Information Processing and Management 51 (2015) 329–339

Contents lists available at ScienceDirect

Information Processing and Management

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

Research patterns and trends of Recommendation System in China using co-word analysis

Jiming Hu [a,*], Yin Zhang [b]

[a] Center for Studies of Information Resources, Wuhan University, Wuhan, China
[b] School of Library and Information Science, Kent State University, Kent, OH, USA

ARTICLE INFO

Article history:
Received 15 March 2014
Received in revised form 21 January 2015
Accepted 9 February 2015

Keywords:
Recommendation System in China
Co-word analysis
Research patterns
Research trends

ABSTRACT

This paper examines the research patterns and trends of Recommendation System (RecSys) in China during the period of 2004–2013. Data (keywords in articles) was collected from the China Academic Journal Network Publishing Database (CAJD) and the China Science Periodical Database (CSPD). A co-word analysis was conducted to measure correlation among the extracted keywords. The cluster analysis and social network analysis revealed 12 theme-clusters, network characteristics (centrality and density) of the clusters, the strategic diagram, and the correlation network. The study results show that there are several important themes with a high correlation in Chinese RecSys research, which is considered to be relatively focused, mature, and well-developed overall. Some research themes have developed on a considerable scale, while others remain isolated and undeveloped. This study also identified a few emerging themes with great potential for development. It was also determined that studies overall on the applications of RecSys are increasing.

1. Introduction

Over the past decade, research on RecSys has progressed in China (Hu, Hu, & Deng, 2013) and RecSys has become a hot topic of study in Computer Science and Information Science (Qin & Liang, 2011; Wang & Liu, 2012; Zhang, Chen, & Fu, 2013). Researchers have also successfully applied RecSys theories, algorithms, technologies, and strategies to other areas (e.g., Hu, Hu, Gao, & Hao, 2013; Li, Xu, & Xu, 2013; Meng, Hu, Wang, & Zhang, 2013; Yin, 2013). For example, Hu et al. (2013) calculated the similarity among friends and tags, and used the results for friend recommendation.

Given the importance of RecSys in China, researchers have reviewed related literature to identify research advances and development within this field. However, much of the effort has only considered particular subfields of RecSys (e.g., Feng & Liang, 2011; Wang & Liu, 2012; Yang & Zhao, 2011; Yi & Deng, 2011; Zhang et al., 2013), and the conclusions were drawn from purely descriptive and qualitative analysis. Yang and Zhao (2011) outlined the recommendation algorithms in recent years, pointing out that the current recommendation algorithms mainly include collaborative filtering, content-based algorithms, graph-based algorithms, and hybrid algorithms; and they further compared the pros and cons of recommendation algorithms. Similarly, after a qualitative study on the failures of the recommendation algorithms, Feng and Liang (2011) concluded that there are two main aspects of RecSys research: alleviating the problem of sparsity (clustering, and dimension

* Corresponding author.
E-mail address: whuhujiming@ymail.com (J. Hu).

http://dx.doi.org/10.1016/j.ipm.2015.02.002
0306-4573/

reduction), and establishing a simulation model (association rules, Bayesian model, neural network, and cloud model). They further pointed out that model-based recommendation algorithms would become a mainstream trend with other recommendation technologies as supplement. Wang and Liu (2012) analyzed three important modules in Recommendation Systems: user modeling, the recommendation object modeling, and the recommendation algorithms, and concluded that future research on RecSys would concentrate on feature modeling (user and object), security, applications of complex network theory, sparsity, cold start, and performance evaluation. After a review of relevant literature on RecSys recommendation techniques, research methods and experimental datasets, Zhang et al. (2013) provided further insights on RecSys and pointed out that future research trends would include context-based recommendation, trust-based recommendation, and social tagging-based recommendation.

The review of the literature on RecSys research in China shows that most studies used qualitative methods, and were conducted according to the personal judgment of a small number of experts. These researchers identified certain key research subfields or branches of RecSys in China; some of which have matured. In addition, several research directions, such as graph-based recommendations and trust-based recommendations have emerged and received more attention from researchers. Lastly, researchers are increasingly beginning to use recommendation methods from other fields, such as book recommendations, tag recommendation, e-commerce recommendation, and friend recommendation.

A review of previous related literature also shows some research gaps and limitations. At this point, little is known about the distribution and structure of RecSys research themes in China. Additionally, previous research findings may be considered limited due to their lack of precise empirical characterization of conceptual contours and the trends over time. There is a need to analyze RecSys in China from a quantitative, empirical, and longitudinal perspective. The major goal of this paper is to address these gaps and limitations by providing a comprehensive exploration and analysis of research advances within RecSys in China over the past 10 years. This study will supplement the previous qualitative work by concentrating on empirical analysis of research structures, correlations among research themes, and research trends.

Co-word analysis has proven to be an effective approach to help identify and illustrate the correlation between research themes (Lee, 2008; Leydesdorff & Welbers, 2011; Stacey, 2013). This study applied this method for shedding light on the entire research structure, the correlation among themes, and the overall evolutionary trends within RecSys studies in China.

2. Methodology

For this study, the method of co-word analysis, with clustering and social network analysis, was chosen to obtain more intuitive and comprehensive results. There were three important stages in this study:

(1) Identifying the research hotspots within the field of RecSys in China based on keyword frequency and associated co-word data.
(2) Conducting cluster analysis to identify research themes using co-word data (a cluster represents a large research theme); and
(3) Using co-word network analysis to capture the current status and trends of research themes of RecSys in China.

2.1. Co-word analysis

The method of co-word analysis was derived from co-citation analysis (Small, 1973; Small & Griffith, 1974). In the 1980s, the bibliometric system "LEXIMAPPE" was developed to apply co-word analysis in mapping the co-word structure; and co-word analysis was first introduced as a method in the book "Mapping of the Dynamics of Science and Technology" (Callon, Law, & Rip, 1986). Co-word analysis uses sets of shared keywords instead of shared citations (Callon, Courtial, & Laville, 1991; Courtial, 1994; Courtial & Law, 1989; Law, Bauin, Courtial, & Whittaker, 1988; Law & Whittaker, 1992), and can directly deal with textual data through the analysis of co-occurrence frequency of paired words. Researchers therefore agree that co-word analysis is able to help identify the relationships among themes (aggregated clusters of words) in a research field, and consequently trace the development trends (Coulter, Monarch, & Konda, 1998; Lee & Jeong, 2008). Researchers further developed new methods and tools of co-word analysis, such as using co-word clustering (Callon et al., 1991), multidimensional scaling (Peters & Vanraan, 1993), social network analysis (Ding, Chowdhury, & Foo, 2001), and the strategic diagram (Stegmann & Grohmann, 2003). In particular, the strategic diagram takes into account both centrality and density, and thus can also describe the dynamics of research themes.

Co-word analysis, as a well-established and effective approach, can reveal the intellectual structure of a research field (Ronda-Pupo & Guerras-Martin, 2012). It assumes that a group of aggregated keywords could reveal underlying themes, and that co-occurrences of keywords could reflect the associations among the underlying themes. Researchers have successfully used co-word analysis to discern research patterns and trends of specific research fields, including acidification research (Law & Whittaker, 1992), European social psychology (Vala, Lima, & Caetano, 1996), plant proteins (DeLooze & Lemarie, 1997), software engineering (Coulter et al., 1998), information retrieval (Ding et al., 2001), robot technology (Lee & Jeong, 2008), stem cells (An & Wu, 2011), digital library (Liu, Hu, & Wang, 2012), library and information science (Hu, Hu, Deng, & Liu, 2013), and renewable energy (Romo-Fernandez, Guerrero-Bote, & Moya-Anegon, 2013).