

**学术学位硕士研究生学位论文开题报告**

论文题目： 基于大数据的创业突发热点分析

|  |  |
| --- | --- |
| 姓 名： | 孔阿栋 |
| 学 号： | 21524005 |
| 学科专业： | 计算机科学与技术 |
| 指导教师： | 冯林 |
| 入学日期： | 2015年9月 |
| 报告日期： |  |
| 报告地点： |  |

研究生院制表

说 明

学位论文开题考核是硕士研究生课程学习结束后开展学位论文工作的基本要求，是保证学位论文质量、工作进度和研究生培养质量的首要环节。

一、考核内容：首先，考查硕士生对本学科专业的基础理论与专业知识的掌握程度、课程学习情况等；其次，考查学位论文工作准备情况，包括论文选题、文献阅读、工作难度、研究思路、研究基础、写作能力和答辩表达能力等；此外，还要考查学术参与学术活动情况、及学习和工作态度等。

二、考核时间：原则上，硕士生的开题报告应在第2学期末进行。

三、报告撰写：开题报告正文字数不少于6000字；参考文献数量不少于20篇，其中，外文资料不少于二分之一，近5年文献不少于三分之一；正文及参考文献等撰写要求参见《大连理工大学硕士学位论文格式规范》。

四、考核办法：开题考核由学部（学院）按学科专业集中组织3-5名本学科领域专家以答辩的方式进行。硕士生进行口头陈述、答辩，研究生口头陈述时间不少于10分钟。专家组给出考核成绩和是否通过的意见。

五、报告保存：开题报告一式两份，签字后分别由学部（学院）和学生保存。

六、信息登录：研究生开题后登录研究生信息管理系统上传开题报告（PDF文档）及考核结果。

**开题报告正文**

撰写大纲：

1. 课程学习情况（附成绩单）、参加科研和学术活动等情况；
2. 学位论文研究背景、目的和意义；
3. 国内外研究现状及发展动态分析；
4. 主要研究内容、研究目标、拟解决的关键问题；
5. 学位论文的研究方法、技术路线、试验手段、关键技术等论述；
6. 年度研究计划；
7. 现有的研究基础；
8. 现有研究条件及可能遇到的困难和问题分析。
9. 参考文献（不占字数）。

# 1 学习及学术活动情况

## 1.1 课程成绩单

|  |  |  |  |
| --- | --- | --- | --- |
| 课程 | 课程学分 | 选修学期 | 成绩 |
| 传感器网络技术 | 2 | 2 | 88 |
| 高性能计算 | 2 | 2 | 94 |
| 高级操作系统 | 3 | 1 | 87 |
| 算法设计与分析 | 3 | 1 | 82 |
| 人工智能 | 2 | 2 | 92 |
| 分布式数据库 | 2 | 1 | 94 |
| 论文写作与学术规范 | 1 | 1 | 94 |
| 中国特色社会主义理论与实践研究 | 2 | 2 | 83 |
| 口语交流 I （基础口语表达） | 1 | 1 | 93 |
| 阅读与写作 I （基础读写技能） | 2 | 2 | 87 |
| 数理统计 | 2 | 1 | 92 |
| 数据仓库技术 | 2 | 2 | P |
| 信息检索与文本挖掘 | 2 | 1 | 95 |
| 分布式对象技术 | 2 | 2 | P |
| 中间件技术 | 2 | 2 | P |
| 创业管理与企业家精神 | 2 | 2 | P |
| 马克思主义与社会科学方法论 | 1 | 2 | P |

## 1.2 参加学术活动情况

2015年7月18日，本人参加了由大连理工大学主办的2015年大数据分析与应用研讨会（The 2015 International Workshop on Big Data Analytics and Applications）。研讨会邀请了包括来自美国伊利诺大学芝加哥分校、美国罗格斯-新泽西州立大学、美国纽约市立大学、中国科学技术大学的4位海内外数据挖掘领域专家出席，吸引了来自大连各高校的140多名师生参与。

本人认真听取了各位老师的报告，使自己对大数据时代各方面的发展有了较为深入的了解，对于机器学习的相关概念和研究方法有了深刻的认识。

## 1.3 本文主要内容和组织结构

本文主要内容安排如下：

第一章介绍学习及学术活动情况。

第二章介绍大数据的创业突发热点分析的研究背景，目的和意义。

第三章分析了国内外研究现状及发展动态。

第四章介绍了主要研究的内容及方法。

第五章介绍现有的研究基础和年度研究计划。

第六章介绍可能遇到的困难和问题分析。

# 2 研究背景、目的和意义

互联网从PC时代进化到互联网时代，然后从移动互联网时代进阶到了大数据时代。PC时代催生了大量的互联网上市企业。移动互联网时代，中国创业热潮风生水起，不仅有大量的移动互联网企业赴美上市，更是诞生了无数个创业奇迹。移动互联网不仅为我们的生活带来了便利，更是把创业热潮推向了历史最高峰。大数据时代的创业热点更是成为创业者、投资者的关注点，进而提出了对于创业热点的分析。

网络舆情是随着网络和信息技术的快速发展普及产生的，是大众就各种社会问题在现代的网络空间中发表的不同意见和态度的信息。互联网时代，网络已成为社会舆论的思想文化信息的集中地。舆论信息反映了民众的思想状况，对于社会的和谐稳定有着重要的作用，因而有效的对网络上的舆情信息进行监控，实时掌握网络舆情信息的动态是非常有必要的。

“大众创业万众创新”的提出，各种互联网应用创新和商业模式创新不断涌现，引发传统行业生态的深刻变革。互联网移动化、产品化进程加快，互联网企业聚力转型，抢占未来产业高地。移动互联网时代，中国创业热潮风生水起，移动互联网不仅为我们的生活带来了便利，更是把创业热潮推向了历史最高峰。

在这个遍地创业的环境中，想要看清当下现状，光靠经验和理智并未能拨开迷雾。大数据时代，通过对大量创业或投资数据的抓取分析，能够及时发现热点话题，对创业者来讲，能够看清当下现状，拨开创业迷雾，抢占先机。对投资者来讲，能够提供更好的投资选择，降低投资风险的同时追求高回报。互联网创业的繁荣是否只是表象？不管创业是获得融资功成名就，还是死在了沙滩上，站在互联网创投火爆的节点看清当下现状都是很有必要的。

目前信息处理领域对舆情监控的研究和成型的产品很多，对网络舆情监控的研究也较为成熟，但还没有对创业突发热点的抓取和分析。如果能够对创业热点信息进行监控，获得较为准确的突发创业热点，那么不管对于想要创业的创业者还是风险投资的天使人都有很重要的参考价值和意义。针对近年创业热潮的分析监控的研究暂时为空白，本文将已有的对舆情监控的研究方法和分析，运用到对创业突发热点的分析上来，为众多创业者投资者提供有参考价值的创业热点数据，避免盲目创业，降低风险投资。

# 3 国内外研究现状及发展动态分析

近年国外对舆情的研究很早，而且有专门的舆情机构进行深入的研究，其中网络舆情随着互联网技术的发展成为重点。国外相关方面最早的研究是1998年美国国防部提出的话题发现与跟踪技术（Topic Detection and Tracking，简写为TDT）[1]，旨在研究在大量新闻数据中发现重要信息，希望在没有人工干预的情况下自动判断新闻数据的主题。TDT的研究主要涉及以下五个方面：

1. 报道切分（Story Segmentation）：将连续的新闻数据流按内容进行分割；
2. 关联发现（Link Detection）：比较新闻报道是否在阐述相关内容；
3. 主题跟踪（Topic Tracking）：判断后续报道属于哪个已存在的主题；
4. 新事件发现（New Event Detection）：识别出报道是对新事件的阐述；
5. 主题发现（Topic Detection）：识别出新主题。

相对来说国内对网络舆情领域的研究起步较晚，这是由于互联网上世纪90年代才开始在我国发展进来。我国网络舆情的研究从2005年开始[2]，经过10年的研究，也取得了不少研究成果，主要包括网络舆情的基础概念、传播特征、管理和引导和利用计算机技术对网络舆情进行监控分析与跟踪。

除了理论研究外，随着信息检索，数据挖掘等技术的兴起，研究者提出了不少网络舆情信息挖掘与分析的模型，相应的网络舆情监控系统也有少被开发出来并使用。中山大学黄晓斌认为网络舆情是以各种文本的形式在网络上产生与传播的，可以把文本挖掘技术应用到网络舆情分析中，通过对文本挖掘的特点进行分析，用具体实例说明如何对网络舆情信息进行文本挖掘，并据此提出了一个网络舆情信息文本挖掘的模型[3]，这是比较典型的基于统计的方法；钱爱兵提出基于主题的概念，设计了基于舆情主题分类的的分析模型[4]，但这种方式需要人工选择舆情主题；对于如何自动发现主题，郭建永等结合多种聚类方法提出一种增量层次聚类算法[5]。王伟通过网页热点簇构建特征向量来进行网页聚类，提高舆情分析的准确度[6]。

主题模型的研究是一个从简单到复杂的过程，最初的主题模型是潜在语义索引（Latent Semantic Index，简称 LSI）[7]，是由Scott等人于1990年提出，它通过对矩阵进行奇异值分解达到降维的作用，同时能够起到同义词和多义词的冗余信息的消除作用。1999 年Hoffman把概率的思想引入到了 LSI 模型当中，用概率方法取代LSI 中的奇异值分解，提出了pLSI（Probabilistic LSI），称为概率潜在语义索引模型，或者称为概率潜在语义模型（Probabilistic Latent Semantic Analysis， pLSA）[8]。pLSA认为主题是可以通过似然函数来估计的单词的概率分布。2003年Blei 等人进一步改进了pLSA，认为不但主题是单词的概率分布，而且文档也是主题的概率分布，并提出了LDA（Latent Dirichlet Allocation）[11]主题模型。LDA 模型一经提出，得到了学者们的普遍认可，而且得到了广泛的应用。同时学者们也对 LDA 模型进行进一步的研究，根据具体的应用提出了不少变种和改进。如CTM（Correlated Topic Model，相关性主题模型）[9]、DTM（Dynamic Topic Model，动态的主题模型）[9]等。国内的Labeled2LDA（附加类别标签的LDA模型）[10]、LDA-CF（协同过滤方法）等。

# 4 主要研究内容及方法

本文以网络舆情监控为基础，通过对大量创业数据的爬取，构建自己的创业信息数据库，先对数据做预处理，进而使用LDA主题模型进行创业热点分析，论文进展后期计划采用多种聚类方法来对创业热点数据进行统计分析，并对各聚类结果进行对比分析。

## 4.1 信息采集

创业数据的采集是创业热点分析处理的第一步。本文的创业数据来源是有关的创业网站如36氪、it桔子、创业邦等。通过对这些网站的网页结构、数据加载等方面分析发现，这些网站并不是简单的静态网页，网页中的数据都是通过JavaScript代码的执行动态加载再渲染出来的，这类基于Ajax脚本的动态网页带来了更好的用户体验，同时也使得对其进行获取和分析面临着新的问题和挑战。只是简单的借助Python的urllib模块已经无法获取到我们所需要的数据信息，本文采用动态爬虫工具Selenium2爬取动态网页的数据，将特征信息创业公司名称、简介、分类、投资信息、招聘职位月薪开支等数据存入MongoDB数据库中，构建自己的创业数据库。

4.1.1 Selenium2

Selenium2（Selenium1+WebDriver）是一个模拟浏览器进行自动化测试的工具，核心是WebDriver，它提供了一组API可以与真实的浏览器内核交互，支持多种浏览器。Selenium使用JavaScript和Iframes在浏览器中嵌入自动化测试引擎，实现了在任何支持JavaScript动态脚本的浏览器中模拟客户端的各种操作，并顺序执行动态脚本语言程序。

Selenium不同于一般的脚本解析引擎，一般的解析引擎是通过拦截浏览器收发的HTTP请求来实现的，并不能实现对浏览器页面元素的操作，也无法记录用户的动作轨迹。Selenium通过在真实浏览器中模拟用户各种操作，实现了动态脚本的执行与记录[13]。Selenium提供了以下功能模块：

(1) 浏览器驱动模块：即利用Selenium来调用浏览器完成用户定义操作，支持的操作包括启动浏览器、查找页面元素、鼠标交互如click、页面及表单交互、支持JavaScript完成页面异步操作、Cookie管理器以及异常处理支持等。

(2) 页面元素解析模块：采用Beautiful Soup来完成，它是一个HTML或XML文件中提取数据的Python库，能够通过用户指定的转换器实现通用的文档浏览、查找和修改。

(3) 验证码识别模块：验证码是用以区分用户是计算机还是人的公共全自动程序，用来防止恶意破解密码、暴力不断登录等。

4.1.2 MongoDB数据库

MongoDB是一个基于分布式文件存储的数据库，是面向文档的开源的NoSQL数据库系统，由C++编写，提供了强大、灵活、可扩展的数据存储方式。它扩展了关系型数据库的众多功能，如辅助索引、范围查询和排序。MongoDB的功能非常丰富，比如内置的对MapReduce式聚合的支持，以及对地理空间索引的支持。MongoDB的主要特性有[14]：

1. 丰富的数据模型

面向文档的方式将原来的“行”的概念换成更加灵活的“文档”，可以将文档或者数组内嵌进来，所以用一条记录就可以表示非常复杂的层次关系。MongoDB没有模式，文档的键不会事先定义也不会固定不变，由于没有模式需要更改，通常不需要迁移大量数据，应用层可以方便地处理新增的键或丢失的键。

1. 容易扩展

MongoDB 从设计之初就考虑到了扩展的问题，它所采用的面向文档的数据模型使其可以自动在多态服务器之间分割数据。它还可以平衡集群的数据和负载，自动重排文档。

1. 功能丰富

MongoDB支持辅助索引、固定集合、文件存储、存储JavaScript和MapReduce等其他聚合工具的独特功能。

1. 卓越的性能

MongoDB使用MongoDB传输协议作为与服务器交互的主要方式，它对文档进行动态填充，预分配数据文件，用空间换取性能的稳定。默认的存储引擎中使用了内存映射文件，将内存管理工作交给操作系统去处理。动态查询优化器会记住执行查询最高效的方式。

1. 简便的管理

尽可能地让服务器自动配置，通过备份机制来提升系统的可靠性。

MongoDB的核心概念是文档(document)，多个键及其相应的值有序地存放在一起组成文档，每个文档都有一个特殊唯一的键“\_id”，文档类似于关系型数据库中的元组。多个文档组成集合(collection)，集合如同关系型数据库中的表。多个集合组成数据库，一个MongoDB的实例可以容纳多个独立的数据库，每个数据库都有自己的集合和权限。

## 4.2 信息预处理

因为本文已经在Selenium爬取数据时借助Beautiful Soup已经对需要的数据信息进行了提取，所以预处理这一环节就只需要去停用词和中文分词这两步处理。

为节省存储空间和提高搜索效率，在处理自然语言数据（或文本）之前或之后会自动过滤掉某些字或词，这些字或词即被称为Stop Words（停用词）。中文语言中的一些功能词语其他词相比没有特别的实际含义，如“的”、“啊”、“呀”等，这些词通常用于连接句子成分或表达感情、强调语调等用途，因此称为停用词。这些停用词都是人工输入、非自动化生成的，生成后的停用词会形成一个停用词表。停用词的存在不仅会降低实验处理结果的正确率，且由于停用词数量较多，导致文本矩阵存在严重的高维稀疏等问题，从而影响实验性能。

分词是将连续的字序列按照一定的规范重新组合成词序列的过程。由于中文文本处理的基本单位是词语，而中文文本并不像英文那样每一个单词之间会用空格分开，中文文本中只有句子单位之间才会有标点符号，因此中文分词也就比英文要复杂的多。中文分词是进行文本挖掘的基础，所以首先需要对中文文本进行分词操作。本文中采用Python中文分词组件jieba中文分词，提供的功能包括有分词、添加自定义词典、关键词提取、词性标注、并行分词等。将经过分词的文本内容去停用词后用于主题模型的分析。

## 4.3 LDA主题模型

传统判断两个文档相似性的方法是通过查看两个文档共同出现的单词的多少，如TF-IDF等，这种方法没有考虑到文字背后的语义关联，可能在两个文档共同出现的单词很少甚至没有，但两个文档是相似的。比如苹果和乔布斯这两个词一起出现时，可以设定属于苹果公司这同一主题。所以在判断文档相关性时需要考虑文档的语义，而语义挖掘的利器是主题模型，LDA就是其中一种很有效的模型。

Latent Dirichlet Allocation（LDA）模型LDA 由Blei, David M.、Ng, Andrew Y.、Jordan 于2003 年提出，是一种主题模型，它可以将文档集中每篇文档的主题以概率分布的形式给出，从而通过分析一些文档抽取出它们的主题（分布）出来后，便可以根据主题（分布）进行主题聚类或文本分类。同时，它是一种典型的词袋模型，即一篇文档是由一组词构成，词与词之间没有先后顺序的关系。此外，一篇文档可以包含多个主题，文档中每一个词都由其中的一个主题生成。

LDA 的这三位作者在原始论文中给了一个简单的例子。比如假设事先给定了几个主题，然后通过学习的方式，获取每个主题Topic 对应的词语。然后以一定的概率选取上述某个主题，再以一定的概率选取那个主题下的某个单词，不断的重复这两步，最终生成一篇文章。

当我们看到一篇文章后，往往喜欢推测这篇文章是如何生成的，我们可能会认为作者先确定这篇文章的几个主题，然后围绕这几个主题遣词造句，表达成文。LDA做的就是根据给定的一篇文档，推测其主题分布。

在LDA模型中，一篇文档生成的方式如下：

（1）从狄利克雷分布中取样成文档i 的主题分布；

（2）从主题的多项式分布中取样生成文档i 第j 个词的主题

（3）从狄利克雷分布中取样生成主题对应的词语分布

（4）从词语的多项式分布中采样最终生成词语

将这个过程重复N次就产生了文档d，图模型如下：

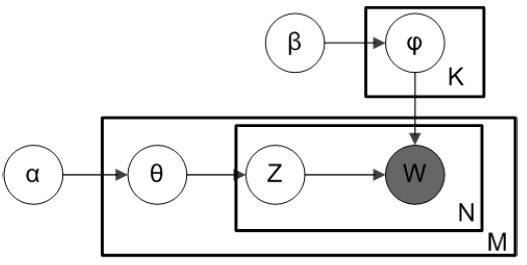


图4.1 LDA模型图

和表示预料级别的参数，也就是每个文档都一样，因此生成过程只采样一次。是文档级别的变量，每个文档对应一个，也就是每个文档产生各个主题z的概率是不同的，生成每个文档采样一次。z和w都是单词级别变量，z由生成，w由z和共同生成，一个单词w对应一个主题z。

从上面的 LDA生成模型的描述中我们知道，在LDA主题模型在生成过程需要两个控制参数，分别是和，有了这两个参数就能够用于生成新的文本，其中表示文档集中各个潜在主题之间的相对强弱，则刻画了各个潜在主题本身的概率分布，即潜在主题中各特征词的概率。通过LDA建模后就可以知道“主题-词语”的分布情况，表示单词在主题中的概率，概率值越大则该词越能表示该主题，因此可以从每个主题中选择前n个概率最大的词作为特征词，这样基于LDA的主题模型的特征词的选取很好的克服了传统的特征词选取中存在的丢失语义的不足。

5 年度研究计划

研二上学期做实验。

研二下学期发表论文。

研三构思并完成毕业设计论文。

6 现有研究基础

最开始采用PySpider爬虫框架，以及PhantomJS无界面浏览器爬取36氪网站的动态资讯文章数据进行主题分析，发现创业资讯文章本身噪声过多，很多无关创业的媒体娱乐文章也会夹杂其中，导致主题提取的效果并不好，所以更换了数据源，改为分析拉勾网上创业公司的信息，从创业公司的投资进程、公司简介、招聘职位月薪开支等方面作为数据分析源。在数据爬取过程中因为没有考虑爬取频率、user-Agent的更换，以及代理IP的使用，导致学院的IP被封，之后发现拉勾网的反爬虫策略有所提升，原来的爬虫策略获取不到动态数据了，于是改用Selenium来驱动有界面浏览器模拟用户行为进行动态数据爬取。

在爬虫代码的编写过程中做了一些优化，包括不加载网页图片，使用抓包分析将爬取网页改为直接请求JSON数据，使用多进程并行爬取，来加快爬取速度。同时也考虑了代码的健壮性，等待所有需要的DOM元素都加载完成再分析，否则就等待一段时间再继续分析；检测框架API的异常情况并捕获，如果异常表明浏览器崩溃就重启浏览器继续爬取。

目前，已阅读一些文献并参考他们在处理文本内容时的数据模型选择及优化。如何将现有的已成熟的网络舆情监控方面的研究应用到抓取分析突发的创业热点上来并且能够进一步优化出更好的实验数据结果是下一步要重点做的。接下来会研读国外的有关英文文献，归纳总结如果对数学模型做改进。

7 可能遇到的困难和问题分析

目前爬虫效率和稳定性还有待提高，数据源较为单一需要进一步扩充，爬到的数据可能有缺失不完整，需要进一步的数据清洗与纠正；算法方面还需要学习多种数学模型，提高创业热点分析结果的有效性和准确性。

参 考 文 献

1. Allan J, Carbonell J G, Doddington G, et al. Topic detection and tracking pilot study final report[J]. 1998.
2. 许鑫, 章成志, 李雯静. 国内网络舆情研究的回顾与展望[J]. 情报理论与实践, 2009 (3): 115-120.
3. 黄晓斌, 赵超. 文本挖掘在网络舆情信息分析中的应用[J]. 情报科学, 2009, 27(1): 94-99.
4. 钱爱兵. 基于主题的网络舆情分析模型及其实现[J]. 现代图书情报技术, 2008, 24(4): 49-55.
5. 郭建永, 蔡勇, 甄艳霞. 基于文本聚类技术的主题发现[J]. 计算机工程与设计, 2008, 29(6): 1426-1428.
6. Wei W, Xin X. 基于聚类的网络舆情热点发现及分析[J]. 现代图书情报技术, 2009, 3(3): 74-79.
7. 林洋港. 概率主题模型在文本分类中的应用研究[D]. 中国科学技术大学, 2009.
8. Hofmann T. Probabilistic latent semantic indexing[C]//Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1999: 50-57.
9. 徐盛. 基于主题模型的高空间分辨率遥感影像分类研究[D]. 上海交通大学, 2012.
10. 李文波, 孙乐, 张大鲲. 基于 Labeled-LDA 模型的文本分类新算法[J]. 计算机学报, 2008, 31(4): 620-627.
11. Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.
12. 杜彬. 基于 Selenium 的定向网络爬虫设计与实现[J]. 金融科技时代, 2016, 24(7): 35-39.
13. 张瑶. 面向 AJAX 脚本网络的网页爬行及解析技术的研究与实现[D]. 东北大学, 2012.
14. Chodorow K. MongoDB: the definitive guide[M]. " O'Reilly Media, Inc.", 2013.
15. 王少鹏,彭岩,王洁.基于LDA的文本聚类在网络舆情分析中的应用研究[J]. 山东大学学报(理学版). 2014(09).
16. 钟荣飞.基于主题模型的网络舆情监控研究[D].南华大学. 2015.
17. 王振振,何明,杜永萍. 基于LDA主题模型的文本相似度计算[J]. 计算机科学. 2013(12).
18. 王伟,许鑫. 基于聚类的网络舆情热点发现及分析[J]. 现代图书情报技术. 2009(03).
19. Zhai Z, Liu B, Xu H, et al.Constrained LDA for grouping product features in opinion mining.Advances in knowledge discovery and data mining. Springer Berlin Heidelberg, pp.448-459, 2011.
20. Bicego, M. , Lovato, P. , Perina, A. , Fasoli, M. , Delledonne, M. , Pezzotti, M. , et al. (2012). Investigating topic models’ capabilities in expression microarray data classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 9,1831–1836 .
21. 何友全,徐澄,徐小乐,唐华姣. 一种基于统计学特征和DOM树的网页去噪技术[J]. 重庆理工大学学报(自然科学版). 2011(01)
22. Ponte J M, Croft W Bruce . A Language Modeling Approach to Information Retrieval[J]. In proceedings of the 21th ACMSIGIR Conference. New York. 1998,275-281.
23. 林洋港.概率主题模型在文本分类中的应用研究[D].合肥:中国科学技术大学,2009.
24. G. Salton,A. Wong,C. S. Yang. A vector space model for automatic indexing[J]. Communications of the ACM . 1975 (11)
25. 王茜仪.基于主成分分析的网络舆情快速预测方法[D]. 江苏大学. 2016.
26. 邹盼湘.网络舆情热点提取与分析[D]. 华南理工大学. 2015.
27. 孙培星.基于情感倾向性的网络舆情分析及演化预测研究[D]. 吉林大学.2016.

**大连理工大学学术学位硕士研究生学位论文开题报告评审意见表**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学 号 | |  | | 学生姓名 | |  | 导 师 | |  | |
| 第一次开题 □ | | | | | | 第二次开题 □ | | | | |
| 导师考核意见（对课程学习情况、参加学术活动情况、学位论文工作及开题报告撰写情况、学习和工作态度等进行考查，给出考核成绩和具体改进意见和建议）：   1. 考核成绩：□ 优秀，□ 良好，□ 中等，□ 及格，□ 不及格 2. 是否通过：□ 通过，□ 不通过 3. 关于开题报告撰写质量及学位论文工作的具体意见（可加页）：   导师签字：  年 月 日 | | | | | | | | | | |
| 评  审  专  家  组 |  | 姓名 | 职称 | | 学科专业 | | | 是否博导 | | 签字 |
| 组长 |  |  | |  | | |  | |  |
| 成员 |  |  | |  | | |  | |  |
|  |  | |  | | |  | |  |
|  |  | |  | | |  | |  |
|  |  | |  | | |  | |  |
|  |  | |  | | |  | |  |
|  |  | |  | | |  | |  |
| 专家组评审意见（对课程学习情况、参加学术活动情况、学位论文工作及开题报告撰写情况、学习和工作态度等进行考查，给出考核成绩和学位论文工作具体改进意见和建议）：   1. 选题是否属于本学科领域（含交叉学科）：□ 是，□ 不是（须重新开题） 2. 考核成绩：□ 优秀，□ 良好，□ 中等，□ 及格，□ 不及格 3. 是否通过：□ 通过，□ 不通过 4. 关于开题报告撰写质量及学位论文工作的具体意见（可加页）：     组长签字：  年 月 日 | | | | | | | | | | |
| 点长意见：  点长签字：  年 月 日 | | | | | | | | | | |