

单位代码: 10359  
学 号: 2012110600

密 级: 公开  
分类号: TP391

合肥工业大学

Hefei University of Technology

硕士学位论文

MASTER'S DISSERTATION

论文题目: 基于特征选择的语音情感识别

学位类别: 学历硕士

专业名称: 计算机应用技术

作者姓名: 李昌群

导师姓名: 安宁 教授

完成时间: 2015 年 04 月

合 肥 工 业 大 学

学历硕士学位论文

基于特征选择的语音情感识别

作者姓名：李昌群

指导教师：安宁 教授

学科专业：计算机应用技术

研究方向：信息系统与智能决策

2015 年 04 月

A Dissertation Submitted for the Degree of Master

**Speech emotion recognition based on feature selection**

By

Li Changqun

Hefei University of Technology

Hefei, Anhui, P.R.China


April, 2015

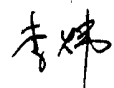
# 合 肥 工 业 大 学


本论文经答辩委员会全体委员审查，确认符合合肥工业大学学历硕士学位论文质量要求。



## 答辩委员会签名（工作单位、职称、姓名）

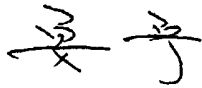
主席： 安徽省软件行业协会 高级工程师 汪炳兴 

委员： 安徽大学 教授 李炜 

合肥工业大学 副教授 郑利平 

合肥工业大学 副教授 邵堃 

合肥工业大学 副教授 沈明玉 

导师： 合肥工业大学 教授 晏宁 

## 学位论文独创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下进行独立研究工作所取得的成果。据我所知，除了文中特别加以标注和致谢的内容外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得合肥工业大学或其他教育机构的学位或证书而使用过的材料。对本文成果做出贡献的个人和集体，本人已在论文中作了明确的说明，并表示谢意。

学位论文中表达的观点纯属作者本人观点，与合肥工业大学无关。

学位论文作者签名：李昌群

签名日期：2015年4月21日

## 学位论文版权使用授权书

本学位论文作者完全了解合肥工业大学有关保留、使用学位论文的规定，即：除保密期内的涉密学位论文外，学校有权保存并向国家有关部门或机构送交论文的复印件和电子光盘，允许论文被查阅或借阅。本人授权合肥工业大学可以将本学位论文的全部或部分内容编入有关数据库，允许采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：李昌群

指导教师签名：夏 3

签名日期：2015年4月21日

签名日期：2015年4月21日

论文作者毕业去向

工作单位：

联系电话：

通讯地址：

E-mail：

邮政编码：

# 致 谢

时光如梭，转眼间三年的时间在不知不觉中即将告一段落。记忆中 2012 年 9 月来到实验室时的情景历历在目。回首走过的三年求学路，心中涌起阵阵的感激之情。

三年的学术生涯中，首先要感谢我的导师安宁老师，无论在学术钻研中还是人生规划上都给了我极大的指导。在他的指导下，使我渐渐的朝着自己喜欢的方向不断去努力。虽然我知道以后的路还很长也许也很难，但是老师给我的教导我会永远记得，并且勇敢的去面对。

同时，由衷的感谢李廉老师，他严谨的学术态度、丰富的学术知识以及对知识孜孜不倦的追求精神深深感染了我，他对我的教导足以让我在接下来的学术路上细细揣摩一辈子。

其次，我要感谢三年来给我帮助很多的杨静老师。是她在我最迷茫的时候给了我一丝希望，渐渐的确立研究生生涯中的学术工作。她的开朗乐观的性格深深影响着我。

再次，我想感谢在整个课题研究期间，给予我很大帮助的王坤霞老师、杨娇云老师、吴玺老师、邵堃老师、刘进军老师以及程文娟老师。

最后，我想感谢三年来给予我很多鼓励与支持的实验室同学们，得到了王爱国、徐培、谷敏敏、陈俊、贵芳，金柳颀等师兄兄弟们的帮助，使我三年来可以在一个和谐的环境中学习生活。同样的想要感谢和我一起走过来的室友们，与你们相处的三年也许会是最自在快乐的三年。

特别想感谢我的父母亲人给我的理解与支持，是他们的坚持让我在这求学的三年可以无忧无虑的尽情展现自己的青春。

再多的话语也无法表达内心最诚挚的感谢。未来的道路中，无论遇到怎样的困难，我都会想起曾经这里的三年，在这里学会的坚持、坚强与努力才是走好人生道路的必要条件。

作者：李昌群

2015 年 04 月 09 日

## 摘要

语音情感识别是人工智能和机器学习领域的一个热点研究问题。研究者们通过发现影响情感的语音特征，利用 HMM、SVM、KNN 等分类器进行语音情感识别。当前的研究在有些时候可能会存在特征过多的问题，影响了识别率以及增加了计算开销等。基于这些问题，本文提出了特征选择的方法，以降低特征维度，防止造成“维度灾难”、提高分类的识别率以及精简结构图。本文的主要研究内容如下：

(1) 将 MMHC 算法应用于特征选择领域，在此基础上提出了 h-MMHC（启发式 MMHC）算法，该算法采用启发式的策略分阶段的添加属性，相比于 MMHC 算法一次性考虑所有属性集提高了计算的性能。

(2) 将 h-MMHC 应用于语音情感识别领域，构建了基于 h-MMHC 算法的语音情感识别框架，对 MFCC 特征集进行特征选择。将原始的 180 个特征降维至 21 个特征，分类的识别率相比于未进行特征选择时提高了 1 个百分点。

(3) 考虑到现实世界中情感类别之间非互斥的特性，本文对 MFCC 特征集中的 7 类情感分别进行特征选择，实验结果表明不同的情感类别之间存在一些共有的特征。该结果为语音情感库的构建提供了理论基础。

通过上述研究，本论文提出启发式的 MMHC 算法用于主特征选择，降低了计算的时间开销。将该算法应用于语音情感识别领域进行特征选择，基于选择出的特征集进行情感识别，识别率提高了一个百分点，同时找出了情感之间所共有的特征，进而验证了各个情感类别之间的非互斥性。

**关键词：**特征选择；情感识别；机器学习；贝叶斯网络；主因素分析；SVM 分类

# Abstract

Speech emotion recognition is a hotspot problem in many areas, such as artificial intelligence, machine learning, etc. By finding speech characteristics that correspond to emotions, researchers use HMM, SVM and KNN classifier for speech emotion recognition. However, there may exist some problems in current studies, i.e. there are too many speech features, which affect the recognition rate and increase the computational time. In order to tackle these problems, this paper presents a feature selection method to reduce the dimension of features, preventing the "dimension disaster" and improving the recognition rate. This dissertation conducts the following research topics:

(1) MMHC algorithm was applied for feature selection. Based on MMHC algorithm, a heuristic MMHC (h-MMHC) algorithm was proposed, which adopts a heuristic strategy that gradually adds features. In contrast to MMHC algorithm that handle all features at one time, h-MMHC can reduce the computational time.

(2) h-MMHC was applied in speech emotion recognition problem and a speech emotion recognition framework was constructed based on h-MMHC algorithm. When using the framework to select features among MFCC dataset, the feature dimension was reduced to 21 characteristics from 180 characteristics, besides the classification recognition rate increase by 1% compared to non-feature selection methods.

(3) Considering emotions in real world are not exclusive, we conduct feature selection for each emotion. The experimental results show that different emotions share a few characteristics. This results provide a theoretical support for the construction of emotional voice library.

Through the above research, this dissertation proposed a heuristic MMHC algorithm for feature selection with the result of reducing computational time. When applying this algorithm to feature selection for speech emotion recognition problem, the recognition rate increased by one percent and shared characteristics of different emotions were found, which verified the non-exclusive characteristic of different emotions.

**KEYWORDS:** Feature selection; Emotion recognition; Machine learning; Bayesian network; Main factor analysis; SVM classification



# 目 录

<b>第一章 引言</b>	1
1.1 研究背景	1
1.2 情感识别的发展历史、方法研究以及应用	1
1.2.1 情感识别的发展历史	1
1.2.2 语音情感识别的过程	2
1.2.3 情感识别的应用	4
1.3 机器学习的定义及其应用	4
1.3.1 机器学习的定义	4
1.3.2 机器学习的应用	5
1.4 特征选择方法综述	5
1.4.1 特征选择方法的定义	5
1.4.2 语音情感识别领域中的特征选择	7
1.4.3 基于贝叶斯网络的特征选择	8
1.5 本文的主要研究工作介绍	9
1.6 本文组织结构安排	9
1.7 本章小结	10
<b>第二章 基本理论及 MMHC 算法介绍</b>	11
2.1 关联规则	11
2.2 关联模型	11
2.3 基本概念	13
2.4 MMHC 算法	13
2.5 本章小结	14
<b>第三章 用于主因素分析的贝叶斯网络构建算法研究</b>	15
3.1 主因素分析的背景及现状	15
3.1.1 主因素分析的研究背景	15
3.1.2 主因素分析的现状	15
3.2 启发式最大最小爬山算法: h-MMHC	16
3.2.1 h-MMHC 算法的提出背景	16
3.2.2 h-MMHC 的理论描述	17
3.3 本章小结	18
<b>第四章 h-MMHC 算法在语音情感识别中的应用</b>	20

4.1 基于 h-MMHC 的语音情感特征选择方法 .....	20
4.1.1 背景描述 .....	20
4.1.2 MFCC 数据集介绍 .....	20
4.2 研究方案与实验设计 .....	21
4.2.1 研究方案 .....	21
4.2.2 实验设计及各模块的实现 .....	23
4.3 SVM 分类器的介绍 .....	24
4.3.1 结构风险最小化 .....	24
4.3.2 最优分类超平面 .....	26
4.3.3 核函数 .....	26
4.4 实验结果分析 .....	29
4.4.1 实验环境及其 LIBSVM 参数设计 .....	29
4.4.2 各情感类别主因素提取结果与分析 .....	30
4.4.3 主因素提取后识别率结果与分析 .....	44
4.5 小结 .....	45
<b>第五章 总结与展望 .....</b>	<b>47</b>
5.1 该研究的工作 .....	47
5.2 下一步的研究工作 .....	48
<b>参考文献 .....</b>	<b>49</b>
<b>攻读硕士学位期间的学术活动及成果情况 .....</b>	<b>53</b>

## 图表清单

图 1.1 情感识别流程图 .....	4
图 1.2 基本的特征选择框架 .....	6
图 1.3 改进的特征选择框架 .....	6
图 1.4 基于贝叶斯网络结构的特征选择方法 .....	9
图 4.1 基于 h-MMHC 算法的语音情感识别特征选择流程图 .....	22
图 4.2 二分类情况的图示 .....	27
图 4.3 线性不可分情况的展示 .....	27
图 4.4 对情感 1 有影响的主因素 2 及其次主因素 .....	31
图 4.5 情感 1 中包含所有主因素及次主因素在内的结构图 .....	31
图 4.6 情感 2 第一阶段找出的特征 .....	32
图 4.7 情感 2 在主特征 110 出现时的结构图 .....	32
图 4.8 情感 2 在主因素 157、158 加入时的结构图 .....	33
图 4.9 情感 2 的完整主因素及其次主因素的网络结构图 .....	34
图 4.10 情感 3 在第一阶段产生的结构图 .....	34
图 4.11 情感 3 中包含主因素 109 的结构图 .....	35
图 4.12 情感 3 中包含所有主因素和次主因素的结构图 .....	35
图 4.13 情感 4 中第一个特征组产生的结构图 .....	36
图 4.14 情感 4 中特征 85 加入时的结构图 .....	36
图 4.15 情感 4 的主特征 109 加入后的结构图 .....	37
图 4.16 情感 4 中主因素 157 加入时的结构图 .....	37
图 4.17 情感 4 中包含全部主因素和次主因素的结构图 .....	38
图 4.18 情感 5 中第一阶段的结构图 .....	38
图 4.19 情感 5 中加入 109 在内的特征组后的结构图 .....	39
图 4.20 情感 5 中加入 157 在内的特征组后的结构图 .....	39
图 4.21 情感 5 中包含所有主因素和次主因素的结构图 .....	40
图 4.22 情感 6 中第一阶段得到的结构图 .....	40
图 4.23 情感 6 中加入 54 在内的特征组后得到的结构图 .....	41
图 4.24 情感 6 中加入包含 161 在内的特征组后得到的结构图 .....	41
图 4.25 情感 6 包含所有主因素和次主因素的结构图 .....	42
图 4.26 情感 7 中第一阶段的结构图 .....	42
图 4.27 情感 7 中加入包含 41 在内的特征组时得到的结构图 .....	43

图 4.28 情感 7 中包含所有主因素和次主因素的结构图 ..... 43

# 表格清单

表 1.1 部分语音情感库说明.....	3
表 3.1 h-MMHC 算法描述.....	17
表 4.1 数据处理前后识别率对比.....	29
表 4.2 不同核函数情况下的识别率对比.....	30
表 4.3 主因素提取方法与未进行主因素提取识别率对比表.....	44
表 4.4 7 个小 SVM 分类器的识别率.....	45

## 第一章 引言

### 1.1 研究背景

我们会经常听到人们说起这样一个词“听话听音”，该词语中的“话”也就是说话人说出的话语，而“音”在语音情感的研究领域则可以解释为我们要识别出说话者说话的情感。要想尽量准确的识别出受测者的语音中所包含的情感，需要借助相应的技术从不同的方面去识别。引申到语音情感识别研究领域也就是利用相应的技术提取出能够判别出情感的特征集，进而综合进行判断，以获得一个客观的情感判断结果。

所谓语音情感识别就是人们利用相关的计算机技术，从测试者的语音信号中提取出影响情感重要的特征，采用相应的分类识别技术来判断该语音信号具体所属的情感状态。不论是人工智能还是机器学习领域中，语音情感识别的研究已成为一个热点为众多研究者所研究。研究者们通过大量的努力不断地寻找着对语音情感具有主要影响的特征。

为了更加有效的从语音信号中判别出语音情感所处的状态，必须从大量的原始数据中进行情感识别，研究者们为此做了大量的研究工作<sup>[1-3]</sup>。本文研究工作是基于前人挑选出的 MFCC 属性集，在不显著降低情感识别率的条件下，利用 h-MMHC 算法进行主因素的学习，找出 MFCC 属性集中对情感类别有主要影响的因素。

### 1.2 情感识别的发展历史、方法研究以及应用

#### 1.2.1 情感识别的发展历史

现实世界中，人与人之间的交流有很多不同的形式，其中语音信号被认为是最快速以及最自然的媒介。这样的事实促使大量的研究者们去思考语音信号是否也是人类与机器之间最快速且最有效的方式。然而，这样的想法要求机器必须具有足够的智能以识别人类的语音。早在古代时期，我国就有关于人类情感识别的研究，例如黄帝内经（大概春秋战国时期）中就有提到人有七情：怒、喜、忧、思、悲、恐、惊。可是不管我们在语音识别方面取得了多大的进步，都无法实现机器与人类之间拥有最自然的交互，其主要原因就是机器根本无法完全识别出说话人在说话当时的情感状态。如此的状况便促使了一个相对来说比较新的领域的产生，也就是我们所说的语音情感识别，它被定义为从说话者的语句中识别出测试者说话时的情感状态。

情感识别的发展经历了两个重要的阶段，分别是它的萌芽时期以及发展时

期。二十世纪的中前期是它的萌芽时期，该阶段的主要工作是概念的提出，也就是“计算机的情感能力”概念的提出，该概念是由美国麻省理工学院的教授 Minsky 提出的。Minsky<sup>[4]</sup>于 1985 年在专著“The Society of Mind”中指出情感识别的关键不是要求进行识别的机器具有何种情感，而是我们希望被使用的机器如何实现智能自动的识别出语音信号中所包含的情感。基于当时社会的发展现状，人们对这一概念未能采取及时的重视，因此这仅仅是为语音情感的发展做了一个前期的铺垫。第二个重要的阶段是在二十世纪的中后期，是由来自美国麻省理工学院媒体实验室的 Picard 教授开启的。1997 年“情感计算”的概念首次出现在 Picard 教授的专著中，由此便开启了“情感计算”这一新的研究领域，为情感识别树立了里程碑，进入其发展阶段。

语音情感识别虽然起步较晚，但是自它正式发展时即在世界范围内受到研究者的重视。1972 年，C.Williams<sup>[5]</sup>等人对语音的基音轮廓进行测试中，发现测试者的情感变化对基音轮廓的变化会形成强烈的影响，该研究工作可以概括为国外最早从工学的角度来进行语音情感的研究。紧接着，1992 年，Y.Kitahara<sup>[6]</sup>等人通过对说话人语音的韵律变化来判断说话人的态度和意图，训练建立了语音理解和语音会话系统的初级模型用于控制人与机器之间会话过程的研究。

虽然国内的语音情感识别相对于国外来说落后了很多年，但是也为情感识别的发展贡献出了自己最大的努力。例如 2003 年 12 月，第一届中国情感计算及智能交互学术会议在北京举行，其主办单位是中科院自动化研究所等相关单位；基于之前的研究工作，中科院自动化研究所将国内的会议发展成了国际性的会议，并于 2005 年 10 在北京举办，该会议旨在将国内外相关领域的研究学者以及专家聚集到一起，共同交流在情感识别领域的研究成果，从而推动了我国情感识别领域的快速发展。除此之外，国内各大高校，例如清华大学成立了人机交互与媒体集成研究所、浙江大学成立了人工智能研究所以及中国科学院成立了语言研究所等等，力求为我国的语音情感识别领域的研究发展贡献自己的一份力量。

### 1.2.2 语音情感识别的过程

从情感识别领域的建立以致情感识别如今的发展，已经有大量研究工作者们对其进行了研究。要想正确的进行情感的识别，必须设计一整套的情感识别系统，该系统主要包括三个方面：（1）选取合适的特征以进行更加有效的情感识别工作。有研究者发现连续性的特征，包括距相关特征、共振峰特征、能量相关的特征以及时序的特征都能够传达很多的情感线索<sup>[2,7]</sup>；除了基于时间的特征以外，还有很多的频谱特征也很有用，例如线性预测系数（LPC）<sup>[8]</sup>、线性预测光谱系数（LPCC）<sup>[9]</sup>以及梅尔频率倒谱系数（MFCC）<sup>[10]</sup>在情感识别工作中占据着重要的影响。（2）设计出一个合适的分类体系。研究者们经过不断地辛勤努力，训练出

了多种多样分类器，例如 HMM（隐马尔可夫模型）、SVM（支持向量机）、ANN（人工神经网络）以及 K-NN（K-近邻网络）等等。对于何种分类器更加有效，研究者们众说纷纭，无法达成一致的意见，但是各个分类器都有其各自的优点以及限制条件。为了将众多分类器的优点集中到一起，整合成一组更加有效的分类器。近来有人已经提出该想法<sup>[11-12]</sup>。纵观前人的大量研究工作，我们可以发现众多分类器中的 HMM 更为突出，因为它几乎应用于所有的情感识别中<sup>[13-14]</sup>。（3）建立一个合适的语音情感库，用于验证前期训练出的分类模型的可执行性。一个好的语音情感库将直接影响情感识别率。如果使用了一个质量很低的情感库，很可能就会得到错误的结论。由于很多的语音情感库是不对外开放的，在此只展现部分的情感库，详见表 1.1：

表 1.1 部分语音情感库说明

Tab 1.1 The description of part of speech emotion database

库名	权限	语言	大小	来源	情感
Berlin emotional database <sup>[15]</sup>	公开、免费	德语	800 句（10 演员*7 类情感*10 句话+一些助手翻译）	专业人士 演员	生气、开心、 悲伤、恐惧、 反感、厌倦、 中性
Natural <sup>[16]</sup>	私有	普通话	388 句，11 个说话人，2 种情感	电话中心	生气、中性
KISMET <sup>[17]</sup>	私有	美式英语	1002 条语句，3 位女性受测者，5 种情感	非专业人士 演员	赞成、倾向性、禁止、宽慰、中性
MPEG-4 <sup>[18]</sup>	私有	英语	2445 条语句，35 个受测者	美国电影	高兴、生气、 反感、恐惧、 悲伤、惊讶、 中性

综上所述，语音情感识别的具体过程可以用如下图 1.1 表示：



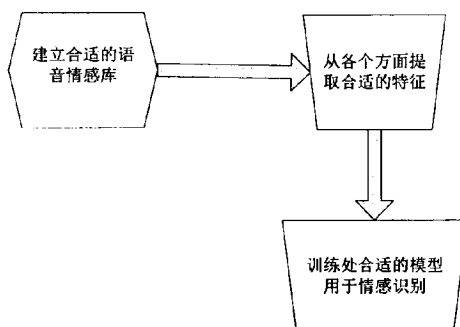


图 1.1 情感识别流程图

Fig 1.1 Flow chart of emotion recognition

### 1.2.3 情感识别的应用

在一些需要最自然的人机交互领域中，语音情感识别受到很大的青睐。例如计算机辅助应用中，机器对于使用该系统的用户做出的回应是建立在其正确识别出使用者情感的基础上<sup>[19]</sup>；车载系统中，语音情感识别系统也得到了很好的应用，该系统可以通过当前驾驶人语音情感的状态来识别此时驾驶人是否处于安全的状态<sup>[19]</sup>；医疗领域中，它还可以作为治疗专家们的诊断工具，用于判断病人人们当前的情感状态，以做出最合适的应答，从而可以避免部分医患冲突；同样的，在自动翻译系统中，正确判断社交场合中说话人当前的情感状态显得尤为重要，准确把握说话人的情感，将可以做出更加合理的翻译；此外，语音情感系统在呼叫中心以及移动通讯领域也得到广泛的应用，用于及时的发现客户当时的情感状态，以便及时应对<sup>[20]</sup>。总之，语音情感识别的研究已经深入到人们生活中的方方面面。然而，要想真正意义上的实现人与机器的自然交互还需要在语音情感识别领域做大量的工作。

## 1.3 机器学习的定义及其应用

### 1.3.1 机器学习的定义

“君子生非异也，善假于物也”<sup>[21]</sup>。人类在原始社会时期就开始了不同形式的学习，例如学习制造石器或防身或狩猎或作为农耕工具、学习驯养猎物，学习农耕等等形式来不断改变着人类生存的社会。如今，“学习”一词是各行各业任何人都经常提起的一个词，学习的能力是人类天生所具有的。但是什么是学习，学习什么，怎么学习却是一个众说纷纭的话题。对此，社会学家们、心理学家们、逻辑学家们甚至其他领域的专家们都持有各自不同的意见。计算机科学家们眼中的学习大抵是利用计算机进行大量实例的训练，从中找出相应的规律，进而为人类所用。机器要想满足人类不断变化的要求，就得通过不断地学习，不断的去模拟人类的学习方式，从而通过大量的训练，学习得出模拟人类思维的模型。

如此便引申出了机器学习这样一个概念。大量的研究者对机器学习的定义不尽相同。

1996 年, Langley 定义机器学习是一门以人工智能为对象的科学, 重点是通过前人的经验学习来研究如何改善具体的算法, 以提高该算法的性能。2004 年, Alpaydin 则认为所谓机器学习, 是使用相应的数据或者是借助以往的经验来优化计算机程序的性能标准。直到目前, 研究者们就机器学习的定义未能给出一个统一的答案, 但是就计算机科学来说, 特别是那些从事人工智能的经验丰富的研究者们, 对于 H.simon 提出的机器学习的定义都比较赞同, 即学习就是一个系统通过执行一个过程以达到改进性能的目的。系统是一个比较宽泛的概念, 它涉及到很多领域, 例如计算机系统、人的系统、神经系统以及控制系统等等, 每一个系统都有其特定的研究领域。即便是在同一个系统里面, 因为我们的目标不一样, 所采取的研究方法以及数据处理的策略都不尽相同。

### 1.3.2 机器学习的应用

机器学习始于 20 世纪 50 年代, 是人工智能领域中一个重要的应用。基于机器学习的技术, 其应用领域也是不断地得到扩展。例如, 人们利用机器学习的方法与技术, 训练出了最成功的语音识别系统, 使人类的通讯变得更加的智能, 更加的方便; 数据挖掘领域中, 研究者们利用机器学习的技术解决了以前很难克服的难题, 例如随着数据量的急剧增长, 特征的维度越来越多, 此时就需要利用机器学习的技术来降维以进行深入挖掘; 同样的, 在模式识别的领域, 模式识别可以被理解成是计算机将不同的事物划分成不同的类别。机器学习中相应的算法被应用于模拟人类智能的模式识别系统中, 来提高其分类的性能。如此, 机器学习的算法便可以在模式识别中得以应用; 生物医学领域, 当然也少不了机器学习的影子, 有研究者将机器学习应用于可穿戴人体外骨骼系统步态识别达到精准和预判的目的, 从而为人类的健康研究再添一笔亮色; 交通领域, 研究者们将图像识别的技术与机器学习结合起来判断驾驶人员是否正常系好安全带, 如此为广人民的人身安全多了一份保障。综上所述, 随着人类社会的不断进步与发展, 计算机技术的发展与普及, 使得机器学习的技术深入到人们生活中的各个方面, 影响并且改变着人们的生活方式。

## 1.4 特征选择方法综述

### 1.4.1 特征选择方法的定义

所谓特征选择, 也就是从事先提取出的大量特征集中按照一定的方法以及原则选取其中更能够描述问题的特征的过程。在事先提取出的原始特征集中, 可能存在大量的冗余以及不相关特征集, 那么进行相应的特征选择, 选取出最有影响

的特征以降低特征维度，从而降低计算的复杂度是很有必要的。特征选择的具体定义没有一个一致的见解，例如 Kira 等人<sup>[22]</sup>认为所谓特征选择就是选取最小的特征子集，该特征子集满足理想情况下、必要的、能够区别目标等条件；John 等人<sup>[23]</sup>则认为特征选择其实就是一个减少特征维度的过程，在该过程中的目标就是要么增加分类精度要么不显著降低分类精度。Dash 等人<sup>[24]</sup>将其定义为在不降低分类精度以及改变类分布的条件下选择尽量小的特征子集。从上述几个人的定义中可以看出进行特征选择的标准大致有不显著降低分类精度、不改变类分布、尽量小的特征子集。Dash 等人于 1997 年提出了基本的特征选择框架，如下图 1.2 所示：

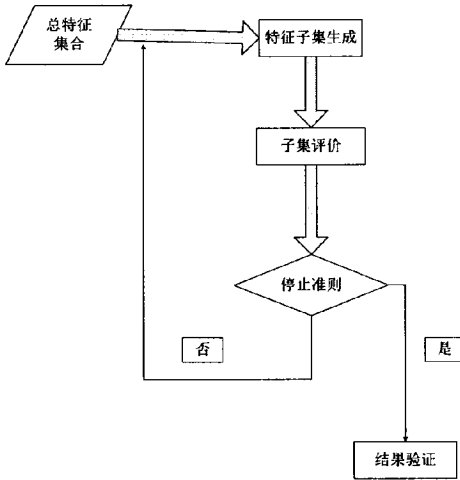


图 1.2 基本的特征选择框架

Fig 1.2 The basic framework of feature selection

从上述的框架中可以看到：在 Dash 等人提出的框架中，相当多的时间是用于子集的搜索，费时费力。基于这样的缺点，Yu 等人<sup>[25]</sup>基于相关和冗余分析，提出了另外一种改进的特征选择框架，避开了原始繁琐的子集搜索，从而可以更加快速的寻找到最优子集。框架如下图 1.3 所示：

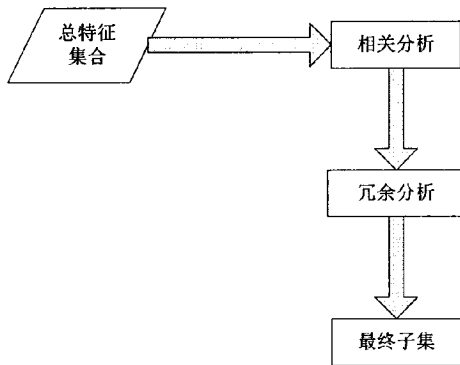


图 1.3 改进的特征选择框架

Fig 1.3 The improved feature selection framework

到目前为止，研究者们已经发现了很多关于特征选择的方法，有基于搜索策

略划分的特征选择方法（包括全局最优、随机搜索和启发式搜索）；也有基于评价准则划分特征选择方法（包括过滤式和封装式）。

### 1.4.2 语音情感识别领域中的特征选择

语音情感识别从提出到得到重视以致蓬勃发展经过了很长一段时间。通常，语音情感的表达会通过很多方面的特征，目前关于具体的影响语音情感的特征还是众说纷纭，未能达到一致的结论。

通过上一节的描述我们知道不同的领域专家们对特征选择的定义是不同的。那么，在语音情感识别领域中，我们将特征选择定义为：在原始特征数据集中，利用相应的特征选择方法从原始特征集中挑选出特征子集，该特征子集与分类类别具有最大相关性以及最小冗余度。挑选出的特征子集与原始特征集对语音情感的分类识别率相比不显著降低或者有所提升。

要想自动的识别出语音中的情感，要做好两方面的工作：一方面是分析研究不同的情感状态会对语音产生什么样的影响，尤其是对哪些特征产生了影响；第二方面则在上一步的工作上分析具体的可以用于语音情感识别的特征。关于第一方面的工作，从语音情感发展至今，不同领域的专家们都各自贡献着自己的努力，以心理学以及韵律学领域的研究为例，从他们的工作中可以发现语音的韵律特征是受情感影响最明显的特征。韵律特征主要包括音强、音速以及语速等。此外，某些情感还会在语音的音质、受测者吐字的清晰度等特征上得以体现。情感识别中，除了韵律特征会影响情感，Liebennan 等人<sup>[26]</sup>设计实验滤除语音中几乎所有的频谱信息，实验发现在未滤除频谱信息时情感的识别率可以达到 85%，滤除后则只有 47%，如此证明了语音的频谱特征也是受情感影响很明显的一个特征。研究者们通过多年的研究找出了一些频谱特征，例如 LPC (Linear predictor coefficients)、LPCC (Linear predictor Cepstral coefficients) 以及 MFCC (Mel-frequency Ceptral coefficients)。

经过研究者们多年的努力，已经有大量的影响语音情感的特征被发现。由于对影响语音情感的具体特征研究者们还未能达成共识，因此他们通常是尽量选取较多的特征来进行情感识别。但是根据模式识别以及数据挖掘领域中的相关知识，我们知道：特征的维数不是越高越好，有时维数过高的话可能会产生“维数灾难”这样的问题。进行相应的特征选择，降低特征维数，在某些领域的某种特定场合下，在不显著降低精确度的情况下，可能会使得计算的效率大大提高。

关于语音情感识别领域中具体的特征选择方法，研究者们就特征选择的方法做了大量的研究工作。具体提出的方法包括利用相关性挑选出主因素的 PCA 方法；基于模糊理论的模糊熵方法；特征数从零开始逐步增加，自下而上的序列前向 SFS 搜索法；特征数从某个数目逐渐减少，自上而下的序列后向选择 SBS 搜索

法以及基于启发式原理的优化选择 PFS 搜索法。

在进行特征选择后，使得参与运算的特征维数极大的减少了，大大减少了计算的复杂度，甚至可以提高识别率。基于特征选择如此多的优点，其应用范围也是极其广泛的。例如，王蕊等人<sup>[27]</sup>将特征选择的方法应用于基于 LFP 的神经解码中以减少神经解码中的计算消耗；张杰慧等人<sup>[28]</sup>将遗传算法（GA）的特征选择算法应用于肺结节的计算机辅助 CT 诊断中，获得了较好的识别率；熊小草等人<sup>[29]</sup>提出了基于图的特征选择方法应用于文本分类中，获得了较好的分类精度。

#### 1.4.3 基于贝叶斯网络的特征选择

贝叶斯网络是一种表示变量之间因果关系的信度网络结构图。整个网络结构包括两部分，网络节点（属性节点）以及连接节点之间的有向边，相对于其他节点来说，每条边描述一个父节点与子节点之间直接的因果联系。贝叶斯网络结构的学习包括全局结构学习和局部结构的学习。可以通过不同的贝叶斯网络结构算法学习出贝叶斯网络结构图。通常，在构造过程中相关性比较大的两个节点之间会存在边，但是具体是何种影响还要根据具体的数据集以及结构算法来定。

在许多的研究领域中，研究者们普遍约定的一个基本原则就是特征最小化原则，也就是用最少的特征数目来最简单化的描述一个问题。可能只有部分的特征集是相关的，用这些相关的特征集合就可以表征一个事物或者一个问题。从上述的描述中可以知道贝叶斯因果模型可以找到与目标节点相关的原因节点，通常情况下，我们将原因节点视为对目标节点有直接影响的节点，那么在利用由众多特征组成的数据集进行分类训练时，首先利用贝叶斯网络结构模型学习出与目标节点相关的节点，进而利用学习出的相关节点进行训练，势必会提高分类的精确度或者不显著降低精确度。具体的基于贝叶斯网络结构进行特征选择的模型可用如下图 1.4 所示：

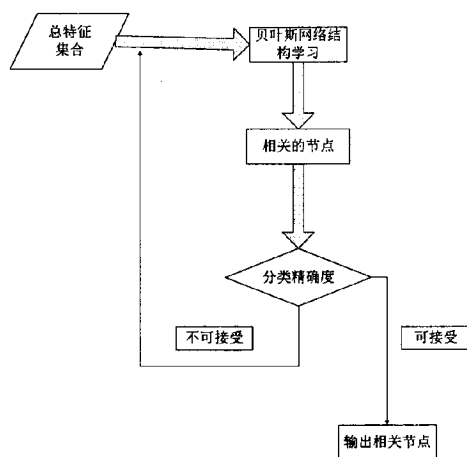


图 1.4 基于贝叶斯网络结构的特征选择方法

Fig 1.4 A feature selection method based on Bayesian network structure

## 1.5 本文的主要研究工作介绍

本文的主要研究工作是对于千人提出的 MMHC 算法进行改进，得到 h-MMHC 算法。h-MMHC 算法相对 MMHC 算法一次性的添加属性来说，他采用启发式的策略分阶段的添加属性。h-MMHC 算法得到网络结构图较 MMHC 来说结构图较为清晰。并且时间较短。

基于提出的 h-MMHC 算法进行特征选择，然后基于选择出的情感特征进行情感识别。使用的数据集是前人提取出的 MFCC 特征集。该特征集是由 535 个样本，180 个特征列，第 181 列是类别标签列所组成。

在实验的刚开始阶段，针对原始的 MFCC 数据集进行数据预处理，包括数据的标准化以及归一化。h-MMHC 算法处理的数据只能是离散化数据，本文对 7 种不同的情感分别进行离散化，并将离散的结果代入主因素提取阶段。各个情感具体的离散化程度取决于各情感类别中主因素提取的个数。

情感识别阶段基于主因素提取阶段得到的主因素以及次主因素进行情感识别。实验分为两个步骤：第一个步骤是针对提取出来的 21 个因素直接进行 7 分类；第二个步骤是建立 7 个独立的 SVM 二分类器，然后组合这 7 个小的二分类器进行 7 分类。

## 1.6 本文组织结构安排

全文共分为五章，具体安排如下：

第一章 引言。主要介绍情感识别的发展历史及其应用、机器学习的定义及其进行特征选择的必要性，由此引出基于贝叶斯网络结构的特征选择。最后，简

要介绍了一下本文的研究工作。

第二章 基本概念及数学基础。主要回顾一下在本文中会用到的相关数学基本概念。

第三章 介绍 h-MMHC 算法。对主因素分析的背景进行阐述，同时就 h-MMHC 的具体算法理论进行详细描述。

第四章 h-MMHC 算法在语音情感识别中的应用。该部分介绍了本文的实验设计与研究方案、基于 h-MMHC 算法的特征选择方法并对实验的结果进行分析，最后利用 h-MMHC 算法选取的主因素构建情感分类模型。

第五章 总结与展望。该部分针对上面提出的主因素方法以及应用做了总结说明，并且就下一步的工作进行展望。

## 1.7 本章小结

在该章节中主要就语音情感识别的发展历史以及研究现状进行回顾，情感识别从提出到如今的发展经历了两个重要的阶段，分别是其萌芽提出时期以及正式发展时期。虽然语音情感识别的起步较晚，但是其国内外的发展却是有目共睹的。要想正确的进行语音情感的识别，必须要设计一整套的情感识别系统。随着人工智能不断的深入人们的生活，语音情感识别的应用也是越来越广泛。在情感识别领域，要想利用机器自动的识别出人们说话时语音中携带的情感，需要借用机器学习的相关理论知识，不断的模拟人类判断识别语音情感的过程，从大量采集的语音样本数据中结合前人总结的对语音情感有影响的特征集进行训练学习。现如今，随着互联网的不断发展，在某些研究领域，例如生物医学以及情感识别领域中的数据特征维数相当的大，如此多的特征在利用机器学习进行训练学习时，会出现所谓的“维数灾难”问题，无形之中使得计算的复杂度极其复杂，甚至还会影响相应的识别精度。为了解决维数过多而造成的分析难题，研究者们不断的尝试进行特征提取和特征选择，特征提取是产生新的特征集合，特征选择是基于原始的特征集进行选择产生新的子集用于训练测试。在不同的领域，根据不同的问题已经提出了不同的特征选择方法。该章节中提出的基于 h-MMHC 来进行特征选择的方法，从特征之间的相关性入手去寻找对目标特征有最大影响的主特征。

## 第二章 基本理论及 MMHC 算法介绍

### 2.1 关联规则

分类规则挖掘以及关联规则挖掘是数据挖掘领域最重要的两大挖掘技术。分类规则挖掘的目标就是在包含规则集的数据库选取部分子集，用以训练形成更加精确的分类器<sup>[30]</sup>；关联规则同样是挖掘相应的规则，该规则所得到的支持度和置信度大于用户给定的最小支持度和最小信任度<sup>[31]</sup>。关联规则的挖掘目标事先是不确定的；而分类规则的挖掘目标却是事先已知的，并且有且仅有一个目标，那就是类标签。关联规则最开始的提出背景是人们对沃尔玛超市的购物篮分析得来的，也就是最经典的“尿布—啤酒”的案例故事。由此，人们开始越来越重视利用数据挖掘的方法去挖掘出事物之间的关联规则。

定义 1 关联规则 (Association Rule)：设  $I = \{i_1, i_2, \dots, i_m\}$  是一组项目集， $W = \{T_1, T_2, \dots, T_n\}$  是一组事务集。用唯一的事务标识符 TID 来表示对应的事务。。关联规则的形式可以表示如为： $X \rightarrow Y$ ,  $X \subset I$ ,  $Y \subset I$  并且  $X \cap Y = \Phi$ 。对于规则  $X \rightarrow Y$  在事务集  $W$  上具有  $c$  的置信度也就是说在  $W$  中有  $c\%$  的事务既包含  $X$  同时包含  $Y$ 。同样的，如果说规则  $X \rightarrow Y$  在事务集  $W$  上具有  $s$  的支持度也就是说在  $W$  中有  $s\%$  的事务包含  $X \cup Y$ 。

在给定事务集  $W$  的情况下，所谓的挖掘出关联规则也就是从给定的数据集中找出支持度和置信度分别大于用户规定的最小支持度和最小置信度的规则。关联规则的挖掘通常分为两个步骤，第一步是挖掘出给定数据集中支持度大于最小支持度的频繁项集，第二步是基于第一步挖掘出的频繁项集来挖掘置信度大于最小置信度的关联规则。

所谓关联模型也就是基于关联规则建立起来的一组模型。常见的挖掘关联规则的算法有：Apriori 算法、Sampling 算法以及 Partition 算法。

### 2.2 关联模型

客观世界中，事物之间是普遍存在联系的，他们之间先后相继、彼此制约的关系使我们常说的因果关系。“因”是指引起某现象发生的现象，而“果”则是由在“因”的作用下产生的现象。不同领域的研究者们对因果关系的表示是不尽相同的。Pearl 和 Spirtes 等人将因果关系与有向无环图建立起联系，并且对此进行形式化的描述<sup>[32]</sup>。使用有向无环图 DAG (Directed Acyclic Graph) 来描述因果关系，结构图比较清晰，使人一目了然，并且还可以根据 DAG 所表示的因果关系进行相应的预测分析以及推理判断。可以用数学模型贝叶斯网络模型和结构等式模型来描述 DAG 的结构，通常，对于离散的问题，我们用贝叶斯网络模型进行描



述。因此，贝叶斯网络结构通常被人们称为因果贝叶斯网络结构。

定义 2 关联结构 (Causal Structure) <sup>[33]</sup>: 给定一组变量  $V = \{X_1, X_2, \dots, X_n\}$ , 基于该变量集得到的关联结构是一个有向无环图, 该图中的每个节点对应变量集中的一个变量, 节点之间的有向边则表示变量集中两变量之间有关联关系。

定义 3 关联模型 (Correlation Model) <sup>[33]</sup>: 给定变量集合  $V = \{X_1, X_2, \dots, X_n\}$ , 通常, 用一个二元组来描述一个关联模型。二元组的形式为:  $M = \langle G, \Theta_G \rangle$ ,  $G$  表示基于变量集  $V$  建立的因果结构,  $\Theta_G$  表示与  $G$  相容的一组参数的集合。 $\Theta_G$  为每个  $X_i \in V$  指定了一个函数依赖  $x_i = f_{x_i}(pa(X_i), \mu_{x_i})$ ,  $Pa(X_i)$  表示节点  $X_i$  在  $G$  上的父节点,  $pa(X_i)$  表示  $Pa(X_i)$  的一种特定取值,  $x_i$  表示  $X_i$  的一种特定取值,  $\mu_{x_i}$  是分布符合  $p(\mu_{x_i})$  的一个随机扰动序列。

在人工智能的研究领域, 贝叶斯网络模型是关联模型的一种特殊形式。图中的节点代表数据集中的特征, 有向边则表示属性之间的关联。贝叶斯网络的学习过程实质上是在结构空间中寻找与样例数据集联合概率分布最一致的网络结构。通常采用基于评分和约束满足的学习方法进行结构的学习。

定义 4 贝叶斯网络 (Bayesian Network, 简记为 BN) <sup>[34]</sup>: BN 是基于变量集  $V = \{X_1, X_2, \dots, X_n\}$  基础上建立起来的一个二元组  $B = \langle G, \theta \rangle$ , 其中  $G = \langle V, E \rangle$  是在变量集  $V = \{X_1, X_2, \dots, X_n\}$  上的有向无环图 DAG,  $V = \{X_1, X_2, \dots, X_n\}$  表示节点集合,  $X_i$  是随机变量,  $E$  对应于 DAG 中边的集合, 同时也反映了随机变量之间的函数依赖关系。 $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  表示网络中每个节点  $X_i$  相对于其父节点集  $Pa(X_i)$  的条件概率分布表, 将  $G$  中父子节点之间的概率依赖程度进行量化表示出来。 $\theta_i = \theta_{x_i|Pa(X_i)} = p(x_i | Pa(X_i))$  表示在  $Pa(X_i)$  确定下节点  $X_i$  取值为  $x_i$  的概率。

定义 5 马尔可夫毯 (Markov Blankets 简记 MB) <sup>[35]</sup>: 在一个网络结构图  $G$  中, 对任意的一个节点  $X_i$ ,  $X_i$  的父节点集  $Pa(X_i)$ 、子节点集以及该子节点集中每个节点的父节点集, 所有的这些集合构成了节点  $X_i$  的马尔可夫毯 MB。

关于贝叶斯网络中的关联性, 主要包括两个方面: 第一个就是搜索目标变量的父节点集以及子节点集 (简记为 PC(T)); 第二个就是搜索目标变量的马尔可夫毯 (简记 MB(T)), MB(T) 集中的变量是 T 的父节点、子节点以及子节点的父节点组成的集合。这两个方面搜索到的变量集皆是与目标变量具有关联的。

通过上述描述可以知道贝叶斯网络结构是基于变量之间的关联性进行构造的。正如 Tsamardinos 等人 <sup>[36-37]</sup> 提出的 MMPC 算法, 该算法分为两个阶段: 第一阶段是前向阶段, 也就是不断的搜索候选的 PC 集 (简记 CPC), 将与目标变量关联度最高的变量放到 CPC 中; 第二阶段是后向阶段, 测试在第一阶段中得到的 CPC 集中的每个变量与目标变量是否可以被 CPC 中的某个子集 D-分离。

## 2.3 基本概念

定义 6 独立 (Independence) [38]: 设  $V = \{X_1, X_2, \dots, X_n\}$  是变量集合, 若相关系数  $\rho(X_i, X_j) = 0$ , 则称  $X_i$  与  $X_j$  (线性) 独立, 记为  $Ind(X_i, X_j)$ 。若  $X_i$  和  $X_j$  独立, 则有  $p(X_i | X_j) = p(X_i)$ , 即变量  $X_j$  的任何变化对变量  $X_i$  的概率分布不产生影响。

实际情况下, 对于某一个变量集合来说, 如果要求不同的变量之间均两两相互独立几乎是不现实的。因此, 在贝叶斯网络中, 经常考虑一个节点关于其他一组节点的独立性。即下所述贝叶斯网络中的条件独立性。

定义 7 贝叶斯网络中的条件独立性 (Conditional independence relation in Bayesian networks) [39]: 令  $Q(X)$  是一个节点  $X$  的所有父节点和子节点的集合, 也就是说,  $Q(X)$  是  $X$  的马尔可夫覆盖, 那么对于其他任何不在  $Q(X)$  中的节点  $Y$ ,  $Y$  与  $X$  关于  $Q(X)$  是条件独立的, 即  $\rho(X, Y | Q(X)) = 0$ , 记作  $Ind_{Q(X)}(X, Y)$ 。

在实际应用中, 通常设置一个阈值  $\delta$ , 当相关系数  $\rho(X, Y | Q(X)) < \delta$  时, 就认为它们是独立的, 这个阈值称为关联阈值。

定义 8 主因素 (Principal factor): 设  $S$  是贝叶斯网络学习任务  $L$  的实例集合,  $V$  是实例中属性的集合,  $d$  是某个特定的目标属性, 与  $d$  的关联值超过某个阈值  $\delta$  的属性称为影响  $d$  的主因素。

注意在此定义中, 主因素可以不止一个。关联阈值  $\delta$  根据问题的需要选取。

## 2.4 MMHC 算法

MMHC 算法又称为最大最小爬山算法, 它基于约束的思想, 结合了局部学习以及搜索评分技术的原则和有效方法。对于小样本的数据集, 该算法也能够表现出很好的性能[40]。

利用 MMHC 算法从训练数据集中学习贝叶斯网络主要经过两个阶段: 阶段一, 找出节点集中各个节点的直接父节点。该阶段利用 MMPC 算法学习得到每个节点的父节点和子节点集合, 学习过程中, 利用条件独立性测试, 如果变量  $X, Y$  关于变量集  $Z$  条件独立, 即  $Ind(X, Y | Z)$  成立, 则可以说明  $X$  和  $Y$  之间没有直接的联系, 相反则说明存在直接的父子节点联系。在整个阶段一, 利用该方法对各个节点做一次计算, 最终得到所有变量的候选父子节点。阶段二, 确定节点联系, 即边的定向阶段。在该阶段中利用爬山算法去确定各个边, 依次计算各个节点与候选父子节点的相关系数, 从中选择出相关系数最大的节点对, 添加边, 即完成了所谓的定向。按照此方法不断的执行, 最终定向出图中所有的边。

MMHC 算法描述:

该算法主要分为约束阶段和搜索阶段。

输入：数据集  $D$

//约束阶段

为每个变量  $X \in V$  计算其候选父节点和子节点集，计算  $PC_x = MMPC(X, D)$ 。

//搜索阶段

在受限的空间里，从一个空的结构图开始，通过执行贪婪的爬山搜索，例如添加边，删除边，逆置边等操作，找到具有最优评分的网络  $G$ ，只有当  $y \in PC_x$  时才执行添加边  $Y \rightarrow X$  操作。

输出：关联贝叶斯网络结构图  $G$

## 2.5 本章小结

本章主要介绍了一些基本的概念以及定理。通过分析可知在数据挖掘领域，关联规则的挖掘是很重要的一个技术，基于此形成的关联模型更是一种很重要的理论模型，它可以应用于其他各种领域，以方便研究者们探索特定领域的关联规则。其次，因果关系是客观事物普遍存在的一种联系，由因及果的探索模式受到了很多学者的关注。Pearl 等人利用图论中的有向无环图 DAG 来对因果关系进行描述，使得属性之间的因果关系更加明确，一目了然。并且还可以根据 DAG 所表示的因果关系进行相应的预测分析以及推理判断。同时，可以用数学模型贝叶斯网络模型以及结构等式模型来对 DAG 的结构进行描述，将因果关系进行量化，从而为人们利用计算机分析事物之间的因果关系开辟了新的里程碑。通常，对于离散的问题，我们采用贝叶斯网络进行描述，而连续的问题则采用结构等式模型。通过贝叶斯网络结构的定义可以清楚地明白所谓的贝叶斯网络结构实质是一个二元组  $B = \langle G, \theta \rangle$ ，分别由基于变量集建立的有向无环图  $G = \langle V, E \rangle$  和每个节点相对其父节点集的条件概率分布表  $\theta$  组成。贝叶斯网络的学习包括结构学习和参数学习，关于结构学习研究者们提出了大量的算法针对不同的领域数据集建立特定的贝叶斯网络结构。MMHC 算法又称为最大最小爬山算法，是贝叶斯网络结构学习中常用到的一种基于离散数据集构建贝叶斯网络结构的算法。该算法的一个突出优点就是对于小样本的数据集它仍能表现出较好的构造特性。

## 第三章 用于主因素分析的贝叶斯网络构建算法研究

### 3.1 主因素分析的背景及现状

#### 3.1.1 主因素分析的研究背景

特征提取问题是数据分析和挖掘的重要研究内容。恰当的算法可在众多的特征属性中，搜索到对于问题影响显著的特征集合，从而更有效的开展对于数据的深入分析。其中一个重要的方法就是主因素分析，也即在确定目标属性后，找出影响目标属性的主要因素，也可理解为基于主因素的综合评价。例如全美红等人<sup>[41]</sup>利用主成分分析法以及熵值法来进行高校教师的绩效评价。主因素分析的方法也称为主成分分析的方法，可以看作是特征提取的一种特殊形式，也即提取对于目标属性影响最大的一个或者若干个特征，从而可以有效的降低特征的维度而不损失评价的效果。

最开始，主成分分析的概念是由英国著名的数学家以及生物统计学家 Karl · Pearson 就非随机变量的问题引入的。随后在 1993 年，美国的数理统计学家 Harold · Hotelling 将该方法推广到随机向量的情况。

基于主因素分析的优点，研究者们将主因素分析的方法应用到具体的研究领域。在具体的领域中结合相应的方法，取得了不错的效果。例如 Turk 等人<sup>[42]</sup>在 1991 年将主成分分析的方法应用到人脸图像识别的领域，效果很显著。由此，该方法在模式识别领域得到极大的重视并且不断的加以创新以适合更多的情形。

#### 3.1.2 主因素分析的现状

主因素分析在一定意义上来说也就是进行特征降维：挑选出对分类有影响的 K 个成分来代替原始的特征，在不显著降低原始分类识别率的情况下。吴恩达<sup>[43]</sup>在深度学习的讲义中介绍到所谓 PCA（主成分分析），其实是一种能够极大地提升无监督特征学习速度的数据降维算法。在该讲义中，吴恩达利用协方差理论的知识推导出了什么是主因素以及如何求出主因素，并且证明出利用求取出的主因素取代原始的数据进行计算，大大降低了原始特征的维度，使得计算的时间大大的缩短，从而提升了效率。

其他的研究者也在利用相应的技术来进行特征选择。例如王爱国等人<sup>[44]</sup>提出了封装的特征子集选择算法，该算法应用于基因领域中高维的基因数据进行特征选择，相对于其他算法来说取得了较好的效果。

基于第二章中介绍的，贝叶斯网络是因果关系的一种数学量化模型，因此可以利用其进行目标变量的“原因”节点的发现研究。贝叶斯网络又称信度网络，

自提出以来一直受到广泛的重视。在社会科学研究中是一种常用的手段，其中包括教育研究领域，利用贝叶斯网络的模型来分析教学质量相关的因素。例如，Choo-Yee Ting 等人<sup>[45]</sup>使用贝叶斯网络来构建一个基于计算机科学探索学习环境概念变化的计算模型；Eva Millán 等人<sup>[46]</sup>使用贝叶斯网络来提高知识评估，使用通用贝叶斯学生模型的整合和评估，用于 Aveiro 大学的计算机测试中；Eva Millán 以及 Tomasz Loboda 等人<sup>[47]</sup>使用贝叶斯网络来设计和完善学生模型。依据该系统可以根据这些信息为学生们提供个性化的服务。同样的，国内学者也在利用数据分析的技术对影响高等教育质量的因素进行探索。马希荣等人<sup>[48]</sup>将数据挖掘中的关联规则分析技术应用到教学评价中，有效地挖掘教育研究中的数据；何万篷等人<sup>[49]</sup>将数学中的可拓方法与数据挖掘技术结合起来，挖掘影响教学质量的关键因素，以及教学质量与教师特征之间的关联规则。所有这些研究都可以归纳为对于某个目标变量的主因素分析。

### 3.2 启发式最大最小爬山算法：h-MMHC

#### 3.2.1 h-MMHC 算法的提出背景

MMHC 算法已经是目前比较流行的方法，并且写进了 Matlab，成为标准算法之一。但是 MMHC 算法有一个缺点，即所用来确定贝叶斯网络结构的属性集合是所有属性，这就造成计算量大的问题。在主因素分析中不尽合理，因为主因素分析并不要求所有的属性之间的关联，而是只需要找出与目标属性具有最大关联的属性。本文结合主因素分析这种特定的应用场合，提出了一种从局部属性集合出发，逐步增加新属性的学习方法，进行网络结构的调整，最终得到主因素。用这样的方法可以提高计算效率。相对于原来的 MMHC 算法，本文提出的方法属于逐步逼近的贝叶斯网络结构学习。其基本思想是，在样例属性集合中，首先取出一部分属性，根据 MMHC 算法，建立有关这部分属性的贝叶斯网络结构，也就是建立这些属性的关联关系。然后将其中与目标属性关联度不大的属性屏蔽，增添新的属性，继续应用 MMHC 算法，构建新的贝叶斯网络。由于每次处理的属性数量较少，因此比起传统的 MMHC 算法，它的复杂度要低，并且容易揭示与目标属性最相关的影响因素。这种方法用于主因素分析中具有明显的优点。由于这种算法使用了对于属性学习的启发策略，因此称为启发式 MMHC 算法，记作 h-MMHC 算法。

MMHC 算法实际上是通过搜索-评分两个步骤确定出一个节点的父节点和子节点，进而形成数据之间概率关联关系的贝叶斯网络。如果相对于节点  $A$ ， $A$  的父节点  $Pa(A)$ ，节点  $B$  关于  $Pa(A)$  与  $A$  是条件独立的，则  $B$  与  $A$  之间没有边，也即  $B$  与  $A$  之间没有直接关联。在这样的前提下，关于属性集  $V = (v_1, v_2, \dots, v_n)$  的联合分布概率

$$P(v_1, v_2, \dots, v_n) = \prod P(v_i | Pa(v_i)) \quad (3.1)$$

由公式(3.1)可以容易计算出对于目标属性的主要影响因素，即主因素。

### 3.2.2 h-MMHC 的理论描述

基于贝叶斯网络可以方便的求出各属性之间的关联程度，因此贝叶斯网络结构的学习经常成为主因素分析算法采用的技术。本文提出的 h-MMHC 算法是一种改进的方法。

h-MMHC 方法具体的流程见下表 3.1:

表 3.1 h-MMHC 算法描述

Tab 3.1 The description of h-MMHC algorithm

算法描述: h-MMHC 算法
<ol style="list-style-type: none"> <li>1. 令 <math>V</math> 为所有属性的集合; <math>d</math> 为目标属性; 关联阈值 <math>\sigma</math>。</li> <li>2. 随机的选取部分属性集 <math>T \in V</math>, <math>d \in T</math>, 利用 MMHC 算法构造关于 <math>T</math> 的贝叶斯网络 <math>B</math>, 即 <math>B = \text{MMHC}(T)</math>。</li> <li>3. 对于所有 <math>z \in B</math>, 并且 <math>z \notin Pa(d)</math>, 若 <math>\rho(z, d) &lt; \sigma</math>, 则删去 <math>z</math>。</li> <li>4. 令删去的属性集为 <math>Z</math>; 添加属性集 <math>M</math>, <math>T = (T - Z) \cup M</math>, 执行步骤 2, 3。</li> <li>5. 直到所有的属性都完成, 得到最终的结构图。</li> </ol>

h-MMHC 算法采用启发式的步骤, 分阶段的添加属性, 而且每次迭代计算时, 只需要对含有  $M$  中的属性变量进行新的搜索和评分, 实际运行表明, 这种方法在很大程度上减少了计算量。这是 h-MMHC 算法区别于 MMHC 算法的独特之处。

该算法的正确性依赖于定理 1 以及定理 2。

定理 1: 令  $S$  为属性的集合,  $d$  是目标属性,  $Z$  和  $M$  是两个属性子集合, 对于任意属性  $X$  和  $Y$ , 如果关于  $Z$ ,  $X \notin pa(d)$ , 那么关于  $Z \cup M$ , 同样有  $X \notin pa(d)$ 。

证明: 根据 MMHC 算法,  $X \notin pa(d)$ , 意味着  $\text{Ind}_Z(X, d)$ , 其中  $Z \subset pa(d)$ , 即  $X$  和  $d$  关于  $Z$  是条件独立的。这时如果增加新的属性子集  $M$ , 则由 MMHC 的最大最小原则, 只要有  $d$  的父节点子集  $Z$  使得  $\text{Ind}_Z(X, d)$ , 则  $X$  也不会添加到新

结构中目标属性  $d$  的父节点集。

定理 1 说明, 当屏蔽一些属性, 在增加一些属性重新计算关联性的时候, 不会把原来非目标属性的父节点变成父节点。但是, 对于已是父节点的属性, 可能会由于新的属性加入而变成不是父节点, 此时, 会有新的父节点替代原来的父节点。因此 h-MMHC 算法在逐步添加属性将会逐步得到目标属性附近的正确结构。

根据 MMHC 算法中的最大最小原则, 在每一步中, 对于目标属性  $d$  而言, 总是把当前关于  $d$  的最小非条件独立属性作为  $d$  的父节点。同时把与  $d$  有最大相关系数的属性作为父节点的候选。如果两个属性  $X, Y$  对于  $d$  都有较大的相关系数, 在算法过程中, 则会比较  $Ind_X(Y, d)$  和  $Ind_Y(X, d)$ , 只有两个都不成立时, 才可能会把  $X$  和  $Y$  都作为  $d$  的父节点的候选, 否则其中一个将通过另一个父节点才连接到  $d$ 。如果  $Ind_Y(X, d)$  成立, 而  $Ind_X(Y, d)$  不成立, 即便由于逐步迭代的原因,  $X$  先被选取, 并进入  $d$  的父节点, 也会在  $Y$  选进来之后, 替代掉  $X$  而成为新的父节点。如果  $Ind_X(Y, d)$  和  $Ind_Y(X, d)$  都成立时, 这时可以比较  $X$  和  $Y$ , 选择对于  $d$  相关系数大者作为  $d$  的父节点的候选。当学习样例集给定时, 这些相关系数都是客观存在的, 因此 h-MMHC 方法可以通过计算这些相关系数得到各属性与目标属性之间的关联性质, 进而准确得到相应的贝叶斯网络结构, 这个过程与初始选取的属性集合无关, 因此, h-MMHC 算法稳定收敛到目标属性附近的结构。这就得到了定理 2 的内容。

定理 2: 给定学习样例集  $S$ , 对于任意初始的包含目标属性  $d$  的属性集合, h-MMHC 算法都能正确收敛到  $d$  的附近的结构。

针对定理 2, 作两点说明。第一, h-MMHC 算法的正确性依赖于学习样例集, 不同的学习样例集可能给出不同的贝叶斯结构。第二, 所谓目标属性附近的结构, 是指距离目标属性最近的那些节点, 至少包含所有的父节点(子节点不必考虑, 因为不属于主因素问题的内容)。具体的界定需要根据问题确定, 因为在实际中, 存在一些节点, 他们对于目标属性最终有较高的关联值, 但却不是父节点。这是 h-MMHC 算法的一个缺陷。这一缺陷可以通过对于主因素再做主因素分析来弥补, 也就是通过二次的主因素分析来得到这些属性, 这样的属性可以称为次主因素。当然, 这种情况是较少发生的, 除非该属性对于某个主因素有较高的单向相关性, 一般而言, 这样的属性设计是较少碰到的。

关于每次需要添加的属性, 既可以根据事先制定的添加策略自动添加, 也可以根据专家先验知识和样例分布特征选择性添加, 因此 h-MMHC 的属性添加具有灵活性, 可以根据问题的需要而裁量考虑。

### 3.3 本章小结

本章主要介绍了主因素分析的背景。主因素分析的技术自 Karl · Pearson 首次

提出以后，在 1991 年被 Turk Dengren 应用到人脸图像识别领域取得较好识别率后，便得到国内外不同领域研究者的极大重视。从此该方法在该领域不断被应用并且不断的被改进以适用于更加复杂的情况。国内外的很多研究者都对主因素的分析做了相当多的工作，如今主因素的分析技术已经相对来说比较成熟。其次，该部分的重点是介绍启发式的 MMHC 算法(heuristic-MMHC)，所谓启发式，是由于这种算法使用了对于属性学习的启发策略，因此称为 h-MMHC 算法，该算法采用分阶段进行属性的添加，相对于 MMHC 算法一次性的添加属性进行网络结构图的构造来说，具有结构图简单，运行时间短等优点。因而在主因素分析中得以应用。关于该算法的正确性，在该部分提出了定理 1 以及定理 2 作为相应的补充说明。



## 第四章 h-MMHC 算法在语音情感识别中的应用

### 4.1 基于 h-MMHC 的语音情感特征选择方法

#### 4.1.1 背景描述

从前面章节的描述中我们可以知道所谓语音情感识别也就是人们利用计算机从测试者的语音中提取出对情感有重要影响的特征，进而根据提取出的特征来判定被测试对象当前所处的情感状态。从人工智能领域中的相关知识我们知道，在进行语音情感识别以及其他分类工作中，并非是特征数目越多越好，有时用于分类的特征集中包含太多不相关的冗余特征，当把这些特征加在一起进行计算的时候，会增加计算机计算的复杂度以及消耗占据大量的内存空间，使得分类工作的时间效率特别低，有时甚至分类的精度也会受到影响。这就是我们经常听到的“维度灾难”现象。基于这样的问题，研究者们开始在前人工作的基础上，着眼于前人挑选出的特征，采取相应的技术进行相应的特征选择，也就是在原始大量的特征集合中利用特征选择的方法选取出部分子集合进行测试。不同的领域研究者们采取的方法不尽相同。例如在情感识别的领域，有研究者采用神经网络的方法结合主因素分析的技术进行特征选择，实验结果表明效果还是很理想的。

通过第二章中描述可知，因果关系是客观世界客观事物之间存在的关系，以及 Pearl 等人利用图论中的有向无环图 DAG 来表示特征属性之间的因果联系，并且以贝叶斯网络结构模型来对离散化的特征进行数学描述，使得特征之间的因果关系得以量化。因此贝叶斯网络通常又被人们称为因果贝叶斯网络，通过其结构图可以清楚地看到结构图中每个特征节点的原因节点以及子结果节点。

从第三章的介绍可以知道主因素分析的技术自提出以来受到学者们的广泛重视，并且成功的应用于很多领域。其中也不乏大量的研究者利用此方法进行特征选择以降低特征的维度。本文中就语音情感识别领域中的 MFCC 情感特征集利用 h-MMHC 算法进行特征选择。h-MMHC 算法采用启发式的策略逐阶段添加属性，与 MMHC 算法一次性考虑所有特征相比减少了运算的时间以及结构图的复杂度。

#### 4.1.2 MFCC 数据集介绍

在语音处理的领域中，我们常用基于声音频率的非线性梅尔刻度的对数能量频谱的线性变化来表示梅尔频率倒谱（Mel-Frequency Cepstral），也就是梅尔频率倒谱系数。MFCC 特征集是 Davis<sup>[50]</sup>于 1980 年首次提出并且应用于语音识别领域。自提出以后便被迅速的应用于语音情感识别中<sup>[51]</sup>。它是继 LPCS 和 LPCC 之后应用于自动语音情感识别的重要方法。

MFCC 的提取过程如下<sup>[52]</sup>：

- (1) 将采集来的一段语音信号分解成多个讯框。
- (2) 通过高通滤波器，来消除语音信号在收集过程中由于发声以及嘴唇等因素所造成的干扰效应。
- (3) 采用傅里叶变换，将时域上难以区分特性的信号转换成频域中的能量分布，从而可以让研究者们通过能量的不同分布来观察语音的不同特性。
- (4) 将第(1)步中得到的频谱通过使用梅尔滤波器得到梅尔刻度。
- (5) 在(4)中获得的梅尔刻度上来提取对数能量。
- (6) 将上述获得的结果采取傅里叶反变换，转换至倒频谱域上去。
- (7) 最后得到的 MFCC 就是上述倒频谱的图的幅度。通常获取其中的 12 个系数，与讯框能量组合得到 13 个系数。

对上述的提取过程有如下说明：

自然界中声音信号是连续变化的，为了将该连续信号进行简化，我们通常假设在一段相当短的时间内音频信号是不变化的。所以在采样时，将多个采样点集成一个单位就构成了所谓的讯框。一般情况下，讯框的长度不宜过短，过短易引起后续频谱计算时的误差，而过长的话则会导致讯框信号的变化幅度太大，不利于后续计算。因此通过研究者们大量的实验得出，通常一个讯框的取值范围是 20-40ms。

通过上述步骤得到的频谱信号中存在一些人耳无法分辨的高频信号，因此采用梅尔滤波器进行过滤。所谓的梅尔滤波器就是用一组 20 个非线性分布的三角带通滤波器组合而成，它可以求出每一个滤波器输出的对数能量。需要注意的是，这 20 个非线性分布的三角带通滤波器在梅尔刻度上是均匀分布的。

考虑到人耳对不同频率的感受差异，梅尔频率通常是根据人类听力的特征来规定的。对于情感识别来说，MFCC 特征集通常包括平均值、最大值、最小值、中值以及标准差。

在该研究中所采用的 MFCC 数据集是从 10 个德国受测者一共 545 条语句中提取的语音信息，该数据集中一共 7 类情感，对应的标签为 1-7，对应的情感分别是 happy, sad, anger, neutral, fear, anxiety 以及 disgust。一共由 180 个属性特征列加上一个情感标签列组成的 535\*181 的数据矩阵。除了最后的列是情感标签列，前面的 180 列对应的是统计意义下的特征，例如平均值、最大值、最小值、中间值以及标准差等统计量，是手工提取出来的统计特征集。

## 4.2 研究方案与实验设计

### 4.2.1 研究方案

基于 h-MMHC 算法的语音情感特征选择的流程图可以见如下图 4.1:

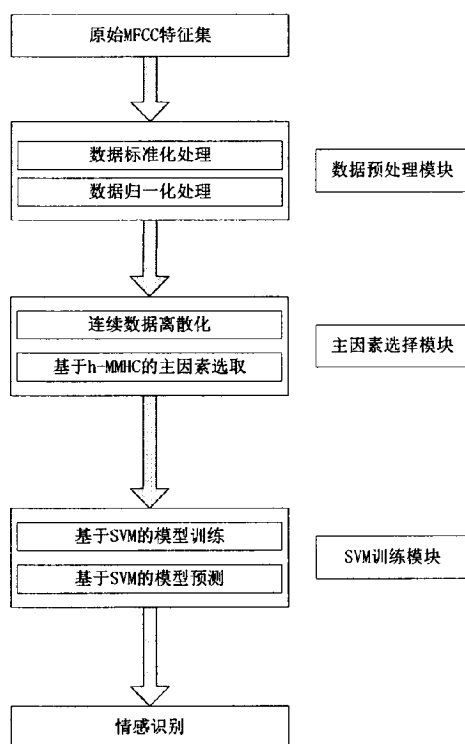


图 4.1 基于 h-MMHC 算法的语音情感识别特征选择流程图

Fig 4.1 The flow chart of speech emotion recognition of feature selection based on h-MMHC algorithm

从上述的流程图中可以看到整体的实验可以分为三大模块，分别是：数据预处理模块、特征提取模块、SVM 情感训练与识别模块。

(1) 数据预处理模块：鉴于 MFCC 特征集中部分特征列之间的数据不规范，无法统一计算，因此执行数据的标准化和归一化操作。语音情感识别领域中，标准化是普遍采用的数据预处理方法之一<sup>[53]</sup>。标准化的目标是消除在语音识别过程中受测者所说的语音与录音过程中所产生的变化所引起的噪音<sup>[54]</sup>。通常标准化的方法有 max-min 标准化方法以及 z-score 标准化方法，在该研究的实验中采用的是 max-min 标准化的方法。

(2) 在将 MFCC 的数据列进行标准化操作后便是对 MFCC 数据的各个样本数据进行归一化操作。所谓归一化操作也就是将原始有量纲的数据转换为无量纲的数据，从而使得数据之间具有可比性。进行归一化操作后不会改变原始数据之间存在的相关关系。在统计学上所谓的归一化是归纳统一样本的统计分布性。

(3) 特征提取模块：在该模块中首先利用在数据预处理模块中得到的结果进行离散化处理。由于本文提出的 h-MMHC 算法只适用于离散的数据，因此必须将 MFCC 中连续的数据进行离散化。离散化的程度则根据利用 h-MMHC 算法选取主

因素的节点数最大化原则而定。在 MFCC 的中的标签列一共有七种，代表 7 类不同的情感。因此在利用 h-MMHC 算法进行特征选取的时候，是分情感分别进行选取的，即找出不同情感中不同的主要影响因素。

(4)SVM 训练模块：该模块是实现 7 个独立情感类别模型的训练与识别。该研究的主要思想是利用主因素分析的思想进行特征的选取，借助贝叶斯网络结构模型的手段找出对情感类别有主要影响的特征。接着利用找出来的主要影响因素进行情感识别，通过实验观察特征选择后的识别率与未进行特征选择后的识别率相比是否有所提升，如果提升的话则说明通过该方法挑选出的特征是有效的，进而说明该研究的工作是有意义的。反之则说明该方法在该领域是不适用的或者此次实验的数据集是存在不合理的地方的。在该模块中的语音情感分类器选用的是基于 SVM 的语音情感分类器。

#### 4.2.2 实验设计及各模块的实现

(一) 数据预处理模块的实现：整个数据预处理模块中分为数据标准化和归一化两种处理方式。对于数据的标准化和归一化处理通过编写 `chuli().m` 函数来实现数据的标准化和归一化操作。将处理后的数据分开存放：不含类标签的前 180 列数据存放在 `datatarget` 矩阵中，而类标签则存放在 `datacategory` 向量中。`Datacategory` 是一个  $526 \times 1$  的向量。526 是源自数据有效化处理后的结果。

(二) 特征选择模块的实现：在该模块中主要实现的功能是数据的离散化以及主因素的选取。对于数据的离散化操作通过编写 `getdatacha().m` 函数来实现对于不同情感类别的数据采取不同的离散程度。将 7 类数据离散后的结果分别存放在 `dt1` 至 `dt7` 中，该结果用于下一阶段也就是主因素选取阶段。当把总的 7 类数据分成 7 类分别存放时，数据相应的处理成  $526 \times 181$ ，其中第 181 列不再是 7 个不同取值，转而代之的是 0,1 二分类标签。这样的处理是为了与本文的研究目标相一致：寻找对每一类情感具有主要影响的因素。通过上一步的数据离散化后，接下来便是将离散化后的结果代入 h-MMHC 算法中去寻找主因素。通过第三章关于 h-MMHC 算法的介绍我们可以知道，h-MMHC 算法是在 MMHC 算法的基础上采用逐步添加属性的方式来构造贝叶斯网络结构图，即由局部到整体的思路。在本文的研究试验中，由于特征集有 180 个，数目相对来说比较大。此次实验的分阶段原则是根据 MFCC 的 12 个倒谱系数来分，每个倒谱系数大概 6 个统计值，因此大概可以分为 30 个阶段依次构造贝叶斯网络结构。每个阶段中目标节点始终保存，下一阶段的结构图构造根据上一阶段的属性进行相应的调整。只有对目标节点有影响的节点才会在下一阶段保留下来，没有影响的节点则被替换为新的节点。该步骤的实现是通过编写函数 `getmaicha().m` 来实现的。

(三) SVM 训练模块的实现: 在该模块中主要是基于 SVM 分类器来进行各个情感的模型训练与识别。SVM 分类器主要包括训练阶段和预测阶段, 因此相应的将数据集分为训练集和测试集, 本文的实验中采用的是十次交叉验证, 也就是整体的会将数据 10 等份, 其中 9 份作为训练数据, 1 份作为预测数据。编写函数 `gettraintest().m` 函数, 用于提取在数据离散化后进行训练集和测试集的选择。基于选择好的训练集和测试集训练相应的分类模型。因为该研究的目的是寻找对每一类情感具有主要影响的因素, 因此我们对 7 类情感分别训练模型。数据离散化的步骤中将每一类情感训练成一个二分类器。关于 SVM 分类器的模型训练是基于 Matlab 中的工具包 Libsvm 中的函数 `svmtrain()` 函数和 `svmpredict()` 函数来实现的。

### 4.3 SVM 分类器的介绍

提到语音情感的识别, 研究者们提出了很多的分类模型, 包括高斯混合模型 (GMM)、人工神经网络 (ANN)、隐马尔可夫模型 (HMM) 以及支持向量机 (SVM), 这些模型都不止一次的被人们反复研究。SVM 充分利用凸二次最优函数, 该函数在解决全局最优的问题中具有很大的优势。在模式识别领域中的一些分类问题, SVM 也表现出了很好的特性, 例如生物信息学、文本分类以及人脸情感识别中都用到该分类模型。同样的, 大量的研究者努力将 SVM 应用于语音情感识别, 实验结果表明基于 SVM 的分类器的分类效果明显的优于其他比较出名的分类模型<sup>[55]</sup>。

#### 4.3.1 结构风险最小化

机器学习的本质是通过学习来逼近问题的真实模型。为了求解一个问题的真实模型, 我们总是借助很多的假设模型去训练学习, 想要找到问题的真实模型, 但是往往问题的真实模型是未知的, 因此我们用“风险”来度量假设模型与真实模型之间的误差 (所谓风险也就是误差的累积)。通常用数学形式可以定义为<sup>[56]</sup>: 变量  $x$  与  $y$  之间存在未知的依赖关系  $F(x, y)$ , 机器学习根据  $n$  个独立同分布的观测样本值  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , 在一组函数  $f(x, w_i)$  中寻找最优函数  $f(x, w_0)$  来对未知函数依赖关系进行估计, 使得真实风险  $R(w)$  最小。真实风险可以用如下数学式表示:

$$R(w) = \int L(y, f(x, w)) dF(x, y) \quad (4.1)$$

公式 (4.1) 中  $\{f(x, w)\}$  是预测函数集, 通常预测函数又称为学习模型,  $w$  是函数的广义参数,  $L(y, f(x, w))$  称为使用  $f(x, w)$  对  $y$  预测所形成的损失。

一般情况下, 真实风险是无从得知的, 只有通过某些可以控制的参数去不断的逼近。随后人们提出用某个变量来表示分类模型在给定样本数据上的分类结果与真实结果之间的差值来表示, 该变量就是后来常用的经验风险  $R_{\text{emp}}(w)$ 。数学

公式表示如下:

$$\text{Remp}(w) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i, w)) \quad (4.2)$$

其中  $n$  表示给定样本的数目。以前的机器学习中通常把经验风险最小化作为学习的目标,但是在后来的研究中发现很多分类函数在给定样本集上达到 100% 的识别率,但是真实分类时却表现的很差,有时甚至是 0% 而宣布失败。经验风险最小化的一个重要的前提是用于训练学习的样本数量要足够的大,大到可以包容各种各样的情况,但是现实中这是很难办到的。

基于这样的考虑,统计学习领域的研究者们引入了“泛化误差界”的概念。该概念指出真实的风险不应该仅仅依靠经验风险,还应该加入置信风险。所谓置信风险表示我们对于分类器在未知样本上的分类结果,能够接受的程度是多少,因此,置信风险不是一个确切的数值,而是一个区间范围。泛化误差界的概念便由此而来。置信风险不仅与样本的大小有关,还与分类函数的 VC 维有关。VC 维 (Vapnik-Chervonenkis Dimension) 是对函数类的一种度量,也可理解为问题的复杂度指标,同时也是机器学习过程一致收敛的速度以及推广性的指标。对于泛化误差的定义,可用如下表达式表示:

$$R(w) \leq \text{Remp}(w) + \Phi(h/n) \quad (4.3)$$

通过公式 (4.3) 可以知道,当给定的样本数量越大的时候,学习的结果可能越准确,相应的置信风险就越小;其次,当分类函数的 VC 维  $h$  越大,其推广能力也就越差,那么相应的置信风险就会随之增大。统计学习的目标自此从追求经验风险最小化转为追求经验风险和置信风险之和最小化,又称为结构风险最小化。在 SVM 中,为了寻求经验风险和置信风险之间的折中,防止“过学习”或“欠学习”以及学习过程中遇到的局部最小化等等不足,采用了结构风险最小化的准则<sup>[57]</sup>。所谓结构风险最小化准则也就是首先将函数集划分为一个个子集,其次使所有子集按照 VC 维的大小进行排序,最后在排好序的各个子集中寻找最小经验风险,并且折衷考虑子集间的经验风险和置信风险,从而使得真实风险达到最小化。

在实际的机器学习中,会出现“欠学习”或者“过学习的现象”。所谓欠学习是指样本数量较小,使得在该样本上的经验风险较大,与之相反的是当训练样本过大,那么在该样本上的置信风险就会增大。为了克服这些不足, SVM 采用了结构风险最小化准则进行优化。

### 4.3.2 最优分类超平面

SVM (Support Vector Machine) 是 Cortes 和 Vapnik 等人于 1992 年在计算学习理论的会议上介绍进入机器学习的领域, 自此各领域的研究者们开始广泛的关注 SVM 并不断的加以创新以适用于该领域的问题。SVM 算法是统计学习理论的一种实现方式, 它既追求训练样本上的准确性又考虑学习空间的复杂性, 适当的将二者进行折中, 是一种很好的分类算法。SVM 分类器对线性可分的情况进行分析, 对于低维空间中线性不可分的情况选用相应的核函数, 映射到高维空间中去, 从而达到在高维空间线性可分的目的。SVM 分类器的分类问题简单的说就是寻找一个最优超平面, 该超平面可以将样本点很好的分开达到分类的目的。通常, 在线性可分的情况下会有多个超平面将样本点分开, 我们选取其中能使两类中离它最近的样本与他距离最大的超平面作为最优超平面。设  $H$  为分类面,  $H_1$ 、 $H_2$  分别是平行于该分类面, 且是过这两类样本中距离分类面最近样本的直线。定义  $H$  与  $H_1$  之间、 $H$  与  $H_2$  之间的间隔为几何间隔  $\sigma_i$ 。

$$\sigma_i = \frac{1}{\|W\|} |g(x_i)| \quad (4.4)$$

公式 (4.4) 中  $x_i$  表示一组线性可分的样本  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ ,  $g(x_i) = w_i x_i + b$ , 其中  $W = (w_1, w_2, \dots, w_n)$ , 所以该式表示样本数据中每个样本点到分类面的距离。通常定义一组样本中某个样本点的几何间隔最小的为该样本的几何间隔。

分类间隔 (Margin) 就是  $H_1$  与  $H_2$  之间的距离。落在  $H_1$  与  $H_2$  上的训练样本点称为支持向量 (Support Vector 简记为 SV)。无论对于  $H$  还是  $H_1$  或  $H_2$  它们就是一个分类函数, 可以将两类样本完全分开。所谓线性函数可以如此理解: 在一维空间中就是一个点, 二维空间中表现为一条线, 三维空间中则是一个平面, 其他可以依次类推下去。如果不考虑空间维数的话, 统一称作超平面。

### 4.3.3 核函数

早在 20 世纪 60 年代时, Minsky 和 Papert 就提出了当时普遍使用的线性分类学习器对于一些线性不可分的样本存在很大的局限性, 首先表现的就是其计算能力是有限的。对于现实中客观存在的复杂样本数据, 我们应该使用一种比线性函数更具有表达能力的假设空间。也就是说目标概念在一般情况下不能通过给定属性的简单的线性函数直接组合产生, 而应该寻找待研究数据中更抽象化的特征进行组合 [58]。基于该思想, 研究者们提出了核这一概念, 也就是将数据映射到高维空间来增加线性学习器的计算能力。

机器学习中, 对于目标函数的不同表示构成了不同目标函数的复杂度, 如此

带来的学习任务的难度也是显而易见的。理想情况下，我们应该针对特定的学习问题选择合适的表达式，该想法便涉及到我们在数据的预处理阶段需要做的一步工作就是数据的空间转换，也就是表达形式的变换： $x=(x_1,\cdots x_n)\mapsto \phi(x)=(\phi_1(x),\cdots \phi_n(x))$ ，该操作也就是将原始输入空间中的  $x$  映射到一个新的特征空间  $F=\{\phi(x)|x\in X\}$  中。

SVM 最初的应用领域是文本的分类，刚开始人们主要做的是二分类的问题。但是随着研究的不断深入，人们已经无法满足二维空间中的线性可分问题，转而将注意力转移到那些在二维空间中线性不可分的问题的解决办法。在经过不断的探索之后，人们发现在二维空间中的分类函数是线性的  $y=ax+b$  的形式，对于二维空间中线性不可分的问题，如下图所示：

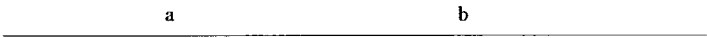


图 4.2 二分类情况的图示

Fig 4.2 Diagram of two classification

在该图中，a,b 之间是一类正例，那么在 a,b 左右的反例通过使用线性的分类函数便无法正确分类。如此人们想到使用另外的分类函数  $y=ax^2+bx+c$  这样的二次函数得到如下图形：

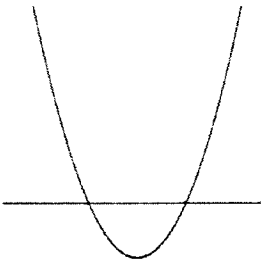


图 4.3 线性不可分情况的展示

Fig 4.3 Presentation of linear inseparable

通过图形可以知道，位于抛物线中间红色表示的为所有正例，所有的正例对应的函数值为负值，而所有的反例对应的函数值为正值，如此便达到分类的目的。如上所述的线性函数以及二元函数也就是我们所说的核函数。具体的核函数的定义如下所示。

定义 4.1<sup>[59]</sup> 所谓的核就是一个函数  $K$ ，对所有的  $x,z\in X$  满足：

$$K(x,z)=\langle \phi(x)\bullet \phi(z)\rangle \tag{4.5}$$

公式 (4.5) 中的  $\phi$  也就是  $X$  到特征空间  $F$  的映射。通常一个函数可以作为核



函数的条件是该函数满足 Mercer 定理。

随着人们对 SVM 认识的不断深入，更多的核函数也不断的被发现。常见的核函数有如下几种：

#### (1) 线性核函数

所谓线性核函数也就是对于线性可分的分类问题常采用该类型的核函数。这是最基础的核函数类型。可以表示为：

$$K(x_1, x_2) = (x_1, x_2) \quad (4.6)$$

#### (2) RBF(径向基)核函数

所谓径向基核函数也就是某种沿着径向对称的标量函数。通常定义为在空间中的任意一个点  $x_1$  到某一中心  $x_2$  之间的欧式距离，其是一个单调的函数，表示

$$\text{为: } K(x_1, x_2) = \exp(-\gamma |x_1 - x_2|^2) \quad (4.7)$$

通常将  $x_2$  称为 RBF 函数的中心， $\gamma$  称为该函数的宽度参数，用于控制该函数的径向作用的范围。高斯核函数是常见的一种径向基核函数。

#### (3) 多项式核函数

所谓多项式核函数也就是选取的映射函数是多项式的形式，其数学表达式可见如下：

$$K(x_1, x_2) = (s(x_1, x_2) + c)^d \quad (4.8)$$

公式 (4.8) 中的  $d$  用来设置该多项式核函数的最高项指数； $s$  用来设置核函数中的  $\gamma$  参数， $c$  用来设置核函数中的  $\text{coef}()$ ，当  $c=0$  时，该核函数是齐次的多项式核函数，否则为非齐次的多项式核函数。

#### (4) Sigmoid 核函数

该函数的特点是连续且光滑，并且严格单调，可以作为阈值函数的首选。其数学表达式如下：

$$K(x_1, x_2) = \tanh[\rho(x_1, x_2) + b] \quad (4.9)$$

公式 (4.9) 中参数  $\rho$  用来进行  $\gamma$  参数设置，默认为  $1/K$ ，其中  $k$  为属性个数， $b$  用来设置  $\text{coef}()$  参数，通常默认为 0。

实际的研究工作中，关于具体的核函数的选择还没用统一的标准。通过大量的研究者的实验结果表明，在一些问题中该类型的核函数效果比较好，而在另一些问题中却效果很差。一般情况下，使用 RBF 径向基核函数得到的效果都不会太差。通常在实际中，一开始研究者们都是根据经验进行选取，在大量的实验基础上再选取效果较好的核函数。

4.4 实验结果分析

4.4.1 实验环境及其 LIBSVM 参数设计

此次实验是在一台 CPU 是 Inter(R)2.94GHz，内存为 4GB，操作系统为 Windows7 的电脑上进行的。部分代码采用的是 Causal-Explorer Matlab\_R14 工具包、FULLBNT 工具包以及 LIBSVM 工具包实现。

LIBSVM 工具包是由台湾大学林志仁教授等人的小组开发设计的，该工具包具有简单、易于使用和快速有效的特点。该软件包针对 Windows 的系统环境不仅提供了编译好的可执行文件，还提供了相应的源代码以方便其他研究者改进、修改或者移植到其他操作系统中进行使用。该软件包提供了大量的默认参数，研究者们基于这些参数就可以获得较好的分类效果，对于 SVM 所涉及的参数调节相对比较少。除此之外，该工具包还提供了交叉检验的功能，此次的实验中我们采取十次交叉验证来完成实验。

在该实验中基于 LIBSVM 工具包进行实验主要包括如下几个步骤：

(1) 在 LIBSVM 的工具包中，在进行 SVM 分类训练之前，通常将原始的数据进行相应的预处理。按照该软件包要求的格式处理数据。该研究中的实验针对原始的数据，一共 535 条语音样本数据，其中 9 个样本包含了无效的数据(NaN: not a number)，因此试验中将这数据去掉，剩下的 526 条数据作为 SVM 分类数据。针对 526 条样本数据，按照十次交叉验证的方式，选取其中的 90%数据集作为 svmtrain 的训练数据以获取 SVM 训练模型，然后剩下的 10%的数据集用于 svmpredict 以进行测试。

(2) 观察数据集中的数据形式可以发现不同属性的数据之间差异较大，为了试验效果考虑，在训练之前对去掉无效样本行后的数据进行标准化和归一化的操作，使得数据之间变成无量纲化的数据从而更方便处理。而试验的结果也证实了在未进行数据处理时的识别率要明显低于数据处理后的识别率。具体的对照可见如下表所示：

表 4.1 数据处理前后识别率对比

Tab 4.1 Recognition rate before and after data processing

数据格式	识别率
未进行处理的数据	65%
经过处理后的数据	71%

(3) 核函数以及交叉验证最佳参数的选择。基于上一节关于 SVM 理论知识的介绍可以知道目前关于具体的核函数的选择通常是基于经验，还没有具体的指

导原则。在该试验中，根据以往的实验经验我们选用线性核函数作为此次实验的核函数类型。通过实验结果选取其中的参数分别为:[ “-v 10    -t 0” ), 根据具体的情况选择适合该模型的最佳惩罚因子。其他的SVM参数则采用其默认设置。对于原始的数据集，该参数设置下不同的核函数对应的识别率如下表 4.2 所示：

表 4.2 不同核函数情况下的识别率对比

Tab 4.2 comparison of different kernel functions in case of recognition rate

t=	识别率
0 (线性核)	72%
1 (多项式核)	24%
2 (RBF 核)	68%
3 (sigmoid 核)	70%

从该表可以知道，在此次的实验中我们选取线性的核函数。

基于上面步骤的考虑之后，就获得了SVM的训练模型。该阶段的目标就是获取最大限度拟合目标函数的参数 c 值以及 g 值，基于该研究的特殊性，g 值选取默认的参数设置。

(4) 利用学习出来的 SVM 训练模型进行语音情感的测试与预测。

4.4.2 各情感类别主因素提取结果与分析

多数场合中，有时主因素起到主要的影响作用，但是次主因素可能也会起到较好的辅助作用，使得整体的识别率会有所提高。本文研究中，在情感1到情感7的主因素提取出来后，整体的识别率略低于未进行属性选择的识别率，但是当加入了次主因素后，整体的识别率确实在原来基础上提高了，并且较未进行属性选择时的识别率有所提高。

通过上文的介绍可以知道，MFCC 特征集是一个 535\*181 的样本集，也就是 180 个属性列，第 181 列是情感类别标签列，包括 7 种情感类别（用数字 1~7 这 7 个数字进行数字化表示），7 种情感类别分别是 happy, sad, anger, neutral, fear, anxiety 以及 disgust。除了最后的列是情感标签列，前面的 180 列对应的是统计意义下的特征，例如平均值、最大值、最小值、中间值以及标准差等统计量，是手工提取出来的统计特征集。MFCC 是梅尔频率倒谱系数，一共是 12 个倒谱系数，后来的研究者将 12 个倒谱系数分别求取其均值、最大值、最小值、中间值以及标准差等，如此便扩展成 12\*5 个属性列，同时对这些系数使用高通滤波器进行过

滤，使用参考阈值系数为 0.97。此方法将 12 个倒谱系数扩展成了 13 个系数。后来又有人对这 13 个属性求取其一阶差分、二阶差分，如此便形成了本文中的 MFCC 特征集。

本研究中提出的 h-MMHC 算法是启发式的 MMHC 算法，也就是在原始 MMHC 一次性添加属性的基础上将属性分阶段添加进来。在该研究中按照各个倒谱系数的统计值为原则来依次构造各个情感状态下寻找主因素的贝叶斯网络结构图。基于篇幅的限制，该部分仅仅展示具有主特征在内的结构图。

(1) 情感类别 1 (Happy) 中的主因素

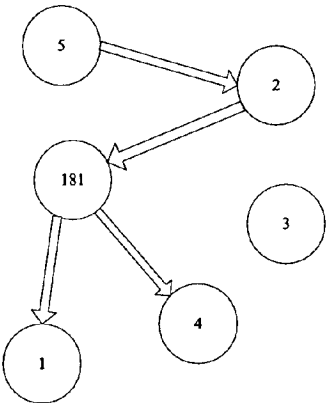


图 4.4 对情感 1 有影响的主因素 2 及其次主因素

Fig 4.4 The main factor 2 effecting emotion 1 and its secondary main factors

图 4.4 表明特征 2 是影响情感 1 的主因素，基于对次主因素的考虑，特征 5 是影响 2 的主因素，因此下一阶段的实验中依然保留特征 5，h-MMHC 算法的核心就是分阶段的添加特征组，每一次的添加都是基于上一阶段的基础上，保留我们需要的因素，不断添加，从而更深层次的挖掘出对分类类别有影响的特征。

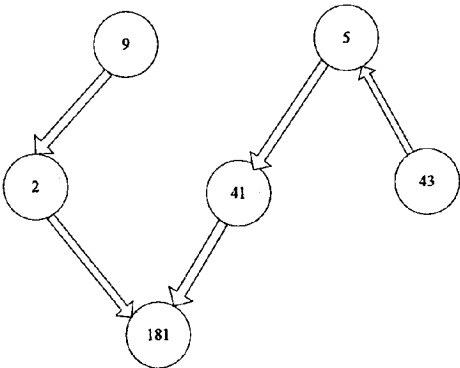


图 4.5 情感 1 中包含所有主因素及次主因素在内的结构图

Fig 4.5 The structure of emotion 1 contains all main factors and secondary factors

图 4.5 可以说明的是，在刚开始的阶段找出了特征 2 是目标节点 181 的主要影

响因素，特征 5 在一开始对特征 2 是有影响的，但是随着后来其他节点的不断加入，会改变特征节点之间的影响关系，正如该结构图展示的，节点 2 在一开始被找到，所以后续的阶段被保留下来，最后直到实验结束，找出了另一个特征 41 对 181 也是有影响的，相应的还找出了他们的次主因素。该研究中，正因为考虑到次主因素的影响所以在刚开始便保留次主因素。

## (2) 情感类别 2 (sad) 的主因素分析

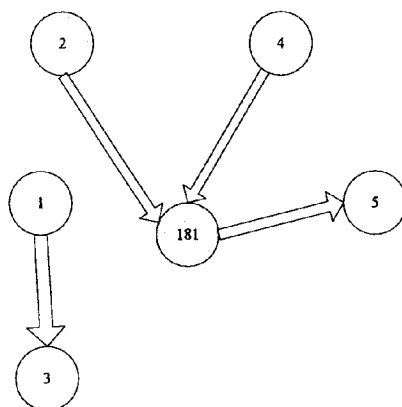


图 4.6 情感 2 第一阶段找出的特征

Fig 4.6 Characteristics was found at the first stage of emotion 2

图 4.6 表示在刚开始的阶段特征 2 和 4 在第一阶段时对分类节点存在影响，那么该关系提示我们在接下来的阶段中需要保留这俩特征，h-MMHC 算法告诉我们：前一阶段的主因素，在随着节点不断加入的时候可能会改变特征的影响关系。如此便得到了如下的图 4.7：

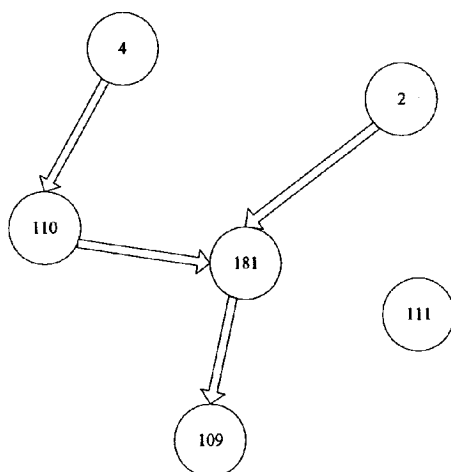


图 4.7 情感 2 在主特征 110 出现时的结构图

Fig 4.7 The structure of emotion 2 after main factor 110 was added

图 4.7 所示的是在不断加入节点后，结构图依次变化，直到加入到 110 这组的特征后得到如上图所示的结构图。通过图中可以看到在当前阶段时，特征 4 已经成为第 110 特征的主因素，不再是 181 目标节点的主因素。因此在下一阶段的构图中便保留了特征节点 2,4,110，便得到如下图 4.8:

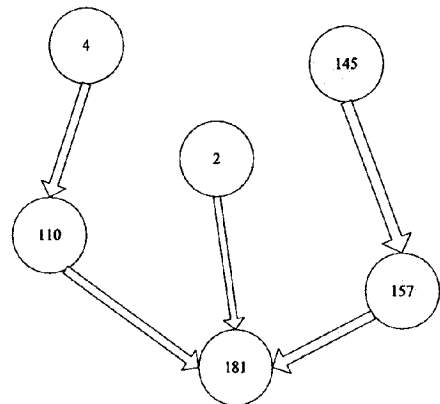


图 4.8 情感 2 在主因素 157、158 加入时的结构图

Fig 4.8 The structure of emotion 2 after main factor 157、158 were added

图 4.8 显示的是在包含 157、158 特征节点在内的特征组加入网络后产生的网络结构图。如图所示，当该组节点加入后，网络的结构图发生了变化，原来对 181 起主要影响的特征 2 在 158 的加入后，成为了影响 158 的主因素，而 158 成为影响 181 的主因素。为了简化结构图，在图中仅展示主因素以及次主因素在内的节点。同样的，对于特征节点 145，在特征 157 的加入后成为 157 的主因素。同样根据上面的结论保留图中的主因素节点以及次主因素节点，继续添加新的特征组，得到如下的结构图 4.9:

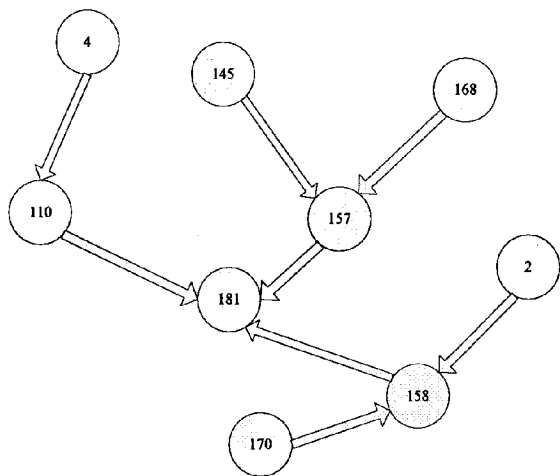


图 4.9 情感 2 的完整主因素及其次主因素的网络结构图

Fig 4.9 The structure of emotion 2 after all main factors and secondary factors were added

图 4.9 所展示的是将一系列的特征组添加后得到的网络结构图，后面的特征组在不断的加入，但是特征之间关系没有发生大的变化，因此得出图 4.10 这样的结构图。该结构图是经过一系列网络图后的精简图，在该图中可以看到对情感 2 有影响的主因素分别是特征 110、157、158，同时图中还展现了他们各自的主因素，也就是我们统称的次主因素。

(3) 情感类别 3 (anger) 的主因素分析

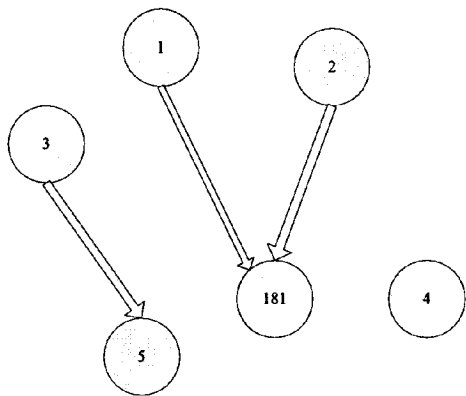


图 4.10 情感 3 在第一阶段产生的结构图

Fig 4.10 The structure chart of emotion 3 on the first stage

图 4.10 表示的是当第一个特征组加入到该算法中时所产生的网络结构图，通过该图可以清楚的看到对第三类情感 (anger) 有影响的主因素，也就是在结构图中有直接的边与 181 特征节点相连的分别是特征节点 1 和 2，节点 4 以及 3 和 5 对

181 是没有直接影响的，因此在下一阶段中只保留节点 1 和节点 2，如此经过反复的训练得到如下图 4.11:

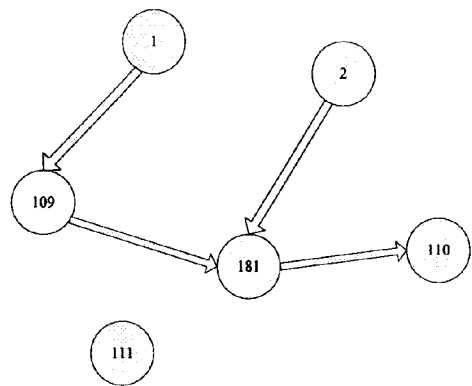


图 4.11 情感 3 中包含主因素 109 的结构图

Fig 4.11 The structure of emotion 3 contains main factor 109

如图 4.11 所示的是在包含 109 所示的特征组加入的情况下所得到的网络结构图，从图中可以看到，在 109 出现之前，特征节点 1 是 181 的主因素节点，可是当 109 出现的时候，却成为了 109 的主因素。这也从另一个方面展示了特征节点 109 对 108 节点具有更强的影响。而节点 2 却一直表现出它对节点 181 的主要影响。当接下来的特征组一次被添加进来的时候，得到图 4.12。

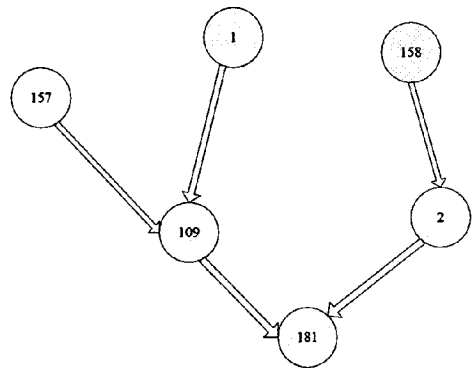


图 4.12 情感 3 中包含所有主因素和次主因素的结构图

Fig 4.12 The structure of emotion 3 contains all main factors and secondary factors

如图 4.12 所示的是剩下的特征组依次添加进来后总结得到的结构图。通过该图可以看到特征节点 1 和 157 依次成为 109 的主要影响因素，158 则成为特征 2 的主要影响因素。而特征节点 2 和 109 则成为 181 的主要影响因素。

(4) 情感类别 4 (neutral) 的主因素分析



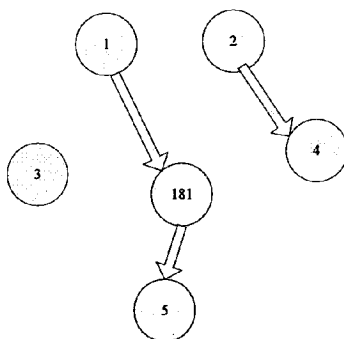


图 4.13 情感 4 中第一个特征组产生的结构图

Fig 4.13 The structure chart of emotion 4 with the first set of feature

如图 4.13 所示的是将 neutral 所属的第一个特征组添加进来后所形成的结构图。在第一个特征组中，可以通过结构图看到特征节点 1 在该阶段是 181 特征节点的主要影响因素，节点 2 是节点 4 的主要影响因素，6 是 3 的主因素节点。尽管在该图中可以看到 181 节点对 5 是有影响的，但是我们在该研究中探讨的是对 181 有影响的特征，因此在下一阶段添加特征组时仅仅保留特征 1。如此便得到图 4.14。

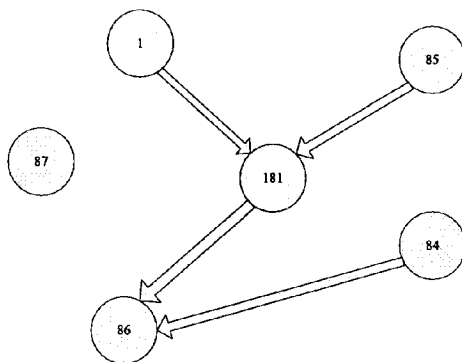


图 4.14 情感 4 中特征 85 加入时的结构图

Fig 4.14 The structure of emotion 4 after feature 85 was added

如图 4.14 所示可以看到当 85 所在的特征组添加进来的时候，特征 85 成为影响 181 的主因素。而节点 1 并未改变影响。特征 84 对 86 产生影响，虽然 181 对 86 在当前阶段也有影响，但是基于研究的主题是研究对 181 有主要影响的因素，因此继续添加特征组，仅保留对 181 有影响的主因素 1 和 85，如此构成了如下图 4.15。

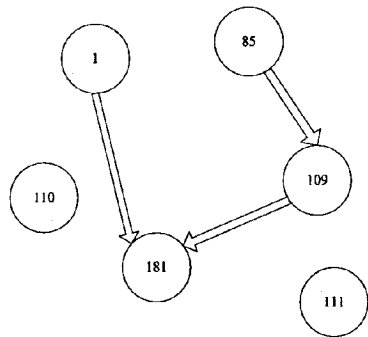


图 4.15 情感 4 的主特征 109 加入后的结构图

Fig 4.15 The structure of emotion 4 after main factor 109 was added

如图 4.15 显示的是在包含 109 特征在内的特征组添加进来后所得到的结构图，通过该图可以看到原先的特征 85 是 181 的主要影响因素，但在 109 特征添加进来后变成了 109 的主要影响因素，转而是 109 变成了 181 的主要影响因素，但是特征 1 却一直未变化，仍然影响着特征 181。而特征节点 110 和 111 则是孤立的，表示在此次的添加阶段这俩特征节点既不对其他特征产生影响也不受其他特征的影响。基于研究的目的，仅仅保留特征节点 1 和 85 以及 109，从而得到如下的结构图 4.16。

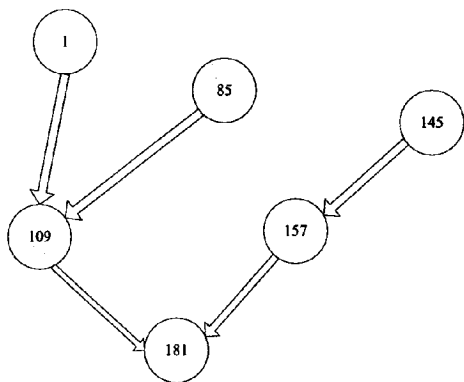


图 4.16 情感 4 中主因素 157 加入时的结构图

Fig 4.16 The structure of emotion 4 after main factor 157 was added

如图 4.16 显示的是继包含 109 特征在内的特征组添加后依次加入不同特征组所出现的情况。基于结构图清晰的考虑，在此仅保留其中有关系的相关特征，而忽视其他非影响的因素。通过该图可以看到特征 157 也是 181 的主因素，而 145 则是 157 的主因素。保留图中的特征，依次添加下面的特征组，得到如下的结构图 4.17：

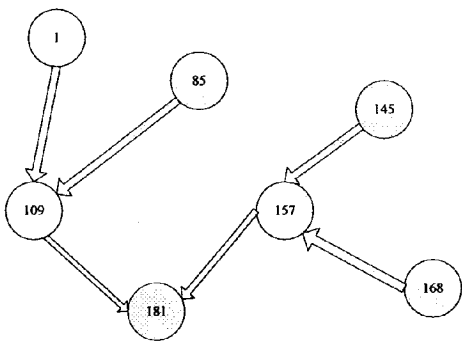


图 4.17 情感 4 中包含全部主因素和次主因素的结构图

Fig 4.17 The structure of emotion 4 contains all main factors and secondary factors

如图 4.17 所示的是继特征 157 之后依次添加接下来的特征组得到的网络结构图。通过该图可以看到特征 168 是特征 157 的主要影响因素。该图是一个包含全体主因素的结构图，可以看到对情感 4 有影响的特征分别是特征 109 和 157，而对特征 109 有影响的则是特征 185，对 157 有影响的则是特征 145 和 168。

(5) 情感 5 (fear) 的主因素分析

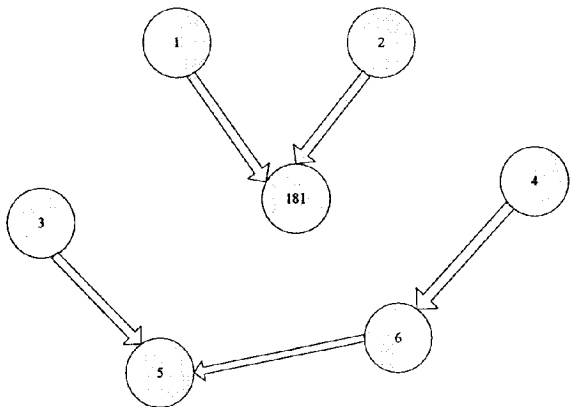


图 4.18 情感 5 中第一阶段的结构图

Fig 4.18 The structure chart of emotion 5 on the first stage

如图 4.18 所示的是第一个特征组添加进来时所得到的结构图，该图中可以看到特征节点 1, 和 2 分别是实行 181 的主要影响因素，而其他四个特征则是彼此之间有影响，但是并不对 181 产生影响，因此保留特征 1 和 2，继续的添加特征组，得到如下的结构图 4.19。

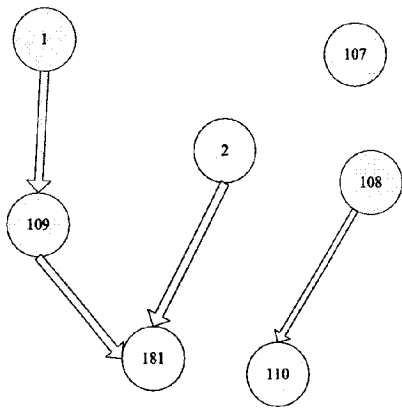


图 4.19 情感 5 中加入 109 在内的特征组后的结构图

Fig 4.19 The structure chart of emotion 5 after the set of feature contain 109 were added

如图 4.19 所示的是在加入包含 109 特征在内的特征组后得到的贝叶斯网络结构图。通过该图可以看到当 109 特征被加入的时候，特征 1 改变了其影响性，由原来的对 181 为主要影响转而成为特征 109 的主要影响因素。而特征 2 却没有改变其影响性。同样按照上面各阶段同样的处理方式，保留节点 1 和 2，依次添加接下来的特征组，得到如下的结构图 4.20。

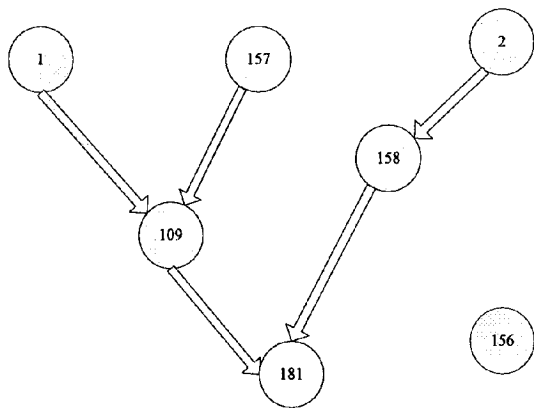


图 4.20 情感 5 中加入 157 在内的特征组后的结构图

Fig 4.20 The structure chart of emotion 5 after the set of feature contain 157 were added

如图 4.20 所示的是在加入包含特征 157 在内的特征组后所得到的网络结构图，通过该图可以看到在该图中对特征 181 有主要影响的因素分别是特征 109 和特征 158，而对特征 109 有影响的分别是特征 1 和特征 157。特征 2 从刚开始对特征 181 有影响，在加入该特征组后开始变成 158 的主要影响因素。特征 156 在该环节是孤立的节点。于是在下一阶段删去特征 156，继续添加特征组，得到如下的结构图 4.21。

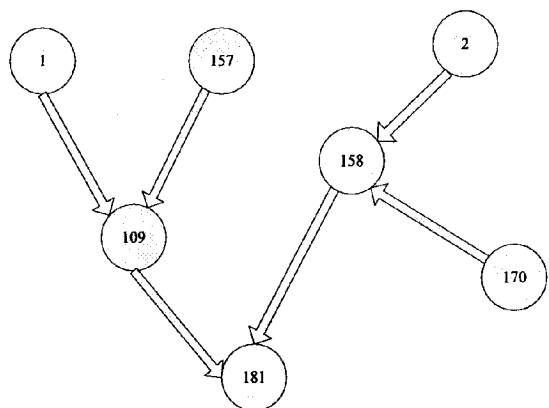


图 4.21 情感 5 中包含所有主因素和次主因素的结构图

Fig 4.21 The structure of emotion 5 contains all main factors and secondary factors

如图 4.21 所示的是将所有特征组添加进来后所得到的结构图。在该图中特征 170 成为特征 158 的主要影响因素，而相应的其他节点却并没有改变位置。于是通过该总体的结构图可以知道对情感 5 有影响的分别是特征 109 和 158，而对特征 109 有影响的是特征 1 和 157，对特征 158 有影响的分别是特征 2 和 170。根据上文的定义可以知道特征 1 和 157、2 以及 170 被称为类别 181 的次主因素。

(6) 情感 6 (anxiety) 的主因素分析

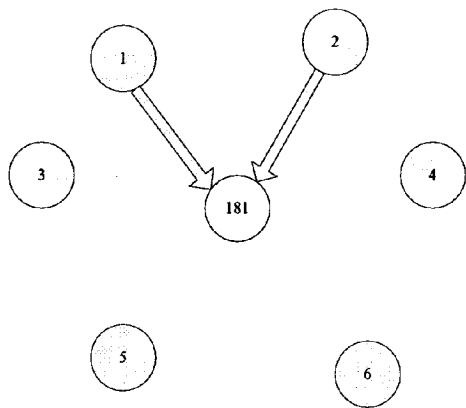


图 4.22 情感 6 中第一阶段得到的结构图

Fig 4.22 The structure chart of emotion 6 on the first stage

通过图 4.22 可以看到在第一组特征添加进来后，特征 1 和 2 分别显示了其影响性，表现在结构图中就是和 181 节点之间存在直接的边表示。而其余的 4 个节点则分别互不影响，因此在接下来的阶段只保留特征节点 1 和 2。如此可以得到结构图 4.23。

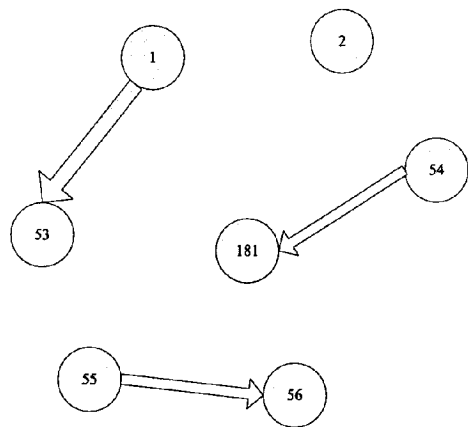


图 4.23 情感 6 中加入 54 在内的特征组后得到的结构图

Fig 4.23 The structure chart of emotion 6 after the set of feature contain 54 were added

如图 4.23 所示的是在加入包含特征 54 后所得到的结构图，该图中可以看到在 54 出现后特征 1 和 2 不再是 181 的主因素，取而代之的是 54 节点成为特征 181 的主要影响因素。该结构图也正说明了原先有影响的因素，在其他特征加入的情况下可能会变成非主因素。因此在接下来的阶段替换掉其他不具影响的因素，保留特征 54，得到结构图 4.24。

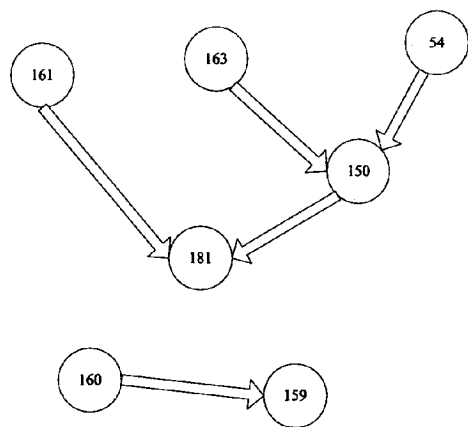


图 4.24 情感 6 中加入包含 161 在内的特征组后得到的结构图

Fig 4.24 The structure chart of emotion 6 after the set of feature contain 161 were added

如图 4.24 所示的是继上一阶段依次加入特征后出现的结构图。在加入包含 150 在内的特征组后，特征 150 成为特征 181 的主因素，接着继续考察 150 的影响，于是继续添加特征组，在添加到包含 160 在内的特征组时得到的结构图如上，该结构图显示了特征 161 和特征 150 分别是 181 的主因素，而特征 163 和特征 54 则成为 150 的主因素。保留该阶段的主因素以及次主因素，继续添加特征，得到如下的结构图 4.25。

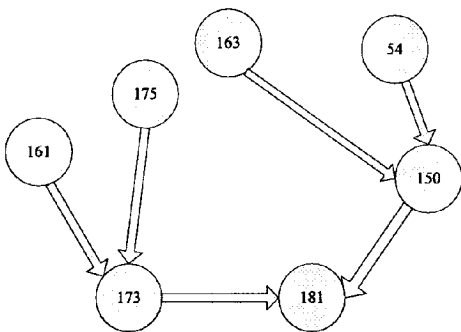


图 4.25 情感 6 包含所有主因素和次主因素的结构图

Fig 4.25 The structure of emotion 6 contains all main factors and secondary factors

如图 4.25 所示的是当所有特征组添加进来后得到的结构图。通过结构图可以看到当包含 173 在内的特征组添加进来后，部分特征的影响关系发生了变化，其中以 161 影响最大，由原先的作为 181 的主因素成为 173 的主因素，而 175 特征也是 173 的主因素。该图展现了情感 6 的主因素分别是 173 和 150，次主因素是 161, 175 以及 163 和 54。

(7) 情感 7 (disgust) 的主因素分析

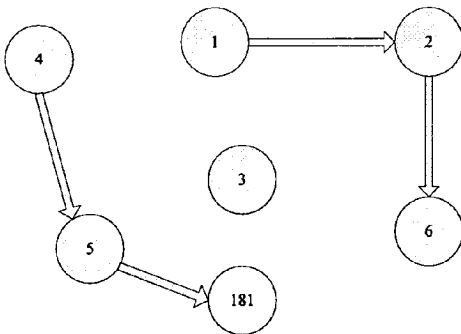


图 4.26 情感 7 中第一阶段的结构图

Fig 4.26 The structure chart of emotion 7 on the first stage

如图 4.26 所示的是第一组特征添加时所得到的结构图。通过结构图可以发现特征 5 是 181 的主因素，特征 4 则是特征 5 的主因素，其余的 4 个节点中，特征 2 是特征 6 的主因素，特征 1 则是特征 2 的主因素，特征 3 是一个孤立的特征。基于上文的思想，保留主因素和非主因素，依次添加下面新的特征组，得到如下结构图 4.27。

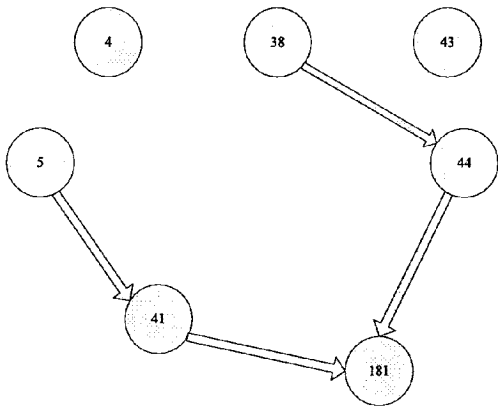


图 4.27 情感 7 中加入包含 41 在内的特征组时得到的结构图

Fig 4.27 The structure chart of emotion 7 after the set of feature contain 41 were added

如图 4.27 所示的是在加入包括特征 41 等特征在内的特征组时所得到的结构图。通过该图可以看到特征 5 在开始时是 181 的主要影响因素，当特征 41 出现的时候改变了其影响性，取而代之的是特征 41 成为 181 的主要影响因素，而特征 5 则是特征 41 的主要影响因素。同样的可以看到特征 44 是 181 的主要影响因素，特征 38 则是特征 44 的主要影响因素。另一个改变较大的是特征 4，从刚开始的特征 5 的主要影响变成了孤立特征，因此保留该阶段图中的主因素和次主因素，得到了如下结构图 4.28。

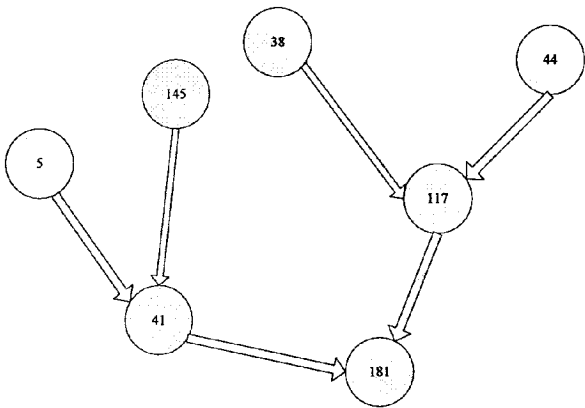


图 4.28 情感 7 中包含所有主因素和次主因素的结构图

Fig 4.28 The structure of emotion 7 contains all main factors and secondary factors

如图 4.28 所示的是所有特征添加后的结构图，通过该图可以发现当特征 117 添加进来后，改变了特征 44 以及 38 的影响性。由原先的作为 181 的主因素转而为 117 的主因素，而特征 117 则成为 181 的主因素。通过该图可以发现对情感 7 有影响的主要影响因素分别是特征 41 和 117，相应的次主因素分别是特征 5、145、38 以及 44。



4.4.3 主因素提取后识别率结果与分析

该研究中选取的 MFCC 特征集包含 7 类情感，也即研究进行的是一个七分类的问题实现。在此次的研究中我们假定各个情感类别之间是互斥的，在进行主因素提取之后进行分类识别的识别率与未进行主因素提取时的识别率对照如下表 4.3:

表 4.3 主因素提取方法与未进行主因素提取识别率对比表

Tab 4.3 Comparison of recognition rate before and after used the main factor extraction method

处理方法	识别率
主因素提取的方法	72%
未进行主因素提取	71%

表 4.3 展示的是两种方法对应识别率的对比。采用主因素提取的方法所采用的特征是基于上述各情感类别下主因素及次主因素合成后的结果。如果单独利用提取出的 15 个主因素进行分类的话。识别率较原来未进行主因素提取的识别率下降五个百分点，此时，当加入各个主因素的次主因素后，构成了 21 个特征，再次进行分类识别时，识别率达到了 72，较原始的方法提高了一个百分点，基于表 4.3 的结果可以说明在 MFCC 特征集中进行特征选择是有必要的。

然而，在现实情况中的情感并不是严格互斥的。有时一条语音样本中所包含的情感可能不止一种。基于这样的考虑，我们又实现了吴恩达所说的另外一种方法，即建立多个分类器。该样本数据中一共有 7 类情感，因此我们需要建立 7 个小的二分类器。在建立 7 个小分类器之前分别提取出对各个情感有影响的主因素。考虑到对于小样本的分类效果来说，SVM 分类器表现出的能力较好，因此实验中对于 7 类情感分别建立 7 个 SVM 小分类器。在该分类器中所采用的数据仅仅是对原始的 MFCC 数据进行标准化和归一化操作。结果如下表 4.4 所示:

表 4.4 7 个小 SVM 分类器的识别率

Tab 4.4 The recognition rate of 7 small SVMclassifier

分类器号	主因素方法的识别率	未进行主因素选择的识别率
1	68.74%	73.4%
2	76%	85.59%
3	85.22%	89.53%
4	75.05%	71.64%
5	77.46%	76.95%
6	62.81%	66.31%
7	70.4%	78.25%

通过表 4.4 可以看到当采用 logistic 回归的方法建立 7 个二分类器后得到的识别率普遍不太理想。在实验部分 4.4.1 中我们针对 7 种情感类别分别提取了他们对应的主因素，从分析中可以看到各类情感之间的影响因素并不是独立的，存在相应的重叠，例如特征 2 对情感 1 有影响，同样的对情感 3 和 5 也是存在影响的；特征 109 则不仅对情感 2 有影响，还对情感 4 和 5 都存在影响。依次可以发现还有其它特征也是存在这样的现象。

回头再观察数据的格式时，发现主要原因可能是因为此次的数据无法支持我们去实现用多个二分类器来进行多分类。数据中并未展现出同样的样本可以属于多个情感类别的，也就是数据本身并未展现出互斥的性质。因此在以后的工作中可以考虑对于新的样本数据可以采用这样的方法进行测试。

4.5 小结

在该部分主要介绍此次研究中的实验结果及其分析。首先基于此次实验中用到的 MFCC 特征集进行介绍，依次介绍该特征集产生的背景以及该实验中该特征集的组成。接着介绍此次的实验设计以及实现。针对小样本数据集的分类问题，研究者们普遍青睐于 SVM 分类器。

4.3 小节中针对 SVM 的理论进行详细的分析。通过分析也可以发现对于常见的小样本问题采用 SVM 分类器不失为一种好的方法。最后针对各个情感类别的主要影响因素的实验结果进行分析，以及对于采用主因素的方法进行分类识别和未进行主因素的方法进行分类的识别率进行对比。通过对比结果可以知道，对于采用一个多分类器的情况下，采用主因素分析的方法所产生的分类情况略好于未进行主因素分析的方法。然而，对于另外一个情况，基于现实情况考虑各类情感类

别不是互斥的角度来考虑的话，我们建立多个二分类器来进行多分类，实验的结果是不理想的。造成不理想的原因可能是此次的数据集问题。针对该问题，我们可以在一批新的可以反映互斥关系的数据集上来采用同样的方法进行测试。

## 第五章 总结与展望

### 5.1 该研究的工作

自 1997 年 Picard 教授首次在其专著中提出“情感计算”这个概念后便开启了“情感计算”这一新的研究领域。语音情感识别是“情感计算”领域一项重要的研究工作。语音情感识别起步较晚，但是自从起步开始，国内外研究者们就投入了大量的热情。他们从影响语音的各个方面不断探索提取出大量的特征来进行语音数据的识别。基于这样大量的特征，在计算时势必会带来所谓的“特征灾难”，因此现阶段进行特征维度选择是必不可少的工作环节。

该研究基于贝叶斯理论中因果发现的思想，在不显著降低原始识别率的情况下力求找出影响情感的主要因素，从而达到特征维度减少的目的。具体的工作如下：

第一章 总体的背景知识介绍。在该部分主要介绍了此次研究的背景，也就利用本文提出的 h-MMHC 算法来进行语音情感数据特征集 MFCC 的特征降维。总结了情感识别的发展历史、过程以及应用；由于情感识别是机器学习领域的一个重要应用，因此该部分就机器学习的定义及其应用进行了描述；最后，就本文的重点——特征选择来总结介绍一下具体的特征选择的定义、语音情感识别中的特征选择以及基于贝叶斯网络如何来进行特征选择。

第二章 概括介绍了本文中用到的基本数学概念。无论是情感识别还是机器学习甚至特征选择，都需要充分的数学理论依据。该部分就机器学习以及特征选择中的一些数学理论进行描述，使得本文的研究工作建立在充分的理论基础之上。

第三章 介绍了该研究中进行特征选择所采用的 h-MMHC 算法。阐述该算法提出的背景以及意义。该算法是在 MMHC 算法的基础上改进而来，采用分阶段添加特征的思想，基于贝叶斯网络中的因果发现，逐步找出对于目标特征节点具有主要影响的因素。

第四章 本章是该研究工作的具体应用点。首先基于实验中用到的 SVM 分类器，总结介绍了 SVM 的理论知识以及该实验中具体的 Libsvm 参数的设置；其次，描述该研究工作的实验设计方案以及如何在工具中实现；最后就实验的结果进行展示，分别展示了各情感类别中的主因素提取的过程，由于 MFCC 的特征维度较高，采用 h-MMHC 算法进行分阶段构造结构图时，数目较多，因此仅仅总结展示和主因素相关的结构图。该研究的主要工作是进行主因素的提取，但是验证提取的主因素是否有效的一个标准就是应用提取出的主因素进行分类识别，在不

显著降低识别率的情况下的特征降维才是有意义的。通过实验结果展示中可以看到，利用提取出的主因素及其次主因素的融合情况下，可以让分类的识别率有所提升，由此证明该研究的工作是有意义的。

## 5.2 下一步的研究工作

本文基于贝叶斯理论知识结合 MMHC 算法进行主因素的提取，虽然取得了一定的成果，但是工作中出现的问题还需要在接下来的工作中不断的加以完善，以期获得更好的成果。具体的改进如下：

(1) 本文是将主因素的方法应用于情感识别领域的 MFCC 特征集得到的实验结果。基于语音情感领域中还有其他用于识别语音情感的特征集。因此，下一步的工作就是将该方法应用于其他特征集中以验证该方法的有效性。

(2) 实验中对于分类类别不是互斥的情况应该建立多个分类器来实现多分类。该研究中考虑到实际情况中情感类别之间应该不是互斥的关系，建立了 7 个独立的二分类器来实现情感的 7 分类，可是实验效果不理想。通过实验中的分析可以知道可能是此次使用的数据存在不足所引起的，因此，在接下来的工作中，应该重新建立一个语音库，该库中的数据应该可以直观的反应类别之间的关系：互斥还是非互斥。建立好之后再利用该研究中的方法进行测试，效果应该会比较

## 参考文献

- [1] Calvo R A, D'Mello S. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications[J]. Affective Computing IEEE Transactions on, 2010, 1(1):18 - 37.
- [2] Busso C, Lee S, Narayanan S. Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection[J]. Audio Speech & Language Processing IEEE Transactions on, 2009, 17(4):582 - 596.
- [3] Yujin Y, Peihua Z, Qun Z. Research of speaker recognition based on combination of LPCC and MFCC[C]. //IEEE International Conference on Intelligent Computing and Intelligent Systems. IEEE, 2010:765 - 767.
- [4] Minsk ML. The Society of Mind [M]. New York: Touchstone, 1985:85—86.
- [5] CE W. Emotions and speech: some acoustical correlates.[J]. Journal of the Acoustical Society of America, 1972, (4):1238-1250.
- [6] Kitahara Y, Tohkura Y. Prosodic Control to Express Emotions for Man-Machine Speech Interaction[J]. Ieice Trans, 1992, (2):155-163.
- [7] Vayrynen E, Kortelainen J, Seppanen T. Classifier-based learning of nonlinear feature manifold for visualization of emotional speech prosody[J]. Affective Computing IEEE Transactions on, 2013, 4(1):47 - 56.
- [8] Rabiner L R, Schafer R W. Digital Processing of speech signals[C]. //Pearson Education Isbn. 1978.
- [9] Atal B S. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification.[J]. Journal of the Acoustical Society of America, 1974, 55(6).
- [10] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. Acoustics Speech & Signal Processing IEEE Transactions on, 1980, 28(4):357 - 366.
- [11] Schuller B, Lang M, Rigoll G. Robust acoustic speech emotion recognition by ensembles of classifiers[J]. FORTSCHRITTE DER AKUSTIK, 2005, 31(1): 329.
- [12] Lugger M, Yang B. Combining classifiers with diverse feature sets for robust speaker independent emotion recognition[C]//Proceedings of EUSIPCO. 2009.
- [13] Nwe T L, Foo S W, Silva L C D. Speech emotion recognition using hidden Markov models[J]. Speech Communication, 2003, 41(3):603—623.
- [14] Cairns D A, Hansen J H L. Nonlinear Analysis and Classification of Speech Under

- Stressed Conditions[J]. Journal of the Acoustical Society of America, 1999, 96(6):3392-3400.
- [15] Burkhardt, F, Paeschke, A, Rolfes, M, Sendlmeier, W, & Weiss, B. A database of german emotional speech[J]. Proceedings of Interspeech Lissabon, 2005:1517--1520.
- [16] Morrison D, Wang R, Silva L C D. Ensemble methods for spoken emotion recognition in call-centres[J]. Speech Communication, 2007, 49(2):98-112.
- [17] Breazeal C, Aryananda L. Recognition of affective communicative intent in robot-directed speech[J]. Autonomous robots, 2002, 12(1): 83-104.
- [18] Schuller, B., Reiter, S., Müller, R., Al-hames, M., Lang, M., & Rigoll, G., et al. Speaker Independent Speech Emotion Recognition by Ensemble Classification[C]. //IEEE International Conference on Multimedia and Expo. IEEE, 2005:864 - 867.
- [19] Schuller B, Rigoll G, Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture[J]. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004, 1:I - 577-80.
- [20] Ma, J., Jin, H., Yang, L. T., & Tsai, J. J. P. Ubiquitous Intelligence and Computing: Third International Conference, UIC 2006, Wuhan, China, September 3-6, 2006[J]. Proceedings (Lecture Notes in Computer Science), Springer-Verlag New York, Inc., Secaucus, NJ, 2006.
- [21] 王先谦（清）. 荀子集解[M]. 中华书局 2010.
- [22] Kira K, Rendell L A. The Feature Selection Problem: Traditional Methods and a New Algorithm[J]. Proceedings of Ninth National Conference on Artificial Intelligence, 1992.
- [23] John G H, Kohavi R, Pflieger K. Irrelevant Features and the Subset Selection Problem[J]. Machine Learning Proceedings of the Eleventh International, 1994:121--129.
- [24] Dash M, Liu H. Feature selection for classification[J]. Intelligent data analysis, 1997, 1(3): 131-156.
- [25] Yu L, Liu H, Guyon I. Efficient feature selection via analysis of relevance and redundancy[J]. Journal of Machine Learning Research, 2004, 5(4):1205--1224.
- [26] Lieberman P, Michaels S B. Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech[J]. Journal of the Acoustical Society of America, 1962, 34(7).
- [27] 王蕊. 特征选择方法在基于 LFP 信号的神经解码中的应用[D]. 浙江大学, 2013.
- [28] 张杰慧. 特征选择算法研究及其在孤立肺结节诊断中的应用[D]. 重庆大学, 2007.

- [29] 熊小草. 文本分类中特征选择的理论分析和算法研究[D]. 清华大学, 2007.
- [30] Brieman, L., Friedman, J. H., Olshen, R., & Stone, C. Classification and regression trees (Belmont, CA: Wadsworth)[J]. 1984.
- [31] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules[J]. Proc.of Intl.conf.on Vldb, 1994.
- [32] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference [M]. San Francisco, CA, USA; Morgan Kaufmann Publishers Inc, 1988.
- [33] Pearl J., 2000. Causality: Models, Reasoning, and Inference [M]. Cambridge: University Press, 2000. Available at: [singapore.cs.ucla.edu/BOOK-2K/](http://singapore.cs.ucla.edu/BOOK-2K/).
- [34] 何德琳, 程勇, 赵瑞莲. 基于 MMHC 算法的贝叶斯网络结构学习算法研究[J]. 北京工商大学学报: 自然科学版, 2008, 26(3): 43-48.
- [35] Margaritis D, Thrun S. Bayesian Network Induction via Local Neighborhoods [J]. Advances in Neural Information Processing Systems, 1999:505--511.
- [36] Aliferis, C. F., Tsamardinos, I., Statnikov, A. R., & Brown, L. E. Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery [J]. Metmbs, 2003.
- [37] Tsamardinos I, Aliferis C, Statnikov A. Time and sample efficient discovery of Markov blankets and direct causal relations[C]. //Proceedings of Can Sigkdd International Conference on Knowledge Discovery & Data Mining. 2003:673-678.
- [38] 杨静, 曹家俭. 连续贝叶斯网络模型在断面调查数据的应用[J]. 计算机工程与应用, 2014, (19). DOI: 10.3778/j.issn.1002-8331.1212-0119.
- [39] Russell, Stuart, and Peter Norvig. "人工智能——一种现代方法 (英文版)." [M] (2002).
- [40] Tsamardinos I, Brown L E, Aliferis C F. The max-min hill-climbing Bayesian network structure learning algorithm[J]. Machine Learning, 2006, 65(1):31-78.
- [41] 全美红, 段富. 基于主成分分析和熵值法的高校教师绩效评价[J]. 计算机应用与软件, 2014, (1): 62-64. DOI: 10.3969/j.issn.1000-386x.2014.01.017.
- [42] Turk M, Pentland A. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1):71 - 86.
- [43] 余凯, 贾磊, 陈雨强等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799-1804.
- [44] Wang, A G., An, N., Chen, G., Li, L., & Alterovitz, G. Accelerating incremental wrapper based gene selection with K-Nearest-Neighbor[C]//Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on. IEEE, 2014: 21-23.



- [45] Ting, Choo-Yee, Sam, Yok-Cheng, Wong, & Chee-Onn. Model of Conceptual Change for INQPRO: A Bayesian Network Approach.[J]. Computers & Education, 2013, 65(7):77-91.
- [46] Millán, E., Descalço, L., Castillo, G., Oliveira, P., & Diogo, S. Using Bayesian networks to improve knowledge assessment[J]. Computers & Education, 2013, 60(1):436–447.
- [47] Millán E, Loboda T, Pérez-de-la-Cruz J L. Bayesian networks for student model engineering [J]. Computers & Education, 2010, 55(4): 1663-1683.
- [48] 马希荣, 孙华志. 数据挖掘技术在教学评价中的应用[J]. 计算机工程与应用, 2003, 39(19): 51-54.
- [49] 方耀楣, 何万篷. 可拓数据挖掘在高校教学质量评价中的应用[J]. 数学的实践与认识, 2009 (4): 82-87.
- [50] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences[J]. Acoustics Speech & Signal Processing IEEE Transactions on, 1980, 28(4):357 - 366.
- [51] Bou-Ghazale S E, Hansen J H L. A comparative study of traditional and newly proposed features for recognition of speech under stress[C]. //IEEE Trans Speech and Audio Process. 2000:429 - 442.
- [52] Lugger M, Yang B. Combining classifiers with diverse feature sets for robust speaker independent emotion recognition[C]//Proceedings of EUSIPCO. 2009.
- [53] Küstner, D., Tato, R., Kemp, T., & Meffert, B. “Towards real life applications in emotion recognition,” Affective Dialogue Systems (ADS’05), E. Andre, L.Dybkaer, W. Minker, and P. Heisterkamp, eds., pp. 25-35, Springer Verlag, May 2004.
- [54] Busso, C., Mariooryad, S., Metallinou, A., & Narayanan, S. Iterative Feature Normalization Scheme for Automatic Emotion Detection from Speech[J]. Affective Computing IEEE Transactions on, 2013, 4(4):386-397.
- [55] Ayadi M E, Kamel M S, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases[J]. Pattern Recognition, 2011, 44(3):572–587.
- [56] 孙刚. 基于 SVM 的入侵检测系统研究[D].北京: 北京邮电大学信息工程学院, 2007.
- [57] Vapnik 著, 张学工译. 统计学习理论的本质[M].北京: 清华大学出版社, 1999.
- [58] Cristianini N, 克里斯蒂亚尼尼, 肖-泰勒, 等. 支持向量机导论[M]. 电子工业出版社, 2004.

## 攻读硕士学位期间的学术活动及成果情况

### 1) 参加的学术交流与科研项目

- (1) 老人福祉信息科技创新引智基地（“111”），项目号：B14025。（编号：B14025），教育部高等学校学科创新引智计划，2014.1-2018.12
- (2) 面向非线性非高斯数据的因果结构学习算法研究（编号：61305064），国家自然科学基金青年项目，2014.1-2016.12

### 2) 发表的学术论文（含专利和软件著作权）

- (1) 李昌群, 杨静等. h-MMHC 算法及其在主因素分析中的应用 [J]. 计算机应用与软件 (已录用 2015-03-11)
- (2) Wang aiguo, An Ning, Lu Xin, Chen Hongtu, Li Changqun, and Levkoff Sue. “A Classification Scheme for Analyzing Mobile Apps Used to Prevent and Manage Disease in Late Life.” JMIR mhealth and uhealth 2, no. 1 (2014): e6.

## 特别声明

本学位论文是在我的导师指导下独立完成的。在研究生学习期间，我的导师要求我坚决抵制学术不端行为。在此，我郑重声明，本论文无任何学术不端行为，如果被发现有任何学术不端行为，一切责任完全由本人承担。

学位论文作者签名：李易群

签字日期：2015年4月23日