

基于科学计量的世界人工智能领域发展状况分析

李悦 苏成 贾佳 许震 田瑞强

(中国科学技术信息研究所 北京 100038)

摘 要 分析了人工智能领域近15年的发展状况,预测了未来的发展趋势,帮助研究人员快速掌握领域概况。运用SciMAT软件进行关键词共现分析,利用生成的主题演化图、战略图揭示发展状况及子领域成熟度,预测未来发展趋势。2002—2016年,人工智能领域发文量及关键词总数总体呈上升趋势,说明该领域发展态势良好。随时间推移,各阶段类团数增多,说明该领域多方面发展。神经网络、智能机器人一直是人工智能领域研究的热点,且研究规模随着时间的推移不断扩大并逐步走向成熟。人工智能正发生从理论到应用的转变;神经网络、智能机器人将会是未来人工智能领域发展的热门。

关键词 人工智能, SciMAT, 共词分析, 主题演进

中图分类号 G301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.12.034

Analysis of Development Status of World Artificial Intelligence Based on Scientific Measurement

LI Yue SU Cheng JIA Jia XU Zhen TIAN Rui-qiang

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract This paper analyzed the field of artificial intelligence in the past 15 years to predict the future development trend, and to help researchers quickly grasp the field profile. SciMAT software was used to carry out the co-occurrence analysis of the keywords, and the development trend and sub-domain maturity were predicted by the generated theme evolution chart, and the strategy map was used to predict the future development trend. The total number of advertisements and the total number of keywords in the field of artificial intelligence are on the rise, indicating that the development of the field is good. The increase in the number of groups at various stages indicates that the development of the field in many ways. In the past 15 years, neural networks and intelligent robots have been hot topics in the field of artificial intelligence, and as time goes on, the scale of the research expand, has been also gradually mature. Artificial intelligence is taking place from theory to application of the change. Neural network and intelligent robot will be the future development of artificial intelligence field.

Keywords Artificial intelligence, SciMAT, Co-word analysis, Theme evolution

2017年,“人工智能”首次被写入政府工作报告:“全面实施战略性新兴产业发展规划,加快新材料、新能源、人工智能、集成电路、生物制药、第五代移动通信等技术研发和转化,做大做强产业集群”^[1]。近年来,人工智能领域飞速发展,百度、腾讯等知名公司纷纷组建人工智能相关实验室。AlphaGo横扫围棋界标志着人工智能又征服了一个新的高峰。许多学者尝试总结和分析该领域的发展状况。胡玉宁等^[2]于2013年运用TDA软件对人工智能领域进行了文献统计分析 & 关键词共现分析,发现当时人工智能研究热点为遗传算法、神经网络和机器学习。曾毅等^[3]于2016年提出未来能为揭示人工智能本质提供更多线索的学科是脑与神经科学、认知科学。

同年,顾险峰^[4]认为,增强学习是当时人工智能研究的热点。对人工智能领域的定量分析,可以系统全面地解释该领域的发展状况,探究各分支间的发展演化关系,有助于研究者对人工智能相关知识进行梳理整合,有助于企业整体把控人工智能发展的成熟度。本研究通过对世界人工智能领域进行科学计量学分析,揭示了该领域的演化关系、发展状况及发展成熟度。

1 数据与理论方法

1.1 数据来源

什么是人工智能?人工智能虽然已经有几十年的发展历史,但尚未有统一的定义。1950年,阿兰·图灵提出了被大

到稿日期:2017-10-24 返修日期:2017-11-24 本文受国家自然科学基金项目;国内外主要学科分类体系的集成映射实证研究(16BTQ077),中国科学技术信息研究所创新研究基金面上项目(MS2017-01),中信所重点工作项目(ZD2017-14)资助。

李悦(1993-),女,硕士生,主要研究方向为科学计量学;苏成(1973-),男,博士,副研究员,主要研究方向为科学计量学;贾佳(1982-),女,博士,主要研究方向为科学计量学;许震(1976-),男,硕士,主要研究方向为信息资源管理;田瑞强(1989-),男,硕士,助理研究员,主要研究方向为科学计量学。

部分人所接受的图灵测试衡量智能机器。一台机器要通过图灵测试需要自然语言处理、知识表示、自动推理、机器学习、机器视觉、机器人技术六大能力。

本研究将国际权威数据库 Web of Science 中的核心合集作为数据来源,依据 EI Compendex 数据库中人工智能的受控词表,选取人工智能的相关词和下位词,制定检索式为:TS=(((((“Adaptive system”) or (“Ambient intelligen *”) or (“Autonomous agent”) or (“Backpropagation”) or (“Biocybernetics”) or (“Brain model *”) or (“Case based reason *”) or (“Cognitive system *”) or (“Deep learning”) or (“Deep neural network”) or (“Expert system”) or (“Formal logic”) or (“Fuzzy system”) or (“Heuristic program”) or (“Intelligent agents”) or (“Intelligent virtual agent”) or (“Learning system”) or (“Local search”) or (“Logic program”) or (“Reinforcement learning”) or (“Robot”) or (“Soft comput *”) or (“System theory”) or (“Theorem proving”) or (“Ubiquitous comput *”) or (“Fuzzy control”) or (“Intelligent computing”) or (“Intelligent system”) or (“Knowledge engineering”) or (“Motion planning”) or (“Multi agent system”) or (“Neural network”) or (“Problem solving”) or (“Robotic”) or (“Swarm”)))).

据王丽雅^[5]研究发现,计算机领域论文的半衰期为 5 年左右,因此本研究时间跨度选取 3 个 5 年段,2002—2016 共 15 年。再对检索结果进行精炼,精炼依据为文献类型: (PROCEEDINGS PAPER OR ARTICLE)。因高被引论文更具有代表性,更能体现领域特征,故取每年被引频次前 5% 的数据作为实验数据。检索日期为 2017 年 3 月 29 日,共得到论文 18376 篇。下载数据保存为 SciMAT 能够直接读取的文本格式。

1.2 理论方法

1.2.1 共词分析法

共词分析法最早是由法国文献计量学家 Callon 等^[6]提出的一种针对文献内容的分析方法。共词分析法通过统计分析关键词在文献中共同出现的频次构建共现矩阵来计算词与词之间的距离,然后借助聚类算法,将距离较近的关键词聚集起来形成概念相对独立的类团。关键词是揭示文献主题信息的词或词组,是文献核心内容、理论、方法的凝练^[7],在共词分析的基础上研究文献的内在联系和科学结构,可以追踪学科发展水平及其内部结构的发展变化^[8]。

共词分析中最常用的词源是关键词^[9],其结果能否准确地反映学科发展情况与数据的质量有着紧密的关系。共词分析的假设前提是作者都认真地选择了论文的关键词,且所选择的关键词能够充分、准确地描述文献的主要内容^[10]。因此本研究利用作者给出的关键词及 Web of Science 数据库加工商提供的关键词进行分析。

1.2.2 SciMAT 简介

SciMAT 由西班牙格拉纳达大学的研究组 Sci²s 开发,它包含了执行科学绘图工作流程所有步骤所需的所有模块,实

现了各种预处理工具,如检测重复和拼写错误的项目、时间切片、数据缩减和网络预处理^[11]。

2 数据清洗与参数选择

2.1 关键词清洗

目前,许多对人工智能的计量研究并没有对关键词进行词形归并,并且没有删除如“model”(模型/建模)、“stability”(稳健性)等指向性弱的关键词。本研究首先利用 WordNet 提取关键词词干进行词形归并,然后用手工作方式把意义相近的关键词进行归并,并将指向性弱的关键词删除。

2.1.1 时间分段

根据计算机相关学科的半衰期特性,本研究以 5 年为周期,将 2002—2016 共 15 年分为 2002—2006 年,2007—2011 年,2012—2016 年 3 个时间段进行研究。

2.1.2 参数选择

因为数据量过大会导致得到的图谱过于庞大杂乱,从而掩盖重要的信息,所以通过阈值设置突出主要信息,去除不重要信息,使得图谱更加清晰。本研究通过对不同阈值进行实验发现设置节点阈值(关键词词频)为 20、连线阈值(关键词共现频次)为 5 时图谱最清晰,聚类类团个数较为合理。

2.1.3 其他参数

本实验选择 Equivalence Index 作为网络的标准化方法, Simple Centers Algorithm 作为聚类方法, Core mapper 作为图谱类型, Jaccard's index 作为进化图的标准方法, Inclusion index 作为重叠图的标准方法。

2.2 战略图

Law 等^[12]于 1988 年提出了用“战略坐标”来描述某一研究领域内部的联系情况和领域间的相互影响情况。战略坐标是在建立主题词的共词矩阵和聚类的基础上,用可视化的形式来表示产生的结果^[13]。

在战略图中, X 轴为向心度,表示领域间相互影响的强度,向心度越大,则该领域与其他领域的联系越强,该领域就趋于中心地位; Y 轴为密度,表示某一领域的内部联系强度,密度越大,类团内部关键词的联系越紧密,发展成类团越高。

在战略图中,处于右上象限说明主题内部联系紧密,发展程度高;处于右下象限表示处于核心位置但成熟度不高,发展潜力大;处于左上象限表示研究成熟度高但处于边缘位置;处于左下象限表示研究成熟度低且处于边缘位置,一般为新的类团或衰退的类团。

3 结果与分析

3.1 发文量统计分析

图 1 为对 2002—2016 年间人工智能领域的论文进行的统计分析,除 2010 年、2011 年和 2016 年发文量有所下降外,总体呈上升趋势。其中,2010 年、2011 年发文量几近相同,人工智能可能进入一个平稳期;而 2011 年, Xavier Glorot 和 Yoshua Bengio^[14]改进了神经网络算法,使得用少量的神经元可以高效、灵活、稳健地表达抽象复杂的概念,掀起了人工智能新的浪潮,2012 年起,该领域发文量持续上升。2015 年之

后,人工智能又有进入新的平稳期的征兆,2016 年发文量有所下降。但由于 2017 年各国政策及市场的刺激,2017 年发文量也许会有所上升。

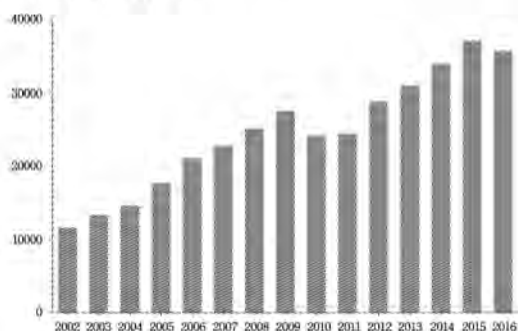


图 1 发文量总数统计

3.2 关键词演化分析

图 2 为 15 年间关键词新生与消亡的情况,从左至右依次为 2002—2006 年、2007—2011 年、2012—2016 年 3 个时间段。图 2 中,向下的箭头表示新生词,向上的箭头表示消亡词,水平箭头表示从上一阶段保留到下一阶段的关键词。可知,各时间段新生词都要大于消亡词,说明该方向蓬勃发展。关键词总数逐年增加,各上一阶段到下一阶段所保留的关键词也越来越多,这说明该方向发展趋向成熟,学科继承性较强。

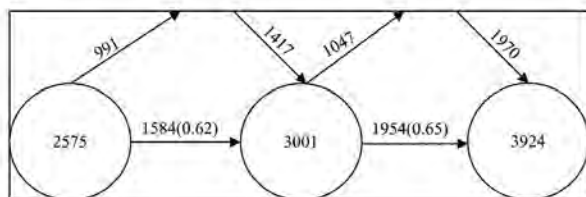


图 2 关键词演化图

3.3 发展状况分析

3.3.1 主题演化分析

图 3 为 2002—2016 年的主题演化图。从左至右依次为 2002—2006 年、2007—2011 年、2012—2016 年 3 个时间段。类团间的实线说明二者有共同的主要关键词,虚线说明类团间有共同的次要关键词。

纵向来看,2002—2006 年研究主题为 NEURAL NETWORK(神经网络)、ROBOT(机器人)、LEARNING(学习)、SCHEDULE(调度)、FUZZY CONTROL(模糊控制)、COGNITION(认知和知觉);2007—2011 年研究主题为 ROBOT(机器人)、NEURAL NETWORK(神经网络)、PARTICLE SWARM(粒子群算法)、FUZZY CONTROL(模糊控制)、SCHEDULE(调度)、LEARNING(学习)、MULTI-AGENT(多智能系统);2012—2016 年研究主题为 ROBOT(机器人)、NEURAL NETWORK(神经网络)、POWER(电力)、BRAIN(大脑)、SUPPORT VECTOR MACHINE(支持向量机)、COGNITION(认知和知觉)、FUZZY CONTROL(模糊控制)、SCHEDULE(调度)、MULTI-AGENT(多智能系统)、DIFFERENTIATION EVOLUTION(差异演化)、IMAGE(图像)、REINFORCE LEARNING(加强学习)、TUTORING(智

能化教学系统)、ESTIMATE(估计)、SIMULATION(模拟)。

横向来看,随着时间推移各阶段类团个数不断增多,表明人工智能领域研究内容不断丰富,当某一研究内容积累到一定程度,就会产生新的分支。NEURAL NETWORK(神经网络)在各个时间段都有出现,类团间实线较粗,表明研究内容比较稳定。在 2012—2016 年间,类团明显增大,表明神经网络的研究规模大幅度扩张。

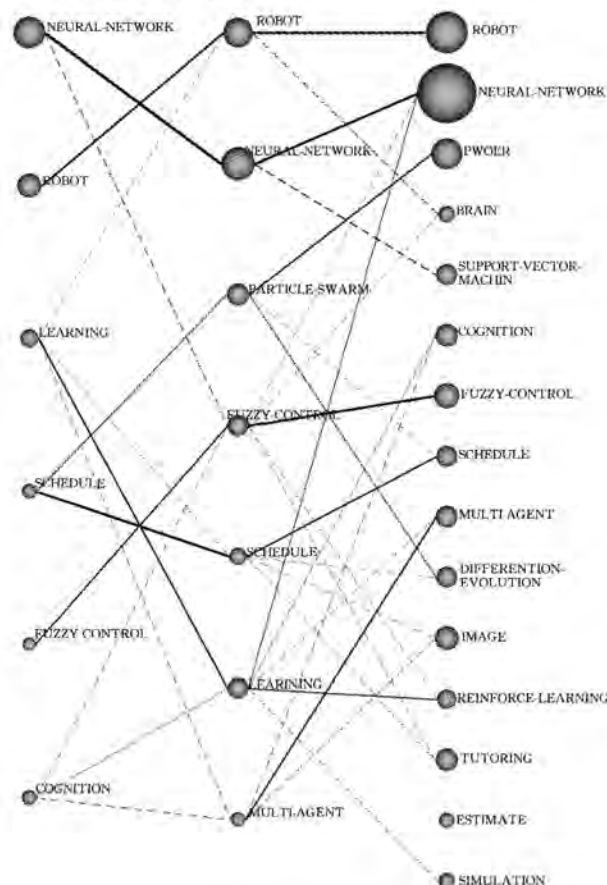


图 3 主题演化图

神经网络一直热度不减,究其原因它是试图模拟人类大脑来解决问题,并且取得了一定成果。同时,研究方向如 ROBOT(机器人)等,也都依赖于神经网络算法来推进发展。因此在未来,神经网络作为人工智能领域许多研究方向的基础动力,依旧会是研究的热门。

ROBOT(机器人)在 2002—2006 年是排名第二的类团,但在后两个阶段上升为排名第一的类团,并且类团逐渐增大,这说明机器人研究的规模不断增大并逐步走向成熟。而 NEURAL NETWORK(神经网络)与 ROBOT(机器人)两个类团的位置在后两个阶段发生了对调,说明人工智能相关算法理论越来越多地应用到实际中,使得机器人相关技术突飞猛进。

LEARNING(学习)这一类团与 ROBOT(机器人)、SCHEDULE(调度)、MULTI-AGENT(多智能系统)都有共同的次要关键词,说明这几个类团与学习算法联系紧密。而实际上这 3 个研究领域也都需要学习算法的支撑。在 2012—2016 年,LEARNING(学习)类团演化为 REINFORCE

LEARNING(加强学习)类团。REINFORCE 这一词语充分说明了学习算法已经到了新的层次。

SCHEDULE(调度)这一类团在 2007—2011 年衍生出新的类团 PARTICLE SWARM(粒子群算法)。粒子群算法是解决和优化调度问题的方法之一。而粒子群算法与 POWER(电力)和 DIFFERENTIATION EVOLUTION(差异演化)有共同的主要关键词,说明粒子群算法已经应用到电力领域,并且与差异演化算法联系紧密。粒子群算法是进化算法的一种。

总的来说,各个类团随着时间推移不断演化出新的类团;各个时间阶段类团联系紧密,除 ESTIMATE(估计)类团外,没有孤立的类团;各个研究领域都不是完全独立的,相互之间都有交叉。

3.3.2 主题发展分析

战略图可以反映主题类团的成熟度,也可以预测未来的发展趋势,其横轴为向心度,纵轴为成熟度。

图 4 为 2002—2006 年战略图,NEURAL NETWORK(神经网络)和 ROBOT(机器人)处于成熟度高、向心度高的位置;LEARNING(学习)这一类团处于成熟度低、向心度高的位置,是未来研究的热点方向;SCHEDULE(调度)处于核心度低、成熟度高的位置;FUZZY CONTROL(模糊控制)处于成熟度、向心度都适中的位置;COGNITION(认知、知觉)处于成熟度低、向心度低的位置,说明它为新兴方向。

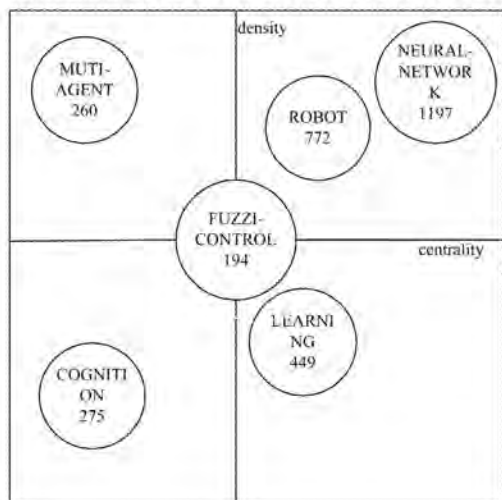


图 4 2002—2006 年战略图

如图 5 所示,NEURAL NETWORK(神经网络),ROBOT(机器人)和 FUZZY CONTROL(模糊控制)处于成熟度高、向心度高的位置。与上一阶段相比,FUZZY CONTROL(模糊控制)的成熟度和向心度都有所提升,说明该方向已经趋向于成熟。

PARTICLE SWARM(粒子群算法)处于向心度高、成熟度低的位置,说明该类团是新生的潜力类团;MULTI-AGENT(多智能系统)处于成熟度高、向心度低的位置;SCHEDULE(调度)和 LEARNING(学习)处于向心度低、成熟度高的位置。相对上一阶段而言,SCHEDULE(调度)和 LEARNING(学习)的向心度大幅下降,但这并不意味着这两

个方向开始衰落。与二者相关的其他方向,如 MULTI-AGENT(多智能系统)、PARTICLE SWARM(粒子群算法)等,仍然处于重要位置。

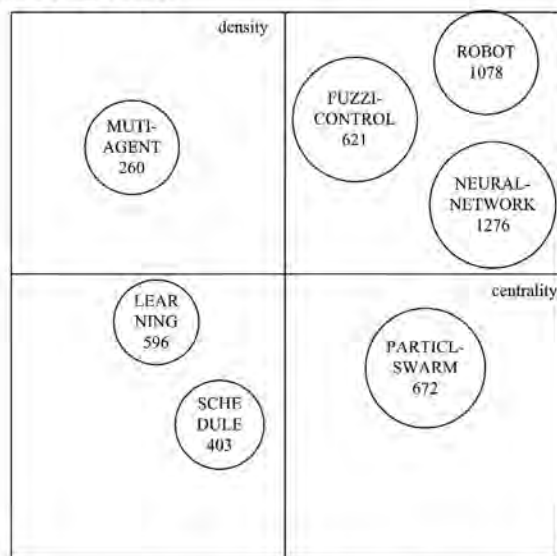


图 5 2007—2011 年战略图

图 6 为 2012—2016 年战略图,ROBOT(机器人)、NEURAL NETWORK(神经网络)、POWER(电力)、FUZZY CONTROL(模糊控制)、MULTI-AGENT(多智能系统)处于成熟度高、向心度高的位置,属于成熟的方向;BRAIN(大脑),COGNITION(认知、知觉)和 ESTIMATE(估计)处于成熟度高、向心度低的位置;IMAGE(图像)、REINFORCE LEARNING(加强学习)、TUTORING(智能化教学系统)和 SIMULATION(模拟)处于成熟度低、向心度低的位置,是新兴方向或边缘;SUPPORT VECTOR MACHINE(支持向量机)、SCHEDULE(调度)、DIFFERENTIATION EVOLUTION(差异演化)处于向心度高、成熟度低的位置,是未来发展的潜力方向。

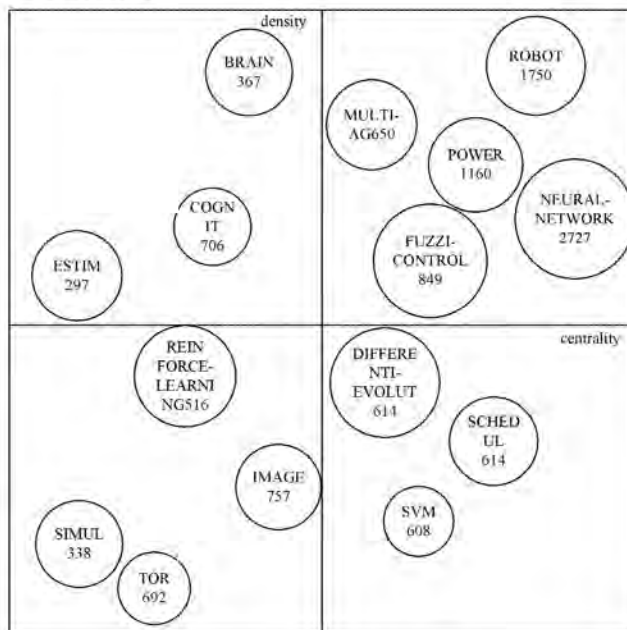


图 6 2012—2016 年战略图

由各阶段战略图可知,NEURAL NETWORK(神经网络)和ROBOT(机器人)一直处于高成熟度、高向心度的位置,说明它们一直都是研究的热门。但2007—2011年、2012—2016年两个阶段,其成熟度有所下降,说明NEURAL NETWORK(神经网络)有新内容出现,可能是由于2006年Geoffrey Hinton^[15]提出由一系列受限波尔兹曼机组成的深度置信网络以及非监督贪心逐层训练使得神经网络突破瓶颈。2012年7月,Hinton^[16]提出通过阻止特征检测器的共同作用来改进神经网络,使得神经网络又向前推进了一步。

从各阶段发展情况来看,人工智能研究的内容越来越丰富,随着时间的推移逐渐由理论研究转向实际应用研究。不仅是机器人领域,人工智能在智能家教系统、图像识别等方面也得到了广泛应用。

同时,人工智能也是多学科交叉的产物,如BRAIN(大脑)、COGNITION(认知、知觉)、DIFFERENTIATION EVOLUTION(差异演化)等类团充分表明其与脑神经科学、认知科学有着密不可分的关系。人类对人脑构造及人脑工作机理的挖掘有助于人工智能的进一步发展。

结束语 人工智能领域发文量及关键词总数总体呈上升趋势,说明该领域发展态势良好。随着时间的推移,各阶段类团数增多,说明该领域多方面发展。15年间,神经网络和智能机器人一直是人工智能领域研究的热点,并随着时间的推移不断扩大研究规模,逐步走向成熟。人工智能正逐渐从理论研究转向应用研究,并不断从其他学科吸收知识。未来,人工智能的研究热点可能依旧是智能机器人、神经网络;而支持向量机、调度算法、差异演化算法是拥有潜力的研究方向。

参考文献

- [1] 李克强. 政府工作报告[EB/OL]. [2017-05-26]. http://www.gov.cn/premier/2017-03/16/co-tent_5177940.htm.
- [2] HU Y N, DING M, CAO G F. Visual analysis of studies on international artificial intelligence in the new century[J]. Chinese Journal of Medical Library and Information Science, 2013, 22(10): 8-13. (in Chinese)
胡玉宁, 丁明, 曹高芳. 新世纪国际人工智能研究领域可视化分析[J]. 中华医学图书情报杂志, 2013, 22(10): 8-13.
- [3] ZENG Y, LIU C L, TAN T N. Retrospect and Outlook of Brain-Inspired Intelligence Research [J]. Chinese Journal of Computers, 2016(1): 212-222. (in Chinese)
曾毅, 刘成林, 谭铁牛. 类脑智能研究的回顾与展望[J]. 计算机学报, 2016(1): 212-222.
- [4] GU X F. Historical review and current development of artificial intelligence[J]. Chinese Journal of Nature, 2016, 38(3): 157-166. (in Chinese)
顾险峰. 人工智能的历史回顾和发展现状[J]. 自然杂志, 2016, 38(3): 157-166.
- [5] WANG L Y. Analysis on Discipline Half-life of Computer Science Based on CNKI[J]. Library and Information, 2015(1): 100-105. (in Chinese)
王丽雅. 基于CNKI的计算机科学学科半衰期分析[J]. 图书与情报, 2015(1): 100-105.
- [6] CALLON M, COURTIAL J P, LAVILLE F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry [J]. Scientometrics, 1991, 22(1): 155-205.
- [7] SU C, CHENG X K, TIAN R Q, et al. A study of visualized presentation of the thematic structure of general practice and its evolution[J]. Chinese Journal of General Practice, 2015, 13(8): 1211-1214. (in Chinese)
苏成, 程薛柯, 田瑞强, 等. 世界全科医学主题结构及其演进可视化研究[J]. 中华全科医学, 2015, 13(8): 1211-1214.
- [8] HE Q. Knowledge Discovery through Co-Word Analysis[J]. Library Trends, 1999, 48(1): 133-159.
- [9] HU C P, CHEN G. Characteristics of Keywords in Scientific Papers and Their Impact on Co-word Analysis[J]. Journal of the Chinese Society for Scientific and Information, 2014, 33(1): 23-32. (in Chinese)
胡昌平, 陈果. 科技论文关键词特征及其对共词分析的影响[J]. 情报学报, 2014, 33(1): 23-32.
- [10] WANG Z Y, LI G, LI C Y, et al. Research on the semantic-based co-word analysis [J]. Scientometrics, 2012, 90(3): 855-875.
- [11] COBO M J, LÓPEZ-HERRERA A G, HERRERA-VIDEIA E, et al. SciMAT: A new Science Mapping Analysis Software Tool [J]. Journal of the American Society for Information Science and Technology, 2012, 63(8): 1609-1630.
- [12] 于海霞, 董颖. 基于图书馆文献数据库共词分析的简单中心算法实证研究[J]. 情报科学, 2013(11): 81-85.
- [13] FENG L, LENG F H. Development of Theoretical Studies of Co-Word Analysis [J]. Journal of Library Science in China, 2006, 32(2): 88-92. (in Chinese)
冯璐, 冷伏海. 共词分析方法理论进展[J]. 中国图书馆学报, 2006, 32(2): 88-92.
- [14] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [J]. Journal of Machine Learning Research, 2011, 15: 315-323.
- [15] HINTON G E, OSINDERO S. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [16] HINTON G E, SRIVASTAVA N, KRIZHEVSKY, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3(4): 212-223.