**(封面A3牛皮纸，内容与该页相同，封面内保留该页内容)**

**(页面设置：A4纸打印，版面上空2.5cm，下空2.0cm,左空2.2cm，右空2.2cm,装订线0.5cm)**

**山 东 科 技 大 学**

**毕 业 设 计（论 文）**

**说 明 书**

**设计题目：** （宋体正文小二号字）

**院 系： 专业： 班级：**

**设 计 人：**

**2020年 月 日**

|  |
| --- |
| **毕业设计任务书**  **专业 班 学生**   1. 毕业设计题目： 2. 设 计 专 题：   三、设计原始资料：      四、设计应解决下列各主要问题：        五、设计说明书应附有下列图纸：      六、命题发出日期： 年 月 日 设计应完成日期： 年 月 日  设计指导人（签章）：  教研室主任（签章）：  院 长（签章）： |

|  |
| --- |
| **特 约 评 阅 人 意 见**    评阅人（签章）：  评阅人职称：  日 期： 年 月 日 |

|  |
| --- |
| **指导教师对毕业设计的评语**    指导教师（签章）：  日 期： 年 月 日 |

|  |
| --- |
| **答辩委员会鉴定意见**    答辩成绩（等次及分数）：  **鉴定意见：**  答辩委员会主任（签章）：  答辩委员会副主任（签章）：  日 期： 年 月 日 |

**摘 要**

随着我国社会经济的迅速发展，传统能源被大量消耗……

建议(不限于)按以下段落来写：

第一段：研究背景及意义。

第二段：研究内容、方法等。

第三段：研究结论及影响。

**关键词：**(缩进两格) 分布式光伏，改进随机潮流，差分进化算法（3~5个）

近年来，随着当代科学技术的发展，各种社会媒体（如：新闻网站，社交网站，电商网站等）迅速崛起，使得网络中涌现出大量的文本数据，这些文本数据的结构与其所属领域特点往往有着较大的联系，于是跨领域的迁移学习在自然语言处理以及机器学习等领域开始得到越来越多的关注。目前传统文本分析的目标多数在同一领域，即对于某一领域的已有标签的文本数据进行机器学习，将训练出来的模型同样应用于本领域。但是在当前网络环境下，各类领域及其繁多复杂，更多的领域的数据是没有标签的，使用人工标注的方式会浪费极大的人力物力，同时，标注结果也会存在很大的主观性，因此亟需一些可实现跨领域文本标注的算法[1]。

在当前时代背景下，海量的文本信息充斥着整个互联网环境，目前传统文本分析的目标多数在同一领域，即对于某一领域的已有标签的文本数据进行机器学习，将训练出来的模型同样应用于本领域。但是在当前网络环境下，各类领域及其繁多复杂，更多的领域的数据是没有标签的，使用人工标注的方式会浪费极大的人力物力，同时，标注结果也会存在很大的主观性，因此亟需一些可实现跨领域文本标注的算法[1]。跨领域的文本特征对齐算法的相关流程主要有文本预处理，文本特征提取，特征对齐，模型检验对齐效果四个过程，下面分别对每一个过程做相关概述。

**Abstract**

With the rapid development of China’s social and economy，the traditional energy is consumed in large quantities...

尽量使用被动语态、一般现在时，若出现“论文”，用"thesis"

（英文采用Times New Roman字体）

**Keywords:** Distributed generation, improved stochastic power flow, differential evolution (专用词汇大写、除第一个关键词首字母大写外其他全部小写、标点符号是“半角逗号+空格”)

**目 录**

**1 绪 论………………………………………………………………………1**

1.1 课题的提出……………………………………………………………………………1

1.2随机潮流计算的研究现状……………………………………………………4

1.3 课题研究目的…………………………………………………………………………9

**2 随机潮流计算方法研究…………………………………………………11**

2.1 概率密度分布模型………………………………………………………………11

2.2 改进随机潮流…………………………………………………………………14

2.3 本章小结……………………………………………………………………………15

**3**

**4**

**5**

**6总结与展望………………………………………………………………70**

6.1 总结………………………………………………………………………………71

6.2 展望………………………………………………………………………………72

**参考文献……………………………………………………………………73**

**致 谢………………………………………………………………………76**

1**绪 论**

**1.1 课题的提出（二级标题1.5倍行距，段前段后为0，对齐网格）**

近年来，随着当代科学技术的发展，各种社会媒体（如：新闻网站，社交网站，电商网站等）迅速崛起，使得人们可以通过越来越多的方式在网络上发表自己的观点，致使当前国内外网络环境中涌现出大量的文本数据。这些文本数据的结构与其所属领域特点往往有着较大的联系，有些词汇甚至可以成为某些领域的代名词，例如：当我们看到“好评”、“物美价廉”等词语时，我们会想到这可能是对于某一商品的评价；当我们看到“吓人”、“振奋人心”等词语时，我们会想到这可能是对于某一消息或新闻的评价。于是一些情感分类算法在机器学习领域应运而生，但是这些情感分类算法的分析目标多数是在同一领域，即对于某一领域的已有标签的文本数据进行机器学习，将训练出来的模型同样应用于本领域，从而完成对未分类的文本的分类任务。但是在当前网络环境下，各类领域及其繁多复杂，而更多的领域的文本数据是没有标签的，同时使用人工标注的方式会浪费极大的人力物力，并且标注结果也会存在很大的主观性，导致标注结果的准确性极大的降低，因此在文本分析中亟需一些可实现跨领域文本标注的算法[1]，来对已有文本进行情感分类。

笔者认为，目前国内经济发展水平呈上升态势，人民群众正在不断谋取更高的物质生活水平与精神生活水平，于是网上购物与新闻知识获取的向往日趋强烈，使得电商领域与新闻领域成为全体网民乐于聚集之地，这两个领域所涵盖的人群数量大且范围广，因此在两个领域中留下了的文本信息较为全面，对这两个领域做迁移学习将有更大的研究价值，例如国家可以通过文本情感分析结果对社会舆论信息进行监控；对于包括新闻网站在内的大多数网站来说，可以对用户评论内容进行关键词提取进而对不符合要求的敏感词进行屏蔽；电商平台可以利用用户评论信息进行产品优化或修改营销计划；一些音视频媒体也可以根据用户对某些影视剧或某些歌曲的评论情况对其他用户进行择优推荐等等。目前面向迁移学习的文本情感分类并没有成文的步骤，笔者在阅读大量文献之后，总结得出做文本情感分类方面的迁移学习应当有如下4个步骤：1、文本预处理，2、文本特征提取，3、特征对齐，4、模型检验。显然，第2与第3步应当是整个过程的重中之重，对于基于特征的迁移学习而言，源域与目标域之间的枢纽特征的匹配对于在两个领域中进行机器学习的效果是至关重要的，因此研究好如何将两个领域中的非枢纽特征更好地转化为枢纽特征，以及如何将两个领域中的枢纽特征做最大程度上的对齐，对于提升迁移学习的准确率是至关重要的，这是选取本课题的重要意义之所在。

本文中，笔者首先对当今时代下跨领域文本特征对齐算法进行全面剖析，而后选取对本课题最为契合的算法，以电商领域为源域，以新闻领域为目标域，完成两个领域之间的迁移学习。

**1.1.1 文本特征提取算法的研究现状**

目前网络环境中，各类文本信息有着多种多样的形式，它们所表达的情感倾向也相差甚大，例如在电商领域与新闻领域中，用户们会发表许多的关于产品或者时政的评论及意见，这些评论及意见通常含有用户的情感极性。Turney[2]等人在2002年第一次将文本情感分类的任务当做学术问题公开提出，之后再行业内引发了众多学者的研究。

目前常用的基于统计的特征选择方法有信息增益方法(IG)、互信息方法(MI)、卡方拟和检验方法(CHI)、文档频率方法(DF)等[3-5]。2010年，在总结已有特征选择方法的基础上，杨凯峰[6]等一种新的基于文档频率(DF)的特征选择方法，此方法主要用于中文文本分类，需要先对每篇文档进行分词处理。他们使用 ICTCLAS(Institute of Computing Technology,Chinese Lexical Analysis System)系统完成文档的分词及词性标注，使用特征词在一个类别中出现的文档数来表示这个特征词与该类别的相关度，他们使用复旦大学整理的语料库进行分类实验，结果显示，在进行特征选择时，补充部分高词频特征词可以提高分类的召回率和准确率。在向量空间模型中，将文本表示单元（特征词）转换为向量通常使用的权重计算方法是TF-IDF(term frequency-inverse document frequency)方法，在一定程度上，这个方法能有效地表示一个特征词在文本中区分文本属性的重要程度，但其理论依据存在不足。2013年路永和[7]等提出了改进TF-IDF算法的文本特征项权值计算方法，改进的核心思想是：每个特征词对每个类的区分能力不同，其重要性用特征选择评估函数反映，同时该词集中分布的类就是其所属的类，即这个特征词要具有这个类的特征性。为避免特征词只集中出现在某篇文档中，要求该特征词在该类别文档中应均匀分布。随后，曹鲁慧[8]等提出一种基于深度学习的中文文本特征提取与分类方法，她们使用卷积神经网络(CNN)进行文本分类，将每个样本用128维的向量进行表示，大大减少特征的维度，加快分类器的训练速度，进而提高分类的准确率。卢晨阳[9]等提出一种基于语义结构的迁移学习文本特征对齐算法，他们使用wordnet模型完成领域独立特征的对齐。他们考虑到，传统的枢纽特征抽取和对齐一般只考虑到句子的结构和词性等信息，而忽略词语特征和类别信息存在一定的相关性，因此在抽取特征时应考虑类别影响。以情感分类为例，在抽取特征训练模型时，首先将源域和目标域的正类样本(积极)和负类样本(消极)分开，将不同词性的词语抽取出来，其中，在情感分类中起主要影响的是名词、动词、形容词、副词，然后对正类样本和负类样本分别抽取名词、动词、形容词和副词，原始的源域和目标域的2份数据变成根据样本类别和词性划分的16 份数据。最后 使用16 份数据作为词向量模型GloVe 模型的输入来进行训练。

另外，学者们对文本特征对齐算法的研究也在不断深入，2006年，John Blitzer[10]等提出一种结构一致学习算法(Structure Correspondence Learning, SCL)，这种算法可以自动归纳来自不同领域的特征之间的对应关系，SCL的核心思想是通过建立不同领域特征与枢纽特征的相关性来识别特征之间的对应关系。SCL的第一步是在两个域的未标记数据上定义一组枢纽特征。然后，使用这些枢纽特征来学习从两个域的原始特征空间到共享的低维实值特征空间的映射。在这个新的空间中，如果内积很高，则表示对齐效果较好。在有监督的训练任务中，同时使用源域中的转换特征和原始特征。在有监督的测试任务中，则同时使用目标域的转换特征和原始特征。如果学习到的映射比较好，那么在源域上学习到的分类器也将在目标域上有效。2013年，孟佳娜[11]等提出一种基于特征变换的跨领域产品评论倾向性分析方法，她们使用一种特征变换算法(Feature Transform Algorithm,FTrA)，首先计算源领域和目标领域的领域独立词，然后，求得源领域的领域依赖词与每个领域独立词之间的关联度值，与某个领域独立词关联度最高的特征，与其相关性越高。同时，求得目标领域的领域依赖词与每个领域独立词之间的关联度值，与某个领域独立词关联度值越高的特征，与其相关性越高。在此基础上，Sinno Jialin Pan[12]等提出了一种谱特征对齐 (Spectral Feature Alignment,SFA) 算法。SFA以一些领域无关词为桥梁，构造了一个二部图，来模拟领域特定词和领域无关词之间的共现关系。其思想是：如果两个特定于域的词与图中更常见的独立于域的词有关联，它们往往会以更高的概率对齐。类似地，如果两个独立于域的词与图中更常见的特定于域的词有连接，则它们倾向于以更高的概率排列在一起。他们采用基于图谱理论[13]的谱聚类算法，在二部图的基础上，将特定于领域的和与领域无关的词联合成一组特征聚类。这样，集群就可以用来减少两个域的特定于域的词之间的不匹配。最后，他们用这些聚类来表示所有的数据实例，并基于新的表示训练情感分类器。与结构对应学习(SCL)[10]等跨域情感分类算法不同，他们提出的SFA算法可以通过在二部图上的联合对齐，充分利用领域无关词和领域特定词之间的关系，学习更紧凑、更有意义的表示图的底层。通过实验表明，在跨领域情感分类的准确性方面，SFA确实比SCL等具有更好的性能。但是Yuhong Zhang[14]等认为，SCL与SFA这类的算法尝试提取出域无关主题词来构造子空间，并在子空间中训练分类器，这样的话，目标域中的域相关特征会被映射为源域的特征，然而在实际应用当中，这些方法无法解决在不同域中具有不同情感的特征词的问题，因为一个主题仅仅能够表达出一种情感，这些特征就会在源域中被错误地训练。针对这一问题，Yuhong Zhang等人提出一种面向跨领域情感分类的公共子空间构造算法[14]，此算法的目标是构造出一个更加精确的公共子空间，然后就可以在这个子空间中为目标域训练出分类器。对于有标签的源域S和无标签的目标域T，他们使用分类比例差[15]( Categorical Proportional Difference, CPD)算法计算源域中特征词的情感倾向，然后基于共现关系预测目标域中这些特征的情感取向，根据不同域之间情感取向的一致性构造公共子空间，于是域相关特征将会被过滤掉，最终使所有的域相关特征都会被投影到公共子空间中来解决域失配问题。跨领域的文本特征对齐是做文本情感分析的迁移学习过程中的重要一步，对与文本特征对齐技术与算法方面的研究，时至今日，仍吸引着众多的学者来专研，这也是笔者以此为课题的原因之一。

**1.1.2 字体要求**

正文模板中所有正文数字及字母应使用新罗马字体。

**1.1.3 公式要求**

公式的编号为分章编号，一般单行居中排版与上下文分开，式号与公式同行居右排版。如：式（2.1）表示第2章第1式。模型表达式如下：

 （2.1）

**1.1.4 图要求**

图+图号+空格+空格+图的名称

图题使用单倍行距

图的编号为分章编号。图的编号和图题应置于图下方，且使用中英文书写。如“图2.1”表示第2章第1图。图号与图题文字置一字空格置于图的正下方，图题用5号字，字体用宋体。图中标注符号文字字号不大于图题的字号。英文图题用Times New Roman字体。

正文中的图若用VISIO绘图，在VISIO中字大小为10号字，且在VISIO中图的宽度不能超过14cm，将VISIO中图直接拷贝到word中，不要缩放，不要插入图片。这样在word中所有图中字大小一致。



图2.1 四阶动态模型框图

**1.1.5 表要求**

表头要求：（单倍行距）

表+表号+空格+空格+表的名称

表的编号为分章编号，如：表2.1是第2章中的第1表。表应有表题，与表号之间空2字，置于表的上方居中，用5号宋体。表中的内容和项目字号不大于图题的字号。

表 2.1 锂离子电池SOC估算方法对比

|  |  |  |
| --- | --- | --- |
| 估算策略 | 优点 | 缺点 |
| 放电实验法 | 准确、可靠 | 须中断，时间长 |
| 安时积分法 | 计算简单 | 不够准确 |
| 开路电压法 | 数值上接近电池电动势 | 需长时间静置 |
| 线性模型法 | 模型简单 | 不够准确 |
| 交流内阻法 | 与SOC关系密切 | 测量困难 |
| 卡尔曼滤波法 | 适合非线性模型 | 需准确的模型算法 |

2**随机潮流计算研究**

**2.1 概率密度分布模型**

参考文献

专著：[序号] 编著者. 书名[M]. 出版地: 出版社, 年代, 起止页码

期刊论文：[序号] 作者. 论文题目[J]. 期刊名称, 年度, 卷(期): 起止页码

学位论文：[序号] 作者. 学位论文名称[D], 发表地: 学位授予单位, 年度

会议内容：[序号] 作者. 论文题目[C]. 会议名称, 出版年, 起止页码

参考文献例子如下：

[1] 王锡凡. 现代电力系统分析[M]. 北京: 科学出版社, 2003, 14-15.

[2] 丁明, 李生虎, 黄凯. 基于蒙特卡罗模拟的概率潮流计算[J]. 电网技术, 2001, 10(11): 10-14.

[3] 郭靖. 原对偶内点法及分枝定界法在无功优化中的应用[D]. 济南: 山东大学, 2004.

[4] Li W, Yang G, Zhou Z, et al. Impact of the distributed photovoltaic on the current protection of 10kV distribution network[C]. Power and Energy Engineering Conference, 2014, 1-4.

参考文献格式要求

1、参考文献格式需要统一，其中**所有标点符号**为半角，而且里边的标点符号后边加空格.

2、参考文献不要大量引用学位论文[D], 多引用近5年的成果论文, 尤其是16 、17年的论文

3、作者先写姓后写名，名字简写后面不加点。例如 Huang Q X,

4、参考文献的类型一定检查仔细，[J]、[C]等

5、参考文献作者多于三人时，列出前三位作者，后面几位用“等”字(英文使用et al)代替；如“张三, 李四, 王五, 赵六.”写成 “张三, 李四, 王五, 等.”

1. 魏晓聪, 林鸿飞. 面向迁移学习的文本特征对齐算法[J]. 计算机工程, 2017, 43(2): 215-219, 226.
2. Turney P D, Littman M L. Measuring praise and criticism: Inference of semantic orientation from association[J]. ACM Transactions on Information Systems(TOIS), 2003, 21(4): 315-346.
3. Galavotti L, Sebastiani F, Simi M. Experiments on the Use of Feature Selection and Negative Evidence in Automated Text Categorization[C]// Research & Advanced Technology for Digital Libraries, European Conference, Ecdl, Lisbon, Portugal, September. Springer-Verlag, 2000.
4. Mladenic D, Brank J, Grobelnik M, et al. Feature selection using linear classifier weights: interaction with classification models[C]// SIGIR 2004: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004. ACM, 2004.
5. 林森, 唐发根. 基于Log似然比的特征选择算法[J]. 计算机工程, 2009(19): 62-64+67.
6. 杨凯峰, 张毅坤, 李燕. Feature Selection Method Based on Document Frequency%基于文档频率的特征选择方法[J]. 计算机工程, 2010, 036(017): 33-35, 38.
7. 路永和, 李焰锋. 改进TF-IDF算法的文本特征项权值计算方法[J]. 图书情报工作, 2013(03): 92-97.
8. 曹鲁慧, 邓玉香, 陈通等. 一种基于深度学习的中文文本特征提取与分类方法[J]. 山东科学, 2019, 32(06): 106-111.
9. 卢晨阳, 康雁, 杨成荣等. 基于语义结构的迁移学习文本特征对齐算法[J]. 计算机工程, 2019, 45(05): 116-121.
10. Blitzer J, Mcdonald R T, Pereira F. Domain Adaptation with Structural Correspondence Learning[C]// EMNLP 2007, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 22-23 July 2006, Sydney, Australia. 2006.
11. 孟佳娜, 段晓东, 杨亮. 基于特征变换的跨领域产品评论倾向性分析[J]. 计算机工程, 2013, 39(10): 167-171.
12. PAN Jialin, NI Xiaochuan, SUN Jiantao, et al. Crossdomain sentiment classification via spectral feature alignment[C]// Proceedings of the 19th International Conference on World Wide Web. New York, USA: ACM Press, 2010:751-760.
13. Chung F R K. Spectral graph theory [M]. American Mathematical Soc, 1997.
14. Zhang Y, Xu X, Hu X. A Common Subspace Construction Method in Cross-Domain Sentiment Classification[C]// International Conference on Electronic Science & Automation Control. 2015.
15. M.Simeon, and R.Hilderman, “Categorical proportional difference: A feature selection method for text categorization,” The Australasian Data Mining Conference. pp. 201–208, 2008.
16. Zhang S, Liu H, Yang L, et al. A Cross-Domain Sentiment Classification Method Based on Extraction of Key Sentiment Sentence[M]// Natural Language Processing and Chinese Computing. Springer-Verlag New York, Inc. 2015.
17. Zhou G, He T, Wu W, et al. Linking Heterogeneous Input Features with Pivots for Domain Adaptation[C]// International Conference on Artificial Intelligence. AAAI Press, 2015.
18. 李鼎宇, 胡学钢. 面向短文本的跨领域情感分类算法[J]. 小型微型计算机系统, 2018, 39(05): 1005-1009.