**北京航空航天大学计算机学院**

**硕士学位论文中期检查报告**

**论文题目**：基于背景热点的微博转发预测研究

**专 业**：计算机技术

**研究方向**：网络信息安全

**研 究 生**：陈 江

**学 号**：ZY1306311

**中心导师**：王丽宏

**校内导师：**巢文涵

**北京航空航天大学计算机学院**

2015年 8 月 31 日

目 录

1 论文工作计划 1

1.1 论文研究目标 1

1.2 论文主要研究内容 1

1.2.1多通道背景热点表示和融合方法 1

1.2.2时序化微博用户兴趣提取 2

1.2.3融合背景热点和微博用户兴趣的微博转发预测模型 2

1.2.4基于背景热点的微博转发预测系统设计与实现方案 2

2 已经完成的工作 3

2.1融合热点话题的用户转发兴趣 3

2.2融合热点话题的微博转发预测研究 4

2.2.1工作概述 4

2.2.2数据集准备 5

2.2.3数据分析 5

2.2.4特征选择 8

2.2.5实验结果及分析 9

3 关键技术或难点 10

3.1 融合热点话题的用户转发兴趣特征 10

3.2 微博转发预测模型构建 10

4 下一阶段工作计划 10

4.1 存在的问题 10

4.2 尚未完成的工作 11

4.3 解决问题的技术思路或措施 11

4.3.1多通道背景热点的表示和提取方法 11

4.3.2基于背景热点的微博转发预测系统设计与实现方案 12

4.4 下一阶段计划 12

5 主要参考文献 12

**基于背景热点的微博转发预测研究**

# 论文工作计划

## 论文研究目标

本文的研究目标是分析背景热点对微博转发的影响，构建基于背景热点的微博转发模型，在此基础上，对微博转发进行预测。其中，背景热点包括微博热点、新闻热点以及搜索热点等。论文的基本假设是，用户转发一条微博客观上受热点事件影响，主观上受个人兴趣影响，当热点事件相关的微博没有出现在用户可视范围之内时，转发主要受个人兴趣的影响；另一方面，当面对一个热点事件时，即使不是用户感兴趣的内容，出于社会因素的影响，用户仍然可能转发微博；同时，当发生的热点事件刚好也是用户感兴趣领域相关的事件时，用户转发相关微博的可能性会更大。故本文的研究主要聚焦在背景热点和用户兴趣，构建基于背景热点及用户兴趣的微博转发模型，从而对微博转发进行预测。

## 论文主要研究内容

### 1.2.1多通道背景热点表示和融合方法

热点是指一定时间、一定范围内，公众最为关心的热点话题、热点事件、热点新闻等的总称。本文考虑的热点是指包括搜索热点、新闻热点、微博热点在内的多通道热点。通过百度搜索风云榜可以获得包括娱乐、人物、小说、游戏、汽车等多领域的搜索热点，并且可以获得每个搜索热点对应的相关新闻；新闻热点则可以通过新浪新闻排行榜、网易新闻排行榜、凤凰新闻排行榜以及今日头条等获得，这些网站同样也提供热点新闻的相关新闻；微博热点则可以新浪博客中的“微博最火”频道获得，此频道同样也提供每篇热门微博的相关微博。

为了得到背景热点简洁且信息量大的表示，本论文将提取背景热点相关新闻的热点词，并以此热点词来表示对应背景热点。本论文的热点词从背景热点相关新闻中提取，能够体现当前的热点事件、重要资讯和流行趋势，这样的热点词具有简洁、时效、信息量大、相关词语之间的互信息大等特点。

### 1.2.2时序化微博用户兴趣提取

用户兴趣是决定用户是否转发一条微博的重要因素，本论文所讨论的用户兴趣是指通过用户所发布历史微博体现出来的用户偏好。在新浪平台，对于一个新注册用户，用户兴趣可以通过用户标签来反映，但随着用户转发微博量的增加，用户所转发或者发布的微博体现出来的用户兴趣虽然能在一定程度上体现标签所反映的用户兴趣，但却不可能是完全吻合的，因为一个人的兴趣虽然具有一定的稳定性，但并不是一直不变的。本论文称这种具有一定稳定性亦具有随时间的波动性的微博用户兴趣为时序化微博用户兴趣。

本论文的用户兴趣模型同时考虑了用户标签和用户历史微博。本论文假设用户标签是用户兴趣的初始状态，随着用户发博数量的增加，本论文将从历史微博中提取用户兴趣，然后结合用户兴趣的初始状态作为用户的兴趣度量。本论文假设用户兴趣可以通过用户标签、用户兴趣空间来标识，在这个用户兴趣标识模型中，随着时间的推进，用户标签和用户兴趣词在标识模型中所占的权重是变化的。用户标签是指用户固有属性里的用户标签，用户兴趣空间是指从用户历史微博中提取的用户兴趣表示。通过用户兴趣模型提取的用户兴趣，将结合其他特征构建微博转发模型。

### 1.2.3融合背景热点和微博用户兴趣的微博转发预测模型

微博转发作为信息传播的一种具体形式，既具有信息传播所体现出来的网络特征，同时也具有信息传播所体现出来的社会学特征。一方面，用户之间的关注和被关注体现了用户之间的关系，亿级数量级的微博用户构成一张庞大的用户关系网络；另一方面，用户之间的微博转发，是用户获取资讯的方式，也是用户之间沟通交流的社会化行为之一。用户转发微博是具有既定的背景知识的，本文假设背景热点就是用户转发微博的背景知识，背景知识刺激用户转发微博是客观因素；主观上，用户转发微博很大程度上受用户兴趣的支配。所以，本论文将融合背景热点及微博用户兴趣构建微博转发模型。

### 1.2.4基于背景热点的微博转发预测系统设计与实现方案

对于一个给定的微博转发数据集，微博是否被转发是既定的事实，一个微博转发预测模型或算法效果的好坏可以通过和数据集的比较得出，转发数据集中被转微博的转发概率为1，预测模型或算法预测的转发概率越接近于1，则预测效果越好。

有了基于背景热点及用户兴趣的微博转发模型之后，本论文将研究该模型对微博转发的预测效果，并对加入背景热点和不加入背景热点的模型预测效果进行比较，研究背景热点对微博转发预测的影响，设计基于背景热点的微博转发预测系统，并实现其功能。

# 已经完成的工作

### 2.1融合热点话题的用户转发兴趣

兴趣是人们对事物喜好或关切的情绪，它表现为人们对某件事物、某项活动的选择性态度和积极的情绪反应。微博用户的兴趣部分地通过用户的转发行为体现出来，我们称之为用户转发兴趣。用户的转发行为受用户对微博的综合认知程度影响，而用户对微博的综合认知程度受多种复杂因素影响，用户转发兴趣及背景热点话题也是影响因素之一。兴趣作为用户的情绪反映，是用户个体行为一种相对稳定表现的形式。而在做转发决策时，个人兴趣、微博内容及当前发生的热点话题共同影响着用户转发决策。

通过分析得出结论：用户兴趣与热点话题越相关，越容易在热点话题期间进行大量的转发，我们提出融合热点话题的用户转发兴趣特征，来计算用户转发兴趣与热点话题的匹配程度。

**热点话题内容表示：**背景热点话题文档级别的表示，根据表示背景热点话题的文档集合，提取关键词，获得背景热点话题词语级别的表示。

**用户转发兴趣表示：**通过用户历史转发微博内容来表示用户转发兴趣。用户历史转发微博表示为,对用户微博进行分词，去除停用词后，形成用户转发兴趣的词语级别的表示：。

**融合热点话题的用户转发兴趣特征计算：**定义为用户转发兴趣与背景热点的匹配程度，计算方法如下：

我们计算了转发微博数量对应用户转发兴趣特征的变化关系。如图3所示，横坐标表示融合热点话题的用户转发兴趣，纵坐标表示与背景热点话题相关的微博转发量。该图反映了用户转发行为与融合背景热点话题的用户转发兴趣之间的关系。由图中可以看出，用户转发兴趣与背景热点话题之间的匹配程度越高（越大），则用户所转发与背景热点话题相关微博的数量越多，表明融合背景热点话题的用户转发兴趣能够有效预测用户的转发行为。

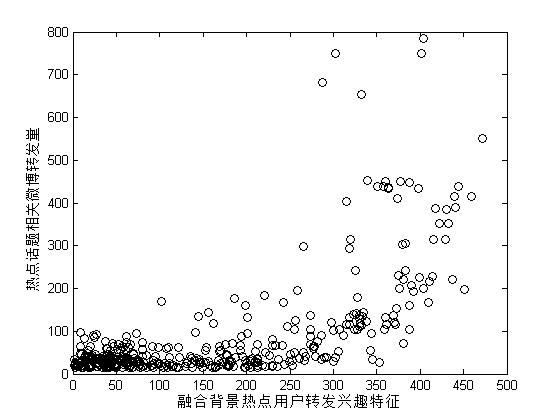


图 1微博转发量对应融合背景热点话题的用户转发兴趣的关系图

### 2.2融合热点话题的微博转发预测研究

#### 2.2.1工作概述

在微博平台中，用户之间通过关注关系构成错综复杂的网络结构，用户通过转发微博传播信息，这种传播方式具有传播快、覆盖广的特点，使得某些微博能够在短时间内形成极大的关注和影响。因此，微博转发研究对话题检测、热点跟踪、舆情监控以及商业营销具有重要价值。

目前针对微博转发的研究主要基于网络结构或基于微博特征。事实上，用户阅读到一条微博时，会根据自己已有知识对微博价值和新颖性进行判断，然后决定是否转发。微博是否会被转发与用户个体行为和用户对微博的背景知识具有紧密相关性，用户所掌握的微博背景知识一方面由历史微博获取，一方面由用户对微博内容的综合认知程度决定，而用户对微博内容的综合认知程度受多种复杂因素影响，社会上发生的热点话题信息是其中重要的影响因素。

前阶段以此为出发点，研究融合背景热点话题的用户转发行为预测方法。提出融合热点话题的微博转发预测方法，对背景热点话题内容和传播趋势对用户转发行为的影响进行量化分析，提出了融合背景热点信息的转发兴趣、转发活跃度、行为模式等特征，并基于分类算法建立了面向热点话题相关微博的转发预测模型。在真实数据上的实验结果表明，本文方法的预测准确性达到96.6%，提升幅度最高达到12.14%。

#### 2.2.2数据集准备

前阶段的研究首先基于百度新闻搜索获取热点话题数据，作为背景热点话题内容。提取热点话题关键词，利用新浪微博检索功能，获得用户1725个，以及用户在2014年5月12日至2014年8月13日期间转发的微博，共计1,210,810条，并对用户和微博之间的转发和非转发关系进行标注。我们按时间将数据集分为训练和测试两个部分，2014年5月12日-2014年7月12日之间作为训练数据，共计895,552条，其中正样例682,324条，负样例213,228条；2014年7月13日-2014年8月13日之间作为测试数据共计315,258条，其中正样例209,999条，负样例105,259条。

在微博数据集的基础上构造矩阵：

其中：，。标记后的数据集元素为一个三元组，当时表示用户转发微博，当时表示用户没有转发微博。

#### 2.2.3数据分析

##### 2.2.3.1热点话题对用户转发的影响研究

以“世界杯”热点话题为例，我们爬取新浪微博2014年4月12日至9月13日期间数据，统计微博总量变化趋势和话题相关微博总量的变化趋势。如图2所示，热点话题期间微博空间的微博总量和话题相关的微博总量都呈现出明显的增长趋势，表明用户转发行为会受到热点话题的影响。

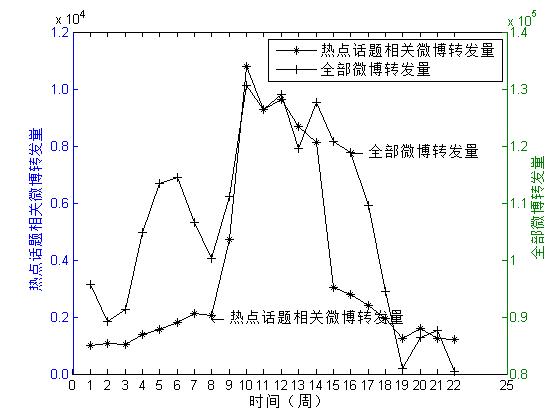


图 2热点话题期间微博转发量变化趋势

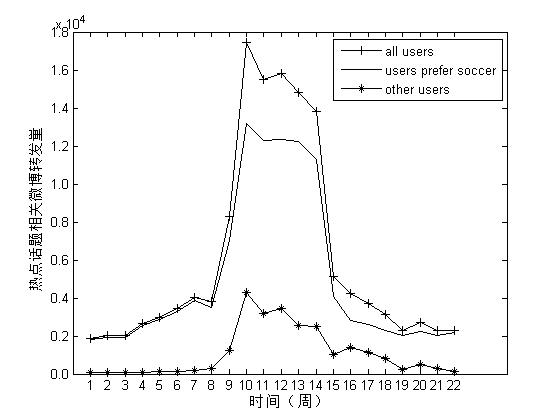


图 3热点话题对各类用户微博转发量的影响

结论：在话题传播周期内，热点话题对各类用户的转发量都有明显的提升，与热点话题相关用户的转发量提升幅度较大。同时，历史上与热点话题不相关的用户也在热点话题期间增加了对热点话题相关微博的转发量，表明仅基于用户历史微博计算的用户兴趣难以有效预测在新的热点话题下的用户转发行为，用户的转发行为会受到当前社会热点话题的影响。

##### 2.2.3.2融合热点话题的用户转发兴趣特征

我们计算了转发微博数量对应用户转发兴趣特征的变化关系。如图3所示，横坐标表示融合热点话题的用户转发兴趣，纵坐标表示与背景热点话题相关的微博转发量。该图反映了用户转发行为与融合背景热点话题的用户转发兴趣之间的关系。由图中可以看出，用户转发兴趣与背景热点话题之间的匹配程度越高（越大），则用户所转发与背景热点话题相关微博的数量越多，表明融合背景热点话题的用户转发兴趣能够有效预测用户的转发行为。

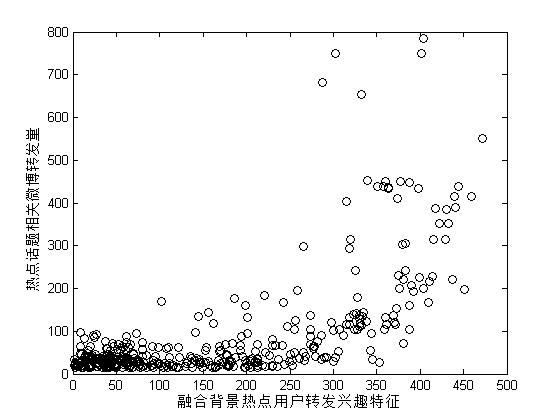


图 4微博转发量对应融合背景热点话题的用户转发兴趣的关系图

##### 2.2.3.3融合热点话题的用户转发行为特征

一段时间内用户转发微博数量的变化趋势可以看做是时间轴上的一个概率分布；我们以一定时间内新闻报道数量变化来衡量背景热点话题的热度变化趋势，也可以看做是时间轴上的一个概率分布。我们通过计算两个分布之间的相似度来计算用户行为与热点话题传播趋势的一致性特征。我们用*KL*距离来反映概率分布和概率分布之间的关系。

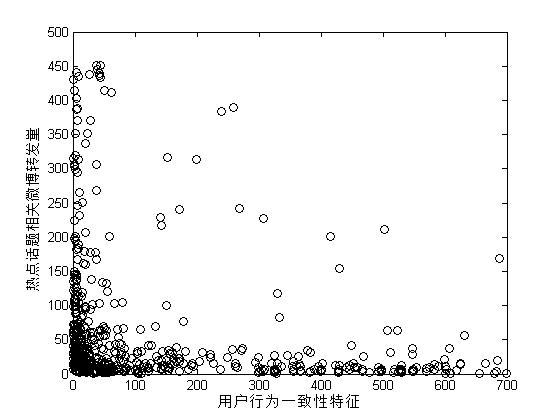


图 5微博转发量对应行为一致性特征的关系图

如图5所示，当*KL*距离增大时，分布之间差异变大，说明用户转发行为与背景热点话题发展趋势之间相关性较小，转发行为具有随机性和非持续性，用户转发的与背景热点话题相关的微博较少；当*KL*距离减小时，分布差异较小，说明用户转发行为与背景热点话题发展趋势之间相关性较大，转发行为和热点话题趋势具有较高的一致性，用户对该热点话题进行了持续关注，用户转发与背景热点话题相关微博较多。这说明，持续关注某一背景热点话题的微博用户对该背景热点话题具有更高的转发兴趣，融合背景热点话题的用户行为一致性特征能够有效检测出热点话题的持续关注用户，同时避免因其总发帖量不高而被忽略的问题。

##### 2.2.3.4融合热点话题的微博内容特征

我们对融合热点话题的微博内容特征值不同的微博获得的转发总量及平均值进行了统计分析。如图5所示,横坐标为值，纵坐标分别对应微博所获得的平均转发量和转发总量。从图中可以看出，以右侧坐标轴为标示的绿色曲线表明微博转发总量随微博内容与背景热点话题相似性的增大而减少，这是因为大部分微博内容简短，所含内容特征较少，高相似性的微博数量大量减少，导致转发总量降低。以左侧坐标轴为标示的蓝色曲线表明，微博获得的平均转发量随微博内容与背景热点话题相似性的增大而提高，表明微博内容与热点话题越相似，越容易受到转发，融合热点话题的微博内容特征能够有效区分微博转发行为。

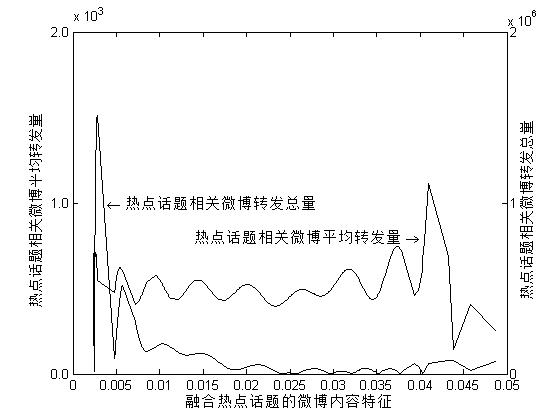


图 6 微博转发对应融合热点话题的微博内容特征对的关系图

#### 2.2.4特征选择

分类特征选择本文提出了融合热点话题的用户转发兴趣特征、用户活跃度、用户行为一致性、微博内容特征，如表1所示特征1到4。

本文采用李英乐[4]等人的方法作为对比试验，该方法使用特征较全面且预测准确性较高，在特征可计算条件下，选择用户影响力、用户转发活跃度、用户发布活跃度、用户兴趣与微博相似度作为对比特征，如表1所示特征5到8。

表 1特征选取

|  |  |
| --- | --- |
| **序号** | **特征含义** |
| 1 | 融合热点话题的用户转发兴趣特征 |
| 2 | 融合热点话题的用户活跃度 |
| 3 | 融合热点话题的用户行为一致性 |
| 4 | 融合热点话题的微博内容特征 |
| 5 | 发布用户影响力（用户粉丝数） |
| 6 | 用户转发活跃度 |
| 7 | 用户发布活跃度 |
| 8 | 用户兴趣与微博相似度 |

#### 2.2.5实验结果及分析

本文采用SVM、朴素贝叶斯、贝叶斯信念网络、决策树等分类算法，来测试所选分类特征应用于转发预测时的效果。

评价方法采用准确率（Precision）、召回率（Recall）和综合评价指标(F-Measure)。

表2实验结果对比

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| classifiers | Precision | | | Recall | | | F-Measure | | |
| baseline | Ours | Combine | baseline | Ours | Combine | baseline | Ours | Combine |
| BayesNet | 0.877 | 0.938 | **0.94** | 0.867 | 0.937 | **0.939** | 0.869 | 0.936 | **0.938** |
| NaiveBayes | 0.802 | 0.756 | **0.813** | 0.736 | 0.714 | **0.784** | 0.74 | 0.643 | **0.789** |
| C4.5 | 0.886 | 0.951 | **0.953** | 0.883 | 0.948 | **0.951** | 0.884 | 0.947 | **0.966** |
| LibSvm | 0.839 | 0.942 | **0.947** | 0.841 | 0.938 | **0.943** | 0.84 | 0.936 | **0.942** |

我们将特征分为三组，分别在四种分类器上进行了对比试验。Baseline方法是对比的基准方法，仅采用了用户和微博自身特征，Ours表示本文所提特征，即融合了热点话题的转发特征，Combine表示将用户和微博自身特征与融合热点话题的转发特征相结合，进行了综合测试。如表2所示，在多个分类器上的测试结果表明本文所提出的融合热点话题的用户转发兴趣特征、用户活跃度、用户行为一致性、微博内容特征能够有效提升转发预测准确性，与传统用户和微博自身特征相结合后，能够进一步提升效果，其中，采用C4.5分类器时的预测效果最好，达到96.6%，对基于SVM分类器的预测模型提升效果最高，达到14.12%，采用NaiveBayes分类器的实验中，ours的实验效果略差于baseline，这是因为我们选取的特征不完全满足朴素贝叶斯的条件独立性假设，导致其在分类准确率上有一定的牺牲，但是从数据上可以看出，本文所提特征在与baseline特征结合后能够提升分类准确率。引入外部热点话题并融合其内容和传播趋势对用户转发行为的影响因素，能够有效提升转发行为的预测准确性。本文的训练集和测试集按照时间先后相互独立，预测准确性的提高也表明了本文所提特征能够很好的刻画用户转发行为模式，具有较好的长期预测效果。

# 关键技术或难点

## 融合热点话题的用户转发兴趣特征

* 背景热点话题对用户转发兴趣的影响I
* 融合背景热点的用户转发兴趣特征的提取

## 微博转发预测模型构建

* 确定热点话题对用户转发行为的影响
* 融合背景热点话题的用户转发行为特征的提取
* 融合背景热点话题的微博内容特征
* 基本预测模型构建
* 同一数据集上多种方法的比较

# 下一阶段工作计划

* 针对背景热点话题与微博内容之间做更多语义相关的工作，以更好地分析微博内容与背景热点的相关性
* 考虑用户兴趣时更多地基于时间序列考虑微博内容特征

## 存在的问题

以用户历史微博的词序列表示用户兴趣和热点相关的微博，存在数据稀疏和用户兴趣不聚合的问题。数据稀疏是由于微博文本短、用词不规范造成的；用户兴趣不聚合是因为词序列表示用户兴趣这种方式并不能聚合文本。

## 尚未完成的工作

* 多通道背景热点的表示和提取方法
* 基于背景热点的微博转发预测系统设计与实现方案

## 解决问题的技术思路或措施

### 4.3.1多通道背景热点的表示和提取方法

**1. 基于多级过滤的候选热点词提取**

本文拟采用中科院自动分词系统（ICTCLAS）先对语料进行分词然后对切分词进行多级过滤：

1. **停用词过滤**

载入一个常用停用词表，用该词表过滤常见的无意义词。该词表中包括130多个英文停用词，500多个中文停用词，以及一些标点符号、数字等特殊符号，共计800余条。

1. **词性过滤**

对于特殊词性的词, 如数词、量词、代词、方位词、叹词等, 它们对于热点词的提取贡献不大, 如“一”、“条”、 “中间”、 “吗”等, 通过编写相应的规则, 判断其词性并进行过滤。

1. **噪音词过滤**

噪音词是语料中普遍存在的与主题无关的无意义词。为了找出噪音词，需要一些标准用语估计词的有效性。本文拟采用联合方差的概念，基于在语料库中出现的频数与包含该词的文档频数分别计算其在语料中发生的概率。联合方差的计算公式如下：

其中， , ,tf为词频，Ztf为全部词条的词频总和，为Ttf的均值，sum表示总的词条数。 , ,df为文档数，Zdf为全部文档数总和，为Tdf的均值，count表示总的文档数。

根据联合方差计算公式进行单个词的联合方差计算，依据联合方差的稳定性选取噪音词。在时间标签的基础上，联合方差越小，变化越小，说明该词的稳定性越好，这个词无意义的概率大，越容易被认为是噪音词。

**2. 热点词提取**

用来进行热点词提取的背景热点相关新闻经过多级过滤后构成热点词集，在多级过滤的时候，把停用词表和实验提取的噪音词表加载到热点词提取算法中，可以把该类词表称为非热点词表，当发现该词的时候将其过滤掉，最后形成热点词集。热点词的特征包括词频（TF）、文档频数（DF）、反文档频率（IDF）。对于热点词的权值计算采用如下公式：

其中，为词的频率；表示在整体样本数据中出现词i的文档数目；M表示所有文档数目；|d|表示文档向量的长度；用来对命名实体赋予不同权重。

根据热点词提取的结果进行排序，确定其是否为热点词。热点词提取的结果进行排序，确定其是否为热点词。按测试语料中的词频、文档频数和权重总结出热点词的加权综合判断公式，更具HK的值进行排序。

综合文献[[[1]](#endnote-1)]的排序方法，得到如下公式：

其中， ，设定为经验值。

### 4.3.2基于背景热点的微博转发预测系统设计与实现方案

## 下一阶段计划

* 基于LDA的短文本用户兴趣提取
* 短文本语义分析与表示

# 主要参考文献

主要参考文献。

1. 1. <http://baike.baidu.com/link?url=Qsdt8nZWb5Q_iTpNaS41WlK2ZxMJeaUC8g9cuHWpK2V01Grlj6wiUx7C4170CTm2988GAfKuQoMHuWdmq1V65C0zVgKyuU1qMYlZ44yMBe_>
   2. Petrovic S, Osborne M, Lavrenko V. RT to Win! Predicting Message Propagation in Twitter[C]//ICWSM. 2011.
   3. Galuba W, Aberer K, Chakraborty D, et al. Outtweeting the twitterers-predicting information cascades in microblogs[C]//Proceedings
   4. of the 3rd conference on Online social networks. 2010, 39(12): 3âAS3.
   5. 李英乐, 于洪涛, 刘力雄. 基于SVM的微博转发规模预测方法[J]. 计算机应用研究, 2013, 30(9): 2594-2597.
   6. 曹玖新, 吴江林, 石伟, 等. 新浪微博网信息传播分析与预测[J]. 计算机学报, 2014, 37(4): 779-790.
   7. Kanavos A, Perikos I, Vikatos P, et al. Modeling ReTweet Diffusion Using Emotional Content[M]//Artificial Intelligence Applications and Innovations. Springer Berlin Heidelberg, 2014: 101-110.
   8. Ma H, Qian W, Xia F, et al. Towards modeling popularity of microblogs[J]. Frontiers of Computer Science Selected Publications from Chinese Universities, 2013, 7(2):171-184.
   9. Ying-Le L I, Hong-Tao Y U, Liu L X. Predict algorithm of micro-blog retweet scale based on SVM[J]. Application Research of Computers, 2013, 30(9):2594-2597.
   10. Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks[J]. Phys.rev.e, 2001, 63(6):138-158.
   11. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks.[J]. Physical Review Letters, 2001, 86(14):3200-3203.
   12. Boyd D, Golder S, Lotan G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter[C]// System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE, 2010:1 - 10.
   13. Suh B, Hong L, Pirolli P, et al. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network[C]// Social Computing (SocialCom), 2010 IEEE Second International Conference on. IEEE, 2010:177-184.
   14. Yang Z, Guo J, Cai K, et al. Understanding retweeting behaviors in social networks[J]. Cikm ’, 2010:1633-1636.
   15. Jiang Y, Counts S. Predicting the Speed, Scale, and Range of Information Diffusion in Twitter[J]. Fourth International Aaai Conference on Weblogs & Social Media, 2010.
   16. Hong L, Dan O, Davison B D. Predicting Popular Messages in Twitter[J]. Www68, 2011.
   17. Zhang Y, Rong L U, Yang Q. Predicting Retweeting in Microblogs[J]. Journal of Chinese Information Processing, 2012, 26(4):109-108.
   18. Bandari R, Asur S, Huberman B A. The Pulse of News in Social Media: Forecasting Popularity[J]. Sixth International Aaai Conference on Weblogs & Social Media, 2012.
   19. Ma Z, Sun A, Cong G. On predicting the popularity of newly emerging hashtags in twitter[J]. Journal of the American Society for Information Science and Technology, 2013, 64(7): 1399-1410.

   Szabo G, Huberman B A. Predicting the popularity of online content[J]. Communications of the ACM, 2010, 53(8): 80-88.

   1. Yang J, Leskovec J. Modeling information diffusion in implicit networks[C]//Data Mining (ICDM), 2010 IEEE 10th International Conference on. IEEE, 2010: 599-608.
   2. <http://baike.baidu.com/subview/45281/8045345.htm#viewPageContent>
   3. <http://baike.baidu.com/link?url=uSpPDwuklxXEAOvRxbkQCpT_pGlLfjTt1dbrHfKS5Iz4Kq8UfNXRmuujk1A77KpPpSLfj8fVbwf5qYcUD2i1a>
   4. 融合热点话题的微博转发预测研究

   [↑](#endnote-ref-1)