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).

2.2. Data collection and pre-processing

In this study, keywords from journal articles were used as the source of data. There are two major full-text journal article databases in China: the China Academic Journal Network Publishing Database (CAJD) developed by the National Knowledge Infrastructure, and the China Science Periodical Database (CSPD) developed by Wanfang Data Corporation Limited. These two databases index almost all Chinese journal articles that represent the current status of scientific research on RecSys in China. In addition, CSPD also includes a foreign literature database in partnership with National Science and Technology Library (NSTL), covering 93% of academic journals collected by Social Citation Index (SCI) and Engineering Index (EI) (Wanfang Data Corporation, 2015). This foreign literature database includes the majority of articles written by Chinese researchers. From the databases, journal articles with a focus on RecSys research in China were gathered by searching keywords "recommendation" in both Chinese and English languages.

These two databases classified journals according to discipline, including the disciplines related to RecSys research such as Computer Software and Application of Computer, Trade Economy, Macro-economic Management and Sustainable Development, Internet Technology, Telecom Technology, Enterprise Economy, and Library Service and Digital Library. The retrieved articles were filtered according to the corresponding disciplines.

The next step was to download and import the bibliographic records of the articles into the Statistical Analysis Tool for Informetrics (SATI) developed by Liu and Ye (2012) for pre-processing. SATI automatically extracted the keywords within the articles and computed their word frequencies. After normalization of keywords, certain keywords were preselected and produced a co-occurrence matrix using SATI. In this matrix, values in the diagonal cells were treated as missing data, and values of non-diagonal cells were co-word frequencies. Next, the co-word matrix was transformed into a Pearson's correlation coefficient matrix with each cell denoting the similarity between the row element (a keyword) and the column element. The co-word correlation matrix was used for the subsequent data analysis.

2.3. Methods for data analysis

This study used cluster and social network analyses for revealing the scientific structure of a specific field effectively (Leydesdorff & Welbers, 2011). In cluster analysis, the hierarchical clustering was adopted using Ward's method as the cluster method, and with Squared Euclidean distance as the distance measure. Thus in general, keywords with high correlation between each other are more likely to be put into the same cluster. Different cutoff steps may be set up in hierarchical clustering to obtain different clustering results, which could provide more interpretations of the correlation between keywords or themes. SPSS 19.0 was used for clustering and obtaining a set of keyword-clusters. Each cluster denotes a possible research theme.

The word co-occurrence matrix may be essentially considered a network, and social network analysis is considered a well-established and effective approach. In this study, several measures of social network analysis were used, such as network centrality, density, core-periphery structure, and strategic diagram. These measures defined below were processed using Ucinet6.0:

- *Centrality* measures the correlation of one sub-network or keyword with the others. Keywords or themes with high centrality could be defined as lying on a central and important position in the cluster, or in the entire research field. Therefore, this value is often used to measure the importance of a node, or a sub-network of a research field (Cambrosio, Limoges, Courtial, & Laville, 1993; Lee, 2008).
- *Density* indicates the strength of correlation between one keyword and another, within its corresponding research themes in a network. Network density reflects the internal coherence or cohesiveness among the nodes in a network. Thus, density of a given cluster can be used as a measure of the capability of a theme that can maintain and develop by itself (Coulter et al., 1998). Specifically, it is reasonable to assume that the higher the centrality, the more centrally a research theme is situated within the whole field; and the higher the density, the more mature or potential a research theme (Hu, Hu, Deng et al., 2013).
- *Core-periphery structure* is made up of some elements (keywords) which are connected with each other. More specifically, the keywords in the core are closely linked, and are important in the research field; while on the contrary, keywords in the periphery are sparsely scattered and are of less importance.
- Furthermore, considering the density and centrality, a *strategic diagram* is mainly used to describe the internal relations within a cluster, as well as the interactions among different fields. In general, the strategic diagram uses a two-dimensional space to plot clusters according to their centrality and density. In a strategic diagram, the *x*-axis stands for centrality, while the *y*-axis stands for density, and the axes' origin is determined by the average or median of centrality and density. Thus, because of different centrality and density, the theme-clusters are located in four quadrants, which could indicate the developing status of research themes. In quadrant I, with high centrality and density, research themes are more mature and central in the overall research field. In quadrant II, research themes are not central but are well-developed. In quadrant III, research themes are peripheral and underdeveloped. And in quadrant IV, with low density and centrality, research themes are central in the field but are undeveloped or immature, or may possibly tend toward maturity (Callon et al., 1991; Munoz-Leiva, Viedma-del-Jesús, Sánchez-Fernández, & López-Herrera, 2012).

3. Results and discussion

3.1. Descriptive analysis of word frequency

The keywords for an article can express its main content, and the keyword frequency of academic journal articles can measure the importance of its related themes in a specific field. In this study, records of 2861 articles were retrieved, and 314 of them were manually excluded or removed because they did not represent regular articles, but notifications, news reports, etc. As a result, records of 2547 articles were selected for subsequent analysis. With the aid of SATI, 10,768 keywords were extracted from these article records (4.22 keywords per article). The keyword frequency distribution displays a power-law distribution with an exponent −1.07, as shown in Fig. 1.

The power-law distribution indicates that the research structure in the field of RecSys in China is unevenly distributed. Specifically, most of the keywords were used only occasionally in the articles while a small number of keywords were used very frequently. This shows that only a few themes were the main focus of RecSys field in China. This is similar to other research fields in China, for example, the Digital Library (Liu et al., 2012).

It was noted that not all the keywords provided by authors were normalized; thus, the extracted keywords were normalized using the Chinese Classified Thesaurus and the English translations. The normalization of these keywords was carried out to ensure consistent treatment of the singular and plural forms of words, unifying synonyms, and clarity of homonyms. For example, *User Preference* was replaced by *User Interest*; *Multi-Agent* was replaced by *Agent*; *Weighted Association Rules* was replaced by *Association Rules*; *College Library* was replaced by *University Library*; *Data Sparsity Problem* was replaced by *Sparsity*; etc. Keywords with a frequency below 10 were merged into broader terms. General terms, such as *Model*, *Application*, *Study*, *Theory*, *Web*, and *Status* were removed. The entire process of normalization helped to determine the final keywords for the analysis, with the aid of two professors specializing in this field, and discussion within the research team. The normalization yielded 150 related keywords with an individual frequency greater than 10, which were used for this study.

Three common keywords were noted: *Recommendation*, *Collaborative Filtering* and *Recommendation System*. These were defined as the domain stop-words because they form a high percentage in the analysis domain. In other words, *Recommendation*, as a research object in this analysis, is meaningless. *Collaborative Filtering* and *Recommendation System* both have frequencies (788 and 590) which are too high, and are probably abnormal values. It was found that these three stop-words could influence the accuracy of clustering and social network analysis, and were therefore excluded from the selected keywords.

Finally, 150 related keywords with a total frequency of 5,963 (about 55% of the frequency of the entire keywords) were determined to represent the main contents of RecSys research in China. Table 1 shows the top 50 keywords.

According to the word frequency, co-word frequency, and co-word correlation coefficient, the overall attributes of keywords or themes could be made more explicit.

First, the top ten keywords with high co-word frequency and co-word correlation coefficient are identified and shown in Table 2. Each of the top 10 keyword lists could be considered as a focus of RecSys field in China to a varying degree. In particular, *E-Commerce*, *Clustering*, and *User Interest*, occurring in the top total co-word frequency list and the top total co-word correlation coefficient list, could be considered as the representative focuses of RecSys subfield in China. These keywords have the most direct connections to other themes. Likewise, the themes of *Data Mining* also play an important role in RecSys, and are more likely the basis of recommendation study.

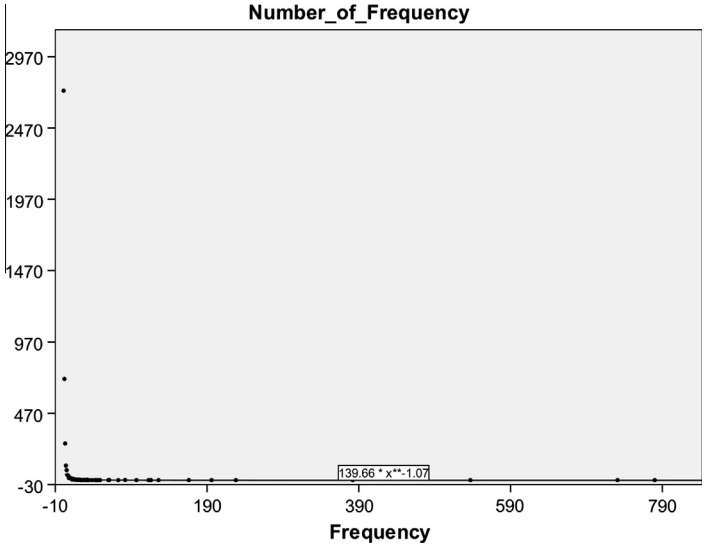

Fig. 1. Distribution of number of keyword frequency of RecSys field in China.

Table 1
The top 50 keywords of RecSys field in China.

No.	Keyword	Frequency	No.	Keyword	Frequency
1	Personalized Recommendation	520	26	Trust Model	54
2	E-Commerce	353	27	Intelligent Recommendation	53
3	Association Rules	218	28	MAE	51
4	Data Mining	205	29	Hybrid Recommendation	49
5	Clustering	159	30	EC Recommendation System	48
6	Similarity	150	31	Item Similarity	47
7	Personalized Recommender System	141	32	user Model	45
8	User Interest	137	33	Personalized Information Recommendation	45
9	Recommendation Algorithm	119	34	Social Networking	44
10	Sparsity	93	35	Interest Measure	43
11	Digital Library	80	36	Social Tagging	43
12	Trust	75	37	Bayesian Network	42
13	User Interest Model	75	38	Similarity Computation	42
14	Web Mining	72	39	Rating Prediction	40
15	Context	67	40	User Behavior	38
16	Agent	64	41	Trustworthiness	36
17	Personalization	59	42	Information Retrieval	36
18	Web Usage Mining	58	43	Search Engine	35
19	Ontology	58	44	Information Recommendation	34
20	Web Log Mining	57	45	Web Service	34
21	P2P Network	57	46	Content Filtering	33
22	Tag	56	47	Social Networks	33
23	Personalized Service	56	48	Cold Start	33
24	Book Recommendation	55	49	Matrix Factorization	33
25	Nearest Neighbor	55	50	Collaborative Recommendation	33

Table 2
The top 10 keywords with high co-word data of RecSys field in China.

No.	Keyword	Total co-word frequency	No.	Keyword	Total co-word correlation coefficient
1	Personalized Recommendation	915	1	Clustering	65.93
2	E-Commerce	598	2	Web Mining	63.71
3	Association Rules	473	3	User Interest Model	61.69
4	Data Mining	410	4	Web Usage Mining	60.11
5	Clustering	314	5	User Interest	60.05
6	Personalized Recommender System	270	6	interest Measure	59.28
7	Similarity	269	7	User Clustering	58.86
8	User Interest	264	8	E-Commerce	58.69
9	Recommendation Algorithm	232	9	User Behavior	58.52
10	Sparsity	177	10	Web Data Mining	56.93

3.2. Cluster analysis

Using the hierarchical clustering method, with the clustering step being set up at one level, 12 clusters (named Cluster 1 to Cluster 12) were obtained. Each of these clusters could be regarded as a research theme. The 150 keywords from the normalization process were divided into 12 theme-clusters. If keywords are grouped into a cluster, they are more likely to have an identical research theme. In this study, five to ten keywords with top frequencies and co-word data (co-word frequency and correlation coefficient) were selected to represent each theme-cluster, because these keywords were most likely to be chosen and used by researchers in each cluster. Table 3 shows the representative keywords within each theme-cluster.

Overall, there were 12 research clusters or themes in the field of RecSys in China. In consultation with the experts in the RecSys field, we believe that this clustering result (12 theme-clusters) is a good representation of the development status of RecSys research in China. The result is in line with previous research. For example, Cluster 1 is related to the research of *Apriori Algorithm* in association rules mining, such as *Frequency Pattern*, *Support Degree*, and so on, as supported in a study by Li and Liu (2012). Cluster 2 includes the applications of recommendation, such as *Personalized Recommendation*, *Book Recommendation*, *Library Service* and *Music Recommendation*, and so on, which was also supported in the findings of a study by Li et al. (2013).

In this study, co-word data includes the total frequency, total co-word frequency, total co-word correlation coefficient, and their average values (shown in Table 4), which could more explicitly demonstrate the characteristics of each research theme-cluster. The comprehensive and proportional average values of co-word data can be designated as very specific indicators to distinguish each research theme.

In order to obtain an accurate and explicit comparison among these 12 theme-clusters, rankings in their respective average frequency, average co-word frequency, and average co-word correlation coefficient are listed in Table 5. The rankings

Table 3
12 clusters of RecSys field in China.

Cluster	Number of keywords	Representative keywords
1	10	Personalized Service; Apriori Algorithm; Frequent Pattern; Support Degree; Decision Tree
2	7	Book Recommendation; Personalized Information Recommendation; Library; Machine Learning; Music Recommendation
3	7	Personalized Recommendation; Personalized Recommender System; Personalization; Intelligent Recommendation; Commodity Recommendation
4	16	Recommendation Algorithm; EC Recommendation System; Fuzzy Clustering; Domain Ontology; Self-Adaptive; Ant Colony Algorithm; Path Analysis; Decision Support; Distributed Computing
5	19	E-Commerce; Association Rules; Data Mining; Clustering; Digital Library; Web Mining; Web Usage Mining; Web Log Mining; Interest Measure; User Behavior
6	9	Similarity; Agent; Hybrid Recommendation; Content Filtering; VSM; Fuzzy Logic
7	8	Sparsity; Nearest Neighbor; MAE; Item Similarity; Rating Prediction; Matrix Factorization
8	13	Ontology; Tag; User Model; Social Networking; Information Retrieval; Random Walk; Markov Model; Complex Network
9	20	User Interest; User Interest Model; Context; Bayesian Network; Similarity Computation; Social Networks; User Clustering; Bipartite Graph; *K*-means; e-Learning
10	17	Information Recommendation; Cold Start; Collaborative Recommendation; Collaborative Filtering Recommendation; Topic Model; User Similarity; User Group; Content Recommendation; Semantic Similarity
11	7	Trust; P2P Network; Trust Model; Trustworthiness; Recommendation Trust
12	17	Web Service; Tag Recommendation; Service Recommendation; Semantic Web; QoS; Microblog; Cloud Computing; Query Recommendation; Knowledge Recommendation; Recommendation Strategy

Table 4
The frequency and co-word data of each theme-cluster.

Cluster	Total frequency	Total co-word frequency	Total co-word correlation coefficient	Average frequency	Average co-word frequency	Average co-word correlation coefficient
1	217	495	370.40	21.70	49.50	37.04
2	180	286	155.99	25.71	40.86	22.28
3	839	1482	270.06	119.86	211.71	38.58
4	450	842	624.53	28.13	52.63	39.03
5	1475	2911	1031.10	77.63	153.21	54.27
6	380	707	456.56	42.22	78.56	50.73
7	353	665	325.67	44.13	83.13	40.71
8	410	780	499.69	31.54	60.00	38.44
9	691	1328	998.08	34.55	66.40	49.90
10	336	648	438.31	19.76	38.12	25.78
11	296	489	118.33	42.29	69.86	16.90
12	336	571	173.58	19.76	33.59	10.21

Table 5
The rankings of 12 clusters according to their average data.

Cluster	Ranking of average frequency	Ranking of average co-word frequency	Ranking of average co-word correlation coefficient	Combined ranking (sorted)
5	2	2	1	5
3	1	1	6	8
7	3	3	4	10
6	5	4	2	11
9	6	6	3	15
11	4	5	11	20
4	8	8	5	21
8	7	7	7	21
1	10	9	8	27
2	9	10	10	29
10	11	11	9	31
12	12	12	12	36

show the relative position and status of related themes in the field of RecSys. Considering these overall indicators, the theme clusters were divided into two levels: high-level clusters with a combined ranking $\leqslant$15 (Clusters 3, 5, 6, 7, and 9) and low-level clusters with a combined ranking >15 (Clusters 1, 2, 4, 8, 10, 11, and 12).

When combining results in Table 3 with Table 4, it was found that several larger research themes or sub-directions exist within the field of RecSys in China during the period under study, such as Cluster 4 relating to *Recommendation Algorithm*, Cluster 5 relating to *Data Mining*, and Cluster 9 relating to *User Interest*. On the other hand, Cluster 2 and Cluster 11, which are

related to *Book Recommendation* and *Recommendation Trust* respectively, and have relatively low indicators, were not found to be the major focus of RecSys.

With high-level indicators, Clusters 3, 5, 6, 7, and 9 appeared to have received more attention from researchers, and could be considered a focus of RecSys field in China. It was found that Cluster 5 had high-level co-word data, indicating that this theme played a central role in the research field. Cluster 3 had the highest average frequency (ranking first) but a relatively low average co-word coefficient, when compared with its frequency (ranked sixth). This case indicates that this theme received the most attention and had moderate association with other themes. Clusters 6 and 9 with top-ranking of co-word data could also be considered bridges that have correlated research subfields.

The low-level indicators of Clusters 1, 2, 4, 8, 10, 11, and 12 suggest that these clusters were not a research focus and had few associations with other themes. Clusters 4 and 8, related to *Recommendation Algorithm* and *Ontology* respectively, included a large number of keywords and had a high co-word correlation coefficient (39.03 and 38.44), suggesting a high association with other themes. Cluster 11 was found to be in the middle among these clusters, and could be seen as a well-developed but independent research subfield of RecSys in China, because of its higher-than-average frequency and lower co-word correlation coefficient. Meanwhile, Clusters 10 and 12 had the lowest-level indicators, and themes in these clusters were unlikely to be a research focus. They might be isolated research themes in RecSys. This is especially true of Cluster 12, which is a larger subfield including relatively more keywords. However, it is the most isolated subfield in the study because themes within Cluster 12 are dispersive on the whole. Therefore, it may be concluded that themes in Cluster 12 are peripheral themes in China at present.

In order to further examine the correlations between these 12 clusters, we set the clustering step at three and obtained a new clustering result. These 12 theme-clusters at clustering step 1 were then merged into six larger clusters (named from Cluster 1′ to Cluster 6′ in Table 6).

According to the 3-step hierarchical clustering, theme-clusters could be merged into a new cluster if they are directly and highly correlated to each other. This means clusters in Table 4, which were merged into a cluster in Table 6, have a high correlation between research subjects. Thus, it may be concluded that there were six large research directions within the RecSys field in China, including *Recommendation Algorithms and Techniques*, *Recommendation Performance and Efficiency*, *Similarity Computing*, *Sparsity and Cold Start*, *User Model*, and *Recommendation based on Network and Trust*.

Clusters 3′ and 5′ are two large subfields of RecSys in China according to the indicators of frequency and co-word data. First, Cluster 3′ is the same as Cluster 5 in one-step clustering. Themes in Cluster 3′ formed a powerful research direction with high internal correlation. Second, Cluster 5′ (including Clusters 8 and 9 in one-step clustering) represents the other large research direction of RecSys; namely, Chinese researchers are increasingly turning their attention to the recommendation based on *Social Networks* by using *Complex Network Theory*. In particular, at present, there would be two significant research subjects: namely, *User Interest* in combination with *Social Networking* (e.g., *Social Networks*, *Social Tagging*, and *User Clustering*); and *Improving Recommendation Algorithm* in combination with *Complex Network* (e.g., *Bayesian Network*, *Bipartite Graph*, and *Community Detection*). Finally, the relatively low-level co-word data of Cluster 5′ also indicates that its association with other themes would be increasing in the future.

3.3. Co-word network analysis

In this study, network analysis was used to reveal the overall correlation structure and the dynamic characteristics of theme-clusters. With this approach, each keyword is considered as a node, and it is connected to another by the co-word correlation. Likewise, co-word correlation is the link between nodes.

Using this rationale, network analysis was conducted on the transformed co-word correlation matrix using Ucinet6.0 to display the current status and trends of research themes of RecSys in China through the measures of network centrality, density, core-periphery matrix, and strategic diagram. In addition, the overall structure and relationship of keywords may be visualized using a proportional correlation network generated by NetDraw embedded in Ucinet6.0.

Table 7 lists the top 10 keywords displaying a high degree centrality and betweenness centrality. This helped identify the position of each keyword in the overall research structure by degree centrality, and also expressed the role played in the association between themes by betweenness centrality.

Table 6
The statistic data of each theme-cluster according to the clustering step at 3.

Cluster in 3-step clustering	Cluster in 1-step clustering	Number of keywords	Total frequency	Total co-word frequency	Total co-word correlation coefficient	Average frequency	Average co-word frequency	Average co-word correlation coefficient
1′	1, 2	17	397	781	526.40	23.71	45.18	29.66
2′	3, 4	23	1289	2324	894.59	73.99	132.17	38.81
3′	5	19	1475	2911	1031.10	77.63	153.21	54.27
4′	6, 7	17	733	1372	782.23	43.17	80.84	45.72
5′	8, 9	33	1101	2108	1497.77	33.04	63.20	44.17
6′	10, 11, 12	41	968	1708	730.22	27.27	47.19	17.63

Table 7
The top 10 keywords with high centrality.

No.	Keyword	Degree centrality	No.	Keyword	Betweenness centrality
1	Clustering	64.93	1	Similarity Computation	11.17
2	Web Mining	62.715	2	User Profile	10.93
3	User Interest Model	60.686	3	Virtual Community	10.80
4	Web Usage Mining	59.106	4	Similarity	10.63
5	User Interest	59.047	5	Web Mining	10.45
6	Interest Measure	58.279	6	Web Usage Mining	10.45
7	User Clustering	57.857	7	Recommendation Algorithm	10.41
8	E-commerce	57.692	8	Sequence Pattern	10.27
9	User Behavior	57.522	9	Social Networks	10.15
10	Web Data Mining	55.927	10	Collaborative Filtering Recommendation	10.12

According to the high degree centrality and from the perspective of the whole network structure, these keywords were noted as the main focuses of RecSys in China: *Clustering*, *Web Mining*, *User Interest Model*, *Web Usage Mining*, and *User Interest*. Meanwhile, because of the high betweenness centrality, these keywords were defined as the important bridges connecting other research themes or subfields, such as *Similarity Computation*, *User Profile*, *Virtual Community*, *Similarity*, and *Web Mining*. Note that research on user interests (e.g., *User Interest*, *Interest Measure*), and data mining (e.g., *Web Mining*, *Web Usage Mining*) revealed that these were the significant research subfields in RecSys; that is to say, a large amount of studies of RecSys were done on these themes.

The core-periphery structure can reflect the overall co-word network. According to the core-periphery structure, more than half of all selected keywords (84 keywords) could be identified as the core keywords. Thus, the tentative conclusion was drawn that research on RecSys in China was mainly focusing on these themes, such as *Personalized Recommendation*, *E-Commerce*, *Association Rules*, *Data Mining*, *Clustering*, *Similarity*, *User Interest*, *Recommendation Algorithm*, *Sparsity*, and *Digital Library*. Finally, it was inferred that research on RecSys was relatively integrated and focused at present.

Based on the co-word correlation matrix, the density of the overall research network was calculated. The value of density (0.238) is relatively high (Hu, Hu, Deng, et al., 2013; Liu et al., 2012). Compared with other domains, this result indicates that research on RecSys in China has gradually become focused and mature. Further, the centrality and density of each theme-cluster in one-step clustering were calculated and presented in Table 8, and the strategic diagram was generated as shown in Fig. 2. The averages of centrality and density were chosen as the origin (34.32, 0.15) of the strategic diagram. The strategic diagram illustrates the current status and dynamic trends of theme-clusters by dividing them into four quadrants.

As shown in Fig. 2, Clusters 5, 7 and 9 with high centrality and high density, are located in quadrant I. These theme-clusters have a high internal correlation between keywords and tend to be mature. This is especially applicable to Cluster 5, representing the research on *Association Rules*, *Data Mining*, *Clustering*, *Interest Measure*, etc., which are the most mature and well-developed research subfields within RecSys in China in recent years. Cluster 7, which includes themes such as *Sparsity*, *Nearest Neighbor*, *MAE*, *Rating Prediction*, is also a relatively mature research subfield. Finally, it was observed that the themes in Cluster 9 are not well-developed, indicating that research on *User Interest*, especially based on *Social Networking*, needs more research.

Displaying high density and low centrality, Cluster 11 is located in quadrant II. This includes research on *Recommendation Trust* (e.g., *Trust Model*, *Trustworthiness*, and *Recommendation Credibility*), which has received a large amount of attention from researchers recently, and has formed an independent and systematic subfield of RecSys in China. Moreover, this cluster of subfields is becoming more and more mature. On the other hand, there are a small amount of connections with other themes. In consultation with experts specializing in this field, it was noted that the studies related to *Recommendation Trust* have grown to a considerable scale with powerful development momentum in China. The results of this study confirm the expert recognition of this research field.

Table 8
Density and centrality of each cluster.

Cluster	Centrality	Density
1	36.04	0.0107
2	21.28	0.0906
3	37.58	−0.0199
4	38.03	0.1267
5	53.27	0.6077
6	49.73	0.1407
7	39.71	0.3387
8	37.44	0.0025
9	48.90	0.1999
10	24.78	0.0060
11	15.90	0.2479
12	9.21	−0.0081

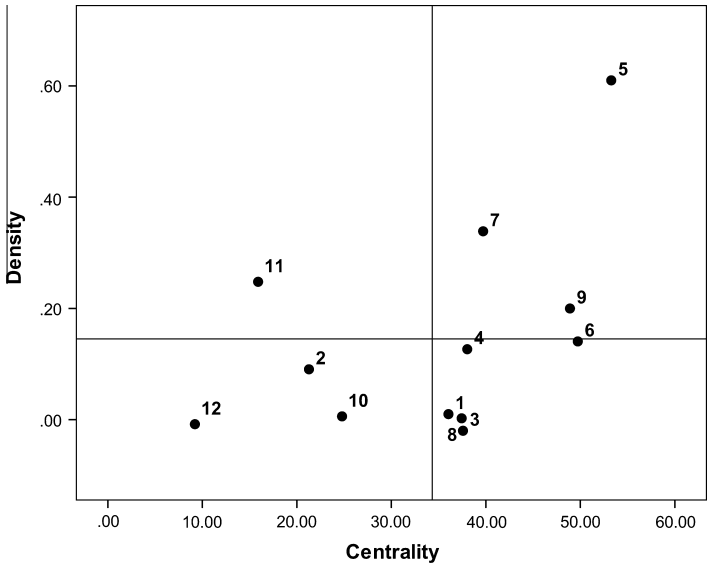

Fig. 2. The strategic diagram of 12 theme clusters of RecSys in China.

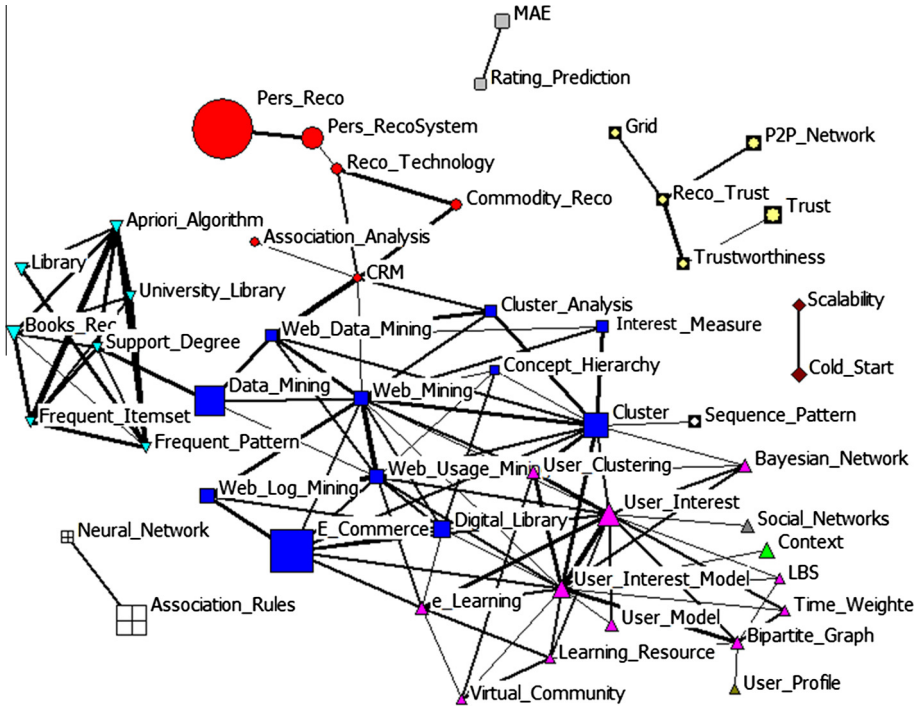

Fig. 3. The network structure of keywords (the Pearson coefficient $\geqslant 0.7$).

With low density and centrality, Clusters 2, 10 and 12 are located in quadrant III. These are relatively peripheral and immature. Cluster 2, which represents the *Book Recommendation* and *Music Recommendation* themes, indicates that techniques of RecSys have been applied in many other fields, but that studies on these themes have not been well-developed. However, it needs to be pointed out that themes in Cluster 2 have good prospects for development. Clusters 10 and 12 are more isolated compared with other themes. In particular, themes such as *Web Service*, *Tag Recommendation*, and *Service Recommendation* in Cluster 12, have been gradually neglected by researchers in China, or have been increasingly marginalized. This finding is consistent with the trend identified by experts from this field in that researchers in China have shifted their research effort to other fields of RecSys and are becoming more concerned with the solutions to specific problems in RecSys, for example, *Cold Start*.

The contradiction in quadrant IV, consisting of Clusters 4 and 6 on one hand and Clusters 1, 3 and 8 on the other, is readily evident. These clusters have low density but relatively high centrality, and it may be considered that themes in them could be central but immature. Primarily, Clusters 1, 3 and 8, located together with similar status, are at the center but are not well-developed in the field of RecSys. They widely connect with other fields, such as *Personalized Recommendation*, *Apriori Algorithm*, and *Ontology*. Remarkably, with a large amount of associations with others, Clusters 4 and 6 are located very close to quadrant I. Thus, themes in them have great potential for development and tend to be mature, for example, *Similarity*, *Agent*, *Hybrid Recommendation*, *Fuzzy Clustering*, and *Domain Ontology*.

Finally, correlation network is a powerful and intuitive approach that can visualize the entire structure of these selected keywords. In the co-word network map (Fig. 3), a connection is presented if its Pearson coefficient is greater than 0.7. It is found that the rest of the keywords were aggregated together into four large theme-clusters and four independent clusters. In addition, clusters with different shapes of nodes were labeled individually in order to distinguish them. In Fig. 3, the size of nodes is proportional to the frequency of keywords, and the thickness of lines is drawn proportionally to the correlation between keywords. The result is similar to and further confirms what has been reported above.

4. Conclusion

In this study, co-word analysis has been used with the aid of clustering and social network analysis methods to reveal the research patterns and trends of RecSys field in China during the period of 2004–2013. Based on the results and discussions reported earlier, the conclusions may be drawn as follows:

(1) In China, research on RecSys is relatively mature and well-developed. The selected 150 keywords represented the themes and research focus areas, measuring over 50% of total frequency of all keywords. Of these, some keywords receive more attention than others. The following appear to be the top keywords: *Personalized Recommendation*, *Data Mining*, *Association Rules*, *User Interest*, *Clustering*, *Sparsity*, *E-commerce*, *Digital Library*, *Trust*, *Agent*, *Ontology*, *Virtual Community*, *Recommendation Algorithm*, *Hybrid Recommendation*, and *Fuzzy Logic*.

(2) Twelve theme-clusters and six larger branches were identified, representing the research subfields of RecSys in China in the past 10 years. Based on their indicators, these clusters were separated into high-level and low-level clusters. First, with relatively more keywords, Cluster 5 is considered as a larger and relatively more important research subfield of RecSys in China. On the other hand, with the least indicators, Cluster 10 and 12 receive the least interest from researchers and are the more isolated research subfields of RecSys in China. In addition, other theme-clusters also have different attributes. For example, Cluster 3, the theme of which is related to *Personalized Recommendation,* is an important interdisciplinary research subfield, and widely connects with other themes. Cluster 11, the theme of which is related to *Recommendation Trust*, is also a well-developed but independent research subfield of RecSys in China.

(3) On the whole, mature and well-developed research subfields of RecSys in China have been formed, and have great potential for further development. For instance, themes in Cluster 5 especially occupy the most central and important position in RecSys field. On the other hand, themes in Clusters 4 and 6 tend to be mature and are increasingly receiving researchers' attention. Overall, research on RecSys in China is focused and mature to some extent, while its development is balanced. This means there are more well-developed research subfields as well as emerging ones under well development. Note that *Recommendation Trust*, *Social Networking*, etc., could be the new research focuses in the future in this area.

(4) On the other hand, some undeveloped or immature research themes continue to persist. The studies on the problems of RecSys, for example, *Cold Start* and *Attack Detection*, need more attention. It was also found that the applications of RecSys remain mainly focused on *E-commerce* and *Digital Library*. It may therefore be inferred that researchers in China are increasingly applying techniques of RecSys in practical applications, such as *E-commerce* based on *Users Behavior* and *Social Networks*, *Book Recommendation*, *Friend Recommendation*, *Knowledge Recommendation*, and *E-learning*. In addition, studies on *Music Recommendation* have begun to appear. It is therefore possible that these themes could emerge as the new focuses in China.

In summary, the contribution of this paper is its use of co-word analysis as an alternative approach and its revelation of the intellectual structure (research patterns and trends) of RecSys in China during the period of 2004–2013. The results and interpretations can contribute to understanding the research on RecSys by Chinese researchers. In general, the approach in this analysis has proved to be effective, but it is necessary to overcome the limitation of the diversity of keywords and the functionality of the analysis software.

This analysis is far from being perfect, and results from this study need to be further verified and recognized using other methods. The researchers of this study intend to conduct another co-word analysis of RecSys field using SCI or another Non-Chinese database. The comparative analysis of domestic and international RecSys research is expected to be carried out in order to find disparities and provide valuable information to promote more research on RecSys in China.

Acknowledgements

This study is supported by National Natural Science Foundation of China Funded Project (71303178), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (13YJC870008), and Initiative Scientific Research Program of Wuhan University (274013).

References

An, X. Y., & Wu, Q. Q. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics, 88*(1), 133–144.
Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics, 22*(1), 155–205.
Callon, M., Law, J., & Rip, A. (1986). *Mapping the dynamics of science and technology: Sociology of science in the real world*. London: Macmillan.
Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics, 27*(2), 119–143.
Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science, 49*(13), 1206–1223.
Courtial, J. P. (1994). A co-word analysis of scientometrics. *Scientometrics, 31*(3), 251–260.
Courtial, J. P., & Law, J. (1989). A co-word study of artificial intelligence. *Social Studies in Science, 19*, 301–311.
DeLooze, M. A., & Lemarie, J. (1997). Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics, 39*(3), 267–280.
Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management, 37*(6), 817–842.
Feng, G. H., & Liang, X. T. (2011). Review of collaborative filtering recommender. *Library and Information Service, 55*(16), 126–130.
Hu, J. M., Hu, C. P., & Deng, S. L. (2013b). Review of research of the information recommendation in the social network environment. *Information and Documentation Services*, 35–39 (in China).
Hu, C. P., Hu, J. M., Deng, S. L., & Liu, Y. (2013a). A co-word analysis of library and information science in China. *Scientometrics, 97*(2), 369–382.
Hu, W., Hu, D., Gao, Y., & Hao, B. (2013c). Friend recommendation algorithm based on association rules and tags. *Computer Engineering and Science, 35*(2), 109–113 (in China).
Law, J., Bauin, S., Courtial, J. P., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics, 14*(3), 251–264.
Law, J., & Whittaker, J. (1992). Mapping acidification research – A test of the co-word method. *Scientometrics, 23*(3), 417–461.
Lee, W. H. (2008). How to identify emerging research fields using scientometrics: An example in the field of information security. *Scientometrics, 76*(3), 503–525.
Lee, B., & Jeong, Y. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics, 77*(1), 3–19.
Leydesdorff, L., & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics, 5*(3), 469–475.
Li, S., Xu, X., & Xu, M. (2013). The measures of books' recommending quality and personalized book recommendation service based on bipartite network of readers and books' lending relationship. *Journal of Library Science in China, 39*(9), 83–95.
Liu, G. Y., Hu, J. M., & Wang, H. L. (2012). A co-word analysis of digital library field in China. *Scientometrics, 91*(1), 203–217.
Liu, Q., & Ye, Y. (2012). A study on mining bibliographic records by designed software SATI: Case study on library and information science. *Journal of Information Resources Management*, 50–58 (in China).
Meng, X., Hu, X., Wang, L., & Zhang, Y. (2013). Mobile recommender systems and their applications. *Journal of Software, 24*(1), 91–108 (in China).
Munoz-Leiva, F., Viedma-del-Jesús, M. I., Sánchez-Fernández, J., & López-Herrera, A. G. (2012). An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality & Quantity, 46*(4), 1077–1095.
Peters, H. P. F., & Vanraan, A. F. J. (1993). Co-word-based science maps of chemical-engineering. 1. Representations by direct multidimensional-scaling. *Research Policy, 22*(1), 23–45.
Qin, G., & Liang, X. (2011). Review of collaborative filtering recommender. *Library and Information Service, 55*(16), 126–130 (in China).
Romo-Fernandez, L. M., Guerrero-Bote, V. P., & Moya-Anegon, F. (2013). Co-word based thematic analysis of renewable energy (1990–2010). *Scientometrics, 97*(3), 743–765.
Ronda-Pupo, G. A., & Guerras-Martin, L. Á. (2012). Dynamics of the evolution of the strategy concept 1962–2008: A co-word analysis. *Strategic Management Journal, 33*(2), 162–188.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.
Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*(1), 17–40.
Stacey, A. (2013). Co-word mapping using correspondence analysis. In *Proceedings of the 12th European conference on research methodology for business and management studies* (pp. 339–346). Academic Conferences Ltd.
Stegmann, J., & Grohmann, G. (2003). Hypothesis generation guided by co-word clustering. *Scientometrics, 56*(1), 111–135.
Vala, J., Lima, M. L., & Caetano, A. (1996). Mapping European social psychology: Co-word analysis of the communications at the 10th general meeting of the EAESP. *European Journal of Social Psychology, 26*(5), 845–850.
Wanfang Data Corporation (2015). Foreign literature database. Available at: <http://g.wanfangdata.com.cn/ResourceDescription/NSTL>.
Wang, G., & Liu, H. (2012). Survey of personalized recommendation system. *Computer Engineering and Applications, 48*(7), 66–76 (in China).
Yang, B., & Zhao, P. (2011). Review of the art of recommendation algorithms. *Journal of Shanxi University (Natural Science Edition), 34*(3), 337–350 (in China).
Yi, M., & Deng, W. (2011). Summary of research on tag-based personalized information recommendation. *Information Studies: Theory & Application, 34*(3), 126–128 (in China).
Yin, F. (2013). Research on method of news recommendation based on content semantic in social network. *Computer Technology and Development*, 253–257 (in China).
Zhang, Y., Chen, W., & Fu, S. (2013). Collaborative filtering recommendation research. *Microcomputer & Its Applications, 32*(6) (in China), 4–6 &10.