



开放科学
(资源服务)
标识码
(OSID)

基于文献的人工智能领域研究的网络特征与演化分析

牛奉高 高旭霞 李志欣

山西大学数学科学学院 太原 030006

摘要: 现阶段人工智能正处于蓬勃发展时期, 挖掘“人工智能”领域的主题特征及演化规律对掌握该领域的发展动态具有重要意义。为了更高效地掌握其发展动态, 进一步了解未来研究趋势, 本文定义了一个新的计量指标: 加权创新系数。本文首先提取“人工智能”领域文献中的关键词, 通过关键词之间的共现关系和共现强度构建加权关键词共现网络(W-KCNs); 其次考虑到每年新出现的关键词与往年已有关键词重要度的不同, 定义了一个度量加权网络创新度的指标: 加权创新系数; 然后着重从节点强度分布、平均加权最近邻度和平均加权聚类系数对W-KCNs进行拓扑特征分析。研究发现: 改进后的加权创新系数更能准确描述每年关键词的创新度; W-KCNs是异配网络, 节点的强度分布近似于幂律分布, 且网络中度数小的关键词加权聚类系数大, 容易形成类簇。

关键词: 人工智能; 加权关键词共现网络(W-KCN); 拓扑特征; 演化规律; 加权创新系数

中图分类号: G350

Network Characteristics and Evolution Analysis of Artificial Intelligence Field Research Based on Literature

NIU Fenggao GAO Xvxia LI Zhixin

School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China

Abstract: At present, artificial intelligence is in the period of vigorous development, thus, it is of great significance to excavate the theme characteristics and evolution law in the field of “artificial intelligence” to grasp its development dynamics. In order to grasp its development dynamics more efficiently and further understand the future research trend, this paper defines a new measurement index: weighted innovation coefficient. This paper first extracts the keywords in the field of “artificial intelligence”,

基金项目: 山西省基础研究项目“加权共现潜在语义向量空间模型及其在文本主题聚类应用中的惩罚性矩阵分解研究”(201801D211002); 山西省高等学校创新人才支持计划“基于潜在语义的文本信息主题深度聚类研究”(2016052006); 国家自然科学基金项目“共现潜在语义向量空间模型及其语义核的构建与应用研究”(71503151)。

作者简介: 牛奉高(1980-), 博士, 副教授, 研究方向: 文本挖掘与复杂网络、信息计量与科学评价, E-mail: nfgao@sxu.edu.cn; 高旭霞(1994-), 硕士研究生, 研究方向: 社会网络分析; 李志欣(1992-), 硕士研究生, 研究方向: 社会网络分析。

constructs the Weighted keyword co-occurrence network (W-KCNs) through the co-occurrence relationship and the co-occurrence intensity between the keywords. Considering the differences in the importance of new keywords from previous years, an index to measure the innovation degree of weighted network is defined: the weighted innovation coefficient. Then the topological characteristics of W-KCNs are analyzed from the point of node strength distribution, average weighted nearest neighbor degree and average weighted clustering coefficient. It is found that the improved weighted innovation coefficient can accurately describe the innovation degree of keywords in each year. W-KCNs is a heterogeneous network, the strength distribution of nodes is similar to the power law distribution, and the keyword weighted clustering coefficients with small degrees in the network are large, and it is easy for clusters.

Keywords: Artificial Intelligence; weighted keyword co-occurrence network(W-KCN); topological characteristics; evolution law; weighted innovation coefficient

引言

人工智能 (Artificial Intelligence, AI) 是一门综合性学科,用于研究人类智能活动的规律,构造具有一定智能的人工系统,研究如何应用计算机的软硬件来模拟人类某些智能行为的基本理论、方法和技术。从软件时代到互联网,再到如今的大数据时代,人工智能逐渐发展为各方关注的焦点^[1]。随着网络技术的不断更新升级以及大数据的深度学习、类脑智能、机器学习等理论与技术的快速发展,人工智能的应用领域越来越广,涉及医疗^[2]、教育^[3]、信息安全^[4]、自然风险^[5]等方面,进而对人类的经济效益^[6]、社会作用^[7]、文化生活^[8]等方面产生了深远的影响。

现如今正是人工智能蓬勃发展的阶段,同时也是将人工智能技术转化为产业应用,以满足实际商业需求、解决现实社会问题的黄金时期。因此,分析“人工智能”领域的拓扑特征及演化趋势,能够深入了解该领域知识结构的发展规律,对于帮助科研人员掌握发展动态,了解未来研究趋势都有重要的参考价值。

近年来,随着信息和数据分析技术的不断发展与普及,计算并分析大规模密集型科学数据已经成为数据挖掘的新趋势,而文本挖掘作为数据挖掘的一个主要方面,也已经成为知识发现的一种新方法^[9]。目前对学科主题演化分析的科学计量方法主要有词频分析法、共引分析法和共词分析法。词频分析法只能显示单个关键词的频率,并不能展示关键词之间的联系;共引分析法以文献作为分析对象,需要庞大的引文索引作为研究基础,并且引文会受语言、心理、学科专业、写作习惯、媒介等因素的影响,存在一定的主观性^[10]。相比于词频分析法和共引分析法,共词分析法能深入文献内部,以文献中的关键词作为分析对象,能够从更微观的角度揭示学科主题的演化规律^[11]。关键词是一个重要的文本元素,可以简要概述文章体系的重要内容和关键信息,HAYM HIRSH^[12]和 Arjun Duvvuru^[13]等人发现分析关键词可以加速文本挖掘。探索拓扑特征和演化动态最常用的两种基于网络的方法分别是共引网络和关键词共现网络(KCN)。共引网络侧重于通过分析引文间的关系来研究科学交流的结构^[14],

而 KCN 侧重于分析文献中关键词之间的关系,进而研究科学、技术领域的知识成分和拓扑特征^[15]。目前的工作主要集中在基于 KCN 的分析上,通过理论与实证研究,已经证明了基于 KCN 的分析与传统文献综述方法相比有很大的实用价值和优势^[16]。

许多学者运用共词分析法对不同领域进行研究,把握这些领域中蕴含的真实情况。刘蓓^[17]等运用共词分析法,研究中国情报学领域中作者合作的紧密程度、团队的稳定性及网络中隐藏的知识结构。殷沈琴^[18]等运用关键词共现分析法,分析不同时间段内主要学术合作群体及其研究的热点话题,从多角度揭示数字图书馆研究的发展趋势和方向。程齐凯^[19]等基于网络社区的演化类型,利用 Z-value 算法和社区相似度算法,构建了一个科研主题演化分析模型来揭示主题发展规律。王一博^[20]等利用关键词间的共现关系构建共现网络,并通过“核心—边缘”结构分析得到了大数据研究领域的核心关键词。陈杨森^[21]等借助共词分析法,发现了国内外对社交媒介研究的区别,进一步了解了社交媒体领域的研究热点及发展趋势。李桃迎^[22]等以电商平台中的网购评语为研究对象,构建高频词共现网络,通过分析网络的结构特性发现了网购评语对销量的交互影响作用。Huajiao Li^[23]等人为了估计某一学术领域文献主题随时间的演化趋势,定义了 KCNs 的创新系数,在该创新系数中,新出现的关键词与已有的关键词有相同的重要性和研究地位。实际上每年新出现的关键词很有可能是未来研究人员将要研究的主要方向,所以该创新系数并不能准确描述学术领域文献主题随时间的演化趋势。

本文通过关键词之间的共现关系和共现强度构建加权关键词共现网络(W-KCNs),引入与加权网络相关的计量指标(强度、平均加权最近邻度和平均加权聚类系数),并在 Huajiao Li^[23]等人定义的创新系数的基础上考虑权重,定义了一个新的计量指标:加权创新系数,结合统计分析和可视化分析深入挖掘“人工智能”研究领域的拓扑特征和演化规律。实验表明,与创新系数相比,加权创新系数更能准确的描述学术领域文献主题随时间的演化趋势。

1 定义测量加权网络创新度的指标: 加权创新系数

Huajiao Li^[23]等人为了估计某一学术领域文献主题随时间的演化趋势,定义了 KCNs 的创新系数:

$$In_c(K) = \frac{\# \left\{ N_K - \left[N_K \cap \left(\bigcup_{t=1}^{K-1} N_t \right) \right] \right\}}{\# \left(\bigcup_{t=1}^K N_t \right)} \quad (1)$$

其中, N_t 是由第 t 年网络中出现的关键词组成的一个集合, $\bigcup_{t=1}^K N_t$ 是由 $t \in [1, K]$ 所有网络中出现过的关键词组成的一个集合, $\left\{ N_K - \left[N_K \cap \left(\bigcup_{t=1}^{K-1} N_t \right) \right] \right\}$ 是由在 $t \in [1, K-1]$ 从未出现过, 在 $t \in (K-1, K]$ 中新出现的关键词组成的一个集合, $\# \left(\bigcup_{t=1}^K N_t \right)$ 是 $\bigcup_{t=1}^K N_t$ 集合中元素的个数。

从上述公式可以看出,新出现的关键词与已有的关键词有相同的重要性和研究地位,但

实际上在 W-KCNs 中每年新出现的关键词很有可能是未来科研人员将要研究的主要方向^[24]。因此理论上这些新出现的关键词对学科领域的创新有更大的贡献作用,所以(1)式并不能准确描述学科领域文献主题随时间的演化趋势。然而,新出现的关键词中频数大于1和等于1对主题发展的影响也不同,因此对频数大于1和等于1的关键词考虑不同的权重,定义加权创新系数,用来估计某一学术领域文献主题随时间的创新度,定义如下:

设 W_{Ki} 是第 K 年网络中新出现的第 i 个关键词, $f(W_{Ki})$ 是第 K 年网络中新出现的第 i 个关键词的频数, 设

$$F_{K1} = \sum_{\substack{i=1 \\ f(W_{Ki}) > 1}}^{\#\{N_K - \bigcup_{i=1}^{K-1} N_i\}} f(W_{Ki}), K \in N^*, \quad (2)$$

$$F_{K2} = \sum_{\substack{i=1 \\ f(W_{Ki}) = 1}}^{\#\{N_K - \bigcup_{i=1}^{K-1} N_i\}} f(W_{Ki}), K \in N^*$$

F_{K1} 与 F_{K2} 分别表示第 K 年频数大于1和频数等于1的关键词的频数和。 λ_1 与 λ_2 分别表示 F_{K1} 和 F_{K2} 对未来主题演化趋势的贡献度,

$\lambda_1 \in (0,1)$, $\lambda_2 \in (0,1)$, 且 $\lambda_1 > \lambda_2$ 。

加权创新系数的计算公式如下:

$$In_c^w(K) = \frac{\lambda_1 F_{K1} + \lambda_2 F_{K2}}{\sum_{n=1}^K \sum_{i=1}^{\#\{N_n\}} f(W_{ni})}, K \in N^* \quad (3)$$

下面对“人工智能”领域的 W-KCNs 运用加权创新系数进行创新度分析,并通过实验结果验证加权创新系数的优越性,同时从强度分布、节点间的连接特性以及聚类情况这几个方面对其进行详细分析,进而深入挖掘该领域网络具有的拓扑特征和演化规律。

2 数据来源与分析

2.1 数据来源

Web of Science 是全球最大、覆盖学科最多的综合性学术信息资源,因此从 Web of Science 数据库中检索主题为“Artificial Intelligence”、“Artificial Intelligences”或“AI”的文献,由于1990年之前的文献几乎没有关键词,所以下载1990—2017年的主题为“人工智能”的文献作为本文研究的样本数据。检索时间为2018年6月25日。

2.2 数据分析

从 Web of Science 中共检索到时间为1990—2017年,主题为“人工智能”的文献14288篇,去掉没有关键词的文献,最终得到10344篇。以年份为单位,提取每篇文献中出现的关键词,构建篇词矩阵和共词矩阵,进而构建 W-KCNs。图1展示了文献数量与关键词数量随时间的变化情况。

根据图1可以看出,1990—2006年“人工智能”领域的文献数量基本持平,关键词数量也基本都在1000以下,但是从2006年开始文献数量和关键词数量增长速度变快,尤其是关键词数量,在2006年以后呈直线增长。因此本文主要分析2006—2017年的 W-KCNs,并将其分为四个时间段,分别为2006—2008年、2009—2011年、2012—2014年以及2015—2017年,主要通过这四个时间段来分析 W-KCNs 的拓扑特征和演化规律。

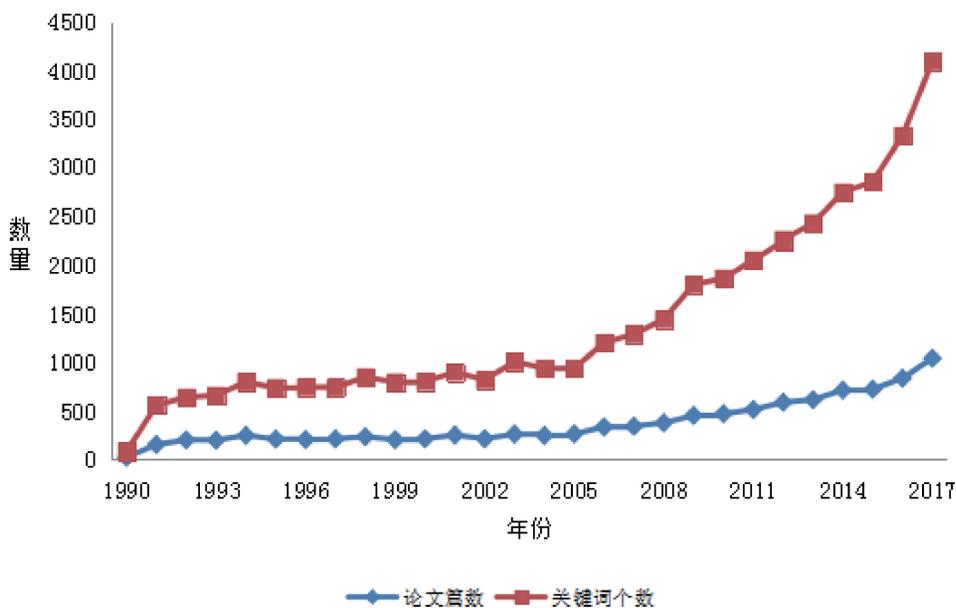


图1 文献数量与关键词数量随时间的变化

3 “人工智能”领域 W-KCNs 的拓扑特征演化分析

3.1 创新系数与加权创新系数的对比及创新度分析

加权创新系数是在创新系数的基础上考虑新出现关键词的频数，并对频数大于1和等于

1的关键词赋不同的权重，运用创新系数和加权创新系数的计算公式分别对实验数据进行计算，通过实验结果分析哪个指标更能体现每年关键词的创新度。

分别将 $\lambda_1=0.6, \lambda_2=0.4$ 、 $\lambda_1=0.7, \lambda_2=0.3$ 以及 $\lambda_1=0.8, \lambda_2=0.2$ 代入加权创新系数中，计算每年关键词的创新度，结果如图2所示。

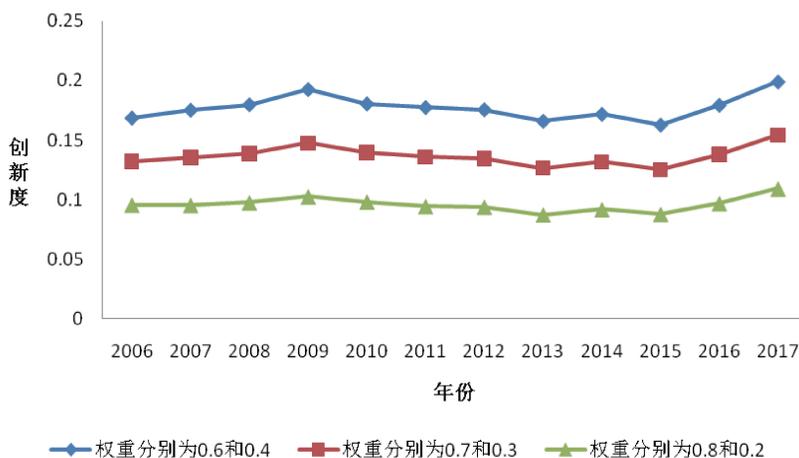


图2 不同权重下关键词的创新度

可以看出，分别对关键词赋予不同的权重，其创新度不同，这是因为，在每年新出现的关

键词中，频数为1的关键词个数大于频数大于1的关键词个数。然而，尽管权重不同，但

这三条折线整体的变化趋势一致。选择 $\lambda_1=0.6$, $\lambda_2=0.4$ 的加权创新系数与原来的创新系数分别计算创新度, 结果如图 3 所示。

观察图 3 可以发现, 从 2006—2017 年加权创新系数 $In_c^w(K)$ 的值都大于创新系数 $In_c(K)$ 的值, 且创新度的整体变化趋势不同。根据创新系数的变化趋势可以看出, 2006 年的 W-KCNs 中关键词的创新度最高, 之后每年的创新度都有所波动, 主要呈下降趋势。这样的变化趋势完全不符合人工智能领域近十几年的发展状况。

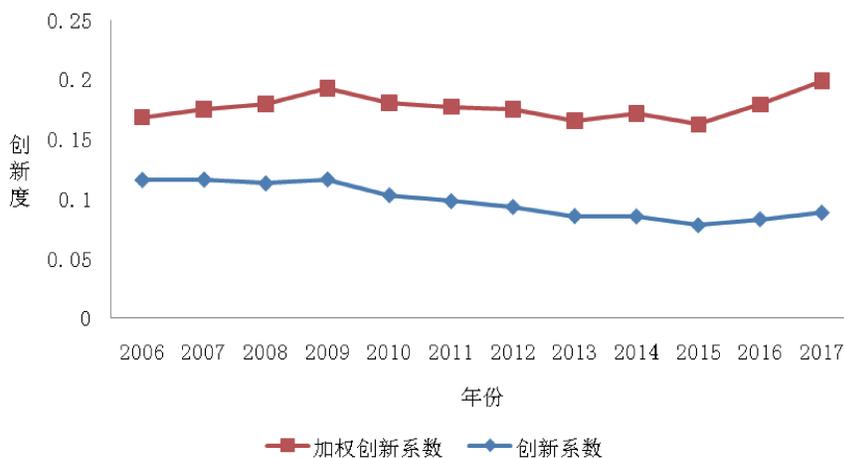


图 3 创新系数与加权创新系数的结果对比

(2) 2010—2015 年创新度在 15%~20% 之间, 加权创新系数略有波动但基本持平, 说明这一阶段对人工智能的研究在稳步进行, 2010 年谷歌制造出无人驾驶汽车; 2013 年人工智能开始从学术研究走向商业化; 2015 年在跨环境抽象、直觉概念理解、虚构画面等方面的突破进一步推动了人工智能的发展。

(3) 2015 年之后, 关键词创新度上升, 且增长速度加快, 这一阶段是人工智能的蓬勃发展期, 深度学习与大数据的兴起带来了人工智能的爆发, 物联网、云计算、大数据的发展提升了人们对人工智能的认知。2016 年谷歌在深

从加权创新系数的变化情况进行分析:

(1) 2006—2009 年的关键词创新度呈上升趋势, 这一阶段是人工智能的稳步发展期, 21 世纪以来互联网的产生和普及推动了人工智能的创新, 2007 年科学家完成了一个精心编写的人工程序, 使西洋跳棋成为迄今为止被计算机彻底掌握了的游戏, 在人机对弈中, 人类很难再战胜计算机; 2009 年人工智能在语音工程、自然语言处理、图像识别、文字识别等方面的研究有了新的突破。

度学习领域取得了重大胜利, 该公司开发的 AlphaGo 掌握了围棋游戏; 微软通过卷积神经网络和递归神经网络开发的虚拟助理, 促进了语音识别技术的发展。2017 年人脸识别、无人车等技术的应用极大地推进了人工智能的发展^[25]。

通过上述分析可以看出由加权创新系数得到的创新度更符合人工智能领域的发展历程, 更能准确描述每年关键词的创新度。实验结果说明新出现关键词的频数及频数的大小对创新度有影响, 对学科领域的创新有更大的贡献作用, 也就说明加权创新系数比创新系数更能描

述学科领域文献主题随时间的演化趋势。

3.2 人工智能加权共现网络的强度分布特征

对于一个加权网络，节点的度数并不能很好的衡量节点在网络中的相对重要性。 w_{ij} 是网络的权重矩阵，指的是节点 i 与节点 j 之间边的权重，其中 $i=1,2,\dots,M; j=1,2,\dots,M$ ， M 是网络中节点的个数。对于无向网络，权重矩阵 w_{ij} 是对称矩阵，即 $w_{ij}=w_{ji}$ 。节点强度定义如下：

$$S_i = \sum_{j \in Q_i} w_{ij} \quad (4)$$

其中 Q_i 是与节点 i 直接相连的点的集合， S_i 是节点 i 的强度，因为强度是节点度和边的权重的复合度量，所以强度比度更能精确地表征节点的重要性。

图4展示了2006—2008年、2009—2011年、2012—2014年和2015—2017年这四个时间段W-KCNs的强度分布情况，横轴代表的是强度值，纵轴代表的是强度的累积分布函数。实线代表的是对数正态拟合，虚线代表的是幂律分布拟合。

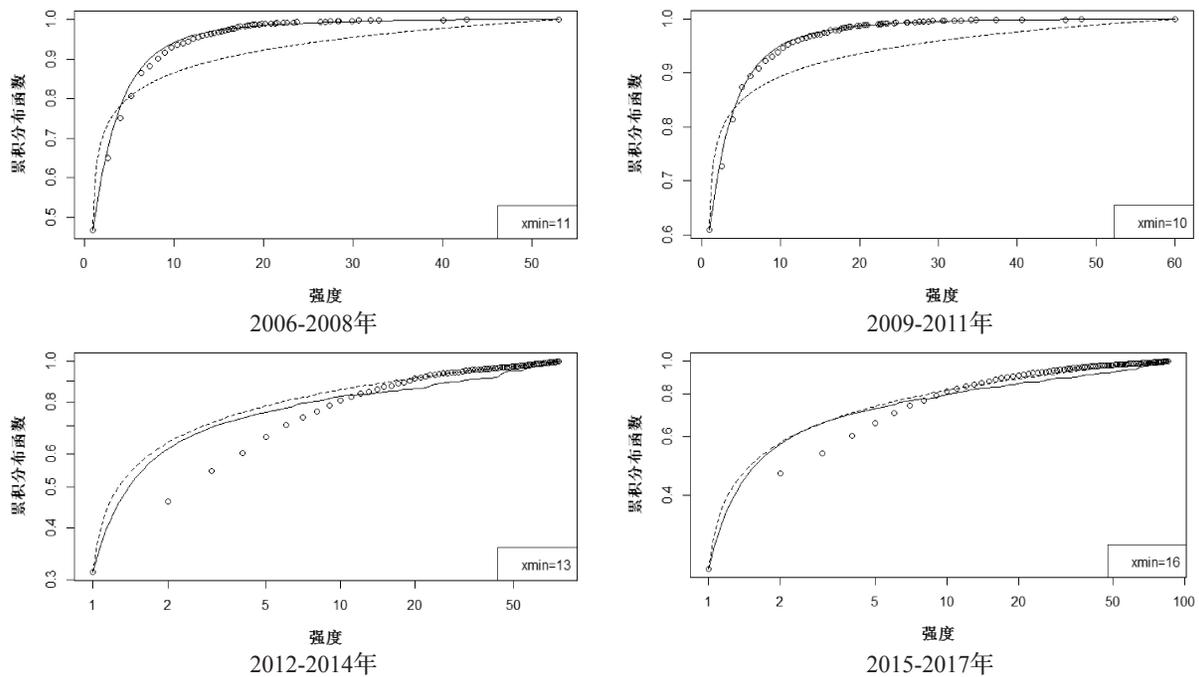


图4 W-KCNs的强度分布

每个图中的 $xmin$ 值表示幂律分布拟合最优时的最小值（2006—2008年： $xmin=11$ ，2009—2011年： $xmin=10$ ，2012—2014年： $xmin=13$ ，2015—2017年： $xmin=16$ ）。通过K-S（Kolmogorov Smirnov）检验确定幂律分布拟合的取值范围（大于 $xmin$ ）和尺度参数 α ， p 值用来判断强度分布能否用幂律分布进行拟合， $p>0.1$ 表示可以用幂律分布拟合强度分布；反之，

若 $p<0.1$ ，表示不能用幂律分布拟合强度分布。

从拟合情况可以看出，2006—2008年和2009—2011年的W-KCNs的强度分布接近于对数正态分布；2012—2014年和2015—2017年的W-KCNs的强度分布近似于幂律拟合。强度分布从对数正态到幂律分布的变化，表明从2009—2011年之后的网络为无标度网络，具有高异质性，即具有高强度值的节点越来越少，

具有低强度值的节点越来越多；极少数关键词有很大的共现次数，而大多数关键词的共现次数都很小。实验结果中拟合的参数值如表1所示。

表1 W-KCNs强度分布的拟合参数值

2006—2008年	$\mu=1.242, \sigma=0.628$
2009—2011年	$\mu=1.27, \sigma=0.703$
2012—2014年	$\alpha=2.019, p=0.708$
2015—2017年	$\alpha=2.006, p=0.939$

从图4可以观察到节点的强度分布从对数正态分布向幂律分布转化，这表明从2009—2011之后的W-KCNs的拓扑结构是无标度网络，具有较高的异质性，即具有高强度值的节点较少，具有低强度值的节点数量较多，也就是说在W-KCNs中只有少数关键词有较大的共现强度，大部分关键词的共现次数都很低。

通过分析节点的强度分布可以了解W-KCNs中的节点属性，然而，仅通过强度分布不足以

衡量节点间的关系及连接属性，因此，为了更全面的掌握网络拓扑特征，接下来我们分别计算平均加权最近邻度和加权聚类系数随节点度数的变化情况，进一步分析网络的结构特征。

3.3 平均加权最近邻度及W-KCNs节点间的连接特性

平均加权最近邻度是节点度的函数，定义如下：

$$k_{nn,i}^w = \frac{1}{S_i} \sum_{j \in Q_i} w_{ij} k_j \quad (5)$$

其中 k_j 是节点 j 的度。

度为 K 的节点的平均加权最近邻度为：

$$K_{nn}^w(K) = \frac{1}{n(K)} \sum_{k_i=K} k_{nn,i}^w \quad (6)$$

$k_{nn,i}^w$ 是一种相关性度量，通过测量其随节点度的变化情况，可以发现节点与具有相似度特征的邻居节点之间连接的趋势。当 $K_{nn}^w(K)$ 随 K 的增大而递增时，网络为同配网络；反之，当 $K_{nn}^w(K)$ 随 K 的增大而递减时，则网络为异配网络。

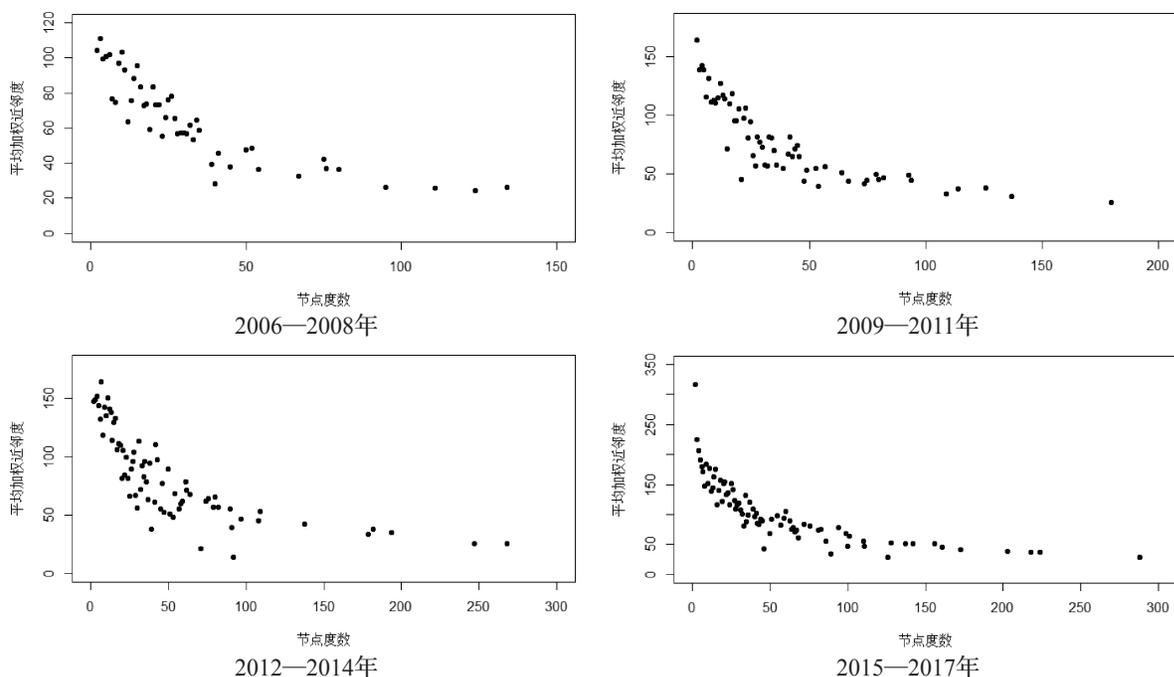


图5 度为 K 的平均加权最近邻度

图 5 分别展示了 2006—2008 年、2009—2011 年、2012—2014 年以及 2015—2017 年这四个阶段 W-KCNs 中度为 K 的节点的平均加权最近邻度，可以看出 $K_m(K)$ 在整体上随 K 的增大而减小，说明 W-KCNs 是异配网络，即度数较小的关键词倾向于与度数较大的关键词建立共现关系。但是平均最近邻度并不是随节点度数的增加呈线性变化，当节点度数 $K < 50$ 时，平均加权最近邻度快速下降， $K > 50$ 之后，平均加权最近邻度下降速度缓慢，这意味着人工智能领域的研究人员在开发新技术、试验新方法的同时，将该领域的研究热点话题和研究前沿技术相结合，围绕高热度主题进行研究，通过对高热度主题的研究，继续探索新的理论方法和技术创新。

为了进一步探讨高度数节点是否连接到低度数节点，以及不同节点间的聚类特征，本文

采用平均加权聚类系数进行分析，平均加权聚类系数与度之间的关系决定了 W-KCNs 中关键词是形成内聚群还是形成不同的类簇。

3.4 加权聚类系数及人工智能领域关键词的聚类特性

加权聚类系数是节点度的函数，用来度量网络的局部凝聚力，反映了节点近邻之间的集聚性。近邻之间关联越紧密，则聚类系数越大；反之，近邻之间关联越松散，则聚类系数越小。

$$C_i^w = \frac{1}{s_i(k_i-1)} \sum_{j \in Q_i, h \in Q_i} \frac{(w_{ij} + w_{ih})}{2} c_{jh} \quad (7)$$

其中 $c_{ij} = \begin{cases} 1, & \text{节点 } i \text{ 与 } j \text{ 之间有边} \\ 0, & \text{节点 } i \text{ 与 } j \text{ 之间无边} \end{cases}$

在 W-KCN 中，聚类系数就是与一个给定的关键词具有共现关系的两个关键词彼此之间恰好也具有共现关系的概率。

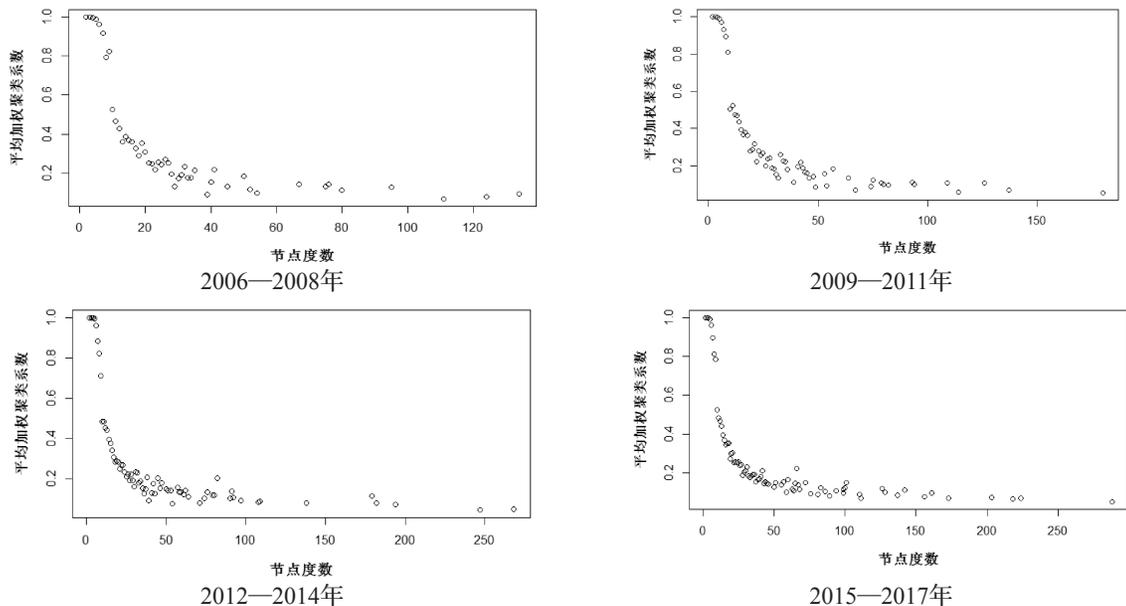


图 6 加权聚类系数随节点度数的变化情况

从实验结果可以看出 2006—2008 年、2009—2011 年、2012—2014 年和 2015—2017

年这四个阶段 W-KCNs 的平均加权聚类系数都是随着节点度数的增大而减小，且在节点度数

大于 10 时, 平均聚类系数都小于 0.5, 从微观角度分析, 这四个 W-KCNs 都存在一个共同特征: 度数较小的关键词更容易与其他度数较小的关键词聚到一类。由于度数较大的关键词与很多关键词都有共现关系, 所以聚到单个的类中的可能性很小。也就是说, 度数较小的关键词构成许多个类簇, 热门关键词与这些类簇中的关键词相连, 构成了连接这些类簇的一个桥梁。

本节对加权关键词共现网络 W-KCNs 进行拓扑特征分析, 利用新定义的加权创新系数衡量每年 W-KCNs 中关键词的创新度, 结果发现加权创新系数比创新系数更优, 更能准确描述人工智能领域的发展历程。W-KCNs 中节点的强度分布从对数正态分布向幂律分布转化, 表明从 2009—2011 之后的 W-KCNs 是无标度网络。通过平均加权最近邻度随节点度数的变化情况, 可以看出 W-KCNs 为异配网络, 度数较小的关键词倾向于与度数较大的关键词建立共现关系。最后, 加权聚类系数表明了高度数节点与低度数节点间的连接特性, 即处于核心地位的关键词作为连接桥梁, 为度数低的、处于网络边缘位置的关键词构建联系。利用上述四个度量指标对 W-KCNs 进行分析, 可以有效揭示不同时间段人工智能领域的研究热点及热点话题的整体演化规律。

4 结论与不足

本文通过关键词之间的共现关系和共现强度构建加权关键词共现网络 (W-KCNs), 并引入与加权网络相关的计量指标 (加权创新系数、强度、平均加权最近邻度和平均加权聚类系数), 结合统计分析和可视化分析更深入的理解“人工智能”研究领域的拓扑特征和演化

规律, 从整体上看, 与创新系数的变化趋势相比, 加权创新系数能更准确的描述学术领域文献主题随时间的演化趋势。主要结论有以下几点:

第一、随着时间的推移, 文献数量与关键词数量都在迅速增加, 尤其从 2006 年开始, 网络变得越来越复杂, 这意味着在“人工智能”领域, 越来越多的学者正在关注和应用已有的科学技术和方法。从整体上来看, 2006 年到 2017 年, 与创新系数的变化趋势相比, 加权创新系数能更准确的描述每年关键词的创新度, 说明新出现关键词的频数对学科领域的创新有更大的贡献作用。

第二、2006—2008 年和 2009—2011 年的 W-KCNs 的强度分布近似于对数正态分布, 2012—2014 年和 2015—2017 年的 W-KCNs 的强度分布近似于幂律分布, 强度分布从对数正态分布转移到幂律分布, 表明在“人工智能”发展过程中, 具有高强度值的关键词越来越少, 具有低强度值的关键词越来越多。

第三、四个阶段的 W-KCNs 中度为 K 的节点的平均加权最近邻度随度数的增加而减小, 表明这四个网络都是异配网络, 即度数高的关键词倾向于与度数低的关键词建立共现关系。对于 $K < 50$ 的关键词, 这种异配关系更明显, 这表明该领域的研究人员正在将现阶段的热点主题与新出现的技术和方法相结合做进一步的研究。

第四、四个阶段的网络平均加权聚类系数都是随着节点度数的增大而减小, $K > 10$ 的关键词平均加权聚类系数都小于 0.5, 表明度数较小的关键词更容易与其他度数较小的关键词聚到一类, 构成多个类簇, 度数较大的关键词是连接这些类簇之间的一个桥梁。

本文在通过 W-KCN 分析“人工智能”领

域文献的演化特征和趋势时仍存在一些问题需要改进。例如,在研究过程中构建的关键词共现矩阵是一个0~1矩阵,没有考虑关键词间的语义相似性,会导致研究结果不能更加准确的反映该领域的发展状况。不同的作者可能会为同一篇论文分配不同的关键词,所以通过关键词构建网络可能会存在一定的主观性。因此,未来的研究应该在构建关键词共现矩阵时考虑词与词之间的语义关系,同时,扩展其他的识别方法,如文献题目、摘要和文章主体等。

参考文献

- [1] 佚名. 论大数据对于人工智能发展的重要性[J]. 软件, 2017(8):67-83.
- [2] 夏黎明, 沈坚, 张荣国, 等. 深度学习技术在医学影像领域的应用[J]. 协和医学杂志, 2018(1):10-14.
- [3] 刘淇, 陈恩红, 朱天宇, 等. 面向在线智慧学习的教育数据挖掘技术研究[J]. 模式识别与人工智能, 2018, 31(1):77-90.
- [4] Xu L, Jiang C, Wang J, et al. Information Security in Big Data: Privacy and Data Mining[J]. IEEE Access, 2017, 2(2):1149-1176.
- [5] Kosmas C. Unveiling soil degradation and desertification risk in the Mediterranean basin: a data mining analysis of the relationships between biophysical and socioeconomic factors in agro-forest landscapes[J]. Journal of Environmental Planning & Management, 2015, 58(10):1789-1803.
- [6] 何玉长, 方坤. 人工智能与实体经济融合的理论阐释[J]. 学术月刊, 2018(5):56-67.
- [7] 戴汝为. 从基于逻辑的人工智能到社会智能的发展[J]. 自然杂志, 2006, 28(6):311-314.
- [8] 贾泽宇. 以人工智能为例浅析计算机科学技术对智能生活的影响[J]. 中国战略新兴产业, 2018(4):191-192.
- [9] Wang L, Tao J, Ranjan R, et al. G-Hadoop: MapReduce across distributed data centers for data-intensive computing[J]. Future Generation Computer Systems, 2013, 29(3):739-750.
- [10] 邱均平, 李晔君, 李江. 共链分析的缺陷及其解决方案研究[J]. 情报理论与实践, 2008, 31(2):170-174.
- [11] 唐果媛, 张薇. 基于共词分析法的学科主题演化研究进展与分析[J]. 图书情报工作, 2015, 59(5):128-136.
- [12] Feldman R, Dagan I, Hirsh H. Mining Text Using Keyword Distributions[J]. Journal of Intelligent Information Systems, 1998, 10(3):281-300.
- [13] Duvvuru A, Radhakrishnan S, More D, et al. Analyzing Structural & Temporal Characteristics of Keyword System in Academic Research Articles [J]. Procedia Computer Science, 2013(20):439-445.
- [14] Tang K Y, Tsai C C. The Intellectual Structure of Research on Educational Technology in Science Education (ETiSE): A Co-citation Network Analysis of Publications in Selected Journals (2008–2013)[J]. Journal of Science Education & Technology, 2016, 25(2):1-18.
- [15] Zhang B, Zhang S, Management S O, et al. Analysis of the Sci-tech Industry Research Hot spots Based on Keywords Co-occurrence Analysis and Social Network Analysis[J]. Science & Technology Management Research, 2016(20):93-98.
- [16] Zhang J, Xie J, Hou W, et al. Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based co-word analysis and social network analysis[J]. Plos One, 2012, 7(4):e34497.
- [17] 刘蓓, 袁毅, Eric B. 社会网络分析法在论文合作网中的应用研究[J]. 情报学报, 2008, 27(3):407-417.
- [18] 殷沈琴, 张计龙, 任磊. 基于关键词共现和社会网络分析法的数字图书馆研究热点分析[J]. 大学图书馆学报, 2011, 29(4):25-30.
- [19] 程齐凯, 王晓光. 一种基于共词网络社区的科研主题演化分析框架[J]. 图书情报工作, 2013, 57(8):91-96.
- [20] 王一博, 郭鑫, 王继民. 基于词共现的大数据研究主题分析[J]. 图书馆论坛, 2014(8):96-102.
- [21] 陈扬森, 陈军. 基于关键词共现与社会网络分析法的国内外社交媒体研究热点分析[J]. 信息资源管理学报, 2017(3).
- [22] 李桃迎, 吕晓宁, 李峰. 网购评语高频词共现网络的结构特征分析[J]. 计算机应用研究, 2019(1):44-45.
- [23] Li H, An H, Wang Y, et al. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network[J]. Physica A Statistical Mechanics & Its Applications, 2016, 450:657-669.
- [24] Lee P C, Su H N. Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis[J]. Innovation, 2010, 12(1):26-40.
- [25] 贺倩. 人工智能技术的发展与应用[J]. 电力信息与通信技术, 2017(9):32-37.