

清华大学学位论文 L^AT_EX 模板

使用示例文档 v5.4.3

(申请清华大学工学硕士学位论文)

培 养 单 位: 计 算 机 科 学 与 技 术 系

学 科: 计 算 机 科 学 与 技 术

研 究 生: 梁 锡 豪

指 导 教 师: 徐 明 星 副 教 授

二〇一九年三月

An Introduction to L^AT_EX Thesis Template of Tsinghua University v5.4.3

Thesis Submitted to
Tsinghua University
in partial fulfillment of the requirement
for the professional degree of
Doctor of Engineering

by
Liang Xihao
(Computer Science and Technology)

Thesis Supervisor : Professor Xu Mingxing

March, 2019

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：(1) 已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；(2) 为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

(保密的论文在解密后应遵守此规定)

作者签名：_____

导师签名：_____

日 期：_____

日 期：_____

摘 要

自 Web2.0 普及后,人们习惯在互联网的各种平台上透过不同的媒体来分享和传递他们的想法和情感。透过对这些媒体进行情感分析,我们可以得知人们的想法和态度。譬如透过分析产品评论了解用户对新产品是否满意,又或者分析社交平台上的舆论了解网民对新政策是否同意,以此快速响应,其中蕴含着丰富的商业价值和政治价值,其相关技术的研究价值在近年也得到了重视。而文本作为较常见的媒体之一,加之部分微博等社交平台推出了用户生产数据的采集服务,提供了充裕的文本数据资源,面向文本的情感识别研究得以更快的速度发展。然而随着在互联网上的交互方式变得丰富,文本的数据结构也变得多种多样,如微博上的短文本、在线实时聊天的两人或多人对话、讨论区里的帖子回复。在处理不同数据结构时,该如何提取当中的信息并加以运用来预测目标结果,成为了重要的研究课题。

本文主要探讨了两个不同数据结构下的情感识别问题,并提出了一个通用的文本情感识别系统设计方案。论文的主要内容如下:

- 面向微博的反讽识别。反讽作为一种特殊的修辞手法,起着把文本字面意思反转的效果。识别文本中反讽修辞的使用对于正确理解发言者的意图起着关键性的作用。本文研究了在微博中识别反讽的出现以及其对应类型,深入了解实验数据当中各个类型的样本的特性,分析单个分类器对各个类的识别能力,给出一个基于集成学习的反讽识别框架,以此在多个独立分类器的基础上实现一个集成系统来达到更好的识别能力。
- 面向三轮对话的情感识别。为了更好地识别一段文本的情感,引入上下文信息被认为有指导性作用。而在文本对话的场景下,要识别某个人在某次发言所表达的情感,可以考虑引入该次聊天的历史记录。本文研究了在两人轮流发言的三轮对话中,识别最后一轮发言所表达的情感。此处采用与面向微博的反讽识别相同的系统设计以评估设计方案的泛用性,另外我们进一步探索了上下文信息的运用方面,透过把不同的运用方面体现在分类器的不同设计方案上,并根据其性能差别和对神经网络权重的分析,了解上下文信息如何有助于识别正文的情感。

关键词: 意图识别; 情感分析; 反讽识别; 深度学习, 集成学习

Abstract

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summarization of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Key words are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 key words, with semi-colons used in between to separate one another.

Key words: Opinion mining; sentiment analysis; irony detection; deep learning; ensemble learning

目 录

第 1 章 引言	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 情感模型	2
1.2.2 情感识别	3
1.2.3 反讽识别	5
1.2.4 社交媒体上的文本意图识别	6
1.3 存在的问题	7
1.4 本论文的内容安排	7
第 2 章 中华人民共和国	8
2.1 其它例子	8
2.1.1 绘图	8
2.1.2 插图	8
插图索引	11
表格索引	12
公式索引	13
参考文献	14
致 谢	16
声 明	17
附录 A 外文资料原文	18
A.1 Single-Objective Programming	18
A.1.1 Linear Programming	19
A.1.2 Nonlinear Programming	20
A.1.3 Integer Programming	21
附录 B 外文资料的调研阅读报告或书面翻译	22
B.1 单目标规划	22
B.1.1 线性规划	22
B.1.2 非线性规划	23

B.1.3 整数规划.....	23
附录 C 其它附录	24
个人简历、在学期间发表的学术论文与研究成果	25

主要符号对照表

HPC	高性能计算 (High Performance Computing)
cluster	集群
Itanium	安腾
SMP	对称多处理
API	应用程序编程接口
PI	聚酰亚胺
MPI	聚酰亚胺模型化合物, N-苯基邻苯酰亚胺
PBI	聚苯并咪唑
MPBI	聚苯并咪唑模型化合物, N-苯基苯并咪唑
PY	聚吡咙
PMDA-BDA	均苯四酸二酐与联苯四胺合成的聚吡咙薄膜
ΔG	活化自由能 (Activation Free Energy)
χ	传输系数 (Transmission Coefficient)
E	能量
m	质量
c	光速
P	概率
T	时间
v	速度
劝学	<p>君子曰：学不可以已。青，取之于蓝，而青于蓝；冰，水为之，而寒于水。木直中绳。輮以为轮，其曲中规。虽有槁暴，不复挺者，輮使之然也。故木受绳则直，金就砺则利，君子博学而日参省乎己，则知明而行无过矣。吾尝终日而思矣，不如须臾之所学也；吾尝跂而望矣，不如登高之博见也。登高而招，臂非加长也，而见者远；顺风而呼，声非加疾也，而闻者彰。假舆马者，非利足也，而致千里；假舟楫者，非能水也，而绝江河，君子生非异也，善假于物也。积土成山，风雨兴焉；积水成渊，蛟龙生焉；积善成德，而神明自得，圣心备焉。故不积跬步，无以至千里；不积小流，无以成江海。骐驎一跃，不能十步；弩马十驾，功在不舍。锲而舍之，朽木不折；锲而不舍，金石可镂。蚓无爪牙之利，筋骨之强，上食埃土，下饮黄泉，用心一也。蟹</p>

六跪而二螯，非蛇鱗之穴无可寄托者，用心躁也。——荀况

第1章 引言

1.1 研究背景与意义

意图识别旨在了解人们对特定事件或实体的想法、态度和情感。自 Web2.0 普及后,大量网民每天在互联网生产着不同媒体类型的数据,其中包含各种信息,如个人生活经历,购买行为,对产品服务的体验评价,对社会时事的看法等等。从人们的日常社交需求来看,这种借由互联网媒体的分享非常便捷,我们可以了解到亲朋好友的近况,也或者随时和不认识的网民交流对具体事件的想法。在商业上,借由对用户的网络行为进行分析,企业可以对应客户或者潜在客户有更深入的了解,对他们的需求和反馈及时作出反应将带来战略性的优势。在社会管理上,政府可以透过对网民在网络上的发言了解人民的想法和舆论的走向,进而作出相应的措施。正因为互联网的普及,才使得以上基于对特定人群的了解来进行决策的做法成为可能。随着数据资源变得丰富,相应的技术在近年有明显的发展,目前市面上已经有公司(如国内的腾讯和阿里巴巴,以及国外的微软和亚马逊等)提供基于大型社交媒体平台(如微博,讨论区等)上的数据进行意图分析相关的规范化服务,然而相应的技术依然有进步空间,研究工作还在不同方向上摸索。

情感在人们的思想表达和交流中起着重要作用^[1],比起了解该想法的细节内容,情感对应该想法的大体倾向。譬如在分析用户对新产品的评论时,从对正负性情感反馈的统计可以得知新产品是否能让大部分的客户满意,或者筛选出表示不满意的用户再进行深入分析,因此情感识别是意图分析中属于非常重要的一环。而由于在互联网上,大部分情况下用户以文本表达想法,面向文本的意图识别成为了近年的最要研究课题之一。相对于人们面对面交流的场景,聆听者可以根据发言者的肢体语言,面部表情以及声调变化等额外提示更好地理解发言者所表达的内容,然而这些信息并不存在于文本当中,这也正是对文本进行情感识别本身的难点之一^[2]。

反讽的修辞手法在意图识别当中起着特殊的影响。Henry Watson Fowler 在《The King's English》一书中描述“即使对反讽的定义有数百种,其中只有包含‘表面意思和实际意思不同’这个概念的才能被接受”。Eric Partridge 在《Usage and Abusage》一书中指出“反讽存在于所表达意思的另一面”。总的多说当反讽在文本中出现,那么文本所表达的意思应该和字面表达的意思完全相反。譬如某人表示“我就喜欢你不断挑战我的底线”,从字面上理解应倾向于正面情感,然而根据常识可知“挑战底线”是一种让人反感的行为,与“喜欢”相矛盾。这段文字实则暗示发言者的“底线”

正被“你”“挑战”，表达的是负面情感。因此在意图识别和情感分析当中，正确检查出反讽的使用能够避免对内容的错误理解。

1.2 国内外研究现状

1.2.1 情感模型

情感计算的基础是对情感作出描述，现有的描述方式可以分成两个大类：范畴观和维度观。范畴观即把不同情感对应到一组离散的情感标签上 Alm 等人^[3]，其中具代表性的有 Plutchik 的情感模型^[4]和 Ekan 的情感理论^[5]和。Plutchik 的情感模型包含十种情感：愤怒、恐惧、悲伤、厌恶、期待、信任、高兴、惊讶；这些情感都各自对应具有重要生存意义的行为，各种复杂的情感都是由这此基本情感构成，另外这十种情感可以分成五组对立的情感对。Ekan 在十年后提出的情感理论和 Plutchik 的相似，但相对地少了期待和信任，是一个六类情感模型。

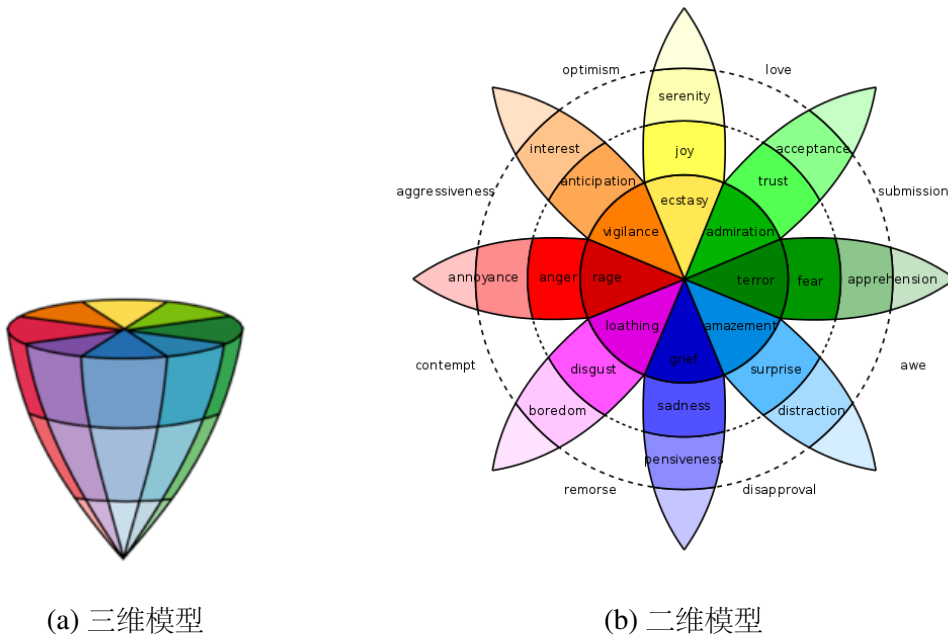


图 1.1 Plutchik^[4] 提出的情感轮模型

以维度观描述情感就是把情感映射到多维空间的点上，而目前维度观情感模型以二维和三维空间的为主。二维情感模型中较有代表性的是 Russell 提出的环状模型 (Circumplex Model)^[6]，其中纵坐标对应情感的激活度 (Arousal)，横坐标对应情感向性 (Valence)，而不同的情感则分布在一个环状的区域内。Bradley 等人^[7]提出的向量模型 (Vector Model) 对横轴和纵轴的定义相似，但其理论假设高激活度的情感应该有较明显的正向或负向，相对地低激活度的情感则偏向中性，故

情感分布在一个回力标形状的区域。Watson 和 Tellegen^[8] 提出的 PANA (positive activation-negative activation) 模型和前两者在理论上则有明显的不同，他们认为情感的正面作用和负面作用是两个独立的成分，所以在模型中纵轴和横轴分别表示情感正面作用和负面作用的强弱，其效果相当于把 Russell 等人提出的环状模型的向量空间旋转 45 度^[9]。

基于三维空间的情感模型中具有代表性的有 Plutchik^[4] 提出的情感轮模型，Plutchik 认为情绪之间包含强度，相似性和两极性三种维度，椎体的顶部和底部分别对应强的情绪和弱的情绪，相似的情感对应椎体中相近的位置，对立的情绪则会对应到椎体中对立的位置上。另外还有 Mehrabian^[10] 提出的 PAD 模型，其三维空间的三个坐标轴分别对应情感愉悦度 (Pleasure)、激活度 (Arousal) 以及优势度 (Dominance)。较近期被提出的是 Hugo^[11] 的情感立方体模型，其三维空间的三个坐标轴分别对应 5-羟色胺 (5-hydroxytryptamine, 5-HT)，多巴胺 (dopamine, DA) 和去甲肾上腺素 (noradrenaline, NE) 三种神经递质所产生信号的强弱，并对空间中一个立方体的八个顶点标记了其对应的情感。

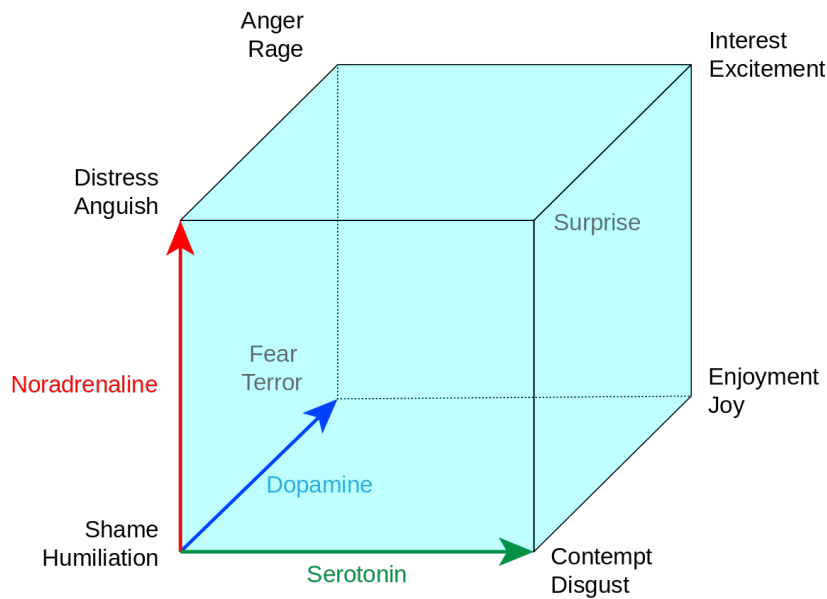


图 1.2 Hugo^[11] 的情感立方体模型

1.2.2 情感识别

对应上述情感模型的分类，情感识别研究可以分成两类。第一类对应范畴观，给定一组情感类型，判断一段文本或针对文本内的某个方面所表达的情感倾向于该组情感中的哪一个，或者是否包含这一组情感中的一个或多个情感。如国际比赛 SemEval2018^[12] 任务一的子任务要求识别一段微博中是否包含愤怒、恐惧、悲

伤等十一种情感中的一种或多种情感。另一类情感识别研究对应维度观,对于给定的情感属性,判断一段文本中该情感属性的强度。其中常见的有情感向性的二分类问题(正性或负性)、三分类问题(正性、中性或负性)、五分类问题(非常正性,正性、中性、负性或非常负性)。其中五分类问题的研究对象一般是互联网上五星评分制的产品评论或者电影评论等。而对目前文本情感识别的研究按照文本的粒度可以大致分成三类:文章级别,句子级别,属性级别。

在文章级别的情感识别中,虽然一篇文章由多个句子组成,均假设其整体存在某种情感偏向,而研究目标则是自动识别出该种情感偏向的类型或强度。**Turney**^[13]利用线上电影评论中"推荐"(大拇指朝上)和"不推荐"(大拇指朝下)的标记研究对电影评论的正负性情感识别。他提出利用点互信息(Pointwise Mutual Information, PMI)来评估单词的语义倾向性,其中PMI透过在大型语料库中统计两个单词的共同出现的情况来评估两个单词的相似度。首先利用PMI值评估各个单词与正向情感的代表性单词"excellent"和负向情感的代表性单词"poor"的相似性,再取这两个PMI值作差得出该单词在语义上倾向于哪一种情感。最近透过计算整段评论的平均语义倾向性评估整体的情感倾向。**Pang**和**Lee**^[14]采用了不同的方法研究相同的问题。他们首先对评论中每个句子的主观程度进行评分,利用评分构造一个带权重的句子关系图,再基于最小割算法结合上下文加强对每个句子主观程度的判断,过滤评论中不带主观情感的句子后再判断整个评论的情感向性,以此加增强了识别能力。

然而一篇文章有可能同时表达了多种观点和情感,因此有另一类研究针对句子或短文本所表达的情感,即句子级别的情感识别。由于句子的文本长度较文章的短,文本内部的逻辑较简单,但相对地所包含的提示信息也较少,课题的难点与前者有所不同。**Khan**等人^[15]研究了对线上评论中句子进行正负性情感识别。他们首先区分出评论中各个句子的主观性和客观性,然后针对带主观情感的句子,利用开源自然语言工具 SentiWordNet 获取各个单词的正负情感属性,再根据句子的词性标注和他们设计的规则计算整个句子的情感向性。他们的研究默认单词的正负情感属性在不同场景下不变,然而**Li**等人^[16]指出部分单词在特定场景下会有不同的情感向性,因此他们提出了一种有监督学习方法,自动学习各单词在指定领域下的情感向性,以此作为文本的特征,并应用于对产品评论的情感识别中。实验结果显示使用 SentiWordNet 提供的全局情感评分和针对领域评估的情感评分相比,后者的识别性能力更好,同时验证了他们的假设。

在一些应用场景当中,我们希望了解发言者对某个特定对象或者它的某个特定属性的想法。譬如新手机推出市场后,厂商需要了解用户对手机的续航能力,拍

照质量,交互体验等各方面的评价,对于评论中同时谈论手机的多个方面并且好评和差评不一时,应该针对各个属性分别识别发言者所表达的感情,因此有了属性级别的情感识别。**Che** 等人^[17]提出一种句子压缩算法,透过对句子进行依存句法分析,过滤与目标属性的情感无关的内容。他们采用了多种语义和语法特征作为输入,以条件随机场 (Conditional random field, CRF) 作为分类器。实验结果显示过滤掉不相关的文本部分后,识别性能有所提升。

Wang 等人^[18]研究了对网上评论中特定实体或属性的情感识别。他们提出了一个基于注意力机制的长短期记忆网络 (Long Short-Term Memory, LSTM), 特点在于只以词向量作为输入,而不采用其他传统的语义和语法特征,另外利用注意力机制自动识别与目标相关的内容。他们实验基于国际比赛 SemEval-2014 任务四中的一个子任务,结果然显示他们的系统性能达到了当时的技术水平。另外经过注意力单元的输出进行分析,验证了注意力机制能有效识别文本中与目标相关的内容。

1.2.3 反讽识别

反讽识别技术作为很多自然语言系统的一部分,其应用场景有线上评论分析,人机对话等,对正确理解评论的情感倾向和话语的意图起着辅助的作用。**Tsur** 等人^{[19][20]}研究了对微博平台 **Twitter** 上的微博以及电商平台亚马逊上的评论进行反讽强度的识别,按明显不含反讽和明显表示反讽分成五级,由人工进行标注。他们提出的 **SASI** 算法分别从文本提取了词频相关的模式特征以及基于标点符号的特征,以 **K** 最近邻算法作为分类器。另外利用在 **Twitter** 按井号标签 **#sarcastic** 自动爬取了额外的反讽语料用于初步的模型训练。对实验结果的比较证明了各种特征的有效性以及添加额外语料对模型训练的帮助。

为了以较低成本获取大量的反讽语料,很多研究从微博平台根据井号标签自动筛选出可能带反讽的微博。**Reyes** 等人^[21]利用 **#irony**, **#education**, **#humor**, **#politics** 在 **Twitter** 上自动获取四组英语微博,并把 **#irony** 对应的微博和另外三组微博两两组成二分类实验。他们提出了四个方面的文本特征以及对应的提取方法,包括:特殊标记(词汇和标点符号等)、不可预期性、表达风格、情感特性;并比较了每项特征在各组微博中的出现情况,显示了与反讽类微博的相关性。另外分别采用朴素贝叶斯和决策树作为分类器,但没有明显的性能区别。

类似的数据收集方法对其他语言同样适用。**Kunneman** 等人^[22]则利用对应的德语井号标签来获取反讽语料,以此研究德语微博中的反讽识别。他们取 **N** 元语法作为输入特征,以 **Balanced Winnow**^[23] 作为分类器,在测试集上达到约 0.85 的

召回率和 0.87 的 AUC 值。但他们进一步经过人工检验评估了基于井号标签自动标注的有效性, 分析结果显示该方法获取的反讽样本中包含约 10

有别于早期以手动设计的特征作为输入和以非神经网络的机器学习方法建模, Poria 等人^[24]首次将神经网络应用于对微博的反讽识别。他们的算法框架主要包含四个卷积神经网络, 并分别利用不同的数据集进行预训练, 分别对应反讽识别、情感极性识别、情感类型识别和性格识别。最后取四个卷积神经网络的中间结果作为特征, 利用支持向量机 (Support Vector Machine, SVM) 进行后融后得出最终预测结果。实验显示引入反讽识别以外的三个语料库提升了系统的识别能力。

1.2.4 社交媒体上的文本意图识别

随着社交媒体的普及, 网民们习惯于在微博, 讨论区等平台上表达个人意见和互相讨论。有别于新闻或学术材料等较正式的文件, 网民可以较随心所欲地发言, 社交媒体因此成为了意图识别的重要研究对象之一。但和一些产品评论或文章等文本类型不同, 社交媒体上的文本普遍较短^[25], 缺少对背景信息的提示, 难以判断其内容所属的领域, 这对正确理解其表达的想法造成困难。另外语言的使用有别于正式的用法, 在中文微博中会出现新的短语或对旧短语有新的解释^[26], 如"锦鲤" 暗示"好运", "灌水" 表示"发表没有意义的内容", 在英文微博中会出现错拼字、非正式缩略语、表情符^{[27][28]}, 如"tnx" 对应英文单词"thanks", " :) " 表示微笑等。虽然没有正式的语言组织对这些新的用法进行整合, 但因为这些新用法更方便或对思想的表达更丰富到位, 随着在网络上的传播而在网民之间达成了共识。传统的文本特征提取建立在单词标准用法和正规的语法分析基础上, 而由于新词汇和新用法的出现, 会出现词汇的意思无法被识别或被错误理解, 这对于透过人工智能来理解文本成为了一大难点。因此有别于传统的文本研究, 在面向社交媒体的文本意图识别时, 需要采用额外手段对文本进行预处理进行规范化, 或透过大量语料尝试自动学习其语言属性。

Angiani 等人^[29]针对英语微博比较了其种常用的文本预处理技术对情感分析的影响, 其中包括单词的规范化、表情符到情感标签的映射、俚语映射、词干提取、停词过滤等。分析显示除了俚语映射以外, 其他技术均对情感识别有正面影响, 其中部分技术有助于统一拼写相似的单汇, 以此关联相同概念的词组。但同时采用所以预处理技术并不保证达到最好的效果, 作者指出依然需要根据应用场景和文本的特性做选择。

1.3 存在的问题

由于近年来各企业或机构对意图识别的需求增加，相关技术的研究备受关注，加上深度学习的快速发展，情感识别和反讽识别的性能水平也在逐年提升。然而相关研究依然存在一些空白部分值得进一步探索：

- **对不同数据结构的处理。**受限于互联网上文本的呈现方式以及数据的收集方式，早期的大部分面向文本的意图识别以文章、微博、评论等形式的数据作为研究对象。随着社交平台功能加强，除了研究的对象本身，我们能获取更多相关的背景信息。以微博平台为例，除了微博的文本内容以外，我们能根据其历史上发布过的内容分析用户的性格和语言使用习惯，另外可以根据微博下的评论加强对微博本文的了解。但相对地，数据结构变得复杂，该如何从中提取特征，该如何对不同信息之间的相互关系建模，这些问题都有值得尝试的空间。
- **针对优化目标的系统调整。**在不同的使用场景下，针对不同的应用目标，我们会采用不同指标来评估系统的识别水平，并以提高该指标为目标对系统进行优化。而对一个指标的加强往往会导致另一个指标的下降，在数据分布不均匀的分类问题当中，这一影响尤其明显。譬如在反讽识别当中，由于微博上包含反讽的微博明显较不含反讽的微博少，若以准确率作为优化指标，整个系统会倾向将微博判断为不含反讽，含反讽的微博召回率会较低。若改以召回率作为优化指标，整个系统会倾向将微博判断为包含反讽，但因为不含反讽的微博比例较高，会把大量不含反讽的微博误判为包含反讽。近年的相关研究的重点以特征提取、深度学习建模、迁移学习等方向为主，在给定在相同的训练数据以及相同的模型框架下，模型的拟合能力是有限的，如何针对不同的优化目标对系统进行调整是个极少被探讨的问题。

1.4 本论文的内容安排

pass

第 2 章 中华人民共和国

2.1 其它例子

在第 1 章中我们学习了贝叶斯公式 (??)，这里我们复习一下：

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} \quad (2-1)$$

2.1.1 绘图

本模板不再预先装载任何绘图包（如 `pstricks`，`pgf` 等），完全由用户来决定。个人觉得 `pgf` 不错，不依赖于 `Postscript`。此外还有很多针对 `LaTeX` 的 GUI 作图工具，如 `XFig(jFig)`, `WinFig`, `Tpx`, `Ipe`, `Dia`, `Inkscape`, `LaTeXPiX`, `jPicEdt`, `jaxdraw` 等等。

2.1.2 插图

强烈推荐《`LaTeX 2ε` 插图指南》！关于子图形的使用细节请参看 `subcaption` 宏包的说明文档。

2.1.2.1 一个图形

一般图形都是处在浮动环境中。之所以称为浮动是指最终排版效果图形的位置不一定与源文件中的位置对应^①，这也是刚使用 `LaTeX` 同学可能遇到的问题。如要强制固定浮动图形的位置，请使用 `float` 宏包，它提供了 `[H]` 参数，比如图 2.1。



图 2.1 利用 `Xfig` 制图

大学之道，在明明德，在亲民，在止于至善。知止而后有定；定而后能静；静而后能安；安而后能虑；虑而后能得。物有本末，事有终始。知所先后，则近道矣。古之欲明明德于天下者，先治其国；欲治其国者，先齐其家；欲齐其家者，先修其身；欲修其身者，先正其心；欲正其心者，先诚其意；欲诚其意者，先致其知；致

^① This is not a bug, but a feature of `LaTeX`!

知在格物。物格而后知至；知至而后意诚；意诚而后心正；心正而后身修；身修而后家齐；家齐而后国治；国治而后天下平。自天子以至于庶人，壹是皆以修身为本。其本乱而未治者否矣。其所厚者薄，而其所薄者厚，未之有也！

——《大学》

2.1.2.2 多个图形

如果多个图形相互独立，并不共用一个图形计数器，那么用 `minipage` 或者 `parbox` 就可以。否则，请参看图 2.2，它包含两个小图，分别是图 2.2(a)和图 2.2(b)。推荐使用 `\subcaptionbox`，因为可以像图 2.2 那样对齐子图的标题，也可以使用 `subcaption` 宏包的 `\subcaption`（放在 `minipage` 中，用法同 `\caption`）或是 `subfigure`、`subtable` 环境，像图 2.3，不要再用 `\subfloat`、`\subfigure` 和 `\subtable`。



(a) 第一个小图形



(b) 第二个小图形，注意这个图略矮些。如果标题很长的话，它会自动换行

图 2.2 包含子图形的大图形 (`subcaptionbox` 示例)



(a) 第一个小图形



(b) 第二个小图形，注意这个图略矮些。`subfigure` 中同一行的子图在顶端对齐。

图 2.3 包含子图形的大图形 (`subfigure` 示例)

古之学者必有师。师者，所以传道受业解惑也。人非生而知之者，孰能无惑？惑而不从师，其为惑也，终不解矣。生乎吾前，其闻道也固先乎吾，吾从而师之；生乎吾後，其闻道也亦先乎吾，吾从而师之。吾师道也，夫庸知其年之先後生於吾乎！是故无贵无贱无长无少，道之所存，师之所存也。



图 2.4 并排第一个图



图 2.5 并排第二个图

嗟乎！师道之不传也久矣，欲人之无惑也难矣。古之圣人，其出人也远矣，犹且从师而问焉；今之众人，其下圣人也亦远矣，而耻学於师。是故圣益圣，愚益愚。圣人之所以为圣，愚人之所以为愚，其皆出於此乎？爱其子，择师而教之，於其身也，则耻师焉，惑焉。彼童子之师，授之书而习其句读者，非吾所谓传其道、解其惑者也。句读之不知，惑之不解，或师焉，或不焉，小学而大遗，吾未见其明也。巫医、乐师、百工之人不耻相师，士大夫之族曰“师”曰“弟子”之云者，则群聚而笑之。问之，则曰：彼与彼年相若也，道相似也，位卑则足羞，官盛则近谀。呜呼！师道之不复，可知矣。巫医、乐师、百工之人。吾子不齿，今其智乃反不能及，其可怪也欤！圣人无常师。孔子师郯子、苌子、师襄、老聃。郯子之徒，其贤不及孔子。孔子曰：“三人行，必有我师。”是故弟子不必不如师，师不必贤於弟子。闻道有先後，术业有专攻，如是而已。

如果要把编号的两个图形并排，那么小页就非常有用：

李氏子蟠，年十七，好古文、六艺，经传皆通习之，不拘於时，学於余。余嘉其能行古道，作师说以贻之。

——韩愈（唐）

插图索引

图 1.1	Plutchik ^[4] 提出的情感轮模型	2
图 1.2	Hugo ^[11] 的情感立方体模型	3
图 2.1	利用 Xfig 制图	8
图 2.2	包含子图形的大图形 (subcaptionbox 示例)	9
图 2.3	包含子图形的大图形 (subfigure 示例)	9
图 2.4	并排第一个图	10
图 2.5	并排第二个图	10

表格索引

公式索引

公式 2-1	8
公式 A-1	19
公式 A-2	19

参考文献

- [1] Banerjee S, Dutta U. Detection of emotions in text: A survey: volume 03[M]. [S.l.: s.n.], 2015
- [2] Chatterjee A, Narahari K N, Joshi M, et al. Semeval-2019 task 3: Emocontext: Contextual emotion detection in text[C]//Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval-2019). Minneapolis, Minnesota: [s.n.], 2019.
- [3] Alm C O, Dan R, Sproat R. Emotions from text: Machine learning for text-based emotion prediction[C]//Conference on Hlt/emnlp. [S.l.: s.n.], 2005.
- [4] Plutchik R, Kellerman H. Emotion: Theory, research and experience[J]. In Theories of emotion, 1980, 11: 399.
- [5] Ekman P. An argument for basic emotions[J]. Cognition & Emotion, 1992, 6(3-4): 169-200.
- [6] Russell J. A circumplex model of affect[J]. Journal of Personality and Social Psychology, 1980, 39: 1161-1178.
- [7] M. Bradley M, Greenwald M, C. Petry M, et al. Remembering pictures: Pleasure and arousal in memory[J]. Journal of experimental psychology. Learning, memory, and cognition, 1992, 18: 379-90.
- [8] Watson D, Tellegen A. Toward a consensual structure of mood[J]. Psychological bulletin, 1985, 98: 219-35.
- [9] Rubin D C, Talarico J M. A comparison of dimensional models of emotion: evidence from emotions, prototypical events, autobiographical memories, and words[J]. Memory, 2009, 17(8): 802-808.
- [10] Mehrabian A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament[J]. Current Psychology, 1996, 14(4): 261-292.
- [11] Hugo L. A new three-dimensional model for emotions and monoamine neurotransmitters[J]. Medical Hypotheses, 2012, 78(2): 341-348.
- [12] Mohammad S, Bravo-Marquez F, Salameh M, et al. Semeval-2018 task 1: Affect in tweets[C]//Proceedings of The 12th International Workshop on Semantic Evaluation. [S.l.: s.n.], 2018: 1-17.
- [13] Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th annual meeting on association for computational linguistics. [S.l.]: Association for Computational Linguistics, 2002: 417-424.
- [14] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts[C]//Proceedings of the 42nd annual meeting on Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2004: 271.
- [15] Khan A, Baharudin B, Khan K. Sentiment classification using sentence-level lexical based[J]. Trends in Applied Sciences Research, 2011, 6(10): 1141-1157.
- [16] Li Y, Lin Y, Zhang J, et al. Constructing domain-dependent sentiment lexicons automatically for sentiment analysis[J]. Information Technology Journal, 2013, 12(5): 990-996.

- [17] Che W, Zhao Y, Guo H, et al. Sentence compression for aspect-based sentiment analysis[J]. IEEE/ACM transactions on audio, speech, and language processing, 2015, 23(12): 2111-2124.
- [18] Wang Y, Huang M, Zhao L, et al. Attention-based lstm for aspect-level sentiment classification [C]//Proceedings of the 2016 conference on empirical methods in natural language processing. [S.l.: s.n.], 2016: 606-615.
- [19] Tsur O, Davidov D, Rappoport A. Icwsm—a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews[C]//Fourth International AAAI Conference on Weblogs and Social Media. [S.l.: s.n.], 2010.
- [20] Davidov D, Tsur O, Rappoport A. Semi-supervised recognition of sarcastic sentences in twitter and amazon[C]//Proceedings of the fourteenth conference on computational natural language learning. [S.l.: Association for Computational Linguistics, 2010: 107-116.
- [21] Reyes A, Rosso P, Veale T. A multidimensional approach for detecting irony in twitter[J]. Language resources and evaluation, 2013, 47(1): 239-268.
- [22] Kunneman F, Liebrecht C, Van Mulken M, et al. Signaling sarcasm: From hyperbole to hashtag [J]. Information Processing & Management, 2015, 51(4): 500-509.
- [23] Littlestone N. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm[J]. Machine learning, 1988, 2(4): 285-318.
- [24] Poria S, Cambria E, Hazarika D, et al. A deeper look into sarcastic tweets using deep convolutional neural networks[J]. arXiv preprint arXiv:1610.08815, 2016.
- [25] Madhusudhanan S, Moorthi D M. A survey on sentiment analysis[J]. Indian Journal of Computer Science and Engineering, 2018, 9(2).
- [26] 谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. 中文信息学报, 2012, 26(1): 73-84.
- [27] Go A, Huang L, Bhayani R. Twitter sentiment analysis[J]. Entropy, 2009, 17: 252.
- [28] Paltoglou G, Thelwall M. Twitter, myspace, digg: Unsupervised sentiment analysis in social media[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2012, 3(4): 66.
- [29] Angiani G, Ferrari L, Fontanini T, et al. A comparison between preprocessing techniques for sentiment analysis in twitter.[C]//KDWeb. [S.l.: s.n.], 2016.
- [30] 薛瑞尼. THUTHESIS: 清华大学学位论文模板[EB/OL]. 2017. <https://github.com/xueruini/thuthesis>.

致 谢

衷心感谢导师 xxx 教授和物理系 xxx 副教授对本人的精心指导。他们的言传身教将使我终生受益。

在美国麻省理工学院化学系进行九个月的合作研究期间，承蒙 xxx 教授热心指导与帮助，不胜感激。感谢 xx 实验室主任 xx 教授，以及实验室全体老师和同学们的热情帮助和支持！本课题承蒙国家自然科学基金资助，特此致谢。

感谢 L^AT_EX 和 THUTHESIS^[30]，帮我节省了不少时间。

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，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

附录 A 外文资料原文

The title of the English paper

Abstract: As one of the most widely used techniques in operations research, *mathematical programming* is defined as a means of maximizing a quantity known as *objective function*, subject to a set of constraints represented by equations and inequalities. Some known subtopics of mathematical programming are linear programming, nonlinear programming, multiobjective programming, goal programming, dynamic programming, and multilevel programming^[1].

It is impossible to cover in a single chapter every concept of mathematical programming. This chapter introduces only the basic concepts and techniques of mathematical programming such that readers gain an understanding of them throughout the book^[2,3].

A.1 Single-Objective Programming

The general form of single-objective programming (SOP) is written as follows,

$$\begin{cases} \max f(x) \\ \text{subject to:} \\ g_j(x) \leq 0, \quad j = 1, 2, \dots, p \end{cases} \quad (123)$$

which maximizes a real-valued function f of $x = (x_1, x_2, \dots, x_n)$ subject to a set of constraints.

Definition A.1: In SOP, we call x a decision vector, and x_1, x_2, \dots, x_n decision variables. The function f is called the objective function. The set

$$S = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, j = 1, 2, \dots, p\} \quad (456)$$

is called the feasible set. An element x in S is called a feasible solution.

Definition A.2: A feasible solution x^* is called the optimal solution of SOP if and only if

$$f(x^*) \geq f(x) \quad (\text{A-1})$$

for any feasible solution x .

One of the outstanding contributions to mathematical programming was known as the Kuhn-Tucker conditions A-2. In order to introduce them, let us give some definitions. An inequality constraint $g_j(x) \leq 0$ is said to be active at a point x^* if $g_j(x^*) = 0$. A point x^* satisfying $g_j(x^*) \leq 0$ is said to be regular if the gradient vectors $\nabla g_j(x)$ of all active constraints are linearly independent.

Let x^* be a regular point of the constraints of SOP and assume that all the functions $f(x)$ and $g_j(x)$, $j = 1, 2, \dots, p$ are differentiable. If x^* is a local optimal solution, then there exist Lagrange multipliers λ_j , $j = 1, 2, \dots, p$ such that the following Kuhn-Tucker conditions hold,

$$\begin{cases} \nabla f(x^*) - \sum_{j=1}^p \lambda_j \nabla g_j(x^*) = 0 \\ \lambda_j g_j(x^*) = 0, \quad j = 1, 2, \dots, p \\ \lambda_j \geq 0, \quad j = 1, 2, \dots, p. \end{cases} \quad (\text{A-2})$$

If all the functions $f(x)$ and $g_j(x)$, $j = 1, 2, \dots, p$ are convex and differentiable, and the point x^* satisfies the Kuhn-Tucker conditions (A-2), then it has been proved that the point x^* is a global optimal solution of SOP.

A.1.1 Linear Programming

If the functions $f(x)$, $g_j(x)$, $j = 1, 2, \dots, p$ are all linear, then SOP is called a *linear programming*.

The feasible set of linear is always convex. A point x is called an extreme point of convex set S if $x \in S$ and x cannot be expressed as a convex combination of two points in S . It has been shown that the optimal solution to linear programming corresponds to an extreme point of its feasible set provided that the feasible set S is bounded. This fact is the basis of the *simplex algorithm* which was developed by Dantzig as a very efficient method for solving linear programming.

Table 1 This is an example for manually numbered table, which would not appear in the list of tables

Network Topology		# of nodes	# of clients			Server
GT-ITM	Waxman Transit-Stub	600	2%	10%	50%	Max. Connectivity
Inet-2.1		6000				
Xue	Rui	Ni	THUThESIS			
	ABCDEF					

Roughly speaking, the simplex algorithm examines only the extreme points of the feasible set, rather than all feasible points. At first, the simplex algorithm selects an extreme point as the initial point. The successive extreme point is selected so as to improve the objective function value. The procedure is repeated until no improvement in objective function value can be made. The last extreme point is the optimal solution.

A.1.2 Nonlinear Programming

If at least one of the functions $f(x)$, $g_j(x)$, $j = 1, 2, \dots, p$ is nonlinear, then SOP is called a *nonlinear programming*.

A large number of classical optimization methods have been developed to treat special-structural nonlinear programming based on the mathematical theory concerned with analyzing the structure of problems.



Figure 1 This is an example for manually numbered figure, which would not appear in the list of figures

Now we consider a nonlinear programming which is confronted solely with maximizing a real-valued function with domain \mathcal{R}^n . Whether derivatives are available or not, the usual strategy is first to select a point in \mathcal{R}^n which is thought to be the most likely place where the maximum exists. If there is no information available on which to base such a selection, a point is chosen at random. From this first point an attempt is made to construct a sequence of points, each of which yields an improved objective function value over its predecessor. The next point to be added to the sequence is chosen by analyzing the behavior of the function at the previous points. This construction continues until some

termination criterion is met. Methods based upon this strategy are called *ascent methods*, which can be classified as *direct methods*, *gradient methods*, and *Hessian methods* according to the information about the behavior of objective function f . Direct methods require only that the function can be evaluated at each point. Gradient methods require the evaluation of first derivatives of f . Hessian methods require the evaluation of second derivatives. In fact, there is no superior method for all problems. The efficiency of a method is very much dependent upon the objective function.

A.1.3 Integer Programming

Integer programming is a special mathematical programming in which all of the variables are assumed to be only integer values. When there are not only integer variables but also conventional continuous variables, we call it *mixed integer programming*. If all the variables are assumed either 0 or 1, then the problem is termed a *zero-one programming*. Although integer programming can be solved by an *exhaustive enumeration* theoretically, it is impractical to solve realistically sized integer programming problems. The most successful algorithm so far found to solve integer programming is called the *branch-and-bound enumeration* developed by Balas (1965) and Dakin (1965). The other technique to integer programming is the *cutting plane method* developed by Gomory (1959).

Uncertain Programming (BaoDing Liu, 2006.2)

References

NOTE: These references are only for demonstration. They are not real citations in the original text.

- [1] Donald E. Knuth. The \TeX book. Addison-Wesley, 1984. ISBN: 0-201-13448-9
- [2] Paul W. Abrahams, Karl Berry and Kathryn A. Hargreaves. \TeX for the Impatient. Addison-Wesley, 1990. ISBN: 0-201-51375-7
- [3] David Salomon. The advanced \TeX book. New York : Springer, 1995. ISBN:0-387-94556-3

附录 B 外文资料的调研阅读报告或书面翻译

英文资料的中文标题

摘要：本章为外文资料翻译内容。如果有摘要可以直接写上来，这部分好像没有明确的规定。

B.1 单目标规划

北冥有鱼，其名为鲲。鲲之大，不知其几千里也。化而为鸟，其名为鹏。鹏之背，不知其几千里也。怒而飞，其翼若垂天之云。是鸟也，海运则将徙于南冥。南冥者，天池也。

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} \quad (123)$$

吾生也有涯，而知也无涯。以有涯随无涯，殆已！已而为知者，殆而已矣！为善无近名，为恶无近刑，缘督以为经，可以保身，可以全生，可以养亲，可以尽年。

B.1.1 线性规划

庖丁为文惠君解牛，手之所触，肩之所倚，足之所履，膝之所倚，砉然响然，奏刀騞然，莫不中音，合于桑林之舞，乃中经首之会。

表 1 这是手动编号但不出现在索引中的一个表格例子

Network Topology		# of nodes	# of clients			Server
GT-ITM	Waxman Transit-Stub	600	2%	10%	50%	Max. Connectivity
Inet-2.1		6000				
Xue	Rui	Ni	THUThESIS			
	ABCDEF					

文惠君曰：“嘻，善哉！技盖至此乎？”庖丁释刀对曰：“臣之所好者道也，进乎技矣。始臣之解牛之时，所见无非全牛者；三年之后，未尝见全牛也；方今之时，臣以神遇而不以目视，官知止而神欲行。依乎天理，批大郤，导大窾，因其固然。技经肯綮之未尝，而况大瓠乎！良庖岁更刀，割也；族庖月更刀，折也；今臣之刀十九年矣，所解数千牛矣，而刀刃若新发于硎。彼节者有间而刀刃者无厚，以无厚

入有间，恢恢乎其于游刃必有余地矣。是以十九年而刀刃若新发于硎。虽然，每至于族，吾见其难为，怵然为戒，视为止，行为迟，动刀甚微，謦然已解，如土委地。提刀而立，为之而四顾，为之踌躇满志，善刀而藏之。”

文惠君曰：“善哉！吾闻庖丁之言，得养生焉。”

B.1.2 非线性规划

孔子与柳下季为友，柳下季之弟名曰盗跖。盗跖从卒九千人，横行天下，侵暴诸侯。穴室枢户，驱人牛马，取人妇女。贪得忘亲，不顾父母兄弟，不祭先祖。所过之邑，大国守城，小国入保，万民苦之。孔子谓柳下季曰：“夫为人父者，必能诏其子；为人兄者，必能教其弟。若父不能诏其子，兄不能教其弟，则无贵父子兄弟之亲矣。今先生，世之才士也，弟为盗跖，为天下害，而弗能教也，丘窃为先生羞之。丘请为先生往说之。”



图 1 这是手动编号但不出现索引中的图片的例子

柳下季曰：“先生言为人父者必能诏其子，为人兄者必能教其弟，若子不听父之诏，弟不受兄之教，虽今先生之辩，将奈之何哉？且跖之为人也，心如涌泉，意如飘风，强足以距敌，辩足以饰非。顺其心则喜，逆其心则怒，易辱人以言。先生必无往。”

孔子不听，颜回为驭，子贡为右，往见盗跖。

B.1.3 整数规划

盗跖乃方休卒徒大山之阳，脍人肝而脯之。孔子下车而前，见谒者曰：“鲁人孔丘，闻将军高义，敬再拜谒者。”谒者入通。盗跖闻之大怒，目如明星，发上指冠，曰：“此夫鲁国之巧伪人孔丘非邪？为我告之：尔作言造语，妄称文、武，冠枝木之冠，带死牛之胁，多辞缪说，不耕而食，不织而衣，摇唇鼓舌，擅生是非，以迷天下之主，使天下学士不反其本，妄作孝弟，而侥幸于封侯富贵者也。子之罪大极重，疾走归！不然，我将以子肝益昼脯之膳。”

附录 C 其它附录

前面两个附录主要是给本科生做例子。其它附录的内容可以放到这里，当然如果你愿意，可以把这部分也放到独立的文件中，然后将其 `\input` 到主文件中。

个人简历、在学期间发表的学术论文与研究成果

个人简历

xxxx 年 xx 月 xx 日出生于 xx 省 xx 县。

xxxx 年 9 月考入 xx 大学 xx 系 xx 专业, xxxx 年 7 月本科毕业并获得 xx 学士学位。

xxxx 年 9 月免试进入 xx 大学 xx 系攻读 xx 学位至今。

发表的学术论文

- [1] Yang Y, Ren T L, Zhang L T, et al. Miniature microphone with silicon- based ferroelectric thin films. Integrated Ferroelectrics, 2003, 52:229-235. (SCI 收录, 检索号:758FZ.)
- [2] 杨轶, 张宁欣, 任天令, 等. 硅基铁电微声学器件中薄膜残余应力的研究. 中国机械工程, 2005, 16(14):1289-1291. (EI 收录, 检索号:0534931 2907.)
- [3] 杨轶, 张宁欣, 任天令, 等. 集成铁电器件中的关键工艺研究. 仪器仪表学报, 2003, 24(S4):192-193. (EI 源刊.)
- [4] Yang Y, Ren T L, Zhu Y P, et al. PMUTs for handwriting recognition. In press. (已被 Integrated Ferroelectrics 录用. SCI 源刊.)
- [5] Wu X M, Yang Y, Cai J, et al. Measurements of ferroelectric MEMS microphones. Integrated Ferroelectrics, 2005, 69:417-429. (SCI 收录, 检索号:896KM)
- [6] 贾泽, 杨轶, 陈兢, 等. 用于压电和电容微麦克风的体硅腐蚀相关研究. 压电与声光, 2006, 28(1):117-119. (EI 收录, 检索号:06129773469)
- [7] 伍晓明, 杨轶, 张宁欣, 等. 基于 MEMS 技术的集成铁电硅微麦克风. 中国集成电路, 2003, 53:59-61.

研究成果

- [1] 任天令, 杨轶, 朱一平, 等. 硅基铁电微声学传感器畴极化区域控制和电极连接的方法: 中国, CN1602118A. (中国专利公开号)

- [2] Ren T L, Yang Y, Zhu Y P, et al. Piezoelectric micro acoustic sensor based on ferroelectric materials: USA, No.11/215, 102. (美国发明专利申请号)

综合论文训练记录表

学生姓名		学号		班级	
论文题目					
主要内容以及进度安排	<div>指导教师签字：_____</div> <div>考核组组长签字：_____</div> <div>年 月 日</div>				
中期考核意见	<div>考核组组长签字：_____</div> <div>年 月 日</div>				

指导教师评语	<div>指导教师签字：_____</div> <div>年 月 日</div>
评阅教师评语	<div>评阅教师签字：_____</div> <div>年 月 日</div>
答辩小组评语	<div>答辩小组组长签字：_____</div> <div>年 月 日</div>

总成绩：_____

教学负责人签字：_____

年 月 日