DOI:10.16652/j.issn.1004-373x.2019.21.025

语音合成技术研究现状与发展趋势的计量分析

热衣扎·哈那提，努尔布力
（新疆大学 信息科学与工程学院，新疆 乌鲁木齐 830046）

摘 要：以Web of Science中近20年收录的1 846篇语音合成领域文献为研究对象，采用文献计量分析方法，利用CiteSpace可视化分析工具绘制知识网络图谱，系统回顾该领域的研究概况及研究热点，理清研究发展脉络。研究发现，语音合成的理论研究已经相对成熟，神经网络成为语音合成领域里使用的新兴技术。另外，在该领域中日本、中国、英国及美国的科研机构具有较强的科研能力。通过上述工作，希望为我国语音合成领域的研究提供进一步的参考和帮助。

关键词：语音合成；文献计量分析；CiteSpace；知识网络图谱；研究现状；发展脉络
中图分类号：TN912.3-34 **文献标识码**：A **文章编号**：1004-373X(2019)21-0116-04

Bibliometric analysis of research status and development trend of speech synthesis technology

HANAT Riza, Nurbol
(College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China)

Abstract: The literatures of 1846 speech synthesis fields collected in the Web of Science in the last 20 years are taken as the research object. The bibliometric analysis method is used. The CiteSpace visual analysis tool is used to draw the knowledge network atlas. The research and research hotspots in the field are systemastically reviewed and the research development context is sorted out. It is found in the study that the theoretical research of speech synthesis has been relatively mature, and the neural network becomes a research hotspot in the speech synthesis field in recent years. In addition, scientific research institutions in the United States, Japan, China, and the United Kingdom have strong scientific research capabilities in this field. Through, It is hoped that the above work can provide further reference and help for the study in the field of Chinese speech synthesis.

Keywords: speech synthesis; bibliometric analysis; CiteSpace; knowledge network atlas; research status; development context

0 引 言

语音合成技术作为人机语音交互的核心技术，被越来越多的研究者给予关注和重视。语音合成技术的发展已有几十年的历史，取得了很多优秀的研究成果。虽然国内很多专家从不同的视角对语音合成进行了总结和综述，但还没有从知识图谱的角度对语音合成领域进行总结分析。鉴于此，本文利用CiteSpace工具对通过Web of Science平台收集到的关于语音合成的核心文献进行计量分析并绘制知识图谱，从宏观角度阐述以下两个问题：国内外近20年来在语音合成领域的研究概况以及主要研究热点。

1 数据来源和研究方法的说明

1.1 数据来源

本文研究的文献来源于信息检索平台Web of Science的核心数据库，数据采用以下的方式收集：

1）标题词检索方法：TI="speech synthesis"OR"text to speech"OR"voice synthesis"OR"concept to speech"OR"intention to speech"OR"text to voice"；

2）时间跨度：1999—2018年；

3）文献类型：期刊（ARTICLE）和会议论文

收稿日期：2018-10-15 **修回日期**：2018-11-20
基金项目：国家自然科学基金项目（61303231）资助；新疆维吾尔自治区重点实验室开放课题（2017D04002）
Project Supported by National Natural Science Foundation of China (61303231), Opening Foundation of the Xinjiang Uygur Autonomous Region Key Laboratory (2017D04002)

（PROCEEDINGS PAPER）。共得到1 846篇关于语音合成领域的核心文献并下载每个文献的28条记录信息，包括标题、作者、摘要、关键词、参考文献等。

1.2　研究方法的说明

本文主要采用计量分析和图谱分析方法，通过它们揭示相关领域的知识来源和发展规律，并把知识结构关系和演化规律用图形的方式呈现出来。可视化工具CiteSpace就是可以用于追踪研究领域热点和发展趋势的文献计量分析工具。本文通过CiteSpace对1 846篇文献进行研究机构的合作网络分析、研究热点的演化分析以及高共被引文献的统计分析。

2　研究概况

2.1　主要研究机构分析

通过对语音合成领域的文献发表量的研究机构进行基本情况统计后发现发文量超过9篇以上的机构有18所。表1列出的是文献量排名前10的研究机构。图1是研究机构直接的合作网络关系图，其中连线代表两个研究机构之间有合作关系；文字大小代表发文量的多少，文字越大发文量越多，文字越小发文量越少。

表1　发文量Top10的研究机构

Table 1　Publications of Top10 research institutions

研究机构	文献数量	国家
爱丁堡大学	80	英国
东京工业大学	59	日本
名古屋工业大学	56	日本
中国科学技术大学	56	中国
中国科学院	45	中国
剑桥大学	33	英国
东京大学	28	日本
西波希米亚大学	27	捷克
卡耐基梅隆大学	27	美国
台湾成功大学	26	中国

通过表1得知，Top10榜单里的研究机构共来自5个国家，分别是日本3所，中国3所，英国2所，捷克和美国各1所。通过对国家发文量的统计，发现日本在语音合成领域里发表的文献量居世界首位，中国和美国的发文量分别排在第二位和第三位。

2.2　主要作者分析

根据基本统计分析，研究文献共涉及到的作者中，发文量超过10篇的作者有58位，发文量超过20篇的作者有16位。发文量排名前10的作者如表2所示。

通过表2的首次发文年份的分布来看，高产作者的首次发文年份最早是从2003年开始的。发文量最多的作者是Yamagishi J，表3列出的高被引文献里该作者的文献有3篇，该3篇文献都与隐马尔科夫模型有关，并结合他的其他文献分析发现，该作者的研究重点主要集中在基于隐马尔科夫模型的语音合成，而从他近几年的文献分析发现他现在的研究重点转向神经网络的研究，该作者在2018年与Wang X等人合著的一篇文献主要研究了深度神经网络在统计参数语音合成中的性能[1]，特别是深层网络能否更好地产生不同声学特征的问题。排在第二位的是作者Tokuda K，该作者在2018年发表的文献[2]里提出了一种基于梅尔倒谱的量化噪声整形方法，提高了基于神经网络的语音波形合成系统的合成语音质量。作者Kobayashi T发文量排在第三位，文献[3]是他近几年与Nose T等人合作的一篇文献，该文献里提出了一种用于语音合成和韵律平衡的紧凑记录脚本的句子选择技术，与传统的句子选择技术相比，该技术所生成的语音参数更接近自然语音的语音参数。

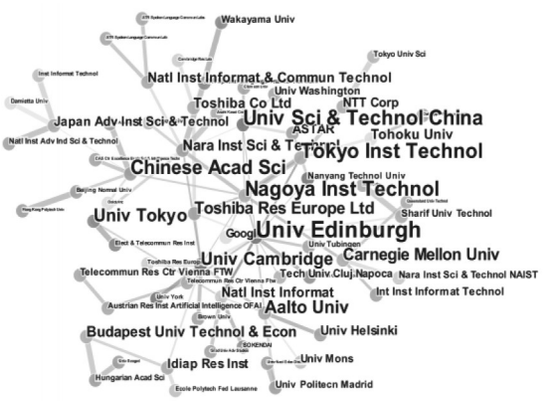

图1　研究机构合作网络图

Fig. 1　Co-research network graph of institutions

表2　高频作者

Table 2　High frequency authors

作者	发文数量	首次发文年份
Yamagishi J	72	2003
Tokuda K	62	2003
Kobayashi T	55	2003
Ling ZH	53	2008
King S	48	2003
Nose T	39	2008
Tao JH	39	2004
Dai LR	36	2008
Nankaku Y	33	2008
Toda T	31	2004

2.3　高被引文献分析

高被引文献是一个研究领域的重要知识来源，反映某一学科的研究水平、发展方向，是探究热点主题、研究演化的重要依据[4]。表3列出的是被引频次较多的10篇文献，被引频次主要来自于本论文研究的数据。

表3 被引频次较多的文献

Table 3 Literatures that has been cited frequently

文 献	第一作者	关注点	频次	年份
Statistical parametric speech synthesis	Zen H	统计参数语音合成	109	2009
Analysis of Speaker Adaptation Algorithms for HMM-Based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm	Yamagishi J	说话人自适应、隐马尔科夫模型	56	2009
A Speech Parameter Generation Algorithm Considering Global Variance for HMM-Based Speech Synthesis	Toda T	语音参数生成、隐马尔科夫模型	47	2007
Details of the NitechHMM-Based Speech Synthesis System for the Blizzard Challenge 2005	Zen H	隐马尔科夫模型	39	2007
A Hidden Semi-Markov Model-Based Speech Synthesis System	Zen H	隐半马尔可夫模型	36	2007
Speech Synthesis Based on Hidden Markov Models	Tokuda K	隐马尔科夫模型	34	2013
Voice Conversion Based on Maximum-Likelihood Estimation of Spectral Parameter Trajectory	Toda T	语音转换、极大似然估计	29	2007
Statistical Parametric Speech Synthesis Using Deep Neural Networks	Zen H	深度神经网络、统计参数	28	2013
Robust Speaker-Adaptive HMM -Based Text-to-Speech Synthesis	Yamagishi J	说话人自适应、隐马尔科夫模型	27	2009
Average-Voice-Based Speech Synthesis Using HSMM-Based Speaker Adaptation and Adaptive Training	Yamagishi J	平均语音、隐半马尔可夫模型	27	2007

作者Zen H等人发表的文献《Statistical parametric speech synthesis》的被引次数最多[5],该文综述了统计参数语音合成中常用的技术,对统计参数语音合成技术和传统的单元选择合成技术进行比较,总结了统计参数语音合成的优点和缺点并对未来工作进行展望。作者Yamagishi J等人发表的文献[6]排在第二位,本文提出新的适应算法约束结构最大线性回归,该方法在语音合成中获得了更好、更稳定的说话人自适应,具有很强的实用性和有效性。文献[7-8]是表3里2013年发表的两篇文献,文献[7]讨论了基于隐马尔科夫模型的语音合成技术在改变说话者身份、情感和说话风格方面的灵活性;文献[8]提出基于深度神经网络的统计参数语音合成方法,使用深度神经网络来解决传统统计参数语音合成方法的一些局限性。

通过表3的关注点来看,基于隐马尔科夫模型的语音合成技术是语音合成领域的重点语音合成技术,说话人自适应技术成为语音合成领域较为重要的研究技术,而深度神经网络是近几年语音合成领域里使用的新兴技术。

3 研究热点

关键词是文献主题内容的高度提炼,对关键词出现的变化进行分析可以了解各时期的研究热点[9]。表4列出的是频次较多、中心性较高、激增值较大的按首次激增年份排序的关键词。

1)频次(Freq)指标计量分析

通过图2,频次较多的关键词“hidden markov model”“text to speech”“unit selection”的首次研究年份集中在1999—2002年,这些研究为语音合成技术的发展奠定了基础。到2005年,关键词“hmm-based speech synthesis”出现,隐马尔科夫模型被用到语音合成研究里面,基于隐马尔科夫模型的语音合成技术从该时期开始研究。到2006年,语音转换技术应用到语音合成领域里,进一步促进了语音合成技术的发展。

表4 关键词Top12的排名统计

Table 4 Rank statistics of keywords in Top12

关键词	频次	中心性	激增值	激增年份
rule	11	0.02	4.162 5	1999
concatenative speech synthesis	23	0.07	3.342 5	2001
text to speech	132	0.08	2.480 9	2002
system	56	0.16	2.706 5	2004
hidden markov model	236	0.09	3.214 7	2006
unit selection	95	0.07	11.397 5	2006
voice conversion	35	0.03	2.950 1	2009
hmm-based speech synthesis	135	0.05	10.286 1	2010
speaker adaptation	66	0.04	4.113 1	2013
deep neural network	38	0.02	9.372 8	2015
recurrent neural network	13	0.02	4.757 8	2015
short term memory	6	0.01	3.141 5	2016

2)中心性(Centrality)指标计量分析

通过表4的关键词的中心性结合图2发现,“system”“hidden markov model”“text to speech”等关键词的中心

性相比其他关键词的中心性较高，首次出现的年份较早，该结果表示系统、隐马尔科夫模型和文本到语音的研究在语音合成领域里研究的时间较长，是较为重要的研究方向。关键词“speaker adaptation”“concatenative speech synthesis”“unit selection”的中心性都大于0.02，说话人自适应是语音合成技术的核心研究部分，级联语音合成受单元选择中使用的单元的库存支配达到高度自然的合成语音质量，单元选择是语音合成领域一个较为重要的研究热点，文献[10]提出的基于隐马尔科夫模型的语音合成方法就用到单元选择。

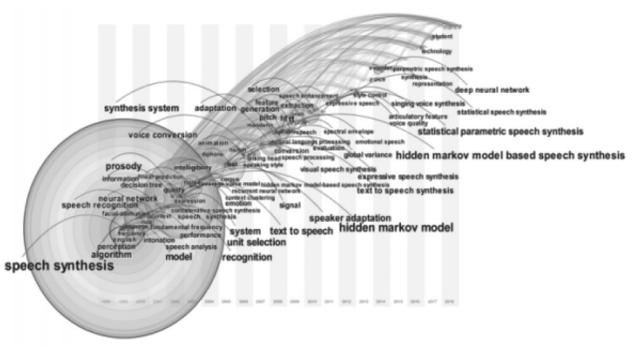

图2　关键词时间分布图

Fig. 2　Time distribution diagram of key words

3）激增（Burst）指标计量分析

激增指数的关注点是单个主题的自身发展变化过程，可以展示热点主题的凸显性。通过表4关键词的激增值和开始激增年份发现，1999—2005年主要的研究主题是围绕规则、文本到语音和语音处理等，该时期的大部分研究工作都在基础的核心部分研究；2006—2014年，研究主题的关注点在语音合成技术的模型，基于隐马尔科夫模型的语音合成技术成为重点，语音转换和说话人自适应技术受到了前所未有的重视；2014—2018年，神经网络成为语音合成领域重要的研究方向，深度学习在语音合成领域的应用进一步促进该领域的快速发展。

通过前文的分析和研究发现，数据可视分析研究的发展分为三个阶段：1999—2005年，初步发展时期；2006—2014年，快速发展时期；2015年—至今，深入发展时期，如表5所示。

4　结　语

国际语音合成领域的研究文献质量不断在稳步提升，日本、中国和英国的一些研究机构在国际上发文量多，与其他研究机构合作关系较密切。基于隐马尔科夫模型的语音合成是该领域的研究重点，而近几年语音合成领域开始使用神经网络技术，解决传统语音合成方法遇到的问题。目前，语音合成领域的研究越来越多，分支越来越细，在未来的发展上，语音合成领域的研究将不断深入，会有越来越多不同领域的技术应用到语音合成领域。

表5　阶段分析表

Table 5　Stage analysis list

时间段	1999—2005年	2006—2014年	2015年—至今
研究热点	规则、级联语音合成、语音处理、文本到语音	隐马尔科夫模型、单元选择、生成过程模型、表达性语音合成	神经网络、深度神经网络、递归神经网络、短期记忆
研究脉络	进行语音技术的探索和研究，推广理念和界定概念，逐渐形成该领域的自身特点	研究方向越来越多，研究分支越来越细，有了成熟的理论知识，并取得了优秀的实践成果	研究越来越深入，应用的技术越来越多，语音合成技术的快速性、有效性变成该时期的研究重点

参　考　文　献

[1] WANG X, TAKAKI S, YAMAGISHI J. Investigating very deep highway networks for parametric speech synthesis [C]// ISCA Speech Synthesis Workshop. [S. l.]: ISCA, 2016: 166-171.

[2] YOSHIMURA T, HASHIMOTO K, OURA K, et al. Mel-cepstrum-based quantization noise shaping applied to neural-network-based speech waveform synthesis [J]. IEEE/ACM transactions on audio speech & language processing, 2018(99): 1.

[3] NOSE T, ARAO Y, KOBAYASHI T, et al. Sentence selection based on extended entropy using phonetic and prosodic contexts for statistical parametric speech synthesis [J]. IEEE/ACM transactions on audio speech & language processing, 2017, 25(5): 1107-1116.

[4] 杨良斌，周新丽，刘益佳，等.近10年来国际网络安全领域研究现状与趋势的可视化分析[J].情报杂志，2017，36(1)：92-100.
YANG Liangbin, ZHOU Xinli, LIU Yijia, et al. The specialty visualization study of current trends and issues of international network security fields in recent 10 years [J]. Journal of intelligence, 2017, 36(1): 92-100.

[5] ZEN H, TOKUDA K, BLACK A W. Statistical parametric speech synthesis [J]. Speech communication, 2009, 51(11): 1039-1064.

[6] YAMAGISHI J, KOBAYASHI T, NAKANO Y, et al. Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm [J]. IEEE transactions on audio speech & language processing, 2009, 17(1): 66-83.

[7] TOKUDA K, NANKAKU Y, TODA T, et al. Speech synthesis based on hidden Markov models [J]. Proceedings of the IEEE, 2013, 101(5): 1234-1252.

（下转第124页）

分析图3可知，本文方法对高校师资培训资源的管理灵活性最大值为98.88%，基于资源签名的自动寻优方法和基于神经网络的资源管理方法的资源管理灵活性最大值依次是91.23%，87.98%。由此可知，本文方法的资源管理灵活性最高，说明在云计算Hadoop平台中，本文方法管理下的高校师资培训数据处理效果好。

3 结 论

本文提出云计算Hadoop平台中基于遗传算法的高校师资培训资源管理方法，和其他资源管理方法相比，该方法不单调度精度高达0.95，而且资源利用率高达0.98，除此之外，资源管理灵活性也未低于97%，可为各大高校师资培训资源管理提供有效帮助。

参 考 文 献

[1] 徐占洋，郑克长.云计算下基于改进遗传算法的聚类融合算法[J].计算机应用，2018，38(2)：458-463.
XU Zhanyang, ZHENG Kezhang. Clustering ensemble algorithms based on improved genetic algorithm in cloud computing [J]. Journal of computer applications, 2018, 38(2): 458-463.

[2] 张淑芬，董岩岩，陈学斌.基于云计算平台Hadoop的HKM聚类算法设计研究[J].应用科学学报，2018，36(3)：118-128.
ZHANG Shufen, DONG Yanyan, Chen Xuebin. HKM clustering algorithm design and research based on Hadoop platform [J]. Journal of applied sciences, 2018, 36(3): 118-128.

[3] 马跃，余骋远，于碧辉.基于资源签名与遗传算法的Hadoop参数自动调优系统[J].计算机应用研究，2017，34(11)：24-27.
MA Yue, YU Chengyuan, YU Bihui. Hadoop parameter automatic tuning system based on resource signature and genetic algorithm [J]. Application research of computers, 2017, 34(11): 24-27.

[4] XIONG Y H, HUANG S Z, WU M, et al. A johnson′s-rule-based genetic algorithm for two-stage-task scheduling problem in data-centers of cloud computing [J]. IEEE transactions on cloud computing, 2019, 7(3): 597-610.

[5] 付晓明，王福林，尚家杰.基于多子代遗传算法优化BP神经网络[J].计算机仿真，2016，33(3)：258-263.
FU Xiaoming, WANG Fulin, SHANG Jiajie. Optimized BP neural network algorithm based on multi-child genetic algorithm [J]. Computer simulation, 2016, 33(3): 258-263.

[6] 李佳，李海波.基于遗传算法的资源服务链构建方法[J].小型微型计算机系统，2016，37(9)：1947-1952.
LI Jia, LI Haibo. Building resource service chain based on genetic algorithm [J]. Journal of Chinese computer systems, 2016, 37(9): 1947-1952.

[7] HAMEED A, KHOSHKBARFOROUSHHA A, RANJAN R, et al. A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems [J]. Computing, 2016, 98(7): 751-774.

[8] 宗思光，刘涛，梁善永.基于改进遗传算法的干扰资源分配问题研究[J].电光与控制，2018，25(5)：45-49.
ZONG Siguang, LIU Tao, LIANG Shanyong. Interference resource allocation based on improved genetic algorithm [J]. Electronics optics & control, 2018, 25(5): 45-49.

[9] 马壮壮，束龙仓，季叶飞，等.基于遗传算法的BP神经网络计算岩溶水安全开采量[J].水文地质工程地质，2016，43(1)：22-27.
MA Zhuangzhuang, SHU Longcang, JI Yefei, et al. Calculation of karst water safe yield by using BP neural network based on genetic algorithm [J]. Hydrogeology and engineering geology, 2016, 43(1): 22-27.

[10] SHAHDI-PASHAKI S, TEYMOURIAN E, TAVAKKOLI-MOGHADDAM R. New approach based on group technology for the consolidation problem in cloud computing-mathematical model and genetic algorithm [J]. Computational & applied mathematics, 2016, 37(1): 693-718.

作者简介：牛志梅(1972—)，女，河南唐河人，硕士研究生，讲师，主要研究方向为数据库、软件工程。

（上接第119页）

[8] ZEN H, SENIOR A, SCHUSTER M. Statistical parametric speech synthesis using deep neural networks [C]// IEEE International Conference on Acoustics, Speech and Signal Processing. [S. l.]: IEEE, 2013: 7962-7966.

[9] 庄少霜.近二十年国外认知语言学领域研究的可视化分析：基于CiteSpaceⅡ的计量分析[J].哈尔滨学院学报，2016，37(8)：97-101.
ZHUANG Shaoshuang. Emerging trends in cognitive linguistics (1996—2015) —a quantitative analysis by CiteSpace Ⅱ [J]. Journal of Harbin University, 2016, 37(8): 97-101.

[10] XIA X J, LING Z H, JIANG Y, et al. Hmm-based unit selection speech synthesis using log likelihood ratios derived from perceptual data [J]. Speech communication, 2014, 63-64(3): 27-37.

作者简介：热衣扎·哈那提(1993—)，女，哈萨克族，新疆伊犁人，硕士生，研究方向为语音合成。
努尔布力(1984—)，男，哈萨克族，新疆阿勒泰人，博士，教授，硕士研究生导师，研究方向为网络安全与数据挖掘、自然语言处理。