016
Available online xxx

Keywords:
Paper recommendation
Knowledge gap
Concept map
Shortest concept paths

ABSTRACT

The massively growing documents make it a challenge for researchers to find high value papers. To solve information explosion, some work on personalized paper recommendation have been proposed. However, the knowledge gap between a researcher's background knowledge and research target is seldom concerned. In this paper, we propose a new method of recommending helpful papers to support researchers by bridging the knowledge gap. First, domain knowledge is extracted as the concept map, which provides a basis of comparing user background knowledge and target knowledge. Then, the knowledge gap is defined with the concept map. To bridge the knowledge gap, the shortest concept paths are searched to explore some suitable knowledge paths, which can help researchers to acquire target knowledge in accordance with their cognition patterns. Finally, experiments are performed to demonstrate the effectiveness of the recommendation method.

1. Introduction

With the development of information technology, great achievements have been made in terms of electronic literatures. Meanwhile, the increasing number of research papers makes it a challenge for researchers to discover helpful knowledge resources. This is commonly called the information overload problem (Drachsler, Hummel, & Koper, 2008; Salehi & Kamalabadi, 2013). To deal with this issue, some information techniques (e.g. information retrieval and information filtering) are adopted by E-libraries and scientific databases to assist knowledge workers. Research paper recommendation needs the technique that suggests helpful papers to researchers via exploring their interests and preferences (Basu, Hirsh, Cohen, & Nevill-Manning, 2001). By actively providing interesting materials, literature recommendation can save much time and efforts for researchers (Pan & Li, 2010).

Content-based filtering and collaborative filtering are the mostly used recommendation techniques in many contexts, including E-commerce, travel, film and music websites, etc. To recommend papers, content-based filtering builds the user profile based on reading history and suggests new papers that well match the profile. In previous work, user profiling is generally built in consideration of the importance of key words; however, it is insufficient to model the user's preference. To refine the preference semantics, several methods have been proposed, such as label-enriched approach (Guan et al., 2010), and ontology-expansion approach (Zhang, Ni, Zhao, Liu, & Yang, 2014). As for the collaborative filtering approach, like-minded research groups are explored first and recommendations are generated based on their similar interests. A key issue in collaborative filtering is how to measure user similarity. To analyze the relationship between researchers, data from

* Corresponding author.
E-mail addresses: wdzhao@fudan.edu.cn (W. Zhao), xiaohanrrr@163.com (R. Wu), 13212010011@fudan.edu.cn (H. Liu).

http://dx.doi.org/10.1016/j.ipm.2016.04.004
0306-4573/

Please cite this article as: W. Zhao et al., Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.04.004

various resources is collected and analyzed, including e-mail logs, co-authors, references, project cooperation and social media (Davoodi, Afsharchi, & Kianmehr, 2012; Durand, Belacel, & Laplante, 2013).

Distinguished from common information resources (e.g. news and tweets), academic literature is an important knowledge resource for researchers. Scientific workers gain their knowledge by reading published literatures. Before navigating an academic database, a researcher usually has one or several focused knowledge goals, which are commonly documented as research proposals, requirement specification or planning statement. The disparity between the knowledge goal and background knowledge is called "knowledge gap". To bridge the gap, researchers go through literature databases and explore papers they need. By learning and digesting these papers, they transform the embedded knowledge into their own. However, the knowledge gap issue is seldom concerned in previous researches. Since a user's historical preference cannot necessarily reveal his/her current knowledge requirements, previous methods can hardly narrow the new knowledge gap. As a result, a large collection of written works still challenge researchers even though only a few well-documented knowledge goals are available. Therefore, continuous efforts are required to support knowledge workers by bridging their knowledge gap.

In this paper, we present an approach called the knowledge gap based recommendation (KGR, for short) to bridge the knowledge gap between a researcher's background knowledge and research target. The method presents domain knowledge in the form of the concept map. First, a set of central concepts are extracted from domain corpus. Then, the strategy builds the links between these concepts according to their associations. A researcher's reading records are analyzed to model his/her background knowledge; and the target knowledge is extracted from the research proposal. In the domain of concept map, the knowledge gap is defined as the shortest paths that connect these two kinds of knowledge. Finally, the concepts in those paths are utilized to discover well-matched papers, which can help bridge the user's knowledge gap.

The paper is structured as follows. Section 2 summarizes recent work related to paper recommendation. In Section 3, a methodology to build the domain concept map is introduced. Section 4 describes how to define the knowledge gap, and proposes the recommendation algorithm based on the gap. Section 5 gives a brief case study and Section 6 evaluates the proposed method. Section 7 draws conclusions.

2. Related work

Many researchers realize that user preference cannot be the only guidance when recommending resources. Therefore, they consider other factors, such as domain knowledge, user background, learning targets and cognitive patterns. The context of literature reading is not well-formed, making it difficult to determine the user's learning targets and cognitive patterns. As a result, previous researches only focus on standardized teaching situations such as e-learning. Zhang et al. (2014) organized disciplines and curriculum information as a knowledge tree, from which association rules between resources and courses were analyzed to recommend teaching resources. Tang and Mccalla, (2004) proposed a paper recommendation method concentrating on teaching characteristics, including the user's knowledge level and knowledge goals. Based on these characteristics, a set of ordered papers are recommended. This paper is devoted to bridging the knowledge gap between the user's background knowledge and research targets.

In some studies, domain knowledge was modeled in the form of domain taxonomy, ontology or concept networks. For example, Liang et al. established a semantic network according to visited documents. The semantic network is composed of several connected semantic trees, and connections in each semantic tree reflect the inheritance relationship between concepts. Then spreading activation was used to semantically expand the user's interest (Liang, Yang, Chen, & Ku, 2008). Spreading activation can achieve beneficial knowledge expansion, enriching original knowledge model with closely related snots. Xu et al. (2012) combined concept networks with social networks, and recommended experts to users with comprehensive utilization of semantic relations between concepts, social cooperation between experts, and professional relationship between concepts and experts. Cantador and Castells established a semantic network composed of domain concepts, and user interests took the form of concepts and user interest levels. The algorithm expanded user interests by spreading activation, and then clustered concepts in semantic networks according to user preferences. Users with similar concept clusters are considered to have similar interests (Cantador & Castells, 2006).

Some approaches extracted and utilized the user's background to build the personalized user profile, which reflected unique background and requirements of each user. Chen et al. built an adaptive ontology for each user according to the user's reading behaviors. User-concept, user-user patterns were extracted from the ontology, and resources were recommended to the user according to similar patterns in the pattern library (Gemmis, Lops, Semeraro, & Musto, 2015). Hawalah and Fasli set up a personalized interest ontology based on user interests and views. Spreading activation was applied to expand user interests, aiming to find relevant concepts the user might be interested in (Hawalah & Fasli, 2011). Some recommendation methods planned learning paths for the user on the basis of dependency between resources, so as to help the user achieve a certain goal. Yu et al. created three ontologies representing the learner's background knowledge, learning resources and domain knowledge respectively. Similar learning resources were found after computing the similarity between users' background and learning resources, then these resources were organized as a learning path in accordance with their prior links (Hawalah & Fasli, 2011). Durand et al. asserted that the learner's knowledge base needed to satisfy the prerequisite when he/she learns resources. Only in this way could the learner gain knowledge from them. On that basis, the strategy (Durand et al., 2013) established a directed graph according to competencies required, and recommended a sequence of learning objects in a well-defined order to the learner. Ordered learning paths could help users to reach the goal in terms of their abilities, so that users can follow some learning paths from the initial set of competencies to the target one.

Please cite this article as: W. Zhao et al., Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.04.004