

· 前沿与综述 ·

基于知网数据的情感识别国内研究情况综述^{*}

胡冬妮^{1 2 3} 王武军³ 王 青³

(¹ 中国科学院声学研究所 语言声学 with 内容理解重点实验室 北京 100190

² 中国科学院大学 北京 100049 ³ 国际关系学院杭州校区 杭州 310015)

摘要: 情感识别研究工作正受到越来越多的关注。为理清国内情感识别研究工作的发展脉络,明晰国内情感识别的研究重点及趋势,特针对知网数据进行了统计分析。通过分析,发现国内情感识别研究可以划分为3个阶段:研究初期、发展期和成熟期;同时明确当前的研究热点仍在语音、生理信号以及多模态情感识别研究,其中语音情感识别研究重点仍是寻找具有更好的分类性能的特征和分类方法,研究难点是维度情感的研究以及标准情感语料库的构建。这些可以为今后的研究工作提供参考。

关键词: 知网数据 情感识别 语音情感识别 分类方法

Review of Domestic Research on Emotion Recognition Based on How-net Data

HU Dongni^{1 2 3}, WANG Wujun³, WANG Qing³

(¹Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China, ²University of Chinese Academy of Sciences, Beijing, 100049, China,

³University of International Relations Hangzhou Campus, Hangzhou, 310015, China)

Abstract: Researchs on Emotion Recognition are getting more and more attention. To clarify the research trend and focus of domestic emotion recognition, we made the statistical analysis based on the how-net data. Through the analysis, we can know that the domestic research on emotion recognition can be divided into three stages: the initial stage, the development period and the mature period. At the same time, the research focus is still in the study of speech, physiological signal and multimodal emotion recognition. And the key parts of speech emotion recognition are still to find the classification methods and the features which has better classification performance. The difficulty lies in the research of dimensional emotion and the construction of standard emotional corpus. And this overview can provide reference for future research work.

Keywords: how-net data emotion recognition speech emotion recognition classification

0 引言

在人工智能发展的如火如荼的今天,如何让机器真正像人一样有情感,并且识别情感,是衡量机器智能化的一个重要内容。

本文于2018-03-82收到,2018-05-29收到修改稿。

^{*} 基金项目: 国家自然科学基金(编号:11590770-4、61650202、U1536117、61671442、11674352、11504406、61601453),国家重点研发计划(编号:2016YFB0801203、2016YFC0800503)和新疆维吾尔自治区重大科技专项(编号:2016A03007-1)。

国外关于情感识别的研究开展的比较早。1972 年,伴随着 C. Williams 等人关于话者情感变化给相应语音的基音轮廓带来影响的研究,开启了语音情感识别的研究先河^[1]。

1988 年,麻省理工学院的 Minsky 教授出版了“The Society of Mind”,书中他首次提出了“计算机的情感能力”这一概念^[2]。

1990 年,麻省理工学院多媒体实验室 Cahn 等人搭建了一套情感识别系统来采集人类各种情感信息(比如生理、面部表情、语音等),以进行情感分析与识别,并让机器针对不同情感做出相应的反应。

1997 年,麻省理工学院的教授 Picard 教授在其著作“Affective Computing”中首次提出情感计算的概念,他认为“情感计算是关于、产生于、影响情感方面的计算”^[3]。

国内对于情感识别虽然起步比较晚,但近年来也产出了一系列重要成果。为理清国内情感识别研究工作的发展脉络,明晰国内情感识别研究的重点与趋势,特针对知网数据进行了统计分析,并对分析结果进行了总结。

本文的结构如下,第一章总结了国内情感识别研究的历史以及发展脉络,第二章通过对语音情感识别历年论文的分析,理清了国内语音情感识别研究重点,第三章为结束语,展望了国内情感识别研究的趋势与方向。

1 研究历史以及发展脉络

在知网上通过对关键词“情感识别”在所有文献中进行“主题”搜索,搜出相关论文共 841 篇,下面按论文资源类型分布、机构分布、会议分布、以及论文研究领域分布等类别分别进行统计,具体统计结果如图 1 - 图 5 所示。

1.1 论文资源类型分布

按论文资源类型将 841 篇论文分为期刊、硕士论文、博士论文、报纸和会议共计 5 个类别。其数量分布如图 1 所示。

从图 1 可以看出,情感识别相关论文中期刊类别最多,占总论文数量的 55.1%,其次是硕士论文,占总论文数量的 34.8%,博士论文占 5.2%,会议论文占 4.5%,报纸类最低,占 0.4%。这几种论文数量随时间变化的情况如图 2 所示。(注:报纸因为数量太少,只有 3 篇,所以没加入图 2 的曲线图中。)

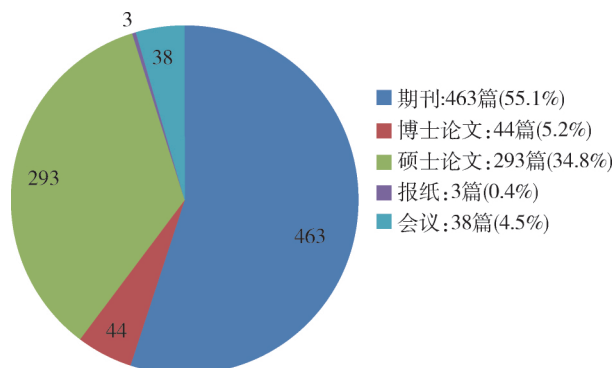


图 1 论文资源类型分布图

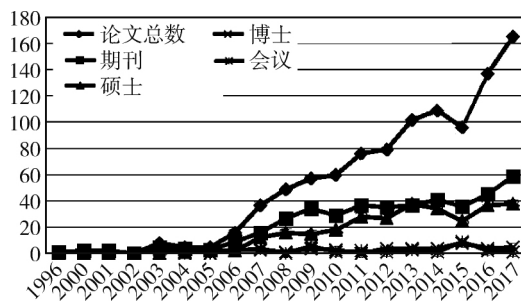


图 2 论文数量随时间变化曲线图

从图 2 可以看出,论文总数以及各类别论文数量整体发展趋势比较一致,呈现中间偶有回落、整体平稳上升的趋势。

论文总数一直稳步增长,在 2003 年达到一个峰值,为 8 篇,随后两年有所回落;在 2006 年激增,由 2005 年的 5 篇增至 16 篇,随后一路上升;在 2014 年达到一个峰值,为 109 篇,然后在 2015 年稍有回落,跌至 96 篇;从 2015 年开始又一路上升。

期刊论文数量变化曲线是平缓上升状态,在 2003 年达到一个峰值,为 5 篇,随后两年有所回落;在 2006

年激增,由 2005 年的 3 篇增至 9 篇,随后一路上升;在 2009 年达到一个峰值,为 34 篇,随后一年稍有回落,跌至 29 篇;从 2010 年开始呈现小范围波动上升状态。

硕士论文数量变化趋势也较为平缓,从 2004 年出现第一篇硕士论文开始,在 2007 年激增,由之前的年均 2 篇激增至 12 篇,随后 3 年在 15 篇左右轻微波动;2011 年达到一个峰值,为 28 篇,随后一年微降至 27 篇;2013 年达到一个峰值,为 38 篇;在 2015 年有明显波谷,降至 25 篇;随后又升至 38 篇。

博士论文数量和会议论文数量曲线与其他曲线稍有区别,博士论文在 2006 年首次出现,会议论文在 2003 年首次出现,然后一直在 5 篇之内轻微起伏波动,但在 2015 年双双达到一个小峰值,分别为 8 篇和 9 篇;随后又降至 5 篇以内。

从图 2 可以看出,国内对情感识别的研究始于 2000 年,东南大学的赵力教授等人发表了相关论文,率先开始了语音情感识别的研究工作;随着 2003 年“第一届中国情感计算机智能交互学术会议”的召开,国内对情感识别的研究工作逐步深入,在该次会议上首次出现关于情感识别的会议论文;2004 年第一次出现关于情感识别的硕士论文;2006 年第一次出现关于情感识别的博士论文。具体论文信息如表 1 所示,汇总了 2000 - 2006 情感识别领域最早的会议、期刊、硕士及博士论文。

表 1 2000 - 2006 情感识别相关论文

作者	文章类别	标题	来源	发表日期
赵力等	期刊	语音信号中的情感特征分析和识别的研究	通信学报	2000
陈建厦	会议	语音情感识别综述	第一届中国情感计算及智能交互学术会议论文集	2003
王治平等	会议	语音信号中情感特征的分析 and 识别	第一届中国情感计算及智能交互学术会议论文集	2003
王青等	会议	基于神经网络的汉语语音情感识别	第一届中国情感计算及智能交互学术会议论文集	2003
王青	硕士论文	基于神经网络的汉语语音情感识别的研究	浙江大学	2004
谢波	博士论文	普通话语音情感识别关键技术研究	浙江大学	2006
刘涛	博士论文	音乐情感认知模型与交互技术研究	浙江大学	2006
林奕琳	博士论文	基于语音信号的情感识别研究	华南理工大学	2006

1.2 情感识别会议分布

841 篇情感识别论文中,会议论文只有 38 篇,占总论文数量的 4.5%。筛选掉跟情感识别研究关联度不高的会议论文后,共有 14 个会议,32 篇论文。其中“全国人机语音通讯学术会议”的论文最多,共有 15 篇,分别发表在 5 届该会议上(2007 - 2017),占总会议论文数量的 46.9%;其次是“中国情感计算及智能交互学术会议”(第 1 届,2003)共有 3 篇,占总会议论文数量的 9.4%,该三篇论文也是知网数据中最早发表情感识别论文的会议;接着是“全国心理学学术会议”在第 16 届(2013)、第 17 届(2014)分别有一篇,“和谐人机环境联合学术会议”在第 3 届(2007)与第 7 届(2011)分别有一篇;最后在“中国控制会议”(第 25 届,2006)、“促进西部发展声学学术交流会”(2007)、“中国西部声学学术交流会”(2012)、“中国高校通信类院系学术研讨会”(2008)、“中国智能计算大会”(2011)、“中国语音学学术会议”(第 11 届,2014)和“中国科协年会”(第 17 届,2015)上各有一篇。

1.3 研究机构分布

从知网上检索出的论文数量上来看,国内高校及科研院所中东南大学是开展情感识别工作较早和较多的,论文中反映出东南大学情感识别主要的研究领域集中在语音情感识别;其次是西南大学,根据论文分布情况可以看到其主要研究领域集中在生理信号情感识别,如图 3 所示。

从图 4 可以看出,参与情感识别研究的高校呈现逐年增长的趋势。这说明随着情感识别研究的兴起,越来越多的高校与科研机构参与到情感识别研究工作中来。并且高校数量的增长呈现了两个小高峰,一个是从 2006 年的 4 所激增至 2007 年的 14 所,增长率为 350%;一个是从 2010 年的 14 所激增至 2011 年 21 所,增长率为 150%。

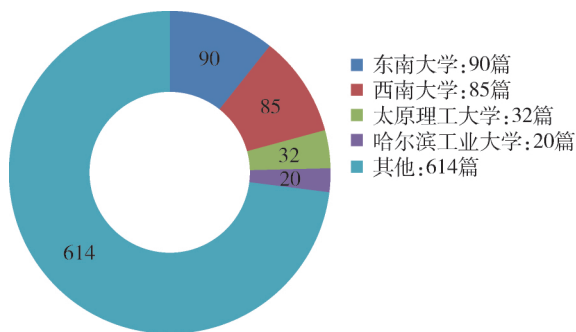


图3 机构分布图

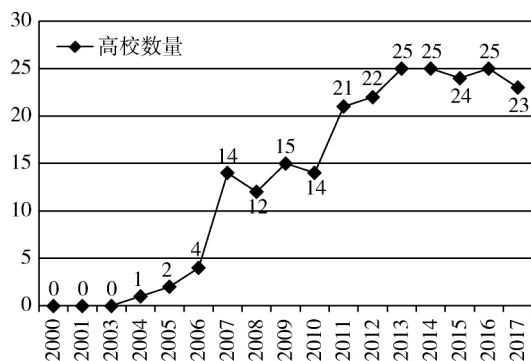


图4 参与情感识别研究的高校数量随时间变化曲线图

1.4 情感识别研究领域分布

图5为情感识别具体研究方向分布图,从图5可以看出,在情感识别领域,几大研究方向研究热度排序为语音情感识别→生理信号情感识别→多模态情感识别→视频图像情感识别→文本情感识别→音乐情感识别。

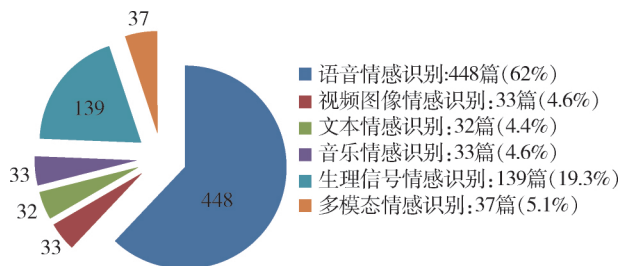


图5 情感识别具体研究方向分布图

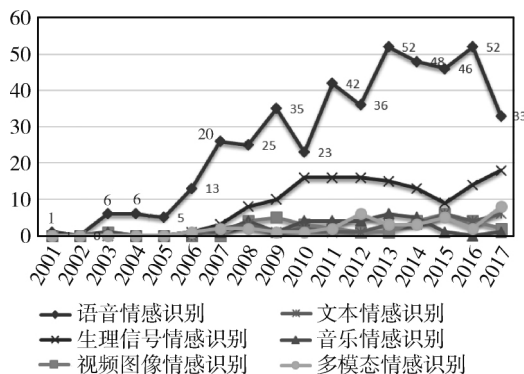


图6 各研究方向论文数量随时间变化曲线图

图6为每一细分的研究方向相关论文数量随时间的变化曲线。从图6可以看出,2006年知网上首次出现了音乐情感识别、生理信号情感识别和多模态情感识别相关的研究文章。具体见表2,列出了2006年以来最早开展音乐情感识别、生理信号情感识别和多模态情感识别相关研究的文章。

表2 最早开展音乐情感识别、生理信号情感识别和多模态情感识别相关研究的论文

作者	文章类别	标题	来源
刘涛	博士论文	音乐情感认知模型与交互技术研究	浙江大学
钱金法等	会议	穿戴式生理参数测量仪研究及在儿童情感识别中的应用	2006 和谐开发中国西部声学学术交流会议论文集
解迎刚等	会议	人脸与眼部检测及其在情感识别中的应用	中国自动化学会控制理论专业委员会·第25届中国控制会

1.5 小结

根据以上分析,发现国内情感识别研究热度总体呈现螺旋式上升状态,中间有小波峰小低谷,但总体仍呈现稳步增长态势。按研究热度变化情况可以将其分为三个阶段。

第一阶段为2000-2006年,研究初步探索期。在这一阶段,国内逐步开始相关研究探索工作,2000年东南大学的赵力教授等人发表了相关论文,率先开始了语音情感识别的研究工作;2003年“第一届中国情感计算机智能交互学术会议”成功召开;2004年第一次出现关于情感识别的硕士论文;2006年第一次出现关于情感识别的博士论文。在这一阶段浙江大学、东南大学和江苏大学等几所高校里逐渐开始情感识别方向

研究。这段时期国内相关研究领域有较多空白,从高校开始逐步开展情感识别研究探索,所涉足的领域也较为狭窄,相关成果也较少。

第二阶段为 2006 年 - 2011 年,研究蓬勃发展期。在这一阶段,情感识别相关研究工作开始蓬勃发展,论文总数从 2006 年的 16 篇增长至 2007 年的 37 篇,期刊论文从 2006 年的 9 篇增长至 2007 年的 16 篇,硕士论文从 2006 年的 2 篇增长到 12 篇,专利的数量在 2010 年实现突破,从之前的一直空白到 2010 年增长到 9 个。另外 2006 年知网上首次出现了音乐情感识别、生理信号情感识别和多模态情感识别相关的研究文章。同时 2007 年参与研究的高校数量激增,由 2006 年的 4 所快速增加到 14 所。这段时间不论从论文数量、参与情感识别研究人员数量、参与研究机构数量、研究细分领域都有较大的增长,说明在这段时期国内关于情感识别的研究进入一个蓬勃发展的状态。

第三阶段为 2011 年 - 至今,研究成熟期。在这一阶段,各类论文数量均得到极大提高,论文总数从 2010 年的 60 篇增长到 2011 年的 76 篇,然后一直稳步增长,在 2014 年前后达到一个峰值 109 篇。体现研究成果转化指标的专利数量也在激增,在 2016 年之前一直在 20 个附近徘徊,2016 年增至 48 个,2017 年增至 62 个。博士论文数量和会议论文数量在 2015 年双双达到一个小峰值,分别为 8 篇和 9 篇。同时参与情感识别研究的高校在 2011 年突增至 21 所。伴随着人工智能的火热应用,情感识别系统应用的论文也逐渐增多,情感识别研究领域也逐步扩大,对音乐情感识别、耳语音情感识别等均有相关研究。这些都说明国内情感识别研究经过十几年的快速发展,已步入稳步上升的成熟期。

2 语音情感识别

从情感识别研究领域分布可以看出,国内情感识别重点集中在语音情感识别领域,共有 440 篇语音情感识别相关文章,占总论文数量的 61%。所以下面具体把国内语音情感识别的情况进行梳理。

2.1 具体研究重点

对这 440 篇文章进行分析,发现研究重点主要集中在 6 大领域,具体分布见表 3,列出了这六大语音情感识别研究领域的论文数量及占比。

表 3 语音情感识别各研究方向论文数量及占比

研究内容	情感特征	分类方法	综述	情感识别应用及系统	情感数据库	耳语音情感识别
论文数量	154 篇	150 篇	19 篇	96 篇	16 篇	5
占比	35%	34.1%	4.3%	21.8%	3.6%	1.2%

其中在语音情感分类方法的研究、情感特征相关技术的研究,以及情感识别的应用及系统搭建三个方面关注较多,在情感识别数据库构建、情感识别综述,以及耳语音情感识别等方面的关注度比较低。

另外,本文中各研究领域的论文数量随时间变化的曲线图绘制出来,以供分析参考。如图 7 所示。

从图 7 中可以看出各研究方向的发展趋势:

(1) 国内关于语音情感识别的研究中最受关注的是语音情感特征分析和各种分类方法的研究,相关研究论文数量变化曲线较为一致,2010 年之前均处于平缓上升状态,在 2010 年时双双达到一个较低值,分别为 3 篇和 12 篇,并同时于 2011 年激增,分别为 12 篇和 21 篇,随后平缓波动,整体呈上升趋势。

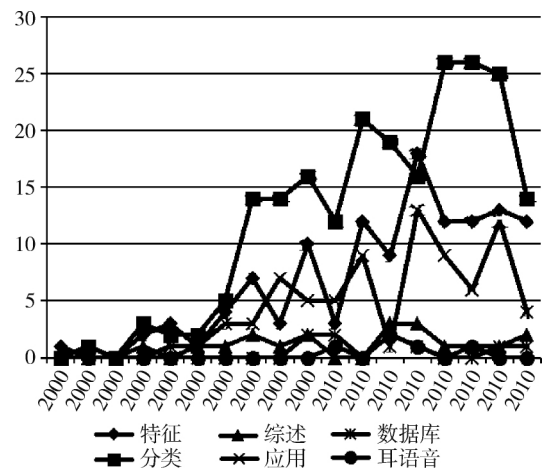


图 7 语音情感识别各研究方向论文数量随时间变化曲线图

(2) 情感识别系统的应用及系统搭建在国内语音情感识别中也有相当多的学者在研究,相关论文数量在 2012 年之后增长迅猛,研究领域最集中的几个应用场景分别为远程教育(24 篇,25%)、智能人机交互(19 篇,19.8%)、呼叫中心(8 篇,8.3%)。

(3) 情感识别综述、相关数据库建立以及耳语音情感识别三部分研究内容相关的论文相对来说数量较少,一直在 5 篇之内波动,说明这三部分内容的研究关注度较低,还有很多空白领域。

2.2 情感分类方法研究

语音情感识别的本质是模式识别问题,因而几乎所有模式识别方法都可以应用到语音情感识别中去^[4]。知网上可以查到的共计 150 篇情感分类方法相关的文章,分别采用了**人工神经网络**(artificial neural network, ANN)、**支持向量机**(support vector machines, SVM)、**隐马尔科夫模型**(Hidden Markov Models, HMM)、**高斯混合模型**(Gaussian Mixture Models, GMM)、**自适应增强**(adaptive boosting, Adaboost) 算法、**K 近邻法**(K-nearest neighbor, KNN) 和**随机森林**(random forest, RF) 等方法。现将主要技术方法汇总如表 4 所示。

表 4 各种分类方法的论文数量汇总

序号	1	2	3	4	5	6	7	8	9
分类方法	ANN	SVM	HMM	GMM	多分类方法组合	自适应	KNN	随机森林	其他
论文数量	50	35	13	9	18	4	9	4	8

每种分类方法都有很多专家学者进行了深入的研究,下面将知网上每种方法典型研究文章汇总如表 5 所示,以供参考。

3 结束语

知网上的数据只是国内研究的一部分,但是“管中窥豹,可见一斑”。通过对知网上情感识别相关数据的梳理,可以还原出国内自 2000 年以来的情感识别相关研究工作的脉络,了解当前的研究重点、热点以及技术进展,明确情感识别相关技术的发展趋势,为今后的研究工作打下基础。

(1) 国内情感识别研究趋势: 总体呈现螺旋式上升状态,中间有小波峰小低谷,但整体变化曲线仍呈现稳步增长态势。

(2) 国内情感识别研究热点: 一是语音情感识别方向,语音情感识别研究的重点仍在寻找更具区分度、更可靠的语音情感特征以及更精准更抗噪的分类器上;二是综合多种单一情感识别的多模态情感识别,这也与当前人工智能的发展趋势一致,情感识别是人工智能里很重要的一部分,而整合声音、图像、生理、行为、文本等各类信息进行综合分析研判更符合人类认识情感的过程。

(3) 语音情感识别难点: 一是仍比较缺乏统一、公开、规范的汉语普通话情感语料库,也缺乏权威的、统一的情感语料库构建标准。导致很多研究工作都是基于自行录制的数据库进行,录制参与人员数量、性别、年龄分布各不相同,包含文本信息也各异,最终产生的语料库规模大小不一,这些导致了基于不同语料库的研究成果难以进行横向比较与分析,一定程度上阻碍了语音情感识别学科的发展;二是关于表达力更强的维度情感的研究还非常缺乏,大部分研究工作都是基于离散情感来开展的,维度情感语料库的建设也极度欠缺。这些方向都需要加强研究,也是语音情感识别非常重要的发展方向。

(4) 情感识别发展趋势: 一是从不断增加的参与高校科研院所数量和相关会议数量上可以看出情感识别研究氛围越加浓厚;二是情感识别数据库的构建正逐步从表演型向引导型转变,从自行录制库向采用公开的、大型的数据库转变,自然度越来越接近真实人机交互环境,这表明情感识别正慢慢的由实验室走向真实应用场景;三是情感识别应用领域愈加广泛,在远程教育、智能人机交互、呼叫中心、可穿戴设备、谎言检测、汽车驾驶、医疗、英语发音质量评测等应用场景都有相关的研究。

表 5 语音情感识别研究典型论文汇总

分类方法	时间	作者	提取特征	情感库	情感类别	结果
ANN	2004	王青[5]	能量参数(平均值、最大/最小值)、基频参数(平均值、最大/最小值、基频范围及平均基频变化率)和语速等	用剪辑法和录音法分别建立了说话人无关和说话人相关两个小规模情感语音数据库	4 种: 愤怒、高兴、悲伤及害怕	平均识别率: 45% - 55% 之间
脉冲耦合神经网络	2008	梁泽等人[6]	语谱图特征时间序列	自行录制, 不包含客观情感的阿拉伯数字串(数字 0 ~ 9 的不规则组合)的汉语发音片段, 共计 120 句	2 种: 平常情感与高兴情感	平均识别率: 90%
混合蛙跳算法神经网络	2011	余华等人[7]	短时能量、基音、发音帧数、共振峰、谱能量比、谐波噪声比、梅尔倒谱系数以及它们的差分、均值、最大值、最小值、方差等 139 个全局统计特征	自行录制, 由 3 男 3 女录制烦躁等 7 种情感语音数据, 每种情感各 400 条, 共 2800 条	7 种: 烦躁、喜悦、愤怒、惊讶、悲伤、恐惧和平静	平均识别率: 81.57%
深度信念网络	2014	黄晨晨等人[8]	发音持续时间、速率、振幅、基频	某呼叫中心的录音语音数据库。该库由 100 男 80 女录制完成, 其中包含 7 种基本情感, 即悲伤、愤怒、惊奇、恐惧、高兴、厌恶和平静, 每种情感录有 200 句参照脚本。	4 种: 高兴、悲伤、惊奇、愤怒	平均识别率: 86.5%
猫群优化神经网络	2015	敖培等人[9]	36 个韵律特征和 38 个音质特征, 共计 74 维特征	自行录制, 由 3 男 3 女分别用 6 种情感进行朗读, 共得到 2160 条情感语料	6 种: 高兴、愤怒、惊奇、恐惧、悲伤、中性	平均识别率: 84.2%
深度信念网络	2016	朱从贤[10]	短时能量、短时幅度及其部分统计特征参数	自行录制, 由 8 男 8 女使用害怕等七种情感朗读 30 个短句, 共 3360 条。	七种: 害怕、厌倦、开心、烦躁、忧虑、伤心和生气	平均识别率: 80%
SVM	2003	王治平等人[11]		自行录制, 共 15 个男性对 4 句中中性语句用 5 种情感(高兴、愤怒、惊奇、悲伤、中性)进行朗读, 共得到 900 句话	4 种: 高兴、愤怒、惊奇、悲伤	平均识别率: 91.38%
Fisher 准则 + SVM	2012	陈立江等人[12]	包括基频等共 288 个备选特征	1. 柏林语音情感库 2. 北京航空航天大学情感语音数据库, 该库由 7 男 8 女对 20 句录音脚本用 7 种情感(包括平静)录制, 每种情感重复 3 次, 共 6300 句情感语料。	6 种: 高兴、愤怒、惊奇、悲伤、恐惧、厌恶	最优组合为 Fisher + SVM
多级 SVM	2017	任浩等人[13]	振幅、基频、共振峰、MFCC 等共计 72 维情感特征	柏林语音情感库	7 种: 恐惧, 厌恶, 快乐, 无聊, 中性, 悲伤和愤怒	平均识别率: 63.74%
HMM	2006	国辛纯等人[14]	基频和能量特征	自行录制, 由 5 男对 50 个语句用 6 种基本情感类型各发音 1 遍, 最后选取其中的 950 句情感数据,	6 种: 欢快、愤怒、惊奇、悲伤、恐惧及中性	平均识别率: 60%

续表

分类方法	时间	作者	提取特征	情感库	情感类别	结果
HMM + ANN	2007	罗毅[15]	10 阶 LPCC、短时能量参数 (包括一阶、二阶差分)、基音 频率参数(包括一阶、二阶差 分)	自行录制,由 3 男 3 女录音,其中 训练样本 55 句,识别样本 88 句	5 种:惊奇、愤 怒、喜悦、悲 伤、厌恶	对厌恶的识 别有较大 提升
GMM + SVM	2011	黄永明等人[16]	MFCC 系数及其一阶差分、二 阶差分系数,组成 39 维特征 向量	1. 柏林语音情感库 2. 自行录制,由 2 男 2 女对 10 个 文本分别用 6 种情感说 3 遍,样 本总数为 720 个。	6 种:悲伤、愤 怒、害怕、高 兴、厌烦、中性	在两个数据 库中的平均 识别率: 1. 80.4% 2. 83.4%
GMM + ANN	2016	袁杰[17]	短时能量、基音频率、1 ~ 12 阶 Mel 倒谱参数和它们的一 阶差分共 28 维情感特征	自行录制,由 8 男 8 女录音而成, 共计 4 种情感 1920 条语句,有效 情感样本数据 1197 条。	4 种:高兴、愤 怒、惊讶和 悲伤	平均识别率: 81.5%
SVM、KNN、 ANN 三分 类器投票 组合	2008	张石清等人[18]	包括韵律特征和音质特征共 计 47 个情感特征	自行录制,由 5 男 5 女分别用 4 种情感对 5 句话各发音 5 遍,共 得到 1000 条情感语料	4 种:生气、高 兴、悲伤和 惊奇	生气 94.7% 高兴 83.2% 悲伤 90.5% 惊奇 81.3%
Adaboost 算法	2014	林思阳[19]		1. EMO - DB 德语语音库 2. CASIA 中文语音库 3. SAVEE 英语语音库		平均识别率: 1. 87.56% 2. 81.75% 3. 76.75%
boosting + HMM	2009	童灿[20]	包括平均振幅、最大振幅、语 句发音持续时间等 10 个特征 参数	自行录制,选取了 5 种情感共 300 个样本	5 种:生气、厌 恶、恐惧、高兴 和悲伤	生气 82.67%、 厌恶 82.53%、 恐惧 80.92%、 高兴 75.58%、 悲伤 87.36%
模糊 KNN + 模糊 SVM	2014	王吉林等人[21]	包括短时能量、短时振幅、共 振峰、基频、MFCC 等共计 42 维情感特征	自行录制,由 5 男 5 女分别用 5 种情感朗读 20 句不同文本数据, 共得到有效语料 624 句。	5 种:高兴、愤 怒、中性、悲 伤、害怕	平均识别率: 83.05%
LDA + ker- nel - KN- NFLC	2015	张昕然等人[22]	包含短时能量、基音、共振峰、 梅尔倒谱系数等 120 个全局 统计特征	自行录制,包含 5 种情感,各 800 条语句	5 种:愤怒、烦 躁、高兴、悲伤 和平静	平均识别率: 82.1%
KNN + RF	2013	刘进华[23]		柏林语音情感库	7 种:生气、焦 虑、烦恼、厌 恶、高兴、悲伤 及中性	平均识别率: 81.09%

综上,说明情感识别仍然是目前的研究热点,并且正受到越来越多的关注,正逐步走向实际应用场景。随着人工智能的研究越加深入,情感识别研究必将迎来更广阔的应用前景。

参 考 文 献

- [1] Williams C E, Stevens. K N. Emotions and Speech: Some Acoustical Correlates [J]. Journal of the Acoustic Society of America, 1972, 52(4B): 1238 - 1250.

- [2] Minsky M. The Society of mind[M]. Simon and Schuster ,1988.
- [3] Picard R W ,Picard R. Affective computing[M]. Cambridge: MIT press ,1997.
- [4] 张晴晴 ,刘勇 ,王智超 ,潘接林 ,颜永红. 卷积神经网络在语音识别中的应用[J]. 网络新媒体技术 ,2014 ,3(06) : 39 – 42.
- [5] 王青. 基于神经网络的汉语语音情感识别的研究[D]. 浙江大学 ,2004.
- [6] 梁泽 ,马义德 ,张恩溯 ,朱望飞 ,汤书森. 一种基于脉冲耦合神经网络的语音情感识别新方法[J]. 计算机应用 ,2008(03) : 710 – 713 + 718.
- [7] 余华 ,黄程韦 ,张潇丹 ,金赞 ,赵力. 混合蛙跳算法神经网络及其在语音情感识别中的应用[J]. 南京理工大学学报 ,2011 ,35(05) : 659 – 663.
- [8] 黄晨晨 ,巩微 ,伏文龙 ,冯东煜. 基于深度信念网络的语音情感识别的研究[J]. 计算机研究与发展 ,2014 ,51(S1) : 75 – 80.
- [9] 敖培 ,赵四方 ,李延强. 基于猫群优化神经网络的语音情感识别[J]. 数字技术与应用 ,2015 ,4(02) : 120.
- [10] 朱从贤. 基于深度学习的语音情感识别方法的研究[D]. 东南大学 ,2016.
- [11] 王治平 ,赵力 ,邹采荣. 基于支持向量机的语音情感识别(英文) [J]. Journal of Southeast University(English Edition) ,2003(04) : 307 – 310.
- [12] 陈立江 ,毛峡 ,Mitsuru ISHIZUKA. 基于 Fisher 准则与 SVM 的分层语音情感识别[J]. 模式识别与人工智能 ,2012 ,25(04) : 604 – 609.
- [13] 任浩 ,叶亮 ,李月 ,沙学军. 基于多级 SVM 分类的语音情感识别算法[J]. 计算机应用研究 ,2017 ,34(06) : 1682 – 1684.
- [14] 国辛纯 ,郭继昌 ,龚修全. 基于 HMM 的语音信号情感识别研究[J]. 电子测量技术 ,2006 ,29(05) : 69 – 70 + 80.
- [15] 罗毅. 一种基于 HMM 和 ANN 的语音情感识别分类器[J]. 微计算机信息 ,2007 ,23(34) : 218 – 219 + 296.
- [16] 黄永明 ,章国宝 ,董飞 ,达飞鹏. 基于两种 GMM – UBM 多维概率输出的 SVM 语音情感识别[J]. 计算机应用研究 ,2011 ,28(01) : 98 – 101.
- [17] 袁杰. 基于 ANN 和 GMM 融合的语音情感识别方法的研究[D]. 东南大学 ,2016.
- [18] 张石清 ,赵知劲. 基于多分类器投票组合的语音情感识别[J]. 微电子学与计算机 ,2008 ,25(12) : 17 – 20.
- [19] 林思阳. 选择性 AdaBoost SVM 语音情感识别算法的研究[D]. 华南理工大学 ,2014.
- [20] 童灿. 基于 boosting HMM 的语音情感识别[C]//2008 年中国高校通信类院系学术研讨会. 中国通信学会青年工作委员会 ,2009.
- [21] 王吉林 ,夏菽兰 ,赵力. 基于模糊 K 近邻的模糊支持向量机的语音情感识别[J]. 微电子学与计算机 ,2014 ,31(04) : 31 – 35.
- [22] 张昕然 ,查诚 ,徐新洲 ,宋鹏 ,赵力. 基于 LDA + kernel – KNNFLC 的语音情感识别方法[J]. 东南大学学报(自然科学版) ,2015 ,45(01) : 5 – 11.
- [23] 刘进华. 基于随机森林的语音情感识别研究[D]. 华南理工大学 ,2013.

作者简介

胡冬妮 (1981 –) ,女 ,工程师 ,主要研究方向为语音信号处理与情感识别。

王武军 (1976 –) ,男 ,高级工程师 ,硕士研究生 ,主要研究方向为语音信号处理。

王青 (1974 –) ,女 ,高级工程师 ,硕士研究生 ,主要研究方向为语音信号处理。