

基于知识图谱的国际协同计算领域可视化分析

姚艳玲^{1,2}, 袁 锋^{1,2}, 王 宁^{1,2}

YAO Yanling^{1,2}, YUAN Feng^{1,2}, WANG Ning^{1,2}

1. 山东管理学院 信息工程学院, 济南 250357

2. 山东管理学院 职工健康数据管理研究所, 济南 250357

1. School of Information Engineering, Shandong Management University, Jinan 250357, China

2. Institute of Staff Health Data Management, Shandong Management University, Jinan 250357, China

YAO Yanling, YUAN Feng, WANG Ning. Visualization analysis on international collaboration computing based on knowledge mapping. Computer Engineering and Applications, 2017, 53(7): 30-40.

Abstract: This paper analyzes 2 297 articles in the field of collaboration computing of SCI from Web of Science database, especially analyzes the time distribution, the research power and cooperation, research direction, the mainstream, the development and the evolution process, the research focus and the frontier. The results show that the international study of collaboration computing is in the rising period; the main countries and regions of this field are the USA, China and Europe, the main research institutions are in the USA and China, although China has formed a certain scientific research strength, but it is lack of high yield researchers, in general, the cooperations among the European countries are the most closely; the research directions are mainly composed of computer science; Foster I has become the most influential mainstream researchers of this field, the mainstream journals include LNCS, Commun ACM Lect and so on, the mainstream area includes 9 parts such as grid computing and so on. The development and evolution process are the preliminary study—distributed computing—grid computing—cloud computing, accompanying with collaborative filtering; cloud computing, grid computing and other key words represent the current hotspots of this field, grid computing, cloud computing, distributed computing, Web services, collaborative filtering and other key words represent the front of this field, and this field has been widely infiltrated and merged into biology, pathology, climatology, education, enterprise information system, social network, multimedia, text analysis, industry-university collaboration and other fields, which reflect that collaboration computing field has strong interdisciplinary characteristics.

Key words: knowledge mapping; collaboration computing; visualization; research strength; evolution process; research hotspots

摘 要:通过CiteSpace对Web of Science的SCI数据库中收录的协同计算领域的2 297篇文献进行分析,具体对国际协同计算领域的发文量时间分布、研究力量及合作、研究方向、研究主流、发展演变过程、研究热点及前沿进行了分析。结果表明该领域的研究正处于上升期;国家、地区以美国、中国和欧洲为核心,其研究机构以美国和中国的机构为核心,虽然我国已经具备了一定的科研实力,但缺乏该领域的高产研究者,总体来看欧洲各国之间的合作最为密切;研究方向的构成以计算机科学为主;Foster I为该领域中最具影响力的主流研究者,该领域的主流期刊(文献)包括Lecture Notes in Computer Science、Commun ACM等,主流分支领域包括网格计算等9部分;发展演变过程为前期相关研究—分布式计算—网格计算—云计算,同时穿插协同过滤的发展路径;云计算、网格计算等关键词代表了当前该领域的热点,网格计算、云计算、分布式计算、网络服务、协同过滤等关键词则代表了当前该领域的前沿,且已经广泛渗透、融合到生物学、病理学、气候学、教育学、企业信息系统、社会网络、多媒体、文本分析、产学研等领域

基金项目:国家社会科学基金(No.16BGL181)。

作者简介:姚艳玲(1983—),通讯作者,女,讲师,主要研究领域为智能信息处理,E-mail:kbyylkyx@163.com;袁锋(1972—),女,博士,教授,主要研究领域为人工智能;王宁(1978—),女,博士,讲师,主要研究领域为云计算。

收稿日期:2016-09-26 **修回日期:**2017-01-04 **文章编号:**1002-8331(2017)07-0030-11

中,体现了协同计算这一领域具有较强的跨学科特点。

关键词:知识图谱;协同计算;可视化;研究力量;演变过程;研究热点

文献标志码:A **中图分类号:**TP393 **doi:**10.3778/j.issn.1002-8331.1609-0363

1 引言

随着管理方式的变革,社会正在逐步由个人工作时代迈入协同工作时代,先进的技术,如计算机网络技术、通讯技术、多媒体技术和群件技术共同构成了协同计算环境,可以使不同地域、不同时间、不同文化背景的人们能够协调一致地为某项任务而共同工作称为协同计算。随着协同计算的发展,该领域的研究力量的分布是怎样的?形成了哪些主流分支领域?其发展演变过程是怎样的?当前的研究热点及前沿又是什么?本文试图通过知识图谱对上述问题进行分析和解答,以期为该领域的研究者提供参考。

2 研究说明

2.1 研究数据来源

本文数据来源于美国科学情报研究所 ISI 出版的 Web of Science 数据库的科学引文数据库 (SCI+ISTP),检索式为 TS=(Collaboration Computing*) or TS=(Collaborative Computing*),共检索得到 2 297 篇文献,数据类型为 Article,数据最终检索时间为 2016 年 4 月 12 日。

2.2 研究内容

本文研究内容包括国际协同计算领域的如下内容:(1)发文量时间分布;(2)研究力量分布及合作分析;(3)研究方向分析;(4)研究主流分析;(5)发展演变过程分析;(6)研究热点及前沿分析。

2.3 研究工具

知识图谱是以科学知识为对象,显示科学知识的发展进程与结构关系的一种图形,借助于知识图谱的帮助,可以透视知识体系中各个领域的结构,预测科学技术知识前沿的发展态势^[1]。常用的知识图谱软件包括 Bibexcel、CiteSpace、SCI²、VOSviewer、TDA、Pajek 等。其中 CiteSpace 软件相较于其他软件具有其自身的优点,可适用于所有的操作系统,能够处理国际通用的 WOS 引文格式以及国内的 CSSCI 引文格式等,其软件自带数据转换工具,具备了丰富的分析功能,包括作者、期刊的耦合分析,作者、国家、机构的合作分析,参考文献的共被引分析及关键词的共词分析等。由于具备上述优点,近年来在国内外被广泛使用,国外应用领域包括天体生物学^[2]、生物能源^[3]、交通运输^[4]、普适计算^[5]、城市研究^[6]、能值研究^[7]、Agent 系统^[8]、知识管理^[9]、知识工程^[10]、纳米材料^[11]、推荐系统^[12]、精神病学^[13]、神经损伤修复^[14]、细胞研究^[15]等以自然科学为主的领域。国内则更

多应用在图情学^[16]、教育学^[17]、体育学^[18]、经济学^[19]、语言学^[20]等人文社会科学领域。在事物之间关系的显示方面,CiteSpace 软件则不如某些专业的社会网络软件功能强大。鉴于此,本文以 CiteSpace 为主要工具,同时以 Pajek 软件为辅助工具,对国际协同计算领域进行可视化研究。

3 结果及分析

3.1 发文量时间分布

该研究领域最早的文献为意大利米兰圣心天主教大学的 Branca、Pontani、Silva 三位学者于 1990 年发表于 *Computers and the Humanities* 期刊上的 *The Interdisciplinary Group for Expression Signs Computing* 一文,该文章开启了协同计算研究的序幕。从图 1 可以看到自 1990 年至今国际协同计算研究领域的总体发文量处于上升状态,说明该领域的研究仍然处于上升期。

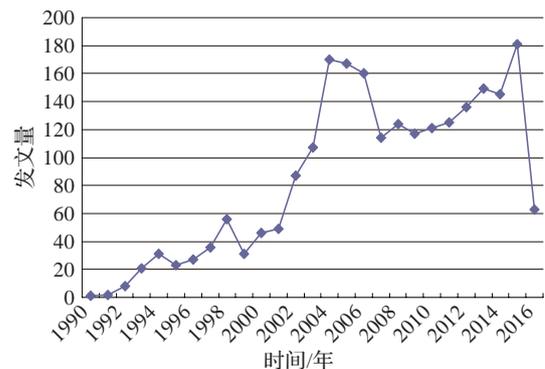


图1 发文量时间分布

3.2 研究力量分布及合作分析

为了分析国际协同计算领域的研究力量分布及其合作,在CiteSpace中分别进行国家/地区(Country/Region)、机构(Institution)、作者(Author)、等分析,设置1990—2016年每2年1个时间片,数据选择标准为每个时间片取top50,同时在Pajek中进行国家(地区)、机构的合作分析,得到图2~图6的研究力量知识图谱。

3.2.1 国家、地区分析

图2和图3表征了国家、地区的发文量及合作知识图谱。从发文量来看,美国发文量(633篇)远远高于其他国家、地区,之后依次为中国(236篇)、英格兰(149篇)、西班牙(102篇),这4个国家、地区的发文量之和约占总发文量的60%,说明美国、中国、英格兰、西班牙在国际协同计算领域的研究处于领先地位。

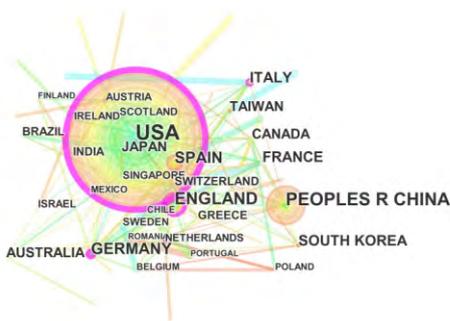


图2 国家、地区知识图谱(CiteSpace)

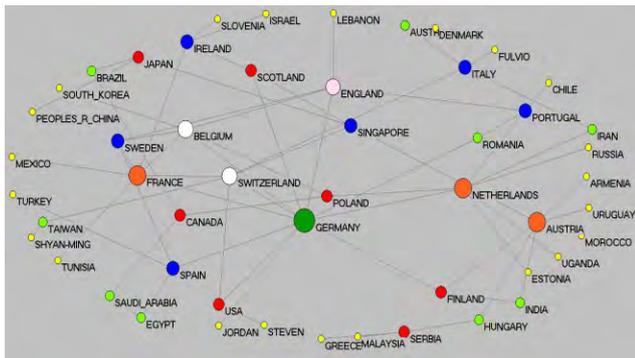


图3 国家、地区合作知识图谱(Pajek)

从合作程度来看,合作最为频繁的国家、地区主要集中在欧洲,如图3所示的中心区域位置,包括德国、法国、英格兰、荷兰、比利时、爱尔兰等。可以看出,图2和图3表征的合作状况较为一致,即欧洲的合作水平最高。本文对中心性排名靠前的10个欧洲国家进行统计,发现其发文量总和为588篇,约占总发文量的25.6%,欧洲各国虽然总体发文量并不突出,但因为欧洲内部各个国家之间较强的合作性,使得欧洲在国际协同计算领域成为一支不可忽视的研究力量。图3所示中国处于边缘位置,说明中国与其他国家、地区在该领域的合作较少,因此我国日后需加强国际间的交流与合作。

3.2.2 研究机构分析

图4和图5表征了机构的发文量及合作知识图谱。从图4的发文量来看,中国的中科院和美国的伊利诺伊大学发文量同为24篇,接下来分别为美国斯坦福大学(20篇)、中国清华大学(19篇)。发文量大于等于10篇的34个机构中,美国占了13家,中国占了10家,西班牙和英国各占了2家,新加坡、加拿大、韩国、智利、澳大利亚、日本各占了1家,此外还有1个国际合作组织——欧洲核子研究组织。可以看出美国和中国在国际协同计算领域高产研究机构数量上占据了绝对的优势。从图4的研究机构合作来看,主要形成了以下几个研究机构合作团体:中国的中科院、国防科技大学、北京航空航天大学、重庆大学组成的机构合作团体,美国的伊利诺伊大

学、加州大学圣地亚哥分校、日本的国立原子物理学研究所、欧洲核子研究组织组成的机构合作团体,中国的上海交通大学、香港大学组成的机构合作团体,中国的清华大学、英国的剑桥大学组成的机构合作团体,智利的智利大学、西班牙的加泰罗尼亚理工大学组成的机构合作团体,美国的佐治亚理工学院、斯坦福大学组成的机构合作团体,美国的田纳西大学、南加州大学组成的机构合作团体等。

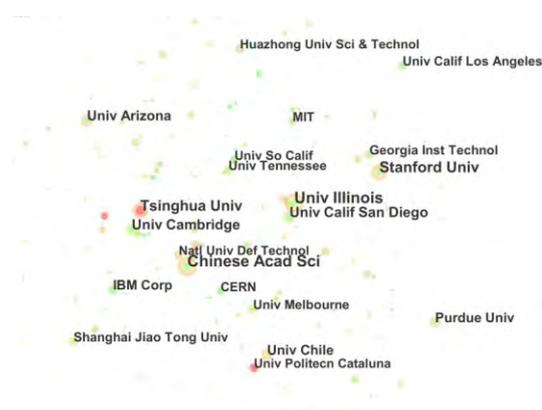


图4 机构知识图谱(CiteSpace)

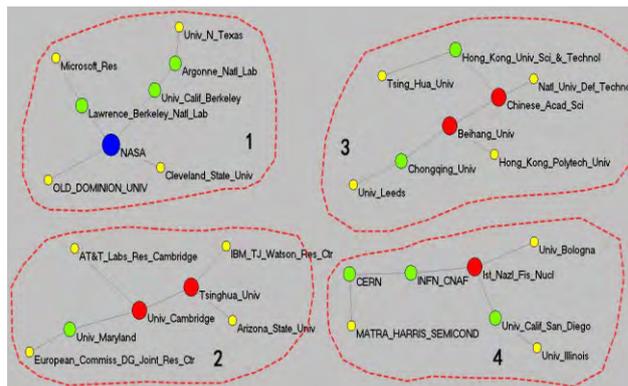


图5 机构合作知识图谱(Pajek)

从图5的研究机构合作来看,主要形成了4个合作群,分别为美国国家航空航天局、美国劳伦斯伯克利国家实验室、美国微软研究院、美国加州大学伯克利分校、美国阿贡国家实验室、美国北德克萨斯大学、美国克利夫兰州立大学、美国奥多明尼昂大学组成的合作群1,中国清华大学、英国剑桥大学、美国马里兰大学、美国TJ华盛顿研究中心、美国亚利桑那州立大学、英国剑桥大学AT&T实验室、欧盟DG联合研究中心组成的合作群2,中国中科院、中国北京航空航天大学、中国重庆大学、中国国防科技大学、英国利兹大学、中国香港科技大学、中国香港理工大学组成的合作群3,日本原子物理学研究所、美国加州大学圣地亚哥分校、美国伊利诺伊大学、意大利博洛尼亚大学、意大利国家核物理研究院、欧洲核子研究组织、法国马特拉哈里斯半导体公司组成的合作群4。可以看出图4和图5在机构合作群体上还是

具有较高的吻合性的。从结果可以看出,我国部分机构已经处于世界合作的核心位置,且基本形成了两大类:中国中科院、中国北京航空航天大学、中国重庆大学、中国国防科技大学、中国香港科技大学、中国香港理工大学等机构更多的是与国内研究机构进行合作交流,而清华大学则更多的是与英国剑桥大学、美国亚利桑那州立大学开展合作。未来,我国部分机构除了与国内研究机构开展合作之外,应该更广泛地与国外的研究机构开展合作,以提升我国核心机构在国际上的竞争力。

3.2.3 作者分析

图6表征了作者发文量及合作的知识图谱。从图6的作者发文量来看,智利大学计算机科学系的 Sergio F. Ochoa、加拿大不列颠哥伦比亚大学电气与计算机工程系的 Victor C.M.Leung 以及英国威尔士大学计算机科学系的 Min Chen 发文量均为8篇,接下来是智利大学经济与工商管理学院的 Gustavo Zurita、智利大学计算机科学系的 Nelson Baloian、美国伊利诺伊大学电子可视化实验室的 Jason Leigh 以及西班牙加泰罗尼亚理工大学的 Fatos Xhafa 发文量均为6篇。发文量大于等于5篇的14位高产作者中,从国家分布来看,美国和智利均为4人,加拿大2人,澳大利亚、韩国、西班牙、英国各有1人;从机构分布来看智利大学共有4人,伊利诺伊大学共有3人,威尔士大学、不列颠哥伦比亚大学、加泰罗尼亚理工大学、埃默里大学、浦项科技大学、墨尔本大学、圣弗朗西斯塞维尔大学各1人。从作者合作来看主要形成了两大作者合作团体,分别为加拿大不列颠哥伦比亚大学电气与计算机工程系的 Victor C.M. Leung、英国威尔士大学计算机科学系的 Min Chen、韩国庆熙大学计算机工程系的 EuiNam Huh、智利大学计算机科学系的 Mohammad Mehedi Hassan、澳大利亚斯温伯恩科技大学信息与通信技术学院的 JinJun Chen、美国加州大学圣地亚哥分校国家显微镜和影像学研究中心的 MH Ellisman 等6位研究者组成的作者



图6 作者知识图谱(CiteSpace)

合作团体,以及智利大学计算机科学系的 Sergio F. Ochoa、Jose A.Pino、西班牙加泰罗尼亚理工大学计算机建筑专业的 Felix Freitag 等3位研究者组成的作者合作团体。可以看出,我国在该领域的研究虽然发文总量居于前列,但缺乏高产作者,今后应培养国际范围内的核心研究者。

3.3 研究方向分析

为了分析国际协同计算领域的研究方向组成结构,本文在CiteSpace中进行研究方向(Category)分析,设置1990—2016年每2年1个时间片,数据选择标准为每个时间片取top50,得到图7、图8的研究方向及其演变知识图谱。



图7 研究方向知识图谱(CiteSpace)

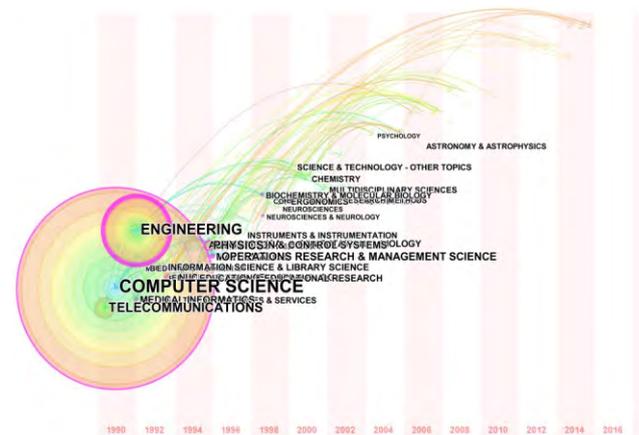


图8 研究方向演变知识图谱(CiteSpace)

从图7的研究方向发文量来看,计算机科学研究方向的发文量最多,为1346篇,约占发文总量的58.6%,可以看出协同计算领域具有计算机科学的本质特征;其次为工程学(462篇)、电信学(210篇)、物理学(104篇)、运筹学与管理学(65篇)、数学(56篇)、自动化与控制系统(51篇)、医学信息学(50篇)等。

为了探究研究方向的演变过程,本文将研究方向按照时区分析进行显示,可以明显看出各个阶段的研究具有不同的热点,如图8所示。可以看出计算机科学

(1990年)、工程学(1992年)、电信学(1990年)这几大主要的研究方向贯穿始终,组成了协同计算领域的主体,也说明了协同计算领域具有非常大比率的计算机科学等的学科属性。之后逐渐出现了医学信息学(1991年)、图情学(1992年)、教育学(1993年)、物理学(1994年)、自动控制系统(1994年)、运筹与管理学(1995年)、数学(1995年)等,之后又出现了化学(2000年)、交叉科学(2001年),近年来,随着协同计算学科交叉性的不断扩展,声学(2010年)、进化生物学(2011年)、护理学(2012年)、冶金工程学(2014年)、植物学(2014年)等学科也不断地与协同计算领域进行融合,进一步扩展了国际协同计算领域的研究与应用。

3.4 研究主流分析

为了分析国际协同计算领域已经形成的研究主流,本文在CiteSpace中分别进行作者共被引(Cited Author)、期刊共被引(Cited Journal)、文献共被引(Cited Reference)等分析,设置1990—2016年每2年1个时间片,数据选择标准为每个时间片取top50,得到图9~图11的研究主流知识图谱。

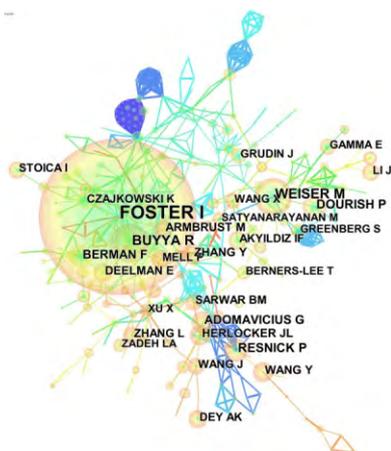


图9 作者被引知识图谱(CiteSpace)

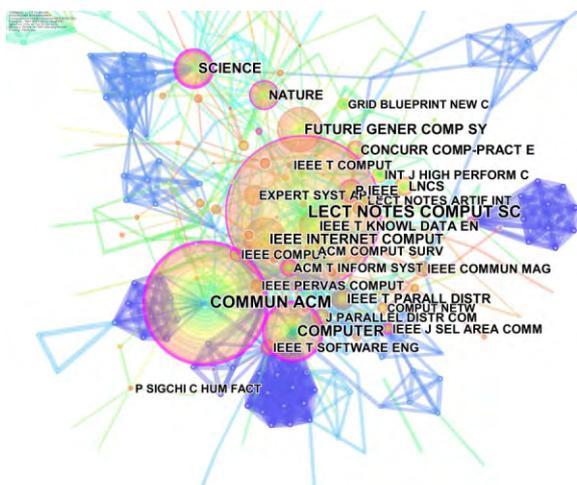


图10 期刊被引知识图谱(CiteSpace)

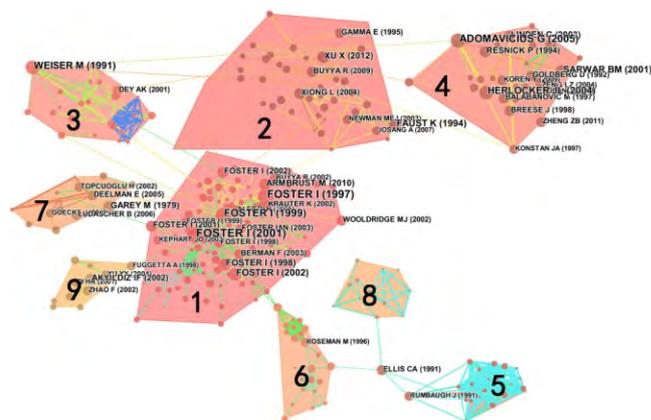


图11 文献共被引知识图谱(聚类)(CiteSpace)

3.4.1 主流作者分析

作者被引次数最多的为美国阿贡实验室数学与计算机部门的 Foster I,作为“网格计算”理论的创始人, Foster I被誉为“网格计算”之父,其被引次数达到了227次,远远超过其他被引作者,占有被引作者总和的6.48%。之后为墨尔本大学计算机科学与软件工程系网格计算与分布式系统实验室的 Buyya R(56次),以及施乐帕洛阿尔托研究中心的首席技术专家 Weiser M(54次)等。

3.4.2 主流期刊分析

从期刊被引次数来看,在计算机科学领域非常知名的计算机科学丛书(LNCS),被引次数最高,为579次。其次被引次数较高的期刊或文献依次为美国计算机协会通讯(Commun ACM)(361次),计算机(Computer)(191次),未来世代计算机系统(Future Gener Comp Sy)(162次),IEEE互联网计算机杂志(IEEE Internet Comput)(153次),科学(Science)(134次),IEEE知识与数据工程汇刊(IEEE T Knowl Data En)(112次),自然(Nature)(102次),工程技术-计算机:理论方法(Concurr Comp-Pract E)(101次)等。

3.4.3 主流分支领域分析

为了研究主流分支领域,本文将被引文献分别按照tf*idf、LLR和MI三种聚类算法进行分类,共得到了9类主要分支领域,按照施引文献的题名进行提取每一类的名称,得到了表1及图11,综合三种算法,最终得到如下结果。

第一类包含106篇文献,内部凝聚度为0.905,平均发表年为2001年,最终命名为“网格计算”,代表文献为Laszewski G于2002发表的“*grid computing: enabling a vision for collaborative research*”一文。在文中,提出了一种能够使协同研究环境成为可能的视角下的网格计算想法,认为通过网格的使用,能够产生协同研究新的模式,同时列出了网格研究中的难题以及待解决的

表1 文献共被引聚类名称

序号	文献量	凝聚度	平均年份	tf*idf	LLR	MI
1	106	0.905	2001	grid	grid computing	broad-network concurrent computing
2	42	0.938	2005	ftm	ftm	collaborative learning
3	31	0.990	1995	representing ubiquitous audio	representing ubiquitous audio	reasoning
4	30	0.983	2001	technique	web computing	workflow management
5	24	0.958	1991	parsi-model approach	parsi-model approach	strategic analysis
6	23	0.986	1996	activity-based support	activity-based support	broad-network concurrent computing
7	17	0.989	2003	grid	iplant	workflow application
8	12	0.988	1994	autonomous healthcare information system	autonomous healthcare information system	strategic analysis
9	11	0.991	2002	mobile agent	wireless sensor network	experience

问题^[21]。

第二类包含42篇文献,内部凝聚度为0.938,平均发表年为2005年,最终命名为“协同计算信任模型”,代表文献为Ahamed S I于2010年发表的“*a formal context specific trust model (FTM) for multimedia and ubiquitous computing environment*”一文,认为在多媒体和普适计算环境下手持设备的自组织网络中资源共享的安全性问题中相互协作非常关键,存储和计算能力的限制问题已经成为资源有效共享的瓶颈。由此,首次提出了针对多媒体和普适计算的信任模型,该模型采用了多跳推荐能力和灵活的行为模型来处理设备之间的交互,其主要作用为避免恶意推荐并进行相应处理^[22]。

第三类包含31篇文献,内部凝聚度为0.990,平均发表年为1995年,最终命名为“音频分析”,代表文献为HIndus D于1993发表的“*capturing, structuring, and representing ubiquitous audio*”一文,认为虽然会话是协作的一个重要组成部分,但很少有计算机支持获取和访问会话内容,其对普适音频进行了研究,认为当时的语音识别技术还不能较好地捕获会话内容,对如何从办公室讨论及电话中进行音频的捕获及其检索进行了研究,同时认为该研究可以进行更广泛的应用,例如桌面音频、移动音频^[23]。

第四类包含30篇文献,内部凝聚度为0.983,平均发表年为2001年,最终命名为“网络计算”,代表文献为Gao Min于2010发表的“*personalization in Web computing and informatics: theories, techniques, applications, and future research*”一文,认为个性化的搜索引擎和推荐系统已被广泛采用,对个性化系统领域进行了回顾,包括用户分析、内容建模、用于推荐的过滤方法等,同时还讨论了不同领域下的个性化系统、技术及其局限性^[24]。

第五类包含24篇文献,内部凝聚度为0.958,平均发表年为1991年,最终命名为“parsi模型”,代表文献为Maxwell T于1999年发表的“*a parsi-model approach*

to modular simulation”一文,认为利用可重用模型组件库可以大大促进复杂模型的开发,在文中描述了一个面向对象的模块规范(MSF),用于在连续空间建模及支持存档模块的实现,MSF可以分级连接创建一个简约的parsi模型,该模型能够为计算机仿真提供服务,例如图形模块的开发和配置,模拟方程的自动微分,模块中的透明计算,完全可配置的时间-空间表示,作者认为将parsi模型引入协同仿真领域具有很大的发展潜力^[25]。

第六类包含23篇文献,内部凝聚度为0.986,平均发表年为1996年,最终命名为“基于作业的协助分析”,代表文献为Bardram J E于2005年发表的“*activity-based support for mobility and collaboration in ubiquitous computing*”一文,提出了一种作业成本计算的设计理念,目的是:(1)通过管理计算机的工作任务集合支持人类作业;(2)支持异构计算环境下的可移植性;(3)支持异步协作,允许多人参加同一作业;(4)支持同步,实时协作,通过“桌面会议”使多个客户端共享作业^[26]。

第七类包含17篇文献,内部凝聚度为0.989,平均发表年为2003年,最终命名为“iplant系统”,代表文献为Lushbough C M于2015年发表的“*life science data analysis workflow development using the bioextract server leveraging the iplant collaborative cyberinfrastructure*”一文,为了处理由高通量实验技术所产生的大量生物数据,设计了BioExtract服务器利用iPlant的协同功能来解决生物信息学领域大数据存储和分析的问题。BioExtract服务器是基于Web的,能够为研究者提供一个用于分析基因组数据的环境^[27]。

第八类包含12篇文献,内部凝聚度为0.988,平均发表年为1994年,最终命名为“自主医疗信息系统”,代表文献为Leisch E于1997年发表的“*a framework for the integration of distributed autonomous healthcare information systems*”一文,基于新推出的病人元记录

(PMR)的概念,同时利用分布式面向对象计算和信息演示的标准,提出了一种分布式的自主医疗信息系统的框架,该框架能够将信息资源的访问与病人记录的内容和结构进行整合^[28]。

第九类包含 11 篇文献,内部凝聚度为 0.991,平均发表年为 2002 年,最终命名为“无线传感器网络”,代表文献为 Chen Min 于 2011 年发表的“*itinerary planning for energy-efficient agent communications in wireless sensor networks*”一文,与传统的无线传感器网络相比,基于移动 Agent 的无线传感器网络可以更加方便、高效节能地进行数据收集与汇总,作者同时考虑了数据聚集和能量效率的影响,提出了首源行程能量最小化(IEMF)算法,然后对 IEMF 算法进行迭代,得到了行程能量最小化算法(IEMA),进一步提出了多 Agent 行程规划(MIP)解决方案的通用框架,结果表明 MIP 与现有的单 Agent 行程规划(SIP)算法相比,其能量效率更高,延迟更低^[29]。

从图 11 可以看到,第一类为整个协同计算领域中最庞大的聚类,说明网络计算为协同计算最为核心的分支领域,其次为协同计算信任模型、音频分析、网络计算、Parsi 模型、基于作业的协助分析等分支领域,其中 iplant 系统、自主医疗信息系统、无线传感器网络这三类分支领域规模最小,属于研究中的边缘分支领域。

3.5 发展演变过程分析

为了分析国际协同计算领域的发展演变过程,本文在 CiteSpace 中进行文献被引(Cited Reference)分析,设置 1990—2016 年每 2 年 1 个时间片,数据选择标准为每个时间片取 top50,然后选择其中被引次数大于等于 10 次的 38 篇文献作为样本进行演变过程的分析,将这 38 篇文献按照发表年份进行排列得到了表 2,由此可以窥探国际协同计算领域的发展演变过程,得到图 12 的发展演变过程知识图谱。

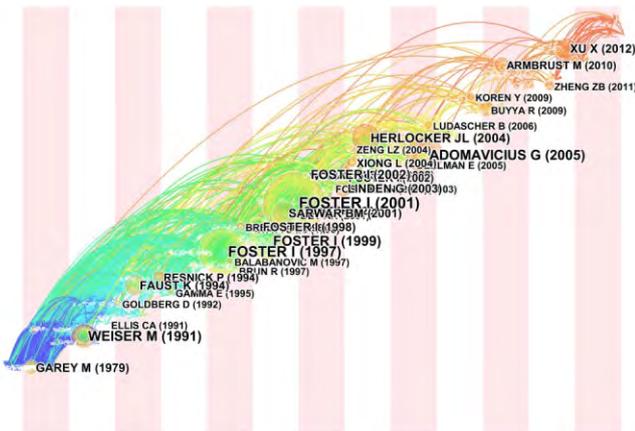


图 12 发展演变过程知识图谱 (CiteSpace)

在这 38 篇文献中,最早的为 1979 年的 Garey M 提出的 NP 完全性理论,该文献为协同计算理论的提出起到了奠基性作用^[30]。1991 年至 1997 年出现了主动定位系统(Active Badge Location System)^[31]、为构建网络应用提供中间件服务和程序库的 Globus 项目^[32]、用于大数据分析的面向对象数据分析框架 Root^[33]。1998 年至 2005 年,出现了网络计算的研究,Foster I 分别就网络计算的设计蓝图^[34-35]、网络计算的安全性^[36]进行了讨论。2006 年,云计算的出现影响了众多产业,使软件更具吸引力,使开发者在部署自己的服务或人力费用时不再需要大量的资本支出^[37-38]。由此可以看出,国际协同计算领域的发展基本上是沿着协同计算早期相关研究—分布式计算—网络计算—云计算,同时穿插协同过滤这样一条路径进行演变的。

3.6 研究热点及前沿分析

为了分析国际协同计算领域的研究热点及前沿,本文在 CiteSpace 中进行摘要(Abstract)分析,设置 1990—2016 年每 2 年 1 个时间片,数据选择标准为每个时间片取 top50,通过 tf-idf 算法得到图 13 的摘要知识图谱。表 3 列出了频次前 20 位的关键词;同时为了更精细化地探究热点,在 Pajek 中对摘要中的高频关键词进行组元分析,共得到 16 个组元,见图 14~图 16 所示。

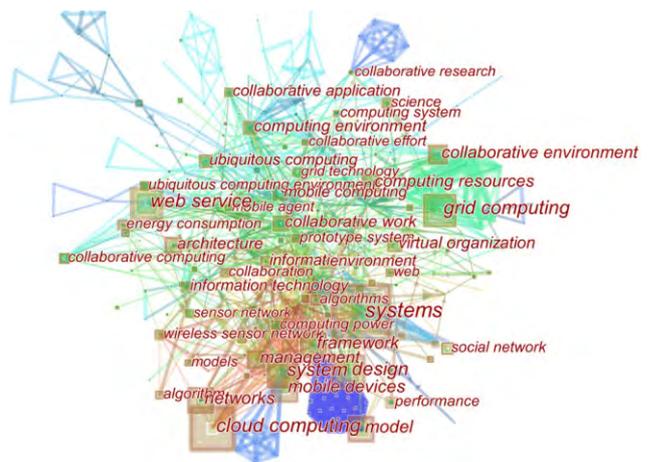


图 13 摘要知识图谱 (CiteSpace)

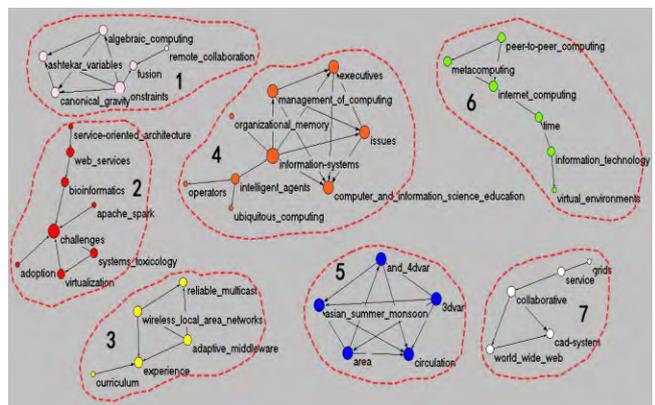


图 14 1~7 号组元分析知识图谱 (Pajek)

表2 高频被引关键文献(被引频次≥10)

频次	作者	年份	类型	英文题名
16	Garey M	1979	图书	<i>Computers and Intractability: A Guide to the Theory of NP-Completeness</i>
27	Weiser M	1991	论文	<i>The computer for the twenty-first century</i>
10	Ellis CA	1991	论文	<i>Groupware-Some Issues and Experiences</i>
11	Goldberg D	1992	论文	<i>Using collaborative filtering to weave an information tapestry</i>
19	Faust K	1994	图书	<i>Social Network Analysis:Methods and Applications</i>
14	Resnick P	1994	论文	<i>GroupLens:an open architecture for collaborative filtering of netnews</i>
11	Gamma E	1995	图书	<i>Design Patterns-Elements of Reusable Object-Oriented Software</i>
39	Foster I	1997	论文	<i>Globus:A metacomputing infrastructure toolkit</i>
12	Balabanovic M	1997	论文	<i>Fab:Content-based, collaborative recommendation</i>
10	Brun R	1997	论文	<i>Root - an object oriented data analysis framework</i>
18	Foster I	1998	图书	<i>The Grid:Blueprint for a new computing infrastructure</i>
12	Breese J	1998	论文	<i>Empirical analysis of predictive algorithms for collaborative filtering</i>
11	Watts DJ	1998	论文	<i>Collective dynamics of small-world networks</i>
10	Foster I	1998	论文	<i>A security architecture for computational grids</i>
31	Foster I	1999	图书	<i>The Grid:Blueprint for a New Computing Infrastructure</i>
47	Foster I	2001	论文	<i>The anatomy of the grid:Enabling scalable virtual organizations</i>
20	Sarwar BM	2001	论文	<i>Item based collaborative filtering recommendation algorithms</i>
14	Foster I	2001	论文	<i>The Anatomy of the Grid:Enabling Scalable Virtual Organizations</i>
10	Dey AK	2001	论文	<i>A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware plications</i>
21	Foster I	2002	图书	<i>The Physiology of the Grid:An Open Grid Services Architecture for Distributed Systems Integration</i>
14	Akyildiz IF	2002	论文	<i>Wireless sensor networks:a survey</i>
14	Foster I	2002	论文	<i>Grid services for distributed system integration</i>
10	Krauter K	2002	论文	<i>A taxonomy and survey of grid resource management systems for distributed computing</i>
10	Wooldridge MJ	2002	图书	<i>An Introduction to MultiAgent System</i>
16	Linden G	2003	论文	<i>Amazon.com recommendations:item-to-item collaborative filtering</i>
11	Berman F	2003	图书	<i>Grid Computing:Making the Global Infrastructure a Reality</i>
11	Foster Ian	2003	图书	<i>The Grid 2:Blueprint for a new computing infrastructure</i>
25	Herlocker JL	2004	论文	<i>Evaluating collaborative filtering recommender systems</i>
13	Xiong L	2004	论文	<i>PeerTrust:Supporting reputation-based trust for peer-to-peer electronic communities</i>
10	Zeng LZ	2004	论文	<i>QoS-aware middleware for web services composition</i>
30	Adomavicius G	2005	论文	<i>Toward the next generation of recommender systems:A survey of the state-of-the-art and possible xtensions</i>
11	Deelman E	2005	论文	<i>Pegasus:a framework for mapping complex scientific workflows onto distributed systems</i>
10	Ludascher B	2006	论文	<i>Scientific workflow management and the Kepler system</i>
12	Buyya R	2009	论文	<i>Cloud computing and emerging IT platforms:Vision, hype, and reality for delivering computing as the 5th utility</i>
10	Koren Y	2009	论文	<i>Matrix factorization techniques for recommender systems</i>
15	Armbrust M	2010	论文	<i>A view of cloud computing</i>
10	Zheng ZB	2011	论文	<i>QoS-Aware Web Service Recommendation by Collaborative Filtering</i>
16	Xu X	2012	论文	<i>From cloud computing to cloud manufacturing</i>

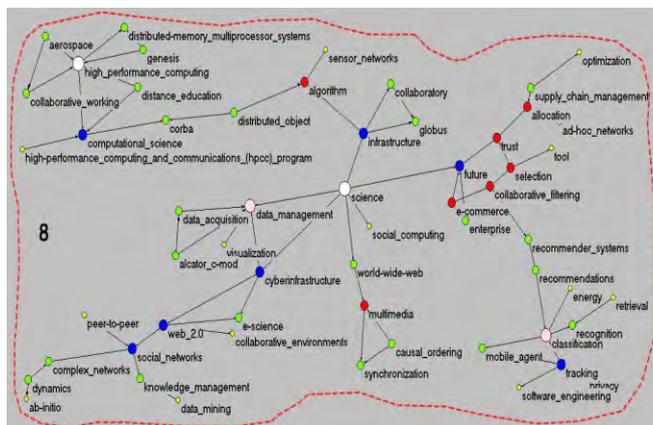


图15 8号组元分析知识图谱(Pajek)

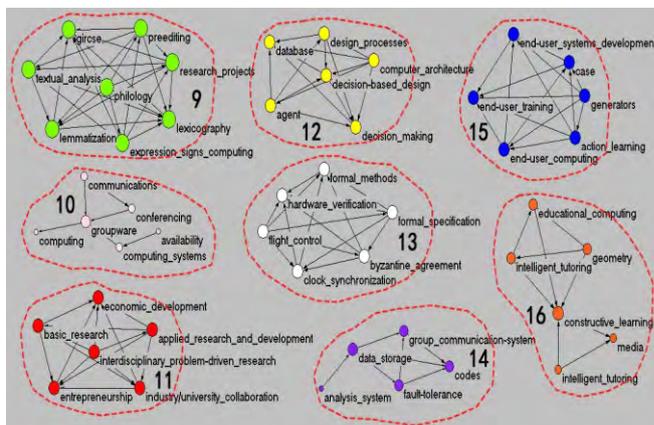


图16 9~16号组元分析知识图谱(Pajek)

表3 高频关键词(前20个)

序号	频次	突现值	关键词	序号	频次	突现值	关键词
1	120	—	systems	11	43	—	collaborative environment
2	83	17.48	cloud computing	12	42	—	management
3	67	14.76	grid computing	13	41	—	computing resources
4	66	—	design	14	39	4.98	computing environment
5	65	—	system	15	38	—	architecture
6	65	3.84	Web service	16	36	—	collaborative work
7	54	—	framework	17	34	4.49	virtual organization
8	53	—	model	18	33	—	information technology
9	53	4.05	networks	19	32	—	mobile computing
10	52	—	mobile devices	20	32	4.47	ubiquitous computing

从图13和表3可以看出,排名前几位的关键词中的Systems、Cloud Computing、Grid Computing、Distributed Computing等核心词汇代表了协同计算领域中典型的模式;System、Network、Framework、Algorithm、Performance、Architecture等词汇则表征了协同计算研究的重要研究层面。

除了Cloud Computing、Grid Computing、Distributed Computing等几个核心词汇外,根据tf-idf算法也探究到了其他几个关键词,如Florida Everglades,即为了支持高性能协同建模而建立的一种开放性的地理建模环境^[39],Threshold Current是在光电集成系统中引入协同计算形成的技术^[40],Treatment Plans则是将协同理念引入医学领域研发的用于三维放射治疗计划的便携式软件工具^[41]。

关键词的突现值可能代表了该领域未来发展的前沿,关键词突现值最高的几个词分别为Cloud Computing(17.48)、Grid Computing(14.76)、Wide Web(7.20)、Computing Environment(4.98)、Sensor Network(4.96)、Grid Service(4.84)、Grid Technology(4.79)、Mobile Agent(4.54)、Virtual Organization(4.49)、Ubiquitous Computing(4.47)、Grid Application(4.42)、High Performance(4.41)等,需要注意的是突现值大的词可能其频次较小,因而在表3中未显示。从突现值的大小来看,云计算和网格计算的突现值最大,代表了未来协同计算发展的两大主流方向,而万维网、传感器网络、网格技术、移动Agent等既是协同计算发展必不可少的实体要素,同时也是该领域中重要的应用技术。

图14~图16中的16个由摘要中高频关键词组成的组元,可以更加精细化地表征当今协同计算领域的研究热点。下面对主要的组元中包含的核心关键词及其代表文献进行分析。

2号组元中的主要关键词包括Bioinformatics、Systems Toxicology等,表征了协同计算应用于生物学、病理学等领域形成的研究热点,如基于工作流的可扩展生物信息学研究^[42]、协同合作的人体毒素基因组Collaboratorium

系统环境的研究^[43]。

4号组元中的核心关键词为Executives、Information Systems,该部分主要为协同计算在企业信息系统中的应用形成的研究热点,如企业网络中基于服务的高级协作研究^[44]。

5号组元中的核心关键词为3dvar、4dvar、Circulation、Asian Summer Monsoon,代表了协同计算应用于气候学领域形成的热点,如台湾天气与气候研究:全球定位系统的应用^[45]。

7号组元中的核心关键词为CAD System,代表了协同计算应用于计算机辅助设计系统形成的热点,如基于组件Agent的协同设计开放式计算机辅助设计系统^[46]。

8号组元为最大的一个组元,以Science为中心,又形成了4个小研究热点,分别为:(1)High Performance Computing关键词为代表的高性能研究,如高性能计算的云计算解决方案^[47];(2)Social Networks、Complex Networks代表的社会网络领域中的协同计算的应用热点,如加强医疗照护环境中的知识流动:一种移动计算的方法^[48];(3)Multimedia代表的多媒体领域中的系统计算的应用热点,如协同无线网络编码的数字视频流研究^[49];(4)Collaborative Filtering、Recommender Systems代表的推荐系统研究热点,如一种保证隐私保护的商务推荐系统^[50]。

9号组元中的核心关键词为Textual Analysis、Expression Signs Computing、Lemmatization、Philology等,表征了协同计算应用于文本分析中形成的研究热点,如用于表达式符号计算的跨学科组研究^[51]。

11号组元中的核心关键词为Industry/University Collaboration,表征了协同计算在产学研中的应用形成的研究热点,如一个三管齐下的战略技术的创造、传递和吸收^[52]。

15号组元中的核心关键词为End-user Systems Development、End-user Computing、End-user Training,表征了终端用户的研究形成的研究热点,如协同应用环境下培训实践的研究^[53]。

16号组元中的核心关键词为 Intelligent Tutoring、Constructive Learning、Educational Computing,代表了协同计算应用于教育领域形成的研究热点,如智能三维协同虚拟学习环境框架的研究^[54]。

除此之外,1号组元中的 Remote Collaboration、3号组元中的 Reliable Multicast、6号组元中的 Meta Computing、10号组元中的 Conferencing、12号组元中的 Decision Based Design、13号组元中的 Clock Synchronization、14号组元中的 Fault Tolerance等关键词也值得关注。

4 结论

(1)从发文量时间分布来看,国际协同计算领域的研究正处于上升期。

(2)从研究力量发文量来看,国际上国家、地区以美国、中国和欧洲等为主,研究机构以美国和中国为主,虽然我国已经具备了一定的科研实力,但缺乏该领域的高产研究者,总体来看欧洲各国之间合作最为密切;研究方向构成上以计算机科学为主,同时工程学、电信学、物理学等也占有一定的比例。

(3)从业已形成的研究主流来看,Foster I为最具影响力的主流研究者;该领域中的主流期刊(文献)包括 LNCS、Commun ACM等;主流分支领域包括网格计算、协同计算信任模型、音频分析、网络计算、parsi模型、基于作业的协助分析、iplant系统、自主医疗信息系统、无线传感器网络9部分;发展演变过程为早期相关研究—分布式计算—网格计算—云计算,同时穿插协同过滤这样一条发展演变路径。

(4)从研究热点来看,云计算、网格计算、分布式计算、移动计算、普适计算等关键词代表了当前该领域的研究热点;网格计算、云计算、分布式计算、网络服务、协同过滤等关键词则代表了该领域的发展前沿,这些前沿词汇代表了该领域未来可能的发展方向。协同计算不仅仅应用于传统的计算机辅助设计、推荐系统、终端用户等方面的研究,同时也广泛渗透、融合到其他领域,如生物学、病理学、气候学、教育学、企业信息系统、社会网络、多媒体、文本分析、产学研等领域,体现了协同计算这一领域较强的跨学科特性。总之协同计算自身既具有较强的应用属性,也具有非常强的跨学科特性,预计随着协同计算领域的不断发展,将会应用到更多领域。

5 不足及今后的研究方向

本文通过知识图谱软件对国际协同计算领域进行了可视化分析,得出了相应的结论,但还存在如下不足:(1)本文采用了以 CiteSpace 为主的可视化软件,受制于该软件自身的特点,所得到的结果具有一定的局限性。(2)本文从发文量时间分布、研究力量分布及合作、研究

方向、研究主流、发展演变过程、研究热点及前沿等角度进行了分析,这些角度并未涵盖该领域的所有方面。今后,应针对该领域从更多视角进行更全面的分析。

参考文献:

- [1] 刘泽渊,陈悦.科学知识图谱方法与应用[M].北京:人民出版社,2008:1-39.
- [2] Tas Z.Collaborative interdisciplinary astrobiology research: A bibliometric study of the NASA Astrobiology Institute[J]. Scientometrics,2015,103(3):1003-1022.
- [3] Qian G, Fong M.Scientometrics analysis on the intellectual structure of the research field of bioenergy[J].Journal of Biobased Materials & Bioenergy,2013,7(2):305-308.
- [4] Liu C, Gui Q.Mapping intellectual structures and dynamics of transport geography research: A scientometric overview from 1982 to 2014[J].Scientometrics,2016,109(1):159-184.
- [5] Zhao R, Wang J.Visualizing the research on pervasive and ubiquitous computing[J].Scientometrics,2011,86(3):593-612.
- [6] Pu B, Qiu Y.A bibliometric analysis on urbanization research from 1984 to 2013[J].Open House International,2015,40(3):37-43.
- [7] Chen D, Liu Z, Luo Z, et al.Bibliometric and visualized analysis of emergy research[J].Ecological Engineering,2016,90:285-293.
- [8] Niazi M, Hussain A.Agent-based computing from multi-agent systems to agent-based models:A visual survey[J]. Scientometrics,2011,89(2):479-499.
- [9] Zhang X, Gao Y, Yan X.From e-learning to social-learning: Mapping development of studies on social media-supported knowledge management[J].Computers in Human Behavior,2015,51:803-811.
- [10] Chen C, Song I Y, Yuan X, et al.The thematic and citation landscape of Data and Knowledge Engineering (1985—2007)[J].Data & Knowledge Engineering,2008,67(2):234-259.
- [11] Chen K, Guan J.A bibliometric investigation of research performance in emerging nanobiopharmaceuticals[J]. Journal of Informetrics,2011,5(2):233-247.
- [12] Kim M C, Chen C.A scientometric review of emerging trends and new developments in recommendation systems[J].Scientometrics,2015,104(1):239-263.
- [13] Wu Y, Duan Z.Analysis on evolution and research focus in psychiatry field[J].BMC Psychiatry,2015,15(1):1-15.
- [14] Gao Y, Qu B, Shen Y.Bibliometric profile of neurogenic bladder in the literature:A 20-year bibliometric analysis [J].Neural Regeneration Research,2015,10(5):797-803.
- [15] Yin Z, Chen D, Li B.Global regulatory T-cell research

- from 2000 to 2015: A bibliometric analysis[J]. Plos One, 2016, 11(9):1-19.
- [16] 卫军朝, 蔚海燕. 基于 CiteSpace II 的数字图书馆研究热点分析[J]. 图书馆杂志, 2011(4):70-77.
- [17] 王娟, 陈世超, 王林丽, 等. 基于 CiteSpace 的教育大数据研究热点与趋势分析[J]. 现代教育技术, 2016, 26(2):5-13.
- [18] 高明, 段卉, 韩尚洁. 基于 CiteSpace III 的国外体育教育研究计量学分析[J]. 体育科学, 2015(1):4-12.
- [19] 衣春波, 许鑫. 自由贸易区领域研究热点与前沿探讨——基于 Citespace II 的计量分析[J]. 上海经济研究, 2014(3):67-78.
- [20] 徐锦芬, 聂睿. 基于 CiteSpace 的国际二语写作研究动态可视化分析(2004—2014)[J]. 外语电化教学, 2015(4):3-9.
- [21] Von Laszewski G. Grid computing: Enabling a vision for collaborative research[M]//Applied Parallel Computing. Berlin Heidelberg: Springer-Verlag, 2002:37-52.
- [22] Ahamed S I. A Formal context specific Trust Model (FTM) for multimedia and ubiquitous computing environment[J]. Telecommunications, 2010(3/4):221-240.
- [23] HIndus D. Capturing, structuring, and representing ubiquitous audio[J]. ACM Transactions on Information Systems, 1993(4):376-400.
- [24] Gao Min. Personalisation in Web computing and informatics: Theories, techniques, applications, and future research[J]. Information Systems Frontiers, 2010(5):607-629.
- [25] Maxwell T. A parsimonious approach to modular simulation[J]. Environmental Modelling & Software, 1999(6):511-517.
- [26] Bardram J E. Activity-based support for mobility and collaboration in ubiquitous computing[J]. Personal and Ubiquitous Computing, 2005(5):312-322.
- [27] Lushbough C M. Life science data analysis workflow development using the bioextract server leveraging the iplant collaborative cyberinfrastructure[J]. Concurrency and Computation-Practice & Experience, 2015(2):408-419.
- [28] Leisch E. A framework for the integration of distributed autonomous healthcare information systems[J]. Medical Informatics, 1997(4):325-335.
- [29] Chen Min. Itinerary planning for energy-efficient agent communications in wireless sensor networks[J]. IEEE Transactions on Vehicular Technology, 2011(7):3290-3299.
- [30] Garey M. Computers and intractability: A guide to the theory of NP-completeness[M]. New York: Freeman, 1979.
- [31] Want R. The active badge location system[J]. ACM Transactions on Information Systems, 1992(1):91-102.
- [32] Foster I. Globus: A metacomputing infrastructure toolkit[J]. International Journal of Super Computer Applications and High Performance Computing, 1997(2):115-128.
- [33] Brun R. ROOT—An object oriented data analysis framework[J]. Nuclear Instruments & Methods in Physics Research Section A: Accelerators Spectrometers Detectors and Associated Equipment, 1997(1/2):81-86.
- [34] Foster I. The grid: Blueprint for a new computing infrastructure[M]. [S.l.]: Morgan Kaufmann, 1998.
- [35] Foster I. The grid 2: Blueprint for a new computing infrastructure[M]. New York: Elsevier Science, 2003.
- [36] Foster I. A security architecture for computational grids[C]//Proceedings of the 5th ACM Conference on Computer and Communications Security (CCS'98), 1998:83-92.
- [37] Michael A. A view of cloud computing[J]. Communications of the ACM, 2010(4):50-58.
- [38] Xu X. From cloud computing to cloud manufacturing[J]. Robotics and Computer-Integrated Manufacturing, 2012(1):75-86.
- [39] Maxwell T, Costanza R. An open geographic modeling environment[J]. Simulation Transactions on the Society for Modeling & Simulation International, 1997, 68(3):175-185.
- [40] Hayashi I. Photo-electronic integrated systems: The basic concept of U-OEICs and their feasibility[J]. Optoelectronics—Devices and Technologies, 1994, 9:1-14.
- [41] Jacky J. Portable software tools for 3D radiation therapy planning[J]. International Journal of Radiation Oncology-biologyphysics, 1994, 30(4):921-928.
- [42] Garza L D L, Veit J, Szolek A. From the desktop to the grid: Scalable bioinformatics via workflow conversion[J]. BMC Bioinformatics, 2016, 17(1):1-12.
- [43] Fasani R A, Livi C B, Choudhury D R. The human toxome collaboratorium: A shared environment for multi-omic computational collaboration within a consortium[J]. Frontiers in Pharmacology, 2015, 6(R86).
- [44] Coutinho C, Cretan A, Silva C F. Service-based negotiation for advanced collaboration in enterprise networks[J]. Journal of Intelligent Manufacturing, 2016, 27(1):201-216.
- [45] Wu C C, Kuo H C, Hsu H H, et al. Weather and climate research in Taiwan: Potential application of GPS/MET data[J]. Terrestrial, 2000, 11(1):211-234.
- [46] Rosenman M, Wang F. A component agent based open CAD system for collaborative design[J]. Automation in Construction, 2001, 10(4):383-397.
- [47] Tomic D, Ogrizovic D, Car Z. Cloud solutions for high performance computing: Oxymoron or realm?[J]. Tehnicki Vjesnik, 2013, 20:177-182.

(下转 53 页)

- 似[J].计算机工程与应用,2009,45(4):165-168.
- [8] 张文修,吴伟志,梁吉业,等.粗糙集理论与方法[M].北京:科学出版社,2001.
- [9] 贾修一,商琳,周献中,等.三支决策理论与应用[M].南京:南京大学出版社,2012.
- [10] 张清华,周玉兰,滕海涛.基于粒计算的认知模型[J].重庆邮电大学学报,2009,21(4):494-501.
- [11] 张文修,吴伟志.粗糙集理论介绍和研究综述[J].模糊系统与数学,2000,14(4):1-10.
- [12] 张文修,梁怡,吴伟志.信息系统与知识发现[M].北京:科学出版社,2003.
- [13] Yao Y Y.An outline of a theory of three-way decisions[C]//Proceedings of the 8th International RSCTC Conference, 2012.
- [14] Yao Y Y.Three-way decisions: An interpretation of rules in rough set theory[C]//Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology, 2009.
- [15] Yao Y Y.Three-way decisions with probabilistic rough sets[J].Information Sciences,2010,180(3):341-353.
- [16] 顾雪峰,杨尔弘,刘杰.动态粒度在实体关系识别的应用[J].山西大学学报,2005,28(A):39-40.
- [17] Liang J Y,Qian Y H,Chu C Y,et al.Rough set approximation based on dynamic granulation[J].Lecture Notes in Artificial Intelligence,2005,4062:701-708.
- [18] 贾修一,李伟,商琳,等.一种自适应三支决策中决策阈值的算法[J].电子学报,2011,39(11):2520-2525.
- [19] Li J H,Mei C L,Xu W H,et al.Concept learning via granular computing: A cognitive viewpoint[J].Information Sciences,2015,298:447-467.

(上接29页)

- [21] Chang L B,Hu B Q,Li K L.Huber's M-estimation-based process uncertainty robust filter for integrated INS/GPS[J].IEEE Sensors Journal,2015,37(2):3367-3374.
- [22] Genz A.Fully symmetric interpolatory rules for multiple integrals over hyper-spherical surfaces[J].Journal of Computational and Applied Mathematics,2003,157(1):187-195.
- [23] Genz A,Monahan J A.Stochastic algorithm for high dimensional integrals over unbounded regions with Gaussian weight[J].Journal of Computational and Applied Mathematics,1999,112(1/2):71-81.
- [24] Tao Sun,Ming Xin.Hypersonic entry vehicle state estimation using high-degree cubature Kalman filter[C]//Proceedings of AIAA Atmospheric Flight Mechanics Conference,2014:2383-2394.
- [25] Jia B,Xin M,Cheng Y.The high-degree cubature Kalman filter[C]//Proceedings of the 51st IEEE Conference on Decision and Control,2012:4095-4100.
- [26] Aravkin Y,Bell B M,Burke J V,et al.An l_1 Laplace robust Kalman smoother[J].IEEE Transactions on Automatic Control,2011,56(12):2898-2911.
- [27] Masreliez J,Martin R D.Robust Bayesian estimation for linear model and robustifying the Kalman filter[J].IEEE Transactions on Automatic Control,1977,22(3):361-371.

(上接40页)

- [48] Oliveira J,Souza D S,de Lima P Z,et al.Enhancing knowledge flow in a health care context: A mobile computing approach[J].Jmir Mhealth & Uhealth,2014,2(4):e17.
- [49] Zhang B,Liu Z,Chan S G,et al.Collaborative wireless freeview video streaming with network coding[J].IEEE Transactions on Multimedia,2016,18(3):521-536.
- [50] Lu Z,Shen H.An accuracy-assured privacy-preserving recommender system for Internet commerce[J].Computer Science & Information Systems,2015,12(4):1307-1326.
- [51] Banca P,Pontani P,Silva M.The interdisciplinary group for expression signs computing[J].Language Resources and Evaluation,1990,24(5):381-392.
- [52] Adrion W R,Mcowen P.A three-pronged strategy for technology creation, transfer and absorption[C]//Proceedings of Conference on Diffusion, Transfer & Implementation of Information Technology,1993:309-320.
- [53] Kang D,Santhanam R.A longitudinal field study of training practices in a collaborative application environment[J].Journal of Management Information Systems,2004,20(3):257-281.
- [54] Laffey J,Schmidt M,Galyen K,et al.Smart 3D collaborative virtual learning environments: A preliminary framework[J].Journal of Ambient Intelligence & Smart Environments,2012,4(1):49-66.