



# Bibliometric Research on Recommender System Research Based on Web of Science

Yafei Di

Literature Information Center, Fudan University, Shanghai, China

## Email address:

16210830002@fudan.edu.cn

## To cite this article:

Yafei Di. Bibliometric Research on Recommender System Research Based on Web of Science. *Science Innovation*.

Vol. 5, No. 5, 2017, pp. 250-255. doi: 10.11648/j.si.20170505.11

Received: June 3, 2017; Accepted: July 4, 2017; Published: July 19, 2017

**Abstract:** Since recommender systems are widely used in many fields, it is important to analyze the present situation of recommender system and its development trends. This paper measured and visualized the related papers between 1996 and 2015 based on Web of Science™ Core Collection. Based on annual publications and times cited, journal distribution and research fields, high-frequency authors and cited references, countries or regions and institutions, high-frequency keywords and bursts keywords, this paper summarized the current situation of recommender system and forecasted the development trends of recommender system. This paper determined that the research hotspot of recommender systems was method for recommender system and the application in the field of electronic commerce, science basic research, electronic learning and knowledge management. The future research hotspot will still be the application of the recommender system.

**Keywords:** Web of Science, Recommender System, Bibliometrics, CiteSpace

## 基于Web of Science的推荐系统研究文献计量研究

狄亚飞

文献信息中心, 复旦大学, 上海, 中国

## 邮箱

16210830002@fudan.edu.cn

**摘要:** 推荐系统现已被广泛应用于很多领域, 本文对推荐系统相关研究进行梳理和分析, 总结相关研究的发展进程并预测其发展趋势。以Web of Science™核心数据库为检索对象, 对1996-2015年间发表的推荐系统相关论文进行文献计量和可视化分析。从年发文量与被引频次、期刊分布与研究领域、高频作者和高被引频文献、国家或地区和机构、高频关键词和突发关键词等角度, 全面梳理推荐系统的发展现状, 预测推荐系统的发展趋势。本文确定了推荐系统研究热点是推荐系统所使用的推荐方法以及推荐系统在电子商务、科学基础研究、在线学习、知识管理等领域中的应用, 未来推荐系统的研究热点仍将是推荐系统的应用。

**关键词:** Web of Science, 推荐系统, 文献计量, CiteSpace

## 1. 引言

互联网技术的迅速发展给用户带来海量信息, 这虽然满足了用户对信息的需求, 却带来了“信息超载”的问题。

解决用户信息超载的问题的一个重要方法就是推荐系统。推荐系统是针对具体项目、满足用户需求的信息过滤系统, 它通过收集用户的偏好信息, 将这些信息用于个性化推荐的制作, 建立起用户与信息产品间的二元关系[1]。推荐系

统现已被广泛应用于很多领域，如电影[2]、音乐[3]、餐馆[4]、购物[5]、新闻[6]、书籍[7]、论文[8]、搜索查询[9]和金融服务[10]等方面。

自20世纪90年代中期推荐系统诞生起，学界对推荐系统进行了大量研究，但对推荐系统相关文献进行计量学研究的文献较少，尚无对推荐系统研究进行全面梳理的文献。因此，本研究采用文献计量分析方法，基于CiteSpace等分析工具，对该领域年发文量和引文量、主要发文期刊、高频作者、国家或地区和机构、热点文献和热点关键词进行分析，全面梳理推荐系统发展现状并对推荐系统研究的发展趋势进行展望。

2. 推荐系统数据来源和分析工具

Web of Science数据库是由美国科技信息所(Institute for Scientific Information, 简称ISI)继发行印刷版的《科学引文索引》(SCI)、机读版SciSearch数据库、光盘版SCI后，于1997年推出的基于Web的最新、最先进的引文索引数据库[11-14]。本文选取Web of Science™核心数据库为数据来源，总检索式为：

TS=(“recommend\* system\*”or“recommend\* model\*”or“recommend\* algorithm\*”or“recommend\* technique\*”or“recommend\* technolog\*”or“recommend\* platform\*”or“recommend\* engin\*”) not ts= (“recommended system\*”or“recommended model\*”or“recommended algorithm\*”or“recommended technique\*”or“recommended technolog\*”or“recommended platform\*”or“recommended engin\*”) AND PY=(1996-2015)

文献类型：Article

检索语言：English

依据上述检索策略，共检索到文献2850篇。利用Excel2016和CiteSpace5.0. R2 SE分析处理1996-2015年间推荐系统相关文献，总结20年间作者、机构和国家或地区的发文情况，对高被引文献进行内容分析，同时总结热点关键词和推荐系统发展趋势。

3. 结果与分析

3.1. 年发文量与被引频次

依据检索策略得出1996-2015年20年间推荐系统文献总量为2850篇，年平均发文量为142.5篇，其变化趋势如图1所示。从图1可知，近20年间，推荐系统论文发文量总体为增加趋势，但有所波动。发文数量的年变化大致可以分为3个阶段，第一阶段为1996-2001年，这段时间发文量较少，总发文量为83篇，占总发文量的2.91%；第二阶段为2002-2007年，这段时间年发文量有所波动，总发文量为591篇，占总发文量的20.74%；第二阶段为2008-2015年，这段时间年发文量增长平稳，发文量也较多，总发文量为2137篇，占总发文量的76.35%。

被引频次是文献计量学中被用来测度学术论文社会显示度和学术影响力的重要指标[15]。1996-2015年20年间2850篇推荐系统文献总被引频次为34292，篇均被引频次为12.03。被引频次的年变化如图1所示，总体为增加趋势。被引频次

的年变化大致可以分为2个阶段，第一阶段为1996-2010年，文献被引频次增长平缓且与发文量成正相关，这表明自1995年推荐系统提出后，对推荐系统研究的热度不断增加；第二阶段为2011-2015年，文献被引频次迅速增长，这表明这一时期推荐系统的研究进入一个快速发展时期。

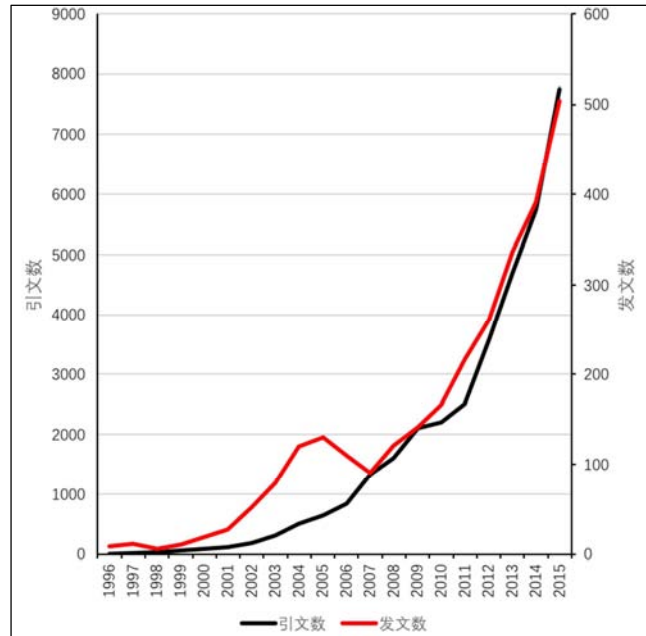


图1 年度发文与被引情况。

3.2. 期刊分布与研究领域分析

1996-2015年发表的2850篇推荐系统文献共分布在835种期刊上，每本期刊平均发文量为3.41篇，发文在10次以上的期刊（含10次）共有56种，并且这些文献分布表现出一定的集中性。其中发文量排名前十的期刊共刊文861篇，刊均发文86.1篇，占据总发文量的30.21%（见表1）。其中发文量排名前五的期刊《Expert Systems with Applications》、《Lecture Notes in Computer Science》、《Lecture Notes in Artificial Intelligence》、《Knowledge Based Systems》和《Information Sciences》共刊文650篇，刊均发文量为130篇，占据总发文量的22.81%。

表1 发文量居前10的期刊。

序号	期刊名称	发文量	占比(%)
1	Expert Systems with Applications	219	7.684
2	Lecture Notes in Computer Science	163	5.719
3	Lecture Notes in Artificial Intelligence	129	4.526
4	Knowledge Based Systems	72	2.526
5	Information Sciences	67	2.351
6	User Modeling and User Adapted Interaction	50	1.754
7	Multimedia Tools and Applications	47	1.649
8	Decision Support Systems	41	1.439
9	Acm Transactions on Intelligent Systems and Technology	37	1.298
10	IEEE Transactions on Knowledge and Data Engineering	36	1.263

基于CiteSpace[16-21]的Category模块对推荐系统文献进行研究领域知识图谱的可视化分析，Links参数选取

Cosine算法, Threshold时间参数区设为: (2, 2, 20), (4, 3, 30), (3, 3, 20), 运行得到1996-2015年推荐系统研究领域分析图谱(如图2), 后文CiteSpace参数设置和算法选择同上。从图2发现推荐系统的研究领域主要集中于计算机科学、人工智能、信息系统、软件编程、通讯、经济学和图书情报学等领域。同时计算机科学处于知识图谱的中心, 且其他较突出的节点也与计算机科学相关, 这表明计算机科学是推荐系统研究中的核心研究领域。

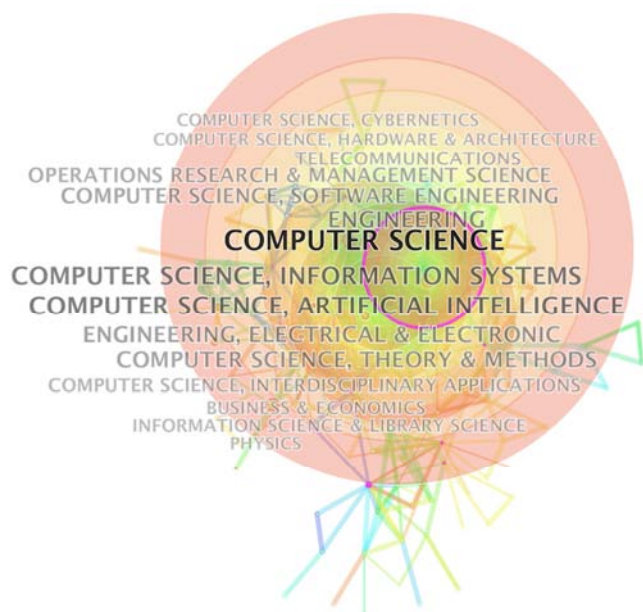


图2 1996-2015年推荐系统研究领域分析图谱。

### 3.3. 高频作者分析

作者发文数量的多少在一定程度上反映了其在该研究领域的影响力大小。1996-2015年20年中发文量在10篇以上(含10篇)的作者共51位。表2列举统计结果中发文量居前10位的作者, 前10位作者共发文253篇, 占比8.88%。基于Web of Science的类别划分, 253篇推荐系统文献排名前五的类别为: computer science artificial intelligence(53篇)、physics multidisciplinary(45篇)、computer science information systems(31篇)、multidisciplinary sciences(19篇)、engineering electrical electronic(18篇), 占前10位作者发文量的65.61%。这表明发文量较高作者的研究领域主要集中于人工智能、物理学、信息系统、交叉学科和电子电气工程。

表2 发文量居前10的作者。

序号	作者	发文量	占比(%)
1	ZHANG YC	41	1.439
2	ZHOU T	35	1.228
3	SMYTH B	32	1.123
4	LIU JG	26	0.912
5	CHEN L	22	0.772
6	ZENG A	20	0.702
7	BLANCO-FERNANDEZ Y	20	0.702
8	ZHANG ZK	19	0.667
9	LOPEZ-NORES M	19	0.667
10	HERRERA-VIEDMA E	19	0.667

### 3.4. 国家或地区 and 机构分析

依据检索结果, 1996-2015年20年间共82个国家或地区发表了推荐系统相关文献, 平均每个国家或地区的发文量为34.76篇, 发文量在100篇以上(含100篇)的国家或地区共计11个。基于CiteSpace的Country模块对推荐系统文献进行国家或地区知识图谱的可视化分析, 运行得到1996-2015年推荐系统国家或地区共现知识图谱(如图3)。

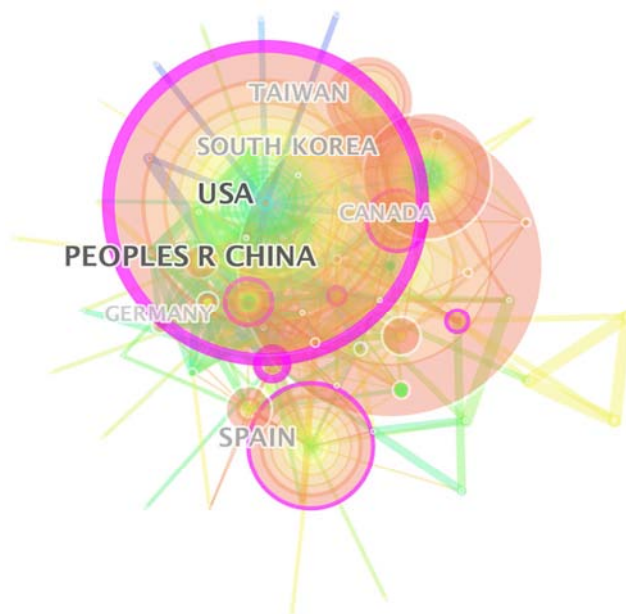


图3 1996-2015年推荐系统国家或地区共现知识图谱。

从图3可知, 美国和中国是推荐系统领域研究最多的两个国家, 其次为西班牙、韩国、台湾地区、加拿大和德国。基于CiteSpace中心度计算得到各个国家或地区的中心度, 对发文量前5位国家或地区进行统计, 美国: 0.52, 中国: 0.03, 西班牙: 0.19, 韩国: 0.02, 台湾: 0。从发文量和中心度发现, 美国在发文量和中心度都领先于其他国家, 表明其在推荐系统领域具有强大的科研实力; 其他国家或地区虽然也在推荐系统领域进行了大量研究, 但影响力较低且缺乏国际合作。

从发文机构来看, 发文量居前10的机构如表3, 总发文量为394篇。其中中国占据4位, 分别是电子科技大学、中国科学技术大学、浙江大学和清华大学, 总发文量为168篇, 占前10位发文机构42.97%。从研究领域看, 中国这4所大学研究方向主要集中于计算机科学和物理学这两大领域, 与国际总体研究领域相似。值得注意的是, 作为推荐系统研究实力最强的美国只有1所研究机构进入前10, 从Web of Science机构拓展来看, 这是因为美国许多大学存在多个分支机构, WOS进行发文机构分析时, 统计的是这些分支机构的发文量, 如整个加州大学系统发文量为47篇, 但WOS统计时以加州大学各个分支机构进行统计。此外欧洲的研究机构对推荐系统研究也较多, 总发文量为169篇, 占前10位发文机构42.89%。



表3 发文量居前10的机构。

序号	机构	地区	发文量
1	弗里堡大学	瑞士	65
2	电子科技大学	中国	58
3	中国科学技术大学	中国	45
4	格拉纳达大学	西班牙	37
5	明尼苏达大学	美国	35
6	浙江大学	中国	35
7	都柏林大学	爱尔兰	34
8	清华大学	中国	31
9	悉尼科技大学	澳大利亚	27
10	马德里理工大学	西班牙	27

3. 5. 高被引文献内容分析

被引频次是个绝对数量指标，是指某期刊论文自发表以来根据特定年份统计的被引用总次数。它直接反映了期刊论文在科学发展和文献交流中所起的作用，从历史的角度反映了期刊论文的学术影响力[22]。从被引次数最高的10篇论文来看（如表4），前10篇高被引论文中的第1、3、6、8和9篇[23-27]都是研究推荐系统所使用的推荐方法，其中第1、8和9篇重点研究了协同过滤推荐方法；第3篇重点研究了组合推荐方法；第6篇重点研究了基于内容的推荐方法。其它5篇文献[28-32]研究了推荐系统的发展、开发和应用，这表明推荐系统的研究热点是推荐系统所使用的推荐方法。

表4 前10位高被引论文。

序号	被引频次	标题
1	1423	Evaluating collaborative filtering recommender systems
2	900	Recommender systems
3	891	Hybrid recommender systems: Survey and experiments
4	700	Matrix Factorization Techniques for Recommender Systems
5	488	The Dynamics of Viral Marketing
6	484	Item-based top-N recommendation algorithms
7	457	Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 Revision
8	411	Latent semantic models for collaborative filtering
9	395	Eigentaste: A constant time collaborative filtering algorithm
10	389	Automatic personalization based on web usage mining - Web usage mining can help improve the scalability, accuracy, and flexibility of recommender systems

3. 6. 热点关键词分析

3. 6. 1. 高频关键词分析

基于CiteSpace的Keyword模块对推荐系统文献进行关键词知识图谱的可视化分析，统计出关键词1562个，出现频次共8192次，平均每个关键词出现频次为5.24次。其中，出现频次大于50次的关键词共30个（如表5），运行得到1996-2015年推荐系统关键词共现知识图谱（如图4）。

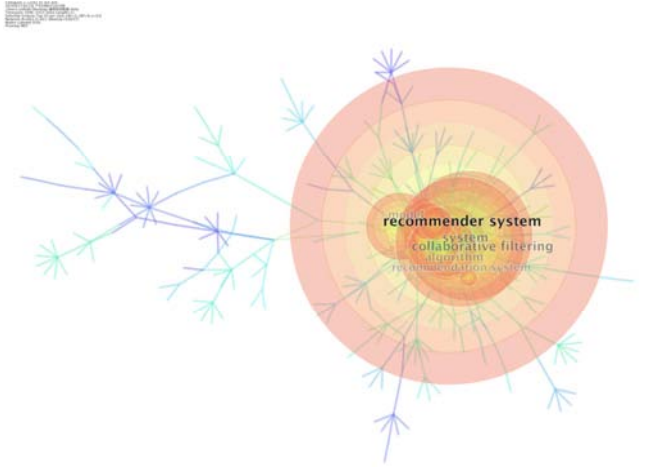


图4 1996-2015年推荐系统关键词共现知识图谱。

这些关键词主要集中在三个方面。第一组：recommender system、system、algorithm、model、recommendation system、recommendation、ontology、framework, 这组关键词主要是推荐系统的简介。第二组：collaborative filtering、information retrieval、design、data mining、classification、experimentation、similarity、retrieval、search、agent、matrix factorization, 这组关键词主要是推荐系统所使用的方法。第三组：web、personalization、information、network、social network、trust、e commerce、internet、performance、management、of the art, 这组关键词主要是推荐系统的评价与应用。这三组关键词表明了推荐系统在1996-2015年的研究热点，从推荐系统的提出，到推荐系统所使用的方法，再到推荐系统的评价和应用。

表5 频次大于50的关键词。

序号	关键词	频次
1	recommender system	1056
2	system	440
3	collaborative filtering	436
4	algorithm	259
5	model	240
6	recommendation system	203
7	web	182
8	personalization	175
9	information	165
10	network	131
11	recommendation	126
12	social network	121
13	trust	109
14	e commerce	83
15	information retrieval	76
16	internet	76
17	data mining	74
18	design	73
19	performance	71
20	ontology	67
21	framework	65
22	classification	61
23	experimentation	63
24	similarity	62
25	management	59
26	of the art	57
27	retrieval	53
28	search	53
29	agent	51
30	matrix factorization	51

### 3.6.2. 突发关键词分析

正如引用的爆发可能表明科学界对发表文章的关注程度，突发关键词的检测也可作为新兴研究趋势的指标[33]。基于CiteSpace的Burst检测功能对关键词进行突发性检测，可以得出前25位突发关键词（如图5），其中Strength值表示关键词突发的强度。以Strength最大的关键词为例分析突发关键词的演变，electronic commerce的Strength值为7.7318，突发时间为2007-2011年。这个时间段是推荐系统算法已经到达成熟的阶段，2005年菲利普·科特勒的“精准营销”[34-35]产生巨大影响后，电子商务也开始大量应用推荐系统，B2C网站[36-37]无论在数量上还是规模上都进入了高速膨胀期，此时该关键词也处于突发期。经过一段时期的发展，推荐系统在电商领域逐渐成熟，突发期也于2011年结束。此时研究热点也发生了转变，其他领域逐渐借鉴了推荐系统电商领域的应用，如科学基础研究、在线学习和知识管理，此时突发关键词转变为science basic research、e learning、knowledge management。

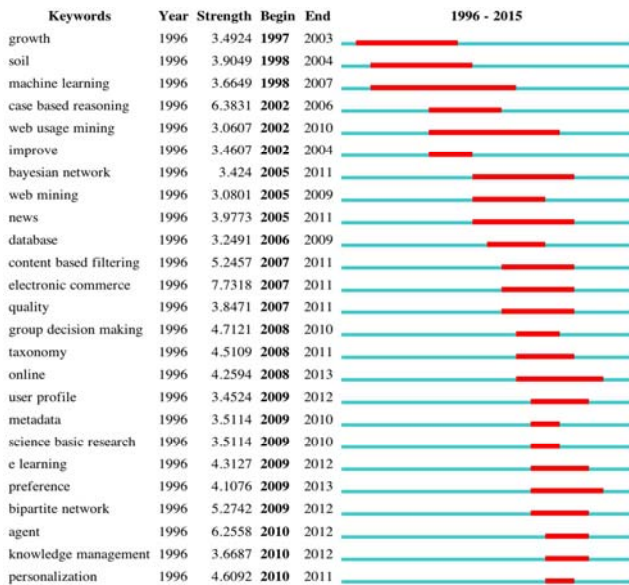


图5 1996-2015年推荐系统文献前25位突发关键词。

不同时期的突发关键词往往表明不同时期新兴的研究趋势，从图5可知推荐系统关键词突发主要分为两个阶段，第一阶段突发的关键词包括：growth、soil、machine learning、case based reasoning、web usage mining、improve，这一阶段是推荐系统自1995年提出后推荐系统最初的发展阶段。第二阶段突发的关键词包括：bayesian network、web mining、news、database、content based filtering、electronic commerce、quality、group decision、taxonomy、online、user profile、metadata、science basic research、e learning、preference、bipartite network、agent、knowledge management、personalization，这一阶段是推荐系统发展过程中各典型推荐方法的提出和推荐系统广泛应用的阶段，如基于内容推荐和群体决策推荐的提出促使推荐系统在科学基础研究、在线学习、知识管理、电商等领域得到广泛应用。从20年的突发关键词的演变过程可以预测未来推荐系统的研究热点是推荐系统在更新兴领域的应用。

## 4. 结论

本文以Web of Science™核心数据库1996-2015年间的推荐系统文献为数据样本，借助CiteSpace等分析工具，通过文献计量和可视化分析对20年间推荐系统领域的发展进行梳理。主要结论有：（1）1996-2015年20年间推荐系统文献总量和被引频次都保持增加趋势，并在近年来增长较快，表明推荐系统现正进入一个快速发展期；（2）20年间发表推荐系统文献的期刊分布存在一定的集中性，发文量排名前五的期刊占据总发文量的22.76%，同时计算机科学是推荐系统研究中的核心研究领域；（3）发文量较高作者的研究领域主要集中于人工智能、物理学、信息系统、交叉学科和电子电气工程，高被引文献的研究热点是推荐系统所使用的推荐方法；（4）从国别和地区来看，美国和中国是推荐系统领域研究最多的两个国家，但中国研究影响力较低；从机构来看，中国有4所大学进入机构发文量前10，中国大学在推荐系统领域实力较强；（5）通过对20年间的高频关键词和突发关键词的分析，得出推荐系统研究热点是推荐系统使用的推荐方法和推荐系统在科学基础研究、在线学习、知识管理、电商等领域中的应用。可以预见，随着个性化服务需求的不断提升，个性化的用户需求必将促进推荐系统在更多领域更广泛的应用。

## 参考文献

- [1] Ricci F, Rokach L, Shapira B, et al. Recommender Systems Handbook [M]. Springer US, 2011.
- [2] Ante Odić, Marko Tkalčič, Jurij F. Tasič, et al. Predicting and detecting the relevant contextual information in a movie-recommender system [J]. Interacting with Computers, 2013, 25(1):74-90.
- [3] Li Q, Myaeng S H, Kim B M. A probabilistic music recommender considering user opinions and audio features [J]. Information Processing & Management, 2007, 43(2):473-487.
- [4] Felfernig A, Isak K, Szabo K, et al. The VITA Financial Services Sales Support Environment. [C]// National Conference on Innovative Applications of Artificial Intelligence. AAAI Press, 2007:1692-1699.
- [5] 张亮, 柏林森, 周涛. 基于跨电商行为的交叉推荐算法[J]. 电子科技大学学报, 2013(1):154-160.
- [6] Chiang J H, Chen Y C. An intelligent news recommender agent for filtering and categorizing large volumes of text corpus [J]. International Journal of Intelligent Systems, 2004, 19(3):201-216.
- [7] Shiratori M. The Book of My Recommendation in Overseas "Mechanical Engineering" [J]. Jama Neurology, 2014, 71(24):1379-1385.
- [8] 段文奇, 惠淑敏. 基于协同过滤的论文推荐一传播平台模型研究[J]. 科学学研究, 2012, 30(10):24-29.

- [9] Smyth B, Coyle M, Briggs P. Communities, Collaboration, and Recommender Systems in Personalized Web Search [J]. *Recommender Systems Handbook*, 2010:579-614.
- [10] Felfernig A, Isak K, Szabo K, et al. The VITA Financial Services Sales Support Environment. [C]// *National Conference on Innovative Applications of Artificial Intelligence*. AAAI Press, 2007:1692-1699.
- [11] Jacso P. As we may search—Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases [J]. *Currentence*, 2005, 89(89):1537-1547.
- [12] Falagas M E, Pitsouni E I, Malietzis G A, et al. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weakness. *FASEB J* [J]. *Faseb Journal*, 2008, 22(2):338-342.
- [13] Archambault, Campbell D, Gingras Y, et al. Comparing bibliometric statistics obtained from the Web of Science and Scopus [J]. *Journal of the Association for Information Science and Technology*, 2009, 60(7):1320-1326.
- [14] 王会梅. 数字图书馆中Web of science数据库的解读与应用 [J]. *农业图书情报学刊*, 2010,22(11):100-103.
- [15] 杨华, 蔡言厚. 论提高被引频次的意义、客观条件和主观努力[J]. *评价与管理*, 2010,02:24-27.
- [16] Börner K, Chen C, Boyack K W. Visualizing knowledge domains [J]. *Annual Review of Information Science & Technology*, 2003, 37(1):179-255.
- [17] Chen C. Mining the Web: Discovering knowledge from hypertext data [J]. *Journal of the Association for Information Science and Technology*, 2004, 55(3):275-276.
- [18] Synnvestedt M B, Chen C, Holmes J H. CiteSpace II: visualization and knowledge discovery in bibliographic databases. [J]. *AMIA. Annual Symposium proceedings / AMIA Symposium*. AMIA Symposium, 2005, 2005:724-728.
- [19] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. *Journal of the Association for Information Science and Technology*, 2006, 57(3):359-377.
- [20] Chen C, Dubin R, Kim M C. Orphan drugs and rare diseases: a scientometric review (2000 – 2014) [J]. *Expert Opinion on Orphan Drugs*, 2014, 2(7):709-724.
- [21] Chen C, Hu Z, Liu S, et al. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. [J]. *Expert Opinion on Biological Therapy*, 2012, 12(5):593-608.
- [22] 张静. 引文、引文分析与学术论文评价[J]. *社会科学管理与评论*, 2008,01:33-38.
- [23] Herlocker J L. Evaluating collaborative filtering recommender systems [J]. *Acm Transactions on Information Systems*, 2004, 22(1):5-53.
- [24] Burke R. Hybrid Recommender Systems: Survey and Experiments [J]. *User Modeling and User-Adapted Interaction*, 2002, 12(4):331-370.
- [25] Leskovec J, Adamic L A, Huberman B A. The Dynamics of Viral Marketing [J]. *Acm Transactions on the Web*, 2005, 1(1):5.
- [26] Hofmann T. Latent semantic models for collaborative filtering [J]. *Acm Transactions on Information Systems*, 2004, 22(1):89-115.
- [27] Goldberg K, Roeder T, Gupta D, et al. Eigentaste: A Constant Time Collaborative Filtering Algorithm [J]. *Information Retrieval Journal*, 2001, 4(2):133-151.
- [28] Resnick P, Varian H R. Recommender systems[J]. *Communications of the Acm*, 1997, 40(3):56-58.
- [29] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems [J]. *Computer*, 2009, 42(8):30-37.
- [30] Leskovec J, Adamic L A, Huberman B A. The dynamics of viral marketing [C]// 2006:228-237.
- [31] Brozek J L, Bousquet J, Baena-Cagnani C E, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. [J]. *Journal of Allergy & Clinical Immunology*, 2010, 126(3):466-476.
- [32] Srivastava J. Automatic personalization based on web usage mining - Web usage mining can help improve the scalability, accuracy, and flexibility of recommender systems [J]. 2000.
- [33] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [M]. John Wiley & Sons, Inc. 2006.
- [34] 刘征宇. 精准营销方法研究[J]. *上海交通大学学报*, 2007(S1):143-146.
- [35] Meng Q, Han X. Research of precise marketing strategy based on data mining [J]. *Metallurgical & Mining Industry*, 2015.
- [36] Mokhtarian P L. A conceptual analysis of the transportation impacts of B2C e-commerce [J]. *Transportation*, 2004, 31(3):257-284.
- [37] Cao M, Zhang Q, Seydel J. B2C e-commerce web site quality: an empirical examination [J]. *Industrial Management & Data Systems*, 2006, 105(5):645-661.