

The Bibliometric Analysis of Current Studies and Developing Trends on Bayesian Network Research

Zhongzheng Xiao¹, Nurbol², Hongyang Liu³

¹College of Information Science and Engineering, Xinjiang University, Urumqi Xinjiang

²Network Center, Xinjiang University, Urumqi Xinjiang

³Xichang Satellite Launch Center, Xichang Sichuan

Email: xzzmailpla@163.com

Received: Feb. 26th, 2020; accepted: Mar. 12th, 2020; published: Mar. 19th, 2020

Abstract

In this paper, 2,930 literatures related to Bayesian network in the recent 10 years in the web of science were taken as the research object. Based on the literature metrological content analysis method, the focus, development rules of research context, existing commonalities and differences, and research status at home and abroad were systematically reviewed. The study found that, as of now, especially in the prevalence of neural networks, Bayesian networks can be deepened and have great potential because of their strong mathematical interpretability. The analysis results are helpful to provide reference for the research status and progress of scholars in the field of Bayesian network research in China.

Keywords

Bayesian Network, Map Analysis, Citespace, Research Context

贝叶斯网络研究现状与发展趋势的文献计量分析

肖中正¹, 努尔布力², 刘宏阳³

¹新疆大学信息科学与工程学院, 新疆 乌鲁木齐

²新疆大学网络中心, 新疆 乌鲁木齐

³西昌卫星发射中心, 四川 西昌

Email: xzzmailpla@163.com

收稿日期: 2020年2月26日; 录用日期: 2020年3月12日; 发布日期: 2020年3月19日

摘要

本文以web of science中近10年2930篇与贝叶斯网络有关的文献为研究对象,基于文献计量内容分析方法系统地回顾了国内外在贝叶斯网络领域的关注点、研究脉络的发展规律、存在的共性与差异性和研究现状。研究发现,截至目前,特别是在神经网络盛行的现在,贝叶斯网络可以凭借其具有较强的数学可解释性,在智能计算领域的贡献不断深化且具有极大的潜力。分析结果有助于为我国贝叶斯网络研究领域的学者提供研究现状及进展的参考。

关键词

贝叶斯网络, 图谱分析, CiteSpace, 研究脉络

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

贝叶斯网络是由途径研究(因果推理链)、因果推断等慢慢衍生而来的。1982年, Pearl等人将贝叶斯网络的概率推理应用于人工智能当中,此后,专家系统中的不确定性表示和推断更多地开始使用贝叶斯网络。和其他人工智能的算法比较,贝叶斯网络因为将图形解释和计算与概率论完美地结合在一起,所以在处理不确定的问题上比较大的优势。这些优势表现在灵活的依赖性拓扑结构;易于理解和解释,有明显的语义强大的不确定性问题处理能力;能有效地进行多源信息融合。随着对贝叶斯网络的不断研究,其逐渐成为数据挖掘、人工智能等领域解决不确定性问题的关键方法之一,是这些年的一个研究热点[1]。

基于此,这篇文章用web of science的核心数据库为数据源,借助CiteSpace工具的文献计量方法对收集到的文献进行统计分析并绘制出知识图谱,解决下面的三个研究问题:

- (1) 国内外近10年来在贝叶斯网络领域的主要国家和研究机构有哪些?
- (2) 国内外近10年来贝叶斯网络领域的研究热点有哪些?
- (3) 国内外贝叶斯网络领域的研究方向的发展有什么趋势?

本文的具体工作如下:首先阐述了数据来源与研究方法,然后对贝叶斯网络分析研究进行描述,得到该领域的主要研究机构及国家、最佳的研究热点以及大致的发展趋势。最后对分析所得到的结果进行总结,希望能用客观以及可视化数据的展示将国内外对于贝叶斯网络研究的现状和大致趋势呈现出来,为我国贝叶斯网络领域的研究人员提供较为清晰的参考。

2. 数据来源和研究方法

2.1. 数据来源

文献来源于Thomson Reuters公司开发的Web of Science信息检索平台。1) 采用的文献检索方法为TS = ((data “Bayesian network”) or (data “probabilistic network”)). 2) 时间跨度 = 2008~2018。3) 索引数据库 = (SCI-EXPANDED), 文献类型 = (PROCEEDINGS PAPER OR ARTICLE), 共得到文献2930篇。经过对比和筛选,删去了与贝叶斯网络无关的文献,最终获得1613条文献题录。这些文献题录包括作者、标题、关键词、摘要、作者地址、引用的参考文献、被引频次、出版日期等属性,供可视化分析。

2.2. 研究方法说明

本文的基本流程如图 1 所示。

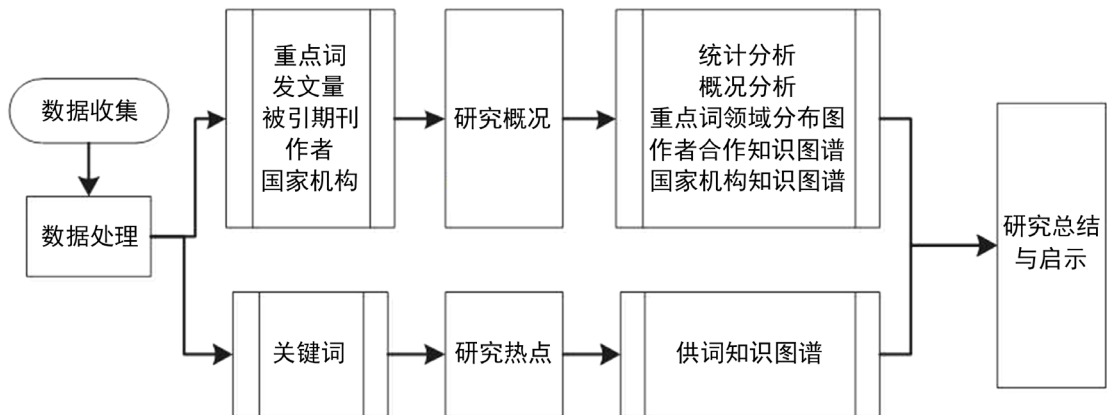


Figure 1. Research method flow chart

图 1. 研究方法流程图

在信息化社会飞速发展的今天，大多数学者选择用数据信息绘制知识图谱对学科知识进行可视化。与其它软件相比，Citespace 软件是一款着眼于分析科学分析中蕴含的潜在知识，并在科学计量学、数据和信息可视化背景下逐渐发展起来的一款引文可视化分析软件，由于是通过可视化的手段来呈现科学知识的结构、规律和分布情况，因此也将通过此类方法分析得到的可视化图形称为“科学知识图谱”，本文以贝叶斯网络的文献数据为基础，分析贝叶斯网络的研究热点、未来趋势等，为相关研究提供依据[2]。

3. 研究概况

首先使用 Carrot2 聚类分析工具，对文章题目与摘要中出现的短语运用 TF-IDF 算法进行词频统计处理得到每篇文章的重点词，接下来运用 Lingo 算法对提取出来的重点词进行聚类处理，得到图 2。以求初步了解在贝叶斯网络的研究过程中都涉及到了哪些领域，并通过气泡大小初步判定哪些分支较为热门哪些分支研究较少。

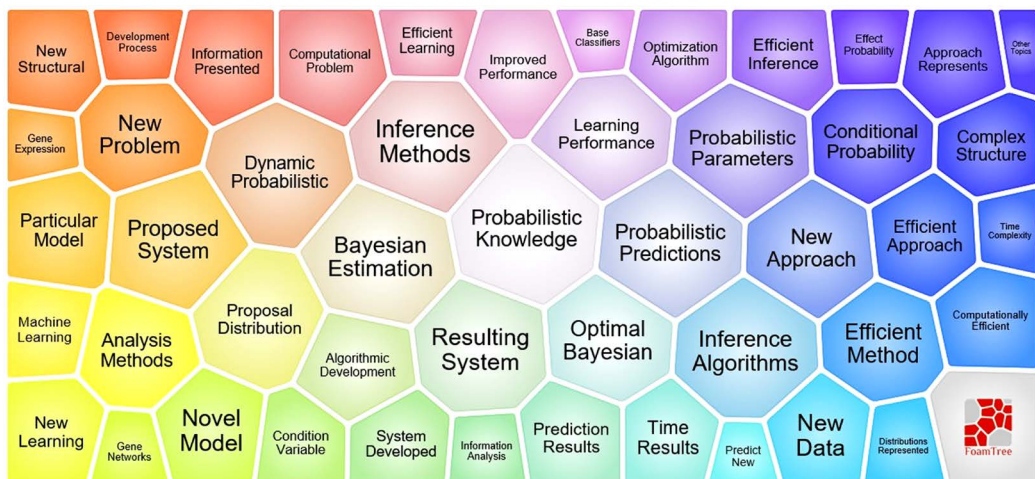


Figure 2. Keyword clustering bubble chart

图 2. 重点词聚类气泡图

将图 2 中的结果, 对比相关领域资料, 对每一个领域进行归纳统计, 得到贝叶斯方法在具体研究和工程应用中涉及的运用方向分支表如表 1 所示。

Table 1. Use direction branch table
表 1. 运用方向分支表

涉及方向	运用简介	涉及重点词	应用场景
数学方向	贝叶斯方法在数学领域运用最为广泛, 主要以统计学分支下的逻辑推理为主。而由逻辑推理可得的结果又间接推动了概率预测、分类方法的发展。进而以贝叶斯在数学领域内的研究成果为基础, 使得其可以进一步运用在计算机和生物医学方面。	Bayesian Estimation (贝叶斯估计)、Probabilistic Knowledge (概率知识)、Dynamic Probabilistic (动态概率论)、Probabilistic Parameters (概率参数)、Inference methods (推理方法)	赛事预测、股票预测、数据分类、风险分析等
计算机方向	贝叶斯在因果推理上表现出的先天优势, 使得贝叶斯方法与目前计算机方向发展热门的人工智能需求不谋而合。同时, 在计算机解决部分只需要近似解问题情况的时, 相比传统算法通过海量计算作为成本得出确定性的结果。将贝叶斯方法运用在算法中可以有效降低整个算法的时间复杂度, 进而优化算法提升效率。	Machine Learning (机器学习)、Algorithmic development (算法开发)、Efficient methods (效率方法)、Optimization methods (最优优化方法)	机器人决策、近似问题算法求解、数据挖掘、特征识别等
生物医学方向	基因与基因之间的潜在关联关系在生物学方向具有重大研究价值。而由于自然界生物种类极其丰富, 基因关联网络的构建具有复杂性高、计算量极大的特点。通过贝叶斯方法, 利用其因果推理的特性可以高效获取具备潜在关联的基因配对信息, 进而进一步揭示不同基因之间的作用机理与功能性信息, 为药医学方向研究疾病的发病原理和应对方法提供一定的指导作用。	Gene Networks (基因网络)、Gene Expression (基因表达)、Complex Structure (结构复杂性)	基因表达图谱分析、调控网络的构建、情绪分析等

3.1. 主要研究国家与机构分析

通过对国家和机构进行分析, 可以帮助了解相关国家或机构对贝叶斯网络领域的研究程度以及它们之间的合作情况。图 3 为设定工具阈值和相关系数后, 绘制的国家及机构图谱, 圆形节点代表国家, 直线分支上的小节点代表机构, 节点大小代表发文量多少; 节点间的连线代表对应的两个国家或机构之间有合作关系; 连线的粗细程度代表两者之间合作的紧密程度[3]。

通过对研究机构聚类分析, 得到知识图谱图 3。分析图 3 和表 2 可以发现, 发文量 Top15 的研究机构隶属于 9 个国家: 中国 3 所, 美国 3 所, 西班牙 3 所, 韩国、法国、荷兰、丹麦、比利时和巴西各 1 所。其中, 丹麦和比利时的国家发文量排名不在国家发文量排名的 Top15 中, 而这两个国家的研究机构的发文量排名却在机构发文量 Top15 中, 对比该机构和国家发文数量可以发现, 这两个国家发表的文献几乎全由丹麦奥尔堡大学、比利时鲁汶大学贡献。英格兰、德国、加拿大、意大利和日本, 虽然这些国家的发文量排在前列, 但其国家的研究机构发文量却不在机构发文量排名前列, 说明这些研究机构的分布相对分散。

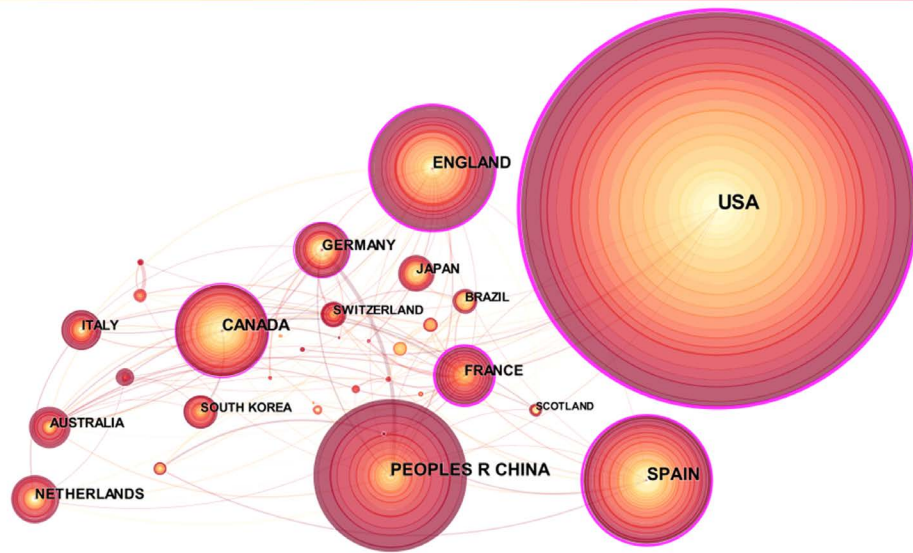


Figure 3. Country and institution knowledge chart
图 3. 国家及机构知识图谱

Table 2. Top 15 research institutions and countries by number of publications
表 2. 按文献发表数量排名 Top15 的研究机构和国家

研究机构	文献数量	十年内首次发 文年份	国家	文献数量	国家中心性
美国伦斯勒理工学院(Rensselaer Polytech Inst)	26	2008	美国	363	0.46
西班牙格拉纳达大学(Univ Granada)	25	2009	中国	214	0.1
荷兰拉德堡德大学(Radboud Univ Nijmegen)	20	2008	英格兰	129	0.21
云南大学(Yunnan Univ)	18	2008	西班牙	118	0.08
西班牙马德里大学(Univ Politecn Madrid)	16	2008	法国	104	0.28
巴西圣保罗大学(Univ Sao Paulo)	15	2008	德国	101	0.1
清华大学(Tsinghua Univ)	15	2012	加拿大	86	0.04
比利时鲁汶大学(Katholieke Univ Leuven)	14	2008	意大利	65	0.04
美国乔治梅森大学(George Mason Univ)	12	2008	荷兰	59	0.01
西班牙阿尔梅里亚大学(Univ Almeria)	12	2012	日本	53	0
法国南特大学(Univ Nantes)	12	2012	澳大利亚	51	0.04
丹麦奥尔堡大学(Aalborg Univ)	11	2013	韩国	49	0.01
中国科学技术大学(Univ Sci & Technol China)	11	2010	瑞士	43	0.03
韩国延世大学(Yonsei Univ)	10	2008	印度	32	0.05
美国卡耐基梅隆大学(Carnegie Mellon Univ)	10	2008	巴西	30	0

中心性是用来表示一个节点和其他节点的关联包括其节点在整个网络中的重要程度。中心性越高的研究机构影响力越大。从中心性的角度看许多国家的中心性是 0，这表示他们对其他的国家对于贝叶斯网络的研究没有产生影响。中心性大于 0 的总共有 28 个国家，其中美国中心性最高，英国和法国在中心性上并列第二，德国排第四，中国排第五。总体来说，美国在国际上的贝叶斯网络研究领域上占着主导

的地位,并具有很强的影响力。日本、巴西两国虽然发文量在 Top15,但其中心性却都为 0,说明日本和巴西对于贝叶斯网络的研究较为封闭,并且对于其他国家的研究并没有产生较大的影响。

3.2. 作者分析

作者的发文量间接反映了作者在某一领域的影响力。据统计,检索的 1613 篇文献中,共得到作者 1723 位。根据普莱斯定律来确定贝叶斯网络研究领域的高产作者,其计算公式为 $M = 0.749 \times \sqrt{N_{\max}}$,其中 N_{\max} 为该研究领域发文量最多的作者所发表的论文篇数, M 为核心作者发文篇数下限,从表 2 中易知 $N_{\max} = 26$,计算 $M = 3.819$,取整为 4,即发表论文 4 篇及 4 篇以上的作者为贝叶斯网络研究领域的核心作者,共计 48 名。

从表 3 可知大多数作者发文量都在 10 及 10 以下,该表列出的首次出现年份是基于本文的研究文献。其中 Ji Q 第一(26 篇),Ji Q 教授主要的研究兴趣是和人工智能应用相关,包括人工神经网络(主要是基于无监督学习),多智能体系统和混合人工智能,专门解决在计算机网络安全领域的问题[4];Leray P 教授(19 篇)的研究兴趣主要集中在贝叶斯网络的构建,特别关注于探索性投影寻踪、自组织映射和核方法[5]。从首次发文年份上看,集中分布在 2008 年至 2012 年,说明 2008 年以前,研究人员对贝叶斯网络的研究正处于逐步摸索的阶段,并在 2008 年至 2012 年渐渐有了研究成果。

Table 3. Number of TOP10 author posts in the past ten years

表 3. 近十年 TOP10 作者发文数量

作者	文献数量	十年内首次发文年份	所属机构
Ji Q	26	2008	美国伦斯勒理工学院
Leray P	19	2008	法国南特大学
Liu WY	18	2008	云南大学
Larranaga P	17	2008	西班牙马德里大学
Yue K	17	2008	云南大学
Bielza C	15	2011	西班牙马德里大学
Lucas PJF	12	2009	荷兰拉德堡德大学
Moral S	10	2011	西班牙格拉纳达大学
Cho SB	10	2008	韩国延世大学
Salmeron A	9	2012	西班牙阿尔梅里亚大学
Gomez-Olmedo M	9	2012	西班牙格拉纳达大学
Cano A	9	2012	西班牙格拉纳达大学
Wang SF	8	2014	中国科学技术大学
Sucar LE	8	2010	美国国家航空航天局
Maua DD	6	2014	巴西圣保罗大学

3.3. 贝叶斯网络研究的主要来源出版物共引分析

通过分析贝叶斯网络学术领域的文献出版物能够比较准确地确定该领域的核心出版物的分布情况。在 Cite Space 中“node type”选择参数“Cited Journal”,得到出版物分布的共引网络图谱和如表 4 所示的贝叶斯网络文献发表数量 TOP15 的被引期刊。

Table 4. Bayesian network literature published in the TOP15 cited journals
表 4. 贝叶斯网络文献发表数量 TOP15 的被引期刊

被引期刊	文献数量	被引期刊中心性
PROBABILISTIC REASON	476	0.12
MACH LEARN	433	0.16
ARTIF INTELL	426	0.14
LECT NOTES COMPUT SC	402	0.03
INT J APPROX REASON	301	0.18
J MACH LEARN RES	295	0.09
LECT NOTES ARTIF INT	269	0.11
IEEE T PATTERN ANAL	246	0.1
J ROY STAT SOC B MET	211	0.06
ANN STAT	184	0.1
PROBABILISTIC GRAPHI	180	0.01
BIOINFORMATICS	172	0.05
J ARTIF INTELL RES	171	0.15
IEEE T KNOWL DATA EN	167	0.07
SCIENCE	163	0.05

其中 PROBABILISTIC REASON 的被引文献数量高达 476 篇，排在第一位。MACH LEARN 的被引文献数量为 433 篇，排名第二，ARTIF INTELL 的被引文献数量为 426 篇，排名第三，随后的还有 LECT NOTES COMPUT SC (402 篇)、INT J APPROX REASON (301 篇)等。从中心性来看，INT J APPROX REASON 以 0.18 排在首位，随后的有 MACH LEARN (0.16)，J ARTIF INTELL RES (0.15)，PROBABILISTIC REASON (0.12)等。前五个出版物不仅被引文献的数量很多，在领域内的影响也很大，尤其 Lecture Notes in Computer Science 占有绝对的领先地位，该丛书每年都出版了诸如 Computer Security — ESORICS、Network and System S。

ecurity 等书籍，涉及了当年的热点研究问题。此外，J ARTIF INTELL RES 虽然被引文献的数量不是特别多，但是中心性很高，近十年的影响因子也很高，说明在领域内影响力较大。综合可以看出，Lecture Notes in Computer Science、Machine Learning、International Journal of approximate Reasoning 和 Artificial Intelligence In Medicine 这些学术出版物是该领域中较为重要的组成部分，SCIENCE 的在自然科学领域虽然有较大影响，但是在贝叶斯网络的研究中其中心性只有 0.05，相对来说在这个领域影响力较低。

3.4. 核心文献分析

文献的被引频次是最能直接体现文献学术影响力及其价值的重要指标。表 5 是被引频次 top10 的文献统计列表，从表中可知 Stegle 发表的“A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies”位居第一位，被引频次高达 132 次，该文研究了一种用于绘制表达数量性状基因座(eQTL)的概率方法，该方法使用贝叶斯建模，它可以模拟基因型的贡献以及已知和隐藏的混杂因子[6]。第二位核心文献是 Ahmadiou 发表的“Enhanced probabilistic

neural network with local decision circles: A robust classifier”一文，被引频次为 120 次，该文探讨了通过使用简单贝叶斯规则的核密度估计，在模式层中计算测试数据对每个训练数据的可能性，并使用局部决策圈(LDC)提出了增强和广义 PNN (EPNN)，以提高其对数据中噪声的鲁棒性[7]。Psorakis 发表的“Adaptive Overlapping Community Detection with Bayesian Non-Negative Matrix Factorization”排在第三位，被引频次为 117 次，该文提出利用非负矩阵分解(NMF)进行群落检测，可以从整个网络相邻矩阵中找到两个非负矩阵，两个矩阵的乘积很好地逼近原始矩阵。通过分解过程中的贝叶斯解释，不仅可以用 Shrinkage 方法在大型网络中捕获最合适的社区数，还可以验证在模糊情况下如何将节点分配给社区的良好阈值。该方法应用于一些真实网络和具有基准的合成网络。重叠群落检测的实验结果表明，该方法能够有效地找到群落数量和重叠度，并且比其他现有的重叠群落检测方法具有更好的性能[8]。结合表 5 的关注点来看，有多篇都与生物学相关，例如 Zoppoli 发表的“Time Delay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach”一文介绍了一种从基于信息理论的算法并利用时间序列测量推断基因调控网络的方法，这种算法的基本思想是通过假设作为基础概率模型的静态马尔可夫随机场来检测表达谱之间的时间延迟的依赖性。使用自动计算的阈值过滤掉信息量较少的依赖关系，保留最可靠的连接[9]。贝叶斯网络模型在生物学领域有着其他机器学习不具备的优势[10]。

Table 5. Top10 cited references
表 5. 被引频次 Top10 的文献

文献	关注点	作者	被引频次
A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies	贝叶斯复杂控制性	Stegle	132
Enhanced probabilistic neural network with local decision circles: A robust classifier	概率神经网络	Ahmadlou	120
Adaptive Overlapping Community Detection with Bayesian NonNegative Matrix Factorization	重叠社团检测	Psorakis	117
TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach	因果网络	Zoppoli	114
Empirically building and evaluating a probabilistic model of user affect	动态贝叶斯网络	Conati	113
Comparison of co-expression measures: mutual information correlation and model based indices	基因监管网络	Song	96
Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons	马尔可夫链	Buesing	89
A Bayesian Approach to Model Checking Biological Systems	系统生物学	Jha	86
A Bayesian regularized artificial neural network for stock market forecasting	贝叶斯正则化人工神经网络	Ticknor	72
A review of recent advances in learner and skill modeling in intelligent learning environments	学习者模型	Desmarais	71

4. 基于共词分析的研究热点

共词分析相比文献的共被引和耦合，其得到的结果是非常直观的。即研究者直接可以通过共词分析的结果，对所研究领域的主题进行分析。图 4 中心性由最外层十字代表。关键字重叠则由同心十字表示，关键词年份不同十字的深浅层次也不同。关键词之间的远近则代表着其之间的关系。其中同心十字与文字大小成正相关，文字越大也表示关键词被引频次越高[11]。表 6 列出的前十五个词是词频较高和中心性较高的关键词。

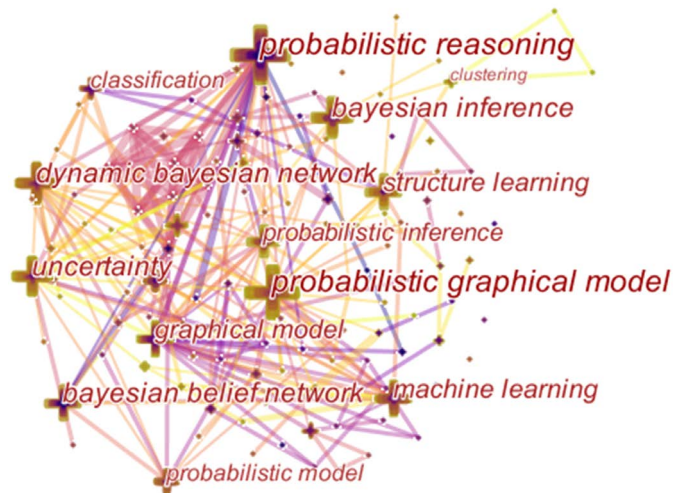


Figure 4. Key word knowledge map
图 4. 关键词知识图谱

Table 6. List of high-frequency words and high centrality keywords for Bayesian network research (Top15)
表 6. 贝叶斯网络研究高频词及高中心性关键词一览表(Top15)

排序	高频关键词			高中心性关键词	
	关键词	词频	排序	关键词	中心性
1	贝叶斯网络(bayesian network)	606	1	选择分类(selection)	0.17
2	模型(model)	173	2	概率(probability)	0.17
3	算法(algorithm)	151	3	识别(recognition)	0.15
4	网络(network)	131	4	贝叶斯网络(bayesian network)	0.1
5	系统(system)	113	5	诊断(diagnosis)	0.1
6	推断(inference)	109	6	模型(model)	0.09
7	概率网络(probabilistic network)	69	7	分类(classification)	0.09
8	概率推断(probabilistic inference)	68	8	不确定性(uncertainty)	0.09
9	分类(classification)	68	9	概率图形化模型(probabilistic graphical model)	0.09
10	不确定性(uncertainty)	62	10	神经网络(neural network)	0.09
11	信念网络(belief network)	61	11	信息(information)	0.09
12	概率图形化模型(probabilistic graphical model)	61	12	最优化(optimization)	0.09
13	动态贝叶斯网络(dynamic bayesian network)	58	13	概率模型(probabilistic model)	0.09
14	贝叶斯推断(bayesian inference)	53	14	鉴定(identification)	0.09
15	神经网络(neural network)	42	15	推断(inference)	0.08

Cite Space V 可以直接使用 WOS 等数据库的文献记录，通过提炼研究的关键词或主题词的频次及不同的关键词或主题词之间共现关系，并采用可视化的形式展现主题词或关键词的频次高低、聚类关系，得出研究热点。将最终精炼整理得到的 1613 篇论文全记录信息导入 Cite Space V，时区分隔(Time Slicing) = (From2008 To2018) (1Years Per Slice)；主题词来源(TermSource)为标题(Title)、摘要(Abstract)、作者关键词(Author Keywords (DE))、扩展关键词(Keywords Plus (ID))，即全部勾选；节点类型(Node Types)选择关键词(Keyword)；选择最小生成树算法(Minimum Spanning Tree)进行剪枝(Pruning)；视图方式(Visualization)保持默认；进行图谱绘制。根据本文的主题，需要对图谱绘制使用的关键词节点进行去重复和合并，合并后“baysian network”的词频达到 606 次，是本研究中贝叶斯网络领域的最重要的主题词。通过最小生成树算法进行聚类分析，得到如表 6 所示的高频关键词、高中心性关键词表。从表 6 中可以看出，在全球范围内来说贝叶斯网络研究领域十分广泛，不仅围绕人工智能、模式识别、机器学习、数据挖掘等传统贝叶斯网络技术应用层面展开，还关注于规划识别、模型诊断、模型诊断串行译码和遥感数据分类等；且在交叉学科如冠心病中医临床诊断、对移动机器人进行避障等领域也有数量可观的论文发表[12]。

5. 研究总结与启示

贝叶斯之父 pearl 1985 年在《Artificial Intelligence》杂志上首先提出了基于贝叶斯理论的贝叶斯网络的构想。由于贝叶斯网络很好地结合了概率论以及图论相关理论，与当时最流行的基于线性规则的推理相比，能够很好地解决复杂、模糊和不确定问题。因而，在以后的不确定场景下，贝叶斯网络得到了很好的应用，并且逐步取代其他机器学习算法。与此同时，贝叶斯网络出现给很多学者在因果推断问题上提供了一种新的思路，因为这种算法能够有效且直观地解决因果关系问题，这也被称为概率图模型，如隐马尔科夫模型(HMM)和动态贝叶斯网络(DBN)等。表 7 列出三个阶段的主要研究热点、研究领域以及研究脉络。



Figure 5. Keywords development context timeline diagram

图 5. 关键词发展脉络时间轴图

Table 7. Phase analysis table
表 7. 阶段分析表

时间	1985~1990	1990~2005	2005~至今
研究热点	贝叶斯理论、链式公式、因果关系、贝叶斯推断等	贝叶斯网络、数据挖掘、结构学习、参数学习等	诊断、模型、大数据、分类、概率图形化模型等
研究领域	数学领域、机器学习领域等	数据挖掘领域、人工智能领域、计算机感知领域[13]等	医学诊断领域、大数据领域、决策信息融合领域等
研究脉络	该时期涉及到的领域较少,主要进行网络基础理论体系和不确定性推理的研究	贝叶斯网络发展进入新阶段,国际上开始致力于网络拓扑结构以及参数等的学习,利用大量数据对贝叶斯网络进行构建	该时期有了系统的知识体系和技术支撑,贝叶斯网络应用趋于成熟化,在医疗中的应用更是独占鳌头[14]

关键词通常是对文章主旨和中心内容的凝练,为检索、阅读等带来很大的便利。Web of Science 核心数据库中的所有文章都新增了补充关键词(keywords plus),这些关键字是通过聚类算法计算得出的。从贝叶斯网络领域每年不同的关注点出发,分析研究贝叶斯网络的发展动向。本文利用 CiteSpace 软件生成关键词,从关键字和相关领域术语的变化过程中,可以推断出发展趋势动向,使用时间轴视图绘制其演化过程,便可得到关键词发展脉络时间轴图。

图 5 中“十字”形节点的大小与此关键词共被引的程度成正比。与关键词节点相连连线的颜色对应关键词出现的年份。

本文简要介绍并分析了国内外近 10 年有关贝叶斯网络的研究概况和研究热点,可供学者从多方面对该领域进行了解。通过以上介绍,我们可以将该领域研究分为三个阶段:(1) 1990 年以前:主要是在贝叶斯网络数学理论方面做研究,在建立贝叶斯网络时也都是以专家先验知识为基础而构建的;(2) 1990~2005 年:主要研究如何在大量数据的基础上,进行贝叶斯网络的自主构建;(3) 2005~至今:主要研究贝叶斯网络的应用,来解决实际问题。结合前文的分析,对全文总结如下:

1) 国内外贝叶斯网络领域的研究文献质量一直在稳步提升,在文献数量上美国和中国最有很大的优势,但在文献影响力上美国、英国和法国处于领先地位,而中国凸显出文献数量多国际影响力低的特点。在这十年内,发表论文的机构主要集中在伦敦勒理工学院、格拉纳达大学、云南大学和中国科学技术大学。其中美国伦敦勒理工学院的 Ji Q. 教授将动态贝叶斯网络广泛应用于计算机视觉、计算机感知,并将理论用于人机交互(特别是人类状态监控),情景意识和决策信息融合[15]。

2) 在近十年内,贝叶斯网络的应用趋于成熟化,它在实际应用中发挥了巨大作用并且体现了深厚的潜力。比如在医学分析领域中有 PATHFINDER 系统,特别事在对“淋巴结”分析和诊断中有比较好的表现;CPCSBN 远程医疗系统,总共有四百多个节点和九百多条弧,相比较于世界上主要的远程诊疗方法,其性能处于领先地位;ALARM 网,具有三十多个节点和四十多条弧,其描述了在手术室中潜在细菌的问题,经常用来做贝叶斯网络在结构学习中检验的标准[16]。Ji Q. 等用动态贝叶斯网络分类器进行人脸识别的研究;Pedro Larranaga 等将贝叶斯网络分类器用于诊断和治疗帕金森综合征;Hudson 等将贝叶斯网络用于模拟军事对抗和预测;C. Bielza 等利用贝叶斯网络确定早期阿尔茨海默病的标志和病发机制。

3) 近年来,对于贝叶斯网络的研究越来越深入,但学者们大都分布在美国、西班牙、法国和中国这几个国家,中国学者的整体论文数量较多,但高水平、影响因子较高的文章却不多,这也跟国内目前比较推崇神经网络和深度学习有关。2018 年“贝叶斯网络之父”Pearl 推出自己的新书《The Book of Why》,81 岁的 Pearl 在新书中描述了一个未来世界真正的人工智能是如何进行思考并且做出决策的。他认为让人工智能真正智能起来的关键因素是进行因果推理。机器不能仅仅把发烧和疟疾联系起来,更要能推断

出疟疾是可以引起发烧。只要这种因果框架被建立起来, 机器就有能力提出反事实的问题——询问在干预介入的情况下因果关系会如何变化, Pearl 认为这才是人工智能科学思考的基础。

参考文献

- [1] 曹杰. 贝叶斯网络结构学习与应用研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2017.
- [2] 乔萨础拉, 努尔布力, 苏芮. 数据可视分析研究现状与发展趋势的图谱分析[J]. 现代电子技术, 2018, 41(14): 161-165+169.
- [3] 陈悦, 陈超美, 刘则渊, 胡志刚, 王贤文. CiteSpace 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(2): 242-253.
- [4] Chaturvedi, I., Ragusa, E., Gastaldo, P., *et al.* (2017) Bayesian Network based Extreme Learning Machine for Subjectivity Detection. *Journal of the Franklin Institute*, **355**, 1780-1797. <https://doi.org/10.1016/j.jfranklin.2017.06.007>
- [5] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. 软件学报, 2018, 29(10): 3068-3090.
- [6] Anton-Sanchez, L., Larrañaga, P., Benavides-Piccione, R., Fernaud, I., De Felipe, J. and Bielza, C. (2017) Three Dimensional Spatial Modeling of Spines along Dendritic Networks in Human Cortical Pyramidal Neurons. *PLoS ONE*, **12**, e0180400. <https://doi.org/10.1371/journal.pone.0180400>
- [7] Zhang, J., Ahlbrand, B., Malik, A., *et al.* (2016) A Visual Analytics Framework for Microblog Data Analysis at Multiple Scales of Aggregation. *Computer Graphics Forum*, **35**, 441-450. <https://doi.org/10.1111/cgf.12920>
- [8] Lau, C.L. and Smith, C.S. (2016) Bayesian Networks in Infectious Disease Eco-Epidemiology. *Reviews on Environmental Health*, **31**, 173-177. <https://doi.org/10.1515/reveh-2015-0052>
- [9] Karimi, I. and Salahshoor, K. (2012) A New Fault Detection and Diagnosis Approach for a Distillation Column Based on a Combined PCA and ANFIS Scheme. 2012 24th Chinese Control and Decision Conference, Taiyuan, 23-25 May 2012, 3408-3413. <https://doi.org/10.1109/CCDC.2012.6244542>
- [10] Cai, B., Huang, L. and Xie, M. (2017) Bayesian Networks in Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, **13**, 2227-2240. <https://doi.org/10.1109/TII.2017.2695583>
- [11] 邱均平, 韩雷. 近十年来我国知识工程研究进展与趋势[J]. 情报科学, 2016, 34(6): 3-9.
- [12] 耿杨, 邵苏杰, 郭少勇, 喻鹏. 基于可见损伤持续时间贝叶斯网络的视频 QoE 评估方法[J]. 通信学报, 2017, 38(6): 136-141.
- [13] 李玉兰. 基于贝叶斯网络的列控车载设备故障诊断研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2016.
- [14] 陈晓艳, 董朝轶. 动态贝叶斯网络结构搜索法辨识生物神经网络连接[J]. 生命科学研究, 2017, 21(6): 527-533.
- [15] 王双成, 高瑞, 杜瑞杰. 具有超父结点时间序列贝叶斯网络集成回归模型[J]. 计算机学报, 2017, 40(12): 2748-2761.
- [16] 赵建喆. 具有认知特性的贝叶斯网络结构学习方法研究[D]: [博士学位论文]. 沈阳: 东北大学, 2015.