生物医学领域人工智能研究热点分析

武文斐 田芳慧 宋 旭 王晓璇 马明峰 张 庆

（济宁医学院医学信息工程学院 日照 276800）

〔摘要〕 对 PubMed 数据库收录的近 3 年生物医学领域人工智能相关文献中出现的高频主题词进行聚类分析，根据聚类结果梳理分析当前该领域研究热点，包括临床医学、生物信息两个方面的 5 个主题，即疾病的计算机辅助诊断、中风康复脑状态分类和阿尔兹海默病检测、生物识别、生物神经系统研究。

〔关键词〕 人工智能；生物医学；研究热点；聚类分析

〔中图分类号〕 R－056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673－6036.2020.02.006

Analysis of Study Hotspots of Artificial Intelligence in Biomedical Field

WU Wenfei, TIAN Fanghui, SONG Xu, WANG Xiaoxuan, MA Mingfeng, ZHANG Qing, School of Medical Information Engineering, Jining Medical University, Rizhao 276800, China

〔**Abstract**〕 A cluster analysis is carried out on the high－frequency subject terms appearing in the literatures related to Artificial Intelligence (AI) in biomedical field in recent three years included in PubMed database. According to the clustering results, the paper analyzes the current study hotspots in this field, including five topics of clinical medicine and bioinformation, namely, computer－aided diagnosis of diseases, classification of brain states for stroke rehabilitation and detection of Alzheimer's disease, biometrics, biological nervous system study.

〔**Keywords**〕 Artificial Intelligence (AI); biomedicine; study hotspots; cluster analysis

1 引言

随着科技的发展和时代的进步，人工智能技术不断发展，已经对各行各业产生深远影响，人们对于人工智能在各个领域的应用也在不断探索中。其中生物医学作为一门极其重要的科学，对目前和未来医学的发展与应用具有推动和引导作用，在医学领域发挥着十分重要的作用。通过人工智能对生物医学领域开展研究，开发一系列与疾病诊断、治疗和康复等方面有关的人工智能相关技术和手段，以促进生物医学研究和应用，为推动生物医学领域的变革和发展做出重要贡献。除生物医学领域外人工智能在生物学的许多方面都有研究，开发了一些生物学的识别和研究分析方法，推动生物学的发展。人工智能对生物医学和生物学领域的研究在临床医学领域也得到实际应用，有助于临床诊断、治疗和康复。本文通过对 PubMed 数据库收录生物学和生物医学领域人工智能相关文献中高频主题词进行共现聚类分析，总结并分析研究热点和现状。

2 资料与方法

2.1 资料来源

在 PubMed 数据库中 Mesh 检索框输入"Artificial Intelligence"，选择"Restrict to MeSH Major Topic"进行加权并检索，限定时间为 2016 年 1 月 1 日－2019 年 5 月 15 日。检索策略为 Search "Artificial Intelligence" [Majr] Filters: Publication date from

〔收稿日期〕 2019－09－18

〔作者简介〕 武文斐，本科生。

2016/01/01 to 2019/05/15。共得到相关文献 9 553 篇，较为全面地涵盖近 3 年生物医学领域人工智能方面的相关文献。

2.2　研究方法

使用 XML 格式套录检索结果，利用书目共现分析系统 BICOMB 软件对文献记录中出现的主要主题词和副主题词进行抽取并统计，将统计结果以出现的频次进行降序排序，设定阈值为≥66，≤2 432，共得到 46 个主题词/副主题词作为高频词。其出现频次最高的前 3 位主题词/副主题词分别为：Neural Networks（Computer）、Machine Learning、Robotics。得到的 46 个高频词占所有与生物医学人工智能相关主题词的累计比例为 15%。利用 BICOMB 软件统计高频词在每篇文献中出现的情况，生成相应的词篇矩阵。将其导入 gCLUTO 软件并对所得到的词篇矩阵进行聚类分析，聚类分析结果可以反映高频词之间的亲疏关系，根据主题词间的语义关系和聚类结果分析人工智能在生物医学领域应用的研究热点。

3　结果

生物医学领域人工智能研究高频主题词聚类可视化结果，见图 1、图 2。其中图 2 的横轴代表文献，纵轴代表聚类的主题词/副主题词。两个词聚到一起的距离越短则代表两者的关系越密切。首先，根据每类高频词的含义及其之间的语义关系总结出每类主题词所代表研究热点，即生物医学领域人工智能研究热点，如主题词机器人/仪器（Robotics/Instrumentation）和外骨骼装置（Exoskeleton Device）距离较近，关系密切，先聚成一类；模型，生物（Models，Biological）再与前面两个词合成一类，依此类推。通过分析这些主题词的语义关系能得出其所代表的类团含义标签，综合各个类别的类标签可以得出该主题的研究热点。其次，利用 gCLUTO 软件对各类成员聚类贡献率指标进行计算，包括描述度和区分度，选择对每类形成贡献最大的来源文献作为表示该类内容的类标签文献[1]。然后再对文献内容进一步分析，进而阐释该类研究方向的具体内容。类成员聚类贡献率指标，见表 1。

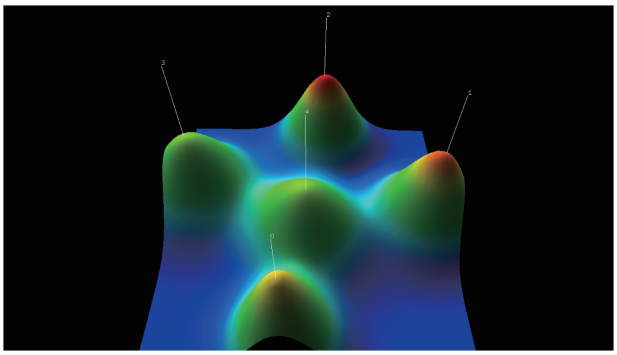

图 1　生物医学领域人工智能研究高频主题词

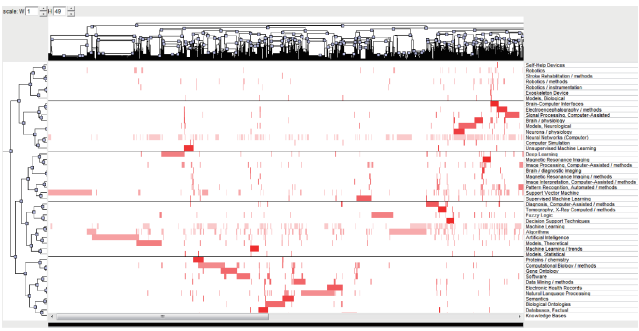

图 2　生物医学领域人工智能研究高频主题词共现聚类

表 1　类成员对聚类贡献率指标

类群	聚类对象数量	类内对象间平均相似度	类间平均相似度	描述度（%）	区分度（%）
0	7	0.200	0.006	27169361 0.8	27169361 0.4
				26478988 0.7	26478988 0.4
				25600057 0.7	25600057 0.4
				27278136 0.7	27278136 0.4
1	9	0.176	0.016	29211653 1.0	29211653 0.6
				26941324 1.0	26941324 0.6
				29250108 0.8	29250108 0.5
				28573983 0.8	28573983 0.5
2	9	0.172	0.017	26193379 0.9	26193379 0.5
				27343830 0.8	27343830 0.5
				29849545 0.7	27776438 0.4
				27776438 0.6	27781951 0.4
3	11	0.148	0.010	27173520 0.7	27173520 0.4
				28371781 0.6	28436885 0.3
				28436885 0.6	28371781 0.3
				28861720 0.6	29186351 0.3
4	10	0.135	0.020	29286609 0.7	29286609 0.4
				30194200 0.5	30194200 0.3
				26839532 0.5	26621986 0.3
				26621986 0.4	28471111 0.3

4 分析

4.1 临床医学领域

4.1.1 计算机辅助诊断 主题词包括 Diagnosis, Computer - Assisted /methods, Tomography, X - Ray Computed/methods, Fuzzy Logic, Decision Support Techniques, Machine Learning, Algorithms, Artificial Intelligence, Models, Theoretical, Machine Learning, Trends, Models, Statistical。在临床诊疗方面，运用计算机对疾病进行辅助诊断是至关重要的，人工智能正在成为医学诊断中的重要工具，可以用于建立诊断模型、计算机断层扫描[2-4]等。如基于模糊局部信息聚类方法的肺癌诊断方法；用卷积神经网络预测脑转移的立体定向放射治疗反应；基于动态不确定因果图模型的黄疸智能诊断等。人工智能技术对疾病诊断的应用在科研和医疗工作中起到重要作用，其不仅为疾病的正确诊疗提供保障，使疾病诊断过程更准确、诊断结果更可靠、后续治疗更高效，还可以为智能诊断奠定基础，提高临床医疗水平，降低医生工作强度，提高医务人员工作效率。

4.1.2 外骨骼和中风康复 主题词包括 Self - Help Devices, Robotics, Stroke Rehabilitation /methods, Robotics / methods, Robotics / instrumentation, Exoskeleton Device, Models, Biological。中风患者在经过治疗后往往需要进行康复训练。通过人工智能结合外骨骼技术或机器人辅助器械可以应用于中风后的康复治疗，包括帮助上肢康复、步态训练干预与步态康复和定制康复方法。可以降低负重行走能量消耗的多关节软外装也为将来在其他应用中增强或恢复步态的辅助设备的研发奠定基础[5-8]。人工智能辅助康复能够为当前中风康复方面的研究提供一定指导，对推进其发展提出建议。

4.1.3 阿尔茨海默病和脑状态分类 主题词包括 Deep Learning, Magnetic Resonance Imaging, Image Processing, Computer - Assisted/methods, Brain/diagnostic imaging, Magnetic Resonance Imaging/methods, Image Interpretation, Computer - Assisted/methods, Pattern Recognition, Automated/methods, Support Vector Machine, Supervised Machine Learning。随着近年来人工智能广泛应用于医疗科技领域，在检测和治疗阿尔茨海默综合症方面，利用人工智能对脑状态的分类和辅助诊断技术已达到新水平，不仅包括对正常脑和病理脑的分类[9-10]，还有对阿尔茨海默病、轻度认知障碍的检测和诊断[11-13]。如磁共振成像扫描中基于支持向量机的 AdaBoost 算法和二维平稳小波变换的病理脑检测，半监督学习稀疏表示分类器的平均系数方法对大脑状态进行解码、分类，多重分形特征与传统静息状态功能磁共振特征相结合的多重分形分析，基于支持向量机、利用磁共振成像多层次特征来区分轻度认知障碍患者的方法可以改善传统基于特征的阿尔茨海默病分类，在阿尔茨海默病研究中有重要作用。以上 3 类主题的共同特征是将人工智能与其他技术相结合，将其应用于疾病的诊断或治疗后的康复活动，在临床上有着较为重要的作用。通过人工智能技术不仅能够为医务工作者提供智能化服务，从而减少工作时间，提高诊断准确性，有效降低误诊率，较大幅度减轻医务人员工作强度，提高诊疗效率，对细化疾病分类也起到推动作用，使诊断更专、更细、更深。此外运用人工智能技术提高治疗后的患者的康复效果，为智能化康复训练和定制康复方法奠定基础。因此，在临床医学领域中人工智能对疾病的检测和治疗、预后和康复起到重要作用，有助于提高医务人员工作质量和效率。

4.2 生物信息方面

4.2.1 生物识别 主题词包括：Computational Biology / methods, Gene Ontology, Proteins / chemistry Semantics, Biological Ontologies, Gene Ontology, Electronic Health Records, Natural Language Processing, Databases, Factual, Knowledge Bases, Data Mining /methods。研究内容为运用计算机生物学、基因本体论、生物本体论、自然语言处理、知识库等方法研究蛋白质之间的语义关系[14-16]、在电子健康记录数据库中识别疾病[17]。如利用基因本体图的信息内容和拓扑性质评估蛋白质之间的语义相似性，利用基因本体注释研究蛋白质序列相似度与语

义相似度的相关性等；又如利用自然语言处理在数据仓库中准确识别脂肪肝病。运用人工智能研究蛋白质间的语义关系及识别疾病有助于科研和医务人员更加高效地开展工作，为疾病诊断提供可靠依据，为疾病治疗奠定良好基础，也为疾病康复提供有效手段。通过对蛋白质语义关系的识别帮助研究人员更好地获取生物信息，充分了解生物表达的内容，为生物信息和生物医学领域研究提供支持，为促进生物学和生物医学发展奠定基础。

4.2.2 生物神经系统研究 主题词包括 Brain - Computer Interfaces, Electroencephalography /methods, Signal Processing, Computer - Assisted, Brain /physiology, Models, Neurological, Neurons /physiology Neural Networks (Computer), Computer Simulation, Unsupervised Machine Learning, Deep Learning, Magnetic Resonance Imaging, Image Interpretation, Computer - Assisted/methods, Pattern Recognition, Automated/methods, Support Vector Machine, Supervised Machine Learning。研究内容是运用计算机辅助、计算机模拟、自动化等人工智能研究方法对生物神经系统在神经网络学习、动物神经网络连接、探索语义关系、听力识别等领域进行研究[18-21]。如加标神经元可以通过聚集标签学习发现预测特征；通过对无脊椎和有脊椎动物神经网络接口连接等方面的研究实现在临床、康复等方面的神经网络应用；通过对事件相关电位的无监督分析以及对神经生理数据的分析来探索语义记忆组织；通过对卷积神经网络的识别和分类，诱发三维脑电输出表示，以识别大脑 - 计算机接口，实现听力识别等。运用人工智能研究生物神经系统，帮助研究人员深入了解神经网络学习、神经网络接口连接等方面的知识，为神经系统方面的研究奠定基础。此外辅助研究语义记忆组织以及实现人工智能的听力识别，实现在学习、临床、康复等方面的人工智能神经网络应用。以上两类主题的共同特征是应用人工智能对生物信息领域进行探索，研究疾病诊断、治疗、康复等方面的方法和手段。通过对疾病识别、蛋白质语义关系和生物神经系统的研究和探索使人工智能在生物医学和生物学领域更加充分、有效地发挥作用，使更便捷的生物医学研究、治疗方法等尽快应用于实际研究、治疗过程中，提高医疗水平，促进生物医学和生物学的研究和发展。

5 结语

随着近年来科技的不断发展，人工智能在生物医学领域得到广泛研究与应用，开发新的研究方法，为生物医学的发展提供可靠的技术保障，从而更好地为疾病预防和诊断、患者康复以及基础研究服务。通过对高频主题词的聚类分析可以总结出在生物医学领域中人工智能主要应用于临床医学、生物信息两个方面，可分为疾病的计算机辅助诊断、生物识别、疾病康复、脑状态分类和阿尔茨海默病的检测诊断及生物神经系统 5 个主题。生物医学研究领域相当广泛，目前的人工智能只是初步应用于该领域，随着科学技术的不断发展必然会不断取得新的突破，创造出更大的价值。未来人工智能会向着更高深和宽泛的方向发展，也会更普遍地应用于更多领域，医疗事业会朝着更加现代化、智能化的方向发展。

参考文献

1 张庆，吕少妮，轩扬．本体在生物医学领域中应用研究热点分析［J］．医学信息学杂志，2019，40（1）：63 - 67.

2 Hao S R, Geng S C, Fan L X, et al. Intelligent Diagnosis of Jaundice with Dynamic Uncertain Causality Graph Model［J］. J Zhejiang Univ Sci B, 2017, 18（5）: 393 - 401.

3 Lavanya M, Kannan P M. Lung Lesion Detection in CT Scan Images Using the Fuzzy Local Information Cluster Means (FLICM) Automatic Segmentation Algorithm and Back Propagation Network Classification［J］. Asian Pac J Cancer Prev, 2017, 18（12）: 3395 - 3399.

4 Cha Y J, Jang W I, Kim M S, et al. Prediction of Response to Stereotactic Radiosurgery for Brain Metastases Using Convolutional Neural Networks［J］. Anticancer Res, 2018, 38（9）: 5437 - 5445.

5 Panizzolo F A, Galiana I, Asbeck A T, et al. A Biologically - inspired Multi - joint Soft Exosuit that Can Reduce the Energy Cost of Loaded Walking［J］. J Neuroeng Rehabil,

2016, 13 (1): 43.

6 Babaiasl M, Mahdioun S H, Jaryani P, et al. A Review of Technological and Clinical Aspects of Robot – aided Rehabilitation of Upper – extremity After Stroke [J]. Disabil Rehabil Assist Technol, 2016, 11 (4): 263 – 280.

7 Louie D R, Eng J J. Powered Robotic Exoskeletons In Post – stroke Rehabilitation of Gait: a scoping review [J]. J Neuroeng Rehabil, 2016, 13 (1): 53.

8 Fukuda H, Morishita T, Ogata T, et al. Tailor – made Rehabilitation Approach Using Multiple Types of Hybrid Assistive Limb Robots for Acute Stroke Patients: a pilot study [J]. Assist Technol, 2016, 28 (1): 53 – 56.

9 Zhang J, Zhang C, Yao L, et al. Brain State Decoding Based on fMRI Using Semisupervised Sparse Representation Classifications [J]. Comput Intell Neurosci, 2018 (2018): 3956536.

10 Nayak D R, Dash R, Majhi B. Stationary Wavelet Transform and AdaBoost with SVM Based Pathological Brain Detection in MRI Scanning [J]. CNS Neurol Disord Drug Targets, 2017, 16 (2): 137 – 149.

11 Long Z, Jing B, Yan H, et al. A Support Vector Machine – based Method to Identify Mild Cognitive Impairment with Multi – level characteristics of Magnetic Resonance Imaging [J]. Neuroscience, 2016 (331): 169 – 176.

12 Ni H, Zhou L, Ning X, et al. Exploring Multifractal – based Features for Mild Alzheimer´s Disease Classification [J]. Magn Reson Med, 2016, 76 (1): 259 – 269.

13 Khedher L, Illan I A, Gorriz J M, et al. Independent Component Analysis – support Vector Machine – based Computer – aided Diagnosis System for Alzheimer´s with Visual Support [J]. Int J Neural Syst, 2017, 27 (3): 1650050.

14 Dutta P, Basu S, Kundu M. Assessment of Semantic Similarity between Proteins Using Information Content and Topological Properties of the Gene Ontology Graph [J]. IEEE/ACM Trans Comput Biol Bioinform, 2018, 15 (3): 839 – 849.

15 Lai P T, Lo Y Y, Huang M S, et al. BelSmile: a biomedical semantic role labeling approach for extracting biological expression language from text [EB/OL]. [2019 – 09 – 16]. https: //dblp. org/rec/journals/biodb/LaiLHHT16.

16 Ikram N, Qadir M A, Afzal M T. Investigating Correlation between Protein Sequence Similarity and Semantic Similarity Using Gene Ontology Annotations [J]. IEEE/ACM Trans Comput Biol Bioinform, 2018, 15 (3): 905 – 912.

17 Redman J S, Natarajan Y, Hou J K, et al. Accurate Identification of Fatty Liver Disease in Data Warehouse Utilizing Natural Language Processing [J]. Dig Dis Sci, 2017, 62 (10): 2713 – 2718.

18 Carabez E, Sugi M, Nambu I, et al. Convolutional Neural Networks with 3D Input for P300 Identification in Auditory Brain – Computer Interfaces [J]. Comput Intell Neurosci, 2017 (2017): 8163949.

19 Van Vliet M, Van Hulle M M, Salmelin R. Exploring the Organization of Semantic Memory through Unsupervised Analysis of Event – related Potentials [J]. J Cogn Neurosci, 2018, 30 (3): 381 – 392.

20 Broccard F D, Joshi S, Wang J, et al. Neuromorphic Neural Interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems [J]. J Neural Eng, 2017, 14 (4): 41002.

21 Gutig R. Spiking Neurons can Discover Predictive Features by Aggregate – label Learning [J]. Science, 2016, 351 (6277): b4113.

关于《医学信息学杂志》启用“科技期刊学术不端文献检测系统”的启事

为了提高编辑部对于学术不端文献的辨别能力，端正学风，维护作者权益，《医学信息学杂志》已正式启用“科技期刊学术不端文献检测系统”，对来稿进行逐篇检查。该系统以《中国学术文献网络出版总库》为全文比对数据库，可检测抄袭与剽窃、伪造、篡改、不当署名、一稿多投等学术不端文献。如查出作者所投稿件存在上述学术不端行为，本刊将立即做退稿处理并予以警告。希望广大作者在论文撰写中保持严谨、谨慎、端正的态度，自觉抵制任何有损学术声誉的行为。

《医学信息学杂志》编辑部