基于SVM的评论垃圾识别研究

摘要

目前商务网站允许用户发表针对产品的一些评论，但其中难免会存在一些垃圾评论，极大地误导了商家和用户辨识信息的真伪。因此评论垃圾识别越来越成为一个值得关注的具有社会价值和应用价值的热点问题。

对于垃圾评论的识别，我们利用可指导的**支持向量机模型（SVM）**，使用评论内容和用户行为两个角度出发构建精确的分类器标注集，最终从评论集合中识别出垃圾评论。

对于问题一，我们仅考虑从产品本身出发的评论内容，支持向量机模型的分类器的构建主要考虑了评论内容的特征。首先利用stanford parser 软件对评论内容的词性进行分析，然后结合评价句的路径匹配模板进行评价语句的提取。然后利用Apriori 信息挖掘算法对频繁出现的名词或非名词短语进行分析，得到能够代表产品特性的**特征词**。之后利用特征词和**情感词**之间的关联寻找情感词。为了确定情感词的感情倾向,在Hownet 词库的基础上，我们建立正面情感词和反面情感词的近义词和反义词词库，通过情感词与词库中的词汇的对比，即可确定情感词的感情倾向，最终计算得到评价语句的感情倾向。在模型执行的每一步中，我们都选取了调研到的资料对所提出的模型进行合理性验证，之后对题目给定的垃圾评论进行检测，得到的结果也较好地趋近于真实情况。最后综合考虑评价语句的所占比例，在问题一部分主要考虑两两指标组合：特征词个数和交互信息，正面和反面评价句比例以及文本的结构，利用LibSVM 对垃圾评论识别系统进行训练和测试，最终得到:综合考虑各种因素的影响时，垃圾评论识别的准确率为0.76，召回率分别为0.69，模型在此条件下比仅考虑单个指标影响时的识别效果好很多。

对于问题二，我们充分考虑了**用户信用度**（包括买家信誉与评价是否匿名）所带来的影响，在此基础上，模型针对调研到的iPhone6产品评价进行了分析，并进行综合指标考量与问题一的结果进行对比，发现支持向量机模型的分类器的构建在同时考虑了信誉等级和评论匿名的情况下，垃圾评论的识别更加全面准确。最终垃圾评论识别的准确率为0.78，召回率为0.77；只考虑信誉等级和匿名时，准确率为0.56，召回率为0.61。

对于问题三，我们从用户行为的角度出发，结合前两问的模型在一般的产品评价集合中识别垃圾评论。考虑到评论内容对垃圾评论识别的不通用性，支持向量机模型的分类器的构建主要考虑了评论者行为对垃圾评论识别的影响，包括用户评分行为、用户评论文本的相似性、用户评分偏差、用户互动行为和用户购买行为等因素，最终给出了一个可执行性较好的模型并给出了基于本文模型对评论识别问题的看法与期望。

整体来看，我们认为在研究方法上,识别评论垃圾的关键是提取表征评论垃圾的特征。而另一方面通过模型的分析看到，评论内容和评论者的特征同等重要, 融合两方面特征的识别模型具有更优的效果。

关键词： *支持向量机模型 Apriori算法 评论内容 评论者行为*

目录

1 引言 **1**

2 问题分析 **2**

3 模型准备 **2**

3.1 评论垃圾的定义与分类 2

3.2 召回率与准确率 3

4 SVM模型建立 **4**

5 问题一：基于评论内容的SVM模型 **5**

5.1 模型建立 5

5.1.1 评价语句的检测与提取 5

5.1.2 典型特征的提取 8

5.1.3 情感词提取 9

5.1.4 感情词汇的倾向识别 9

5.1.5 非典型特性的提取 11

5.1.6 评论语句的倾向预测 12

5.1.7 评论文本结构特征 13

5.2 模型求解 13

6 问题二：基于用户信誉度的SVM模型 **15**

7 问题三：基于用户行为的SVM模型 **17**

7.1 模型建立 17

7.1.1 基于目标产品的垃圾评论者检测模型 17

7.1.2 基于用户评分偏差行为的垃圾评论者检测模型 18

7.1.3 基于用户互动行为的垃圾评论者检测模型 18

7.1.4 基于用户购买行为的垃圾评论者模型 19

7.2 对于识别问题的看法 19

8 模型评价 **20**

1 引言

随着互联网公的发展，公民自由言论、发表意见的权利得到很大体现，但是不真实、或是无效的评论给参考评论的用户带来很大困扰。就网购评论而言，购买商品或消费前，用户往往会查看相关评论信息。如果评价积极，消费者的购买意向可能就大。因而随着网络应用的不断深入，在线“网络口碑”对商品销量及商家名誉的影响力越来越大。某些组织或个人在各种利益的驱动下开始利用网络信息监管的缺失，弄虚作假，制造评论垃圾混淆视听误导用户。清除网络垃圾，净化网络环境，为人们提供一个真实可信的信息获取平台的需求日益迫切。

而对于评论垃圾的判断，我们需要从多个角度来进行分析，提到的对于评论垃圾特征选择与识别方法主要从以下几个角度进行：

表 1: 评论垃圾识别的特征选择和识别方法[1]

相关文献	特征选择	说明	方法
文献[2]	评论特征	内容特征（Content） 情感特征（Sentiment）	Naive Bayes NB-Booster-apping
	评论人特征	产品特征(Product) 评论人的个人特征(Profile) 评论人行为特征（Behavior）	
文献[3]	评论内容与语言特征	词性（POS） 基于LIWC的特征 n-gram特征	Naive Bayes SVM
文献[4]	评论特征 评论人特征 产品特征	评论长度、文本内容相似度、时间等 行为特征（例如发表评论数） 产品的价格、销售率等	Logistic Regression
文献[5]	评论内容 评论人	n-gram特征 基于LIWC的特征 语法特征（Deep Syntax） 词性特征（POS） 评论人行为特征（Spamming Behavior）	SVM
文献[6]	评论人	评论的内容相似度、最大评论数等 评论偏差、评论的重复率等	聚类

同时，在垃圾评论方面也有许多相关学者做出一定的研究。Jindal等[4]在2008年首先定义三种类型的垃圾评论，即不真实的评论(类型 1) 、无关评论(类型 2) 和非评论(类型 3)，之后人工标注部分垃圾评论，以评论、评论者和被评论的商品三个方面的24个特征作为基本特征，使用Logistic回归构建机器学习模型，识别类型1和类型3的垃圾评论，使用Shingle 算法识别重复的评论，并使用识别出的重复评论作为训练集构建机器学习的模型来识别类型1的垃圾评论他们发现使用重复评论作为训练集会遗漏一部分非重复的垃圾评论，于是在2010年他们分析了用户的打分模式[6]，并通过挖掘用户的行为，发现反常的评论模式来分析用户是垃圾评论发表者的可能性。Mukherjee 等[5]在2010年使用三个步骤来检测群体垃圾评论，首先使用Frequent Pattern Mining找出候选的群体，之后计算垃圾信息的指示值，最后使用支持向量机（Support Vector Machines, SVM）算法进行排名，从而检测出群体垃圾评论。Wu等[8]利用正向的Single-tons（评论发表者发表的唯一的一条评论）在一个产品的所有评论中所占的比例和这些

Singletons时间聚集程度来分析评论发表者的可疑行为。何海江[8]在2009 年提出了一种用来衡量评论与文章之间的语义相关程度的向量空间模型cVSM作为评论的特征，采用支持向量机分类算法自动识别垃圾评论，能显著地提高垃圾评论的识别能力之后何海江[9]等在评论识别时，采用基于 Logistic回归的分类器来区分合法评论和垃圾评论，并与支持向量机SVM的性能进行对比。Bhattarai等[10] 在2009年研究了博客垃圾评论的特征，并利用Co-traning思想从已给的数据中主动学习的方法来解决对识别不好或是无法识别的评论的问题。

目前，垃圾评论的识别方法大都建立在文本分类的思想上，将其识别过程视为垃圾评论和非垃圾评论的二分类其具体过程是: 首先构建标准数据集，选择数据集中部分数据作为训练样本，然后针对要分类的对象，找出特征，最后运用Logistic回归、贝叶斯、条件随机场、SVM等算法进行分类通过对之前学者研究的学习，我们最终选择基于SVM算法建立的数学模型来完成对垃圾评论信息的识别研究。

2 问题分析

问题一要求建立合理的数学模型对给定的评论进行识别，并给出算法流程，通过程序验证，给出正确识别率。首先，由于评价内容往往受到多方面因素的影响，我们需要建立一个考虑多指标的数学模型，包括用户评论的内容长度，特征词，情感倾向等等。在算法流程环节，我们需要对每个流程所对应的指标在模型中的可行性进行判断，必须使得每个指标在考虑其单个因素的情况下可以对垃圾信息进行有效检测。然而，单个指标的判断能力往往有限，还应进一步判断多指标结合所对应的检测效果，理论上讲，综合指标的效果应优于单个指标。

问题二要求收集一个更大的关于某件产品的评价集合，建立数学模型和算法进行识别。由于评价集合度提高，评论数量增加，我们在考虑仅由内容决定的评价之外，还应考虑评价者的消费水平，购买信誉以及评论真实度等因素。首先对评价者的属性对评论质量进行判断，之后集合问题一中提出的指标建立综合分析模型。

问题三要求建立更一般的模型，因此我们需要考虑更多的客观因素，例如用户是否有较大的评分偏差，其购买行为产生的影响等，之后在前两问建立的模型基础之上建立一个较为全面、更具普适性的模型。

3 模型准备

在模型建立之前，我们将对部分概念进行定义，并对模型的评价参数进行介绍。

3.1 评论垃圾的定义与分类

对其他消费者产生消极影响或者不产生任何参考价值的评论可以被归为评论垃圾，本文我们将评论垃圾分为两类产生来源，分别从基于产品本身和基于用户行为两方面进行分类。首先，题目所给出的例子大致为从**评价内容**的角度，评论大体可分为：

- 评论并未针对产品，而是关注到品牌、厂商等其他方面
- 评论并未针对被评论的产品，而是谈论其他产品
- 评论无实质性内容，例如变相投放销售广告，无意义的字符组合等
- 评论为无关文本，例如人身攻击等无效评论

而从**用户行为**的角度来看，评论大体可分为：

- 用户的评分行为与评论文本
  若同一用户对同一产品的评分次数越多，且该用户的评分基本上类似，或其评论次数越多，且评论文本相同或者非常相似，则用户可能为垃圾评论者。
- 用户的评分偏差行为
  垃圾评论者为了提升某一个产品或诋毁某一个产品，其与其他人的评分可能会有很大的不同。即如果同一件产品评分较低，但总有一部分评论评分极高，这就极有可能是商家为了销售产品而采取的虚假评论手段。
- 用户互动行为
  一般来说，如果一个用户发表的垃圾评论较多，其评论的平均回复数会比较少，因此，使用用户的平均回复数作为衡量垃圾评论行为得分的指标。
- 用户的购买行为
  当用户对所评论的产品一次都没有购买过，或者只购买了其中很小一部分的时候，该用户为垃圾评论者的可能性就会很大。

3.2 召回率与准确率

在模型建立过程中我们需要对模型的训练结果不断进行调整使其逐渐趋近于真实情况，因此需要一些指标来对模型执行能力进行衡量，而信息检索、分类、识别、翻译等领域两个最基本指标是召回率(Recall Rate) 和准确率(Precision Rate)，召回率也叫查全率，准确率也叫查准率，概念公式:

召回率(Recall)=系统识别到的垃圾评论 / 系统所有垃圾评论总数
准确率(Precision)=系统识别到的垃圾评论 / 系统所有识别到的垃圾评论总数

我们将这样的关系用图1表示：

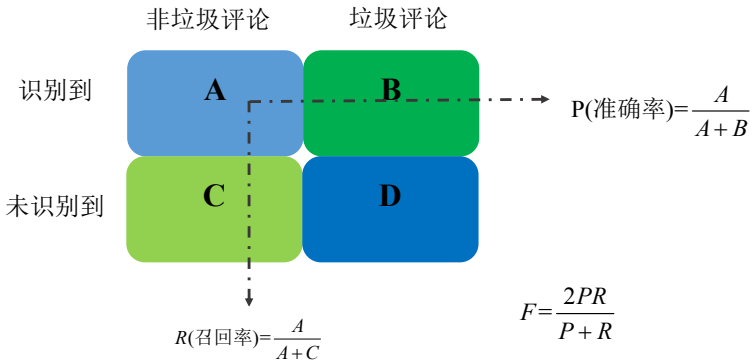


图 1: 准确率、召回率关系图

值得注意的是准确率和召回率是互相影响的，理想情况下肯定是做到两者都高，但是一般情况下准确率高、召回率就低，反之召回率低、准确率高。为了权衡二者的影响，我们引入了F来综合模型的有效性。

$$F = \frac{2PR}{P+R} \tag{1}$$

其中P, R分别为准确率与召回率。

4 SVM模型建立

首先，我们对SVM模型进行简单的介绍，SVM（Support Vector Machine，支持向量机）是一种有监督的机器学习方法，可以学习不同类别的已知样本的特点，进而对未知的样本进行预测。SVM本质上是一个二分类的算法，对于n 维空间的输入样本，它寻找一个最优的分类超平面，使得两类样本在这个超平面下可以获得最好的分类效果。这个最优可以用两类样本中与这个超平面距离最近的点的距离来衡量，称为边缘距离，边缘距离越大，两类样本分得越开，SVM就是寻找最大边缘距离的超平面，这个可以通过求解一个以超平面参数为求解变量的优化问题获得解决。给定适当的约束条件，这是一个二次优化问题，可以通过用KKT条件求解对偶问题等方法进行求解。

对于不是线性可分的问题，就不能通过寻找最优分类超平面进行分类，SVM 这时通过把n维空间的样本映射到更高维的空间中，使得在高维的空间上样本是线性可分的。在实际的算法中，SVM不需要真正地进行样本点的映射，因为算法中涉及到的高维空间的计算总是以内积的形式出现，而高维空间的内积可以通过在原本n维空间中求内积然后再进行一个变换得到，这里计算两个向量在隐式地映射到高维空间的内积的函数就叫做核函数。SVM根据问题性质和数据规模的不同可以选择不同的核函数。

虽然SVM本质上是二分类的分类器，但是可以扩展成多分类的分类器，常见的方法有一对多（one-versus-rest）和一对一（one-versus-one）。在一对多方法中，训练时依次把k类样本中的某个类别归为一类，其它剩下的归为另一类，使用二分类的SVM训练处一个二分类器，最后把得到的k个二分类器组成k分类器。对未知样本分类时，分别用这k个二分类器进行分类，将分类结果中出现最多的那个类别作为最终的分类结果。而一对一方法中，训练时对于任意两类样本都会训练一个二分类器，最终得到k*(k-1)/2个二分类器，共同组成k分类器。对未知样本分类时，使用所有的$k*(k-1)/2$ 个分类器进行分类，将出现最多的那个类别作为该样本最终的分类结果。本文主要使用SVM的将评论分为垃圾评论和非垃圾评论，分类工具采用了LibSVM，核函数采用径向基函数函数，训练时控制SVM对输入量变化的敏感程度γ的值，最终确定γ的值为0.76，其余参数再用默认值，SVM模型应用主体思想如图2所示：

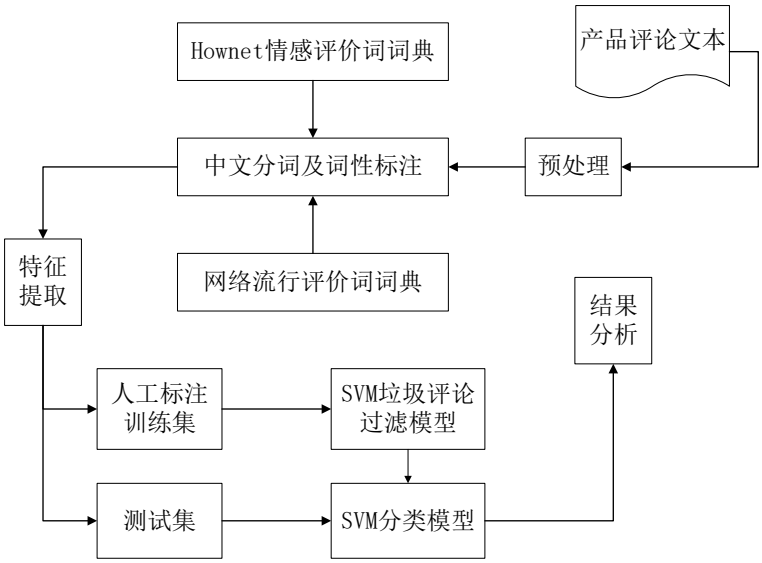


图 2: 研究方法示意图

我们从评论的不同方面选取不同的特征，将这些特征作为SVM模型的输入项进行训练和测试，提取出垃圾评论。 对于评论不同特征的提取，我们将在下面一一介绍。

5 问题一：基于评论内容的SVM模型

5.1 模型建立

我们将从评价句数量，主题词，情感倾向，文本结构，以及评价者属性五个方面对评论垃圾进行判断，其中对应的指标及含义如表2所示：

表 2: 特征分类及其描述和依据[11]

特征分类	特征描述	特征依据
评价句数量	F_1：$\sum Os\,(r_i)$ F_2：$\sum O\,(r_i)/s$	数量越大、占比越高，评论质量就可能越好
主题词	F_3：$PMI\,(S,T_k)$ F_4：主题词个数K	PMI值越小和主题越不相关，垃圾评论概率大
情感倾向	F_5：表达正面情感的短句所占比例 F_6：表达负面情感的短句所占比例	过分褒贬的评论，很可能就是垃圾评论
文本结构	F_7：$\sum l\,(r_i)$ F_8：评论短句个数s F_9：$\sum l\,(r_i)/s$ F_{10}：σ_l F_{11}：$l_c\,(r_i)/s$	中文垃圾评论中，存在较多的非汉字字符，句子长短不一，及评价特征词重复等情况
评价者属性	F_{12}：是否为匿名 F_{13}：信誉、经验或信用等级	信用等级越高，其评论可信度和质量可能越高

5.1.1 评价语句的检测与提取

在给定评论中，产品特征通常是一个名词或者名词短语，因此词性标注显得很有意义。我们用语言分析器 [14]将整句话分为几个小单元，每个小单元即可有对应的此行标注（可能是名词，动词，形容词等等）。这个过程也区分了简单名词和动词词组（语法组块）。我们将词性标注展示在表3中

表 3: 句法分析树标注集

标注	词性	标注	词性	标注	词性
NN	常用名词	NR	固有名词	NT	时间名词
PN	代词	VV	动词	VC	是
CC	不是	VE	有	VA	表语形容词
AS	内容标记（如：了）	VRD	动补复合词	AD	副词

因多数产品评论一般由多个短句组成,断句较随意。因此本文将产品评论分为多个短句进行分析, 为了方便说明首先给出以下相关定义：

定义1： 产品特征词指与评论产品特征相关的词。

定义2： 评价句指构成产品评论文本每个短句中, 包含产品特征或情感倾向的句子。

定义3： 词性搭配结构指评论句子经过分词和词性标注后,句子呈现各种词性按某种顺序组合的特征。

产品特征和评论观点的抽取是垃圾评论识别的两项重要任务，垃圾评论的判断与评论文本中评价句数量的多少有很大关系。因此，如何识别评论中的评价句，是本文特征提取的重点。文献[9-14]利用句法词性搭配结构来提取中文产品评论中的产品特征信息。其中，文献[12]对产品评论词性分析后，选择表达产品特征及评价最为常见的词性，并进行重要程度排序实现对产品特征及语义的抽取。经分析，若评论句子中存在产品特征词，则该句子具有评价句特征的概率很大。在实际评论中，因评论作者的简写或省略，较多评论句子中虽无产品特征词，却有表达对特征评价的a和与表达情感有关的d、v 或I,该类评论句子也应具有评价句特征。故将词性路径模板用于评价句的检测,同时为了提高分词系统对评价词的识别率,在分词系统中加人自定义评价词对评价词的识别率,在分词系统中加人自定义评价词序。最终使用表4所示的词性路径匹配模板集P,按优先级顺序提取评价句。

表 4: 词性路径匹配模板集

优先级/类别	词性路径匹配模板
1	(N)或(VV+D)或(VA)
2	(N+VV)或(VA)
3	(N+D)或(A+VV)或(N)
4	(N+VV+N)
5	(N+D)或(VA)
6	(D)或(VV)或(VA+VA)
7	(D)或(N)

其中N表示名词，为之前介绍的常用名词、固有名词及时间名词的总称。我们举一个例子来对这样的方法进行说明，我们以评价“外观精美，商家态度很好”为例通过Standford parser对评论内容进行分析并给出方法下的拆分方式，其结果如表5所示：

表 5: 词性路径模板匹配的样例

评语	外观	精美	，	商家	态度	很	好	。
词性	NN	VA	PU	NN	NN	AD	VA	PU

为说明其可行性,随机选取天猫商城中的iPhone5、iPhone5S、iPhone6和iPhone6+ 四类商品的评论,评论文本长度每类500条评论,进行人工统计和实验分析对比,结果如图3所示：

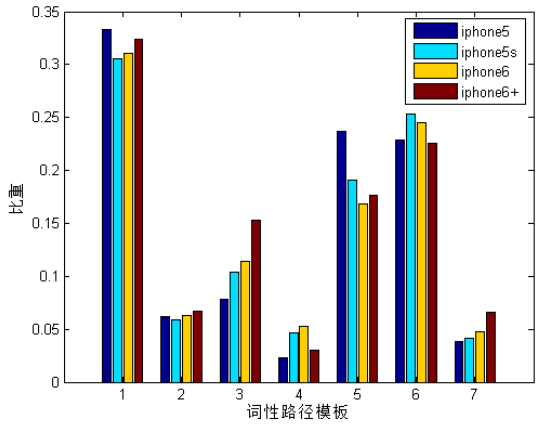


图 3: 7种词性路径模板在不同产品的评论中所占比重

由图3可看出，对于任意一个单个产品，它对应的七个词性路径模板比重之和趋近于1，这说明按照我们给定的词性路径模板基本可以提取出句子中的每个词，可以很好地概括评论语句，这也为后续的其他内容提取工作提供了保障。

在评价语句的提取方面，同样对四类产品分别500条评论内容进行分析，得到提取评价语句过程中的特征词的召回率与准确率，其结果如图4所示：

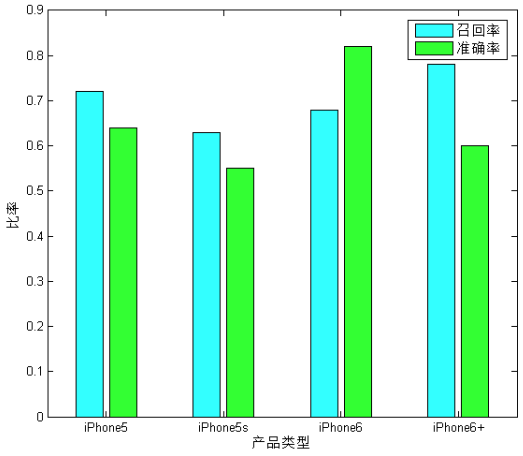


图 4: 评价语句提取后的召回率与准确率

图4说明我们给出的评价语句提取方法可以在很大程度上将评论中可用的评论语句提取出来。在验证了这样的评价语句提取方法的可行性之后，我们给出题目附件中第一类评论中的七条评论进行评价语句提取，得到评论中评价语句所占整个句子的比重如表6所示：

表 6: 7条评论中评价句所占整个评论的比重

评论	1	2	3	4	5	6	7
评价句比重	0.40	0.33	0.50	0.00	1.00	0.50	0.00

事实上，显然题目给出的评论都为与主题无关的无效评论，理论上比重越小越好，而我们所提取出的评价句比重也相对应的维持在较低水平甚至为0，这样的结果也在一定程度上验证了该方法的可行性。

5.1.2 典型特征的提取

典型特征是最能体现评论内容方向性的指标，为区分给定产品评论是否存在广告或与产品无关的信息，根据表达产品特征的词性主要为名词的特点，在评价句中，属于词性路径匹配模板前5 类的，抽取首次其出现的名词集合，作为该评论的主题词。若某评论局较长，出现多个表达产品特征的词语，如连续多个名词、副词或形容词等，则根据汉语修饰词的排序特点，在词性路径匹配模板中，提取最先匹配的名词。

假设经评价句特征提取后得到的k个主题名词集合$T_k = \{t_1, t_2, \cdots, t_k\}$，$1 \leq k \leq K$，预先设定的评论主题特征词集合$S = \{s_1, s_2, \cdots, s_m\}$，判断$T_k$与$S$是否相关，采用点态互信息（Pointwise Mutuat Information, PMI）计算方法 [12]，PMI表示两个词之间的相关性，是一种基于词出现的无监督计算方法，其公式如下：

$$PMI\left(S, T_k\right) = \log_2 \left(\frac{hit\left(SandT_k\right)}{\sqrt{hit\left(S\right) hit\left(T_k\right)}} \right) \tag{2}$$

其中S表示在搜索引擎中查询S返回的页面数，$hit\left("S" and "T_k"\right)$表示在搜索引擎中$S$和$T_k$共同作为关键字搜索返回的页面数。按照我们之前介绍的，PMI表征着评论主题特征词与主题名词之间的相关性大小，PMI值越大，说明二者相关性越大，算法得到的结果就越准确。 由于我们的目的是为了得到人们对于产品的情感倾向，所以提取人们所关注的产品特性是重要的步骤之一。然而，考虑到自然语言理解的困难性，我们很难对句型进行处理。例如，我们举一条有关iPhone6的评论：

拍出的照片很清晰。

在这个句子中，拍出的照片很清晰确实是和手机这件物品相关的有效评价，但是如果仅从手机物品本身的属性出发进行判断，这个评论似乎并无关系，因此很难进行进一步的解读。同时，也有一些隐含的内容也很难通过非人工操作来识别，例如：

手机都装不进兜里。

这位顾客想表达的是手机屏幕较大，不易携带，实际上是对手机尺寸的属性做出的评价，但是整句话没有出现尺寸和大小等特征词汇，因此只关注于名词并作出判断并不合适，有关隐含意义（非典型特征词汇）的提取我们将稍后介绍。

在特征词汇的提取中，我们首先关注于找到出现频率频率较高的典型产品特性。因此，我们选用关联挖掘法来寻找频繁项集，我们所提及的项集为由几个单词或短语简单组合成的集合。

首先，顾客对商品做出的评价有很多都并非直接对产品进行描述，由于不同的顾客有着不同的消费经历与生活背景，因此评论也不尽相同。但是，当顾客评论产品属性时，他们所使用的词汇往往是较容易汇集的。因此我们选用合适的关联挖掘法可以有效地将项集进行汇集，因为由此得到的项集很可能是产品属性。值得注意的是，那些频繁出现的非名词或非名词短语往往更可能是非商品属性。

我们通过基于Apriori算法[13]的信息挖掘法对通过上述步骤产生的非名词或非名词短语进行分析。我们得到的出现频率较高的特征都是潜在的物品特性。本文，我们规定若一个项集在整个句子中出现的频率超过1%，则可认为其为频繁。由该项集衍生出的其他项集也可以认为是可选商品特性。

我们对上文研究过的四款手机的500条评论进行典型特征提取，从评论中提取的特征词的召回率与准确率如表7所示：

表 7: 典型特征提取下的召回率与准确率

产品类型	iPhone5	iPhone5s	iPhone6	iPhone6+
召回率	0.67	0.59	0.73	0.65
准确率	0.55	0.59	0.56	0.57

由表7可看出，四款产品的评价中，召回率与准确率都分别保持在65%和55%左右，可以达到较为满意的提取水平，因此典型特征提取的模型具有较好的适用性，可以用作单独指标进行信息筛选或与其他指标进行组合来判断评论是否为垃圾评论。

表 8: 特征词数及交互信息PMI值

评论	1	2	3	4	5	6	7
特征词数	0	0	0	0	0	2	0
交互信息PMI	0.27	0.22	0.13	0.21	0.18	0.35	0.14

回顾我们之前介绍过的PIM值，它表征着变量之间的关系紧密程度，PMI越小，评论内容与被评论实物联系越小，而特征词数是评论有效内容的重要体现方面，特征次数越多说明评论内容与主题相关性越大。在题目给定的7条评论的分析中，PMI值保持较小，特征次数趋近于0，这样的结果说明通过典型特征来对垃圾评论进行判断是合理的。

5.1.3 情感词提取

评论词语是用来表达主观观点的，之前我们从主观角度出发建立了与形容词出现相关的统计方法。因此形容词的出现对于预测评论是否主观是很有价值的，例如表达一个观点。本文我们用形容词作为观点词汇，同时我们减少对包含一个或多个产品特性的评论中观点词汇的去除，因为我们仅关注顾客在商品特性方面的评价。在这个步骤，我们通过比较有效感情词与特征词的距离来进行情感词提取，具体的流程如图5

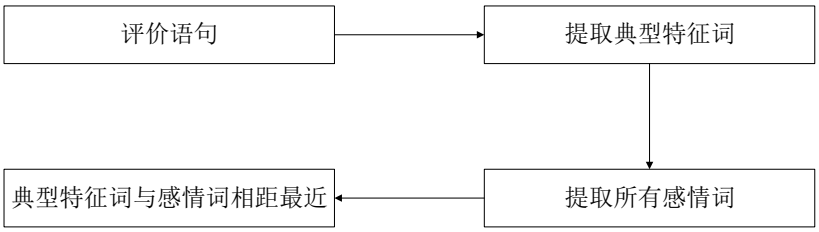


图 5: 情感词提取流程示意图

在这个步骤中，最关键的步骤是提取所有感情词并选定距离典型特征词最近的感情词来进行语义情感推断，由于这个过程是逐词分析，所以需要不断循环进行。

5.1.4 感情词汇的倾向识别

对于每个感情词汇，我们需要识别其语义的方向性，这有助于我们预测每个评价语句的倾向性。单个词语的语义倾向暗示着用户对产品的整个评价方向。例如，“非常好”“完美”等词就是较为积极的词，

用户在使用这些词来进行评价时，往往满意度较高，而“糟糕”、“失败”等词汇则为较消极的词语，用户在使用时往往是处于较差的购物体验。当然并非所有的形容词都具有明显的方向性（例如理论的，数码的等无明显感情色彩的词）。在本文的分析中，我们主要关注积极和消极两个评论词语方向。

Turney在他的研究中提到，我们可以通过给出同一属性的两极，计算给出评价中词组的交互信息值PMI，例如，我们定义一个评价最优的评价为“完美”，最差的评价为“糟糕”，那么当给定一个评价之后，我们分别计算改评价相对于“完美”和“糟糕”的PMI值，哪个值更大，我们则认为改评价更接近于哪个方面。

在其他学者们研究的基础之上，我们提出一个行之有效的方法来进行方向识别，即形容词同义词组成的次于网络来进行形容词方向识别。在词语网络中，以一个词为中心，其反义词为另一个中心，二者衍生出对应的同义词即构成了一个大型网络，那么我们可以以该网络作为数据库来进行语义推断。我们将这样的关系举例如图6表示：

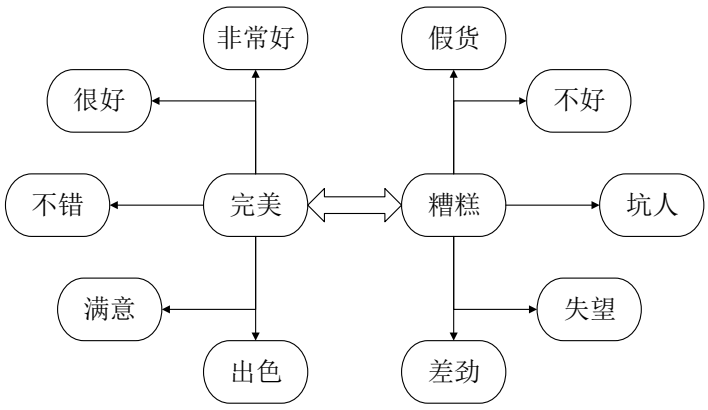


图 6: 近义词网络图示意图

那么在这个网络的基础上，我们就可以认为所有的词汇都可以看作是由一个原始词汇衍生出来的，那么再这样的前提下，给定一个词汇我们只需找到其在网络中对应的原始词汇即可对其进行语义倾向行判断。而且，随着评论的累积，词汇网络不会断得到充实，也保证了这种判断语义倾向的方法的有效性。具体的方法如图7所示：

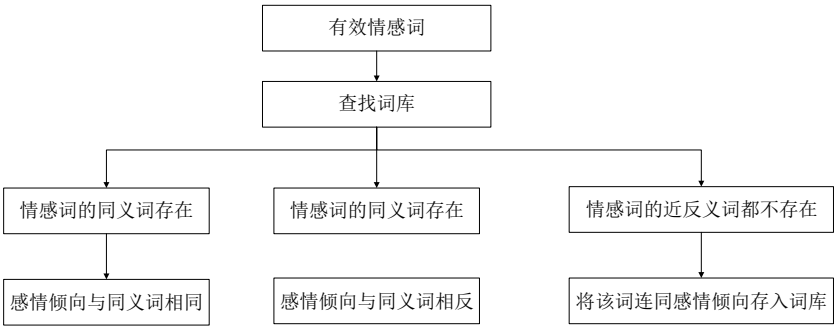


图 7: 情感词感情倾向判断示意图

图7说明我们给出的评价语句提取方法可以在很大程度上将评论中评论者的感情倾向提取出来。在验证了这样的感情倾向提取方法的可行性之后，我们给出题目附件中第一类评论中的七条评论进行情感词

感情倾向提取，得到评论中正面评价语句与负面评价语句所占整个句子的比重如表9 所示：

表 9: 正负面评价语句提取情况

评论	1	2	3	4	5	6	7
正面评价语句所占比	0	0	0	0.5	0	0	0.5
负面评价语句所占比	1	1	1	0.5	1	1	0.5

题中给出的7条评论均为负面评价，因此所得到的结果中负面评价语句所占比例越接近于1，说明模型计算越准确。而模型求解得到的结果中，大多数评论对应的负面评价语句所占比例都为1。

5.1.5 非典型特性的提取

典型特性是顾客评价过程中提及次数较多的商品的特性，但是有些特性只有少部分人提及，并且这些特性对一些潜在客户与产品生产商也是有一定意义的。问题是如果提取出这些非典型特性，我们之前介绍的关联挖掘法并不能实现。我们举个例子来进行分析，例如两句评论：“拍出的照片好极了”，“自带的软件好极了”。两个句子表达的特征词汇都是“好极了”，但是却描述的不同事物，前者是说照片，后者是说软件。由于同一个形容词可能修饰不同的事物，而我们可以认为它是修饰离它最近的的名词或名词短语，因为这也是最有可能的搭配情况，因此我们就可以通过特征词汇来寻找那些通过关联挖掘法无法得到的特性。其流程图如图8所示：

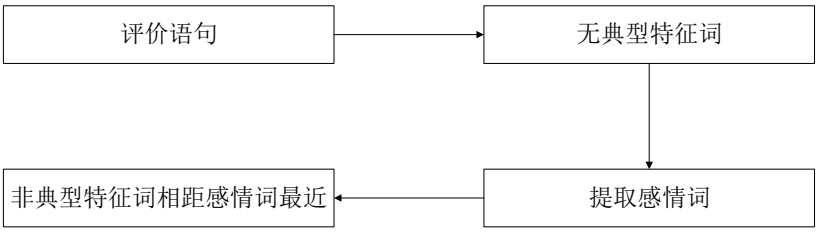


图 8: 非典型特征提取流程示意图

值得注意的是，非典型特性虽然很重要，但是它只是为了我们的结果更客观全面，在评论的筛选过程中，典型特性仍为较重要的部分。

我们对上文研究过的四款手机的500条评论进行典型特征提取，从评论中提取的特征词的召回率与准确率如表10所示：

表 10: 非典型特征提取下的召回率与准确率

产品类型	iPhone5	iPhone5s	iPhone6	iPhone6+
召回率	0.82	0.79	0.76	0.82
准确率	0.75	0.71	0.72	0.69

四款产品对应的评论中的非典型特征的召回率与准确率都较高，这是由于之前我们主要关注的是典型特征，那么其他的词条很大概率上即为非典型特征，这样的结果在肯定非典型特征提取的方案可行的同时，也间接证明了之前的研究方向的正确性。

5.1.6 评论语句的倾向预测

产品评论是表达消费者对产品各种特征的主观感受，其情感倾向特征对其他潜在消费者的购买决策有很大影响。一般地，我们可以通过统计评论中每个句子的正面或负面情感倾向来判断评价倾向，而句子的情感倾向判断采用Hu 等 [13]的方法。先利用极性词词典找出句子中的极性词并判断其极性，再查找句子中去除该极性词后是否还存在“没”、“没有”、“无”和“不”等否定词，若存在就将评价句的极性取反。由于每个评价句较短，所以不必考虑多重否定的情况。若整条评论中，特征$F5$或$F6$中有一个为0，而另一个值相对较大，则可能是过分褒贬的垃圾评论。

在我们之前的介绍中，我们通过对情感词汇统计与估计来推测评论语句的方向性，例如正面或者负面。但是通常我们会选取主要倾向性词汇来判断整个语句的感情倾向，如果统计得到的满意词汇与不满意词汇数量相等，那我们就需要利用有效评价（有效评价即为在评论中与商品特性最接近的评价词）的平均倾向或者先前评论语句的倾向。这个方法的步骤将在下面的图详细介绍：

在用户评价过程中，用户大体有三种评价：明显的好评或差评，评论中有优点有缺点，其他评价。我们对其分析如下：

1. 用户对产品的评价有明显的偏向性。在这种情况下，用户对产品的大多数特性都会表达偏向性明显的评价，而评价词汇很多情况下要么是正面，要么是负面的（取决于用户的主观偏向性）。我们以两个正面评价词汇“很好”，“出色的”为例，二者意思接近，假设用户对手机的成像质量作评价，那么评论内容可能是拍出的照片质量很好，摄像头很出色等评价。

2. 用户在一句评论中涉及到了正面和负面两个方面，而且双方涉及的情感词汇数量是相等的。例如：手机成像质量很好，但是系统自带的相机很一般。

3. 所有其他情况

对于第一种情况，主要倾向很容易判断，这也是评论统计中较为常见的一类评论。对于第二种情况，我们就需要运用有效评论的倾向性来判断用户的评价倾向，所谓有效评论即为与商品相关性最大的评论。对于第三种情况，我们认为它所表达的情感与它之前的句子情感倾向是相同的。我们用这样逐句分析的方法来预测句子的感情倾向，大多数情况下，用户的评论都是连续几个独立语句，因此这样的方式可以得到一个较为全面的结果。

对于含有“但是”、“可是”这类词汇的句子，它们往往意味着句意即将出现转折。那么我们就首先用“但是”之后的句子来进行被描述特性的语义倾向。如果在之后的句子中并没有明显的情感倾向词汇，那么我们就用“但是”之前的句子来进行倾向判断。

对于含有“不”，“没有”这类词汇的句子，它们往往意味着这句话表达的意思与之后的评价词义相反，但是前提条件是“不”必须与所修饰的词离得比较近，如果距离较远的话，这样的方法显然就失去了可靠性。例如

评价1：手机没有我预期得那么好。

评价2：不是我恶意差评，手机真的很差。

评论1中“没有”和“好”距离较近，根据我们的方法推断结果是正确的，第二个评价中“不是”和“很差”之间的距离较远，这样的方法显然就不准确了。整个过程的流程如图9所示：

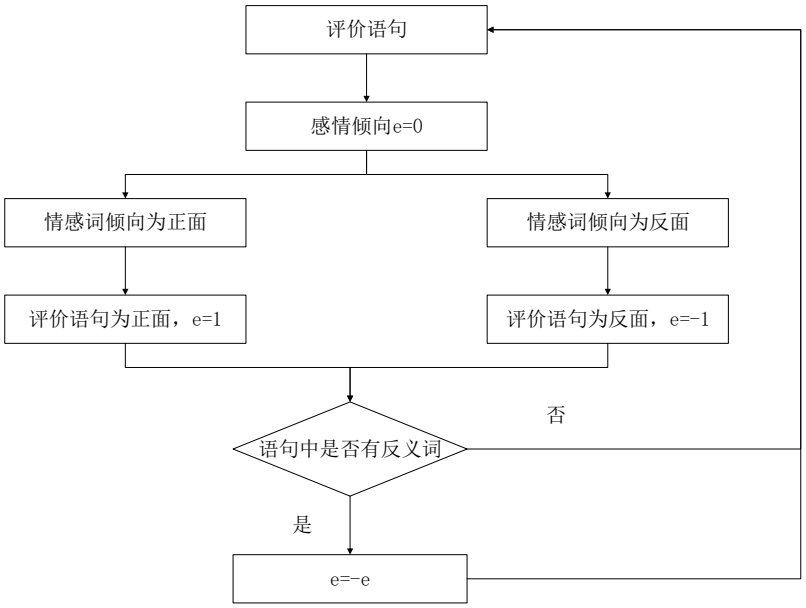


图 9: 评论语句倾向判断示意图

如图9所示，首先我们根据情感词倾向来对评价语句倾向做判断，最后判断语句中是否有反义词，由于一般评论都较为简单，我们不考虑双重否定的情况，认为如果有反义词就与原来表示的意义相反，整个过程对语句中的每个情感倾向词循环操作。

5.1.7 评论文本结构特征

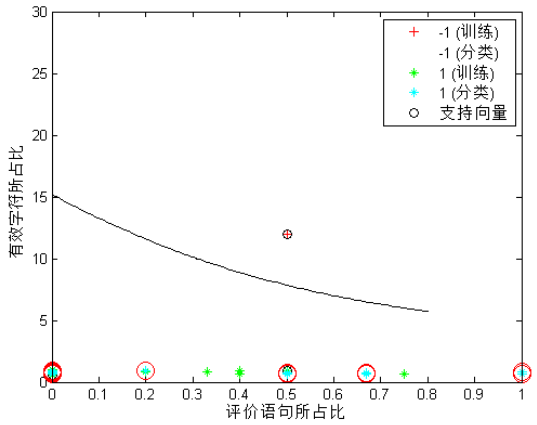
因含广告的垃圾评论一般内容较长、句子长短不一，包括URL链接地址、QQ 号和其它非汉字字符等，故文本参考Chen等[]的研究并结合中文文本的特点，在评论文本结构方面选择$F-12$-$F-12$特征，从而区分含广告或特征词重复的评论。短句平均长度l，短句长度方差σ,中文字符百分比p，其定义分别为两个公式：

$$\begin{aligned} l_m &= \tfrac{1}{s}\sum l_i \\ \sigma_l &= \tfrac{1}{s}\sqrt{\sum (l_i - l_m)^2} \end{aligned} \tag{3}$$

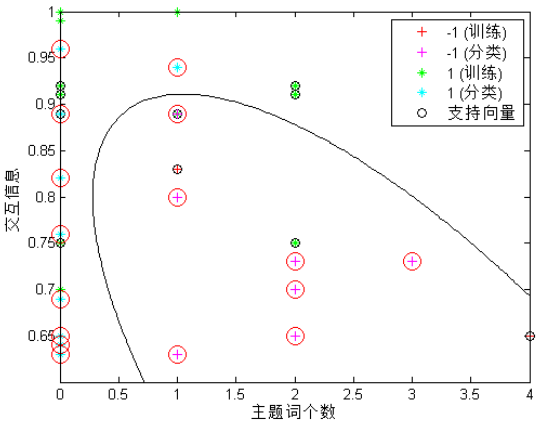
其中l_i为短句长度，s为短句个数，l_m为平均短句长度，σ_l 为短句长度方差。

5.2 模型求解

在之前的研究中，我们考虑了单一因素对信息筛选的影响，但事实上，评论内容是否具有价值应该从多个方面进行考量，接下来，我们将会将前面所介绍的7个因素进行适当组合，探究在组合情况下其对判断评论是否为垃圾评论的影响。在这个过程中，我们首先选取已知的评论进行训练，使得模型不断适应求解目标，得到垃圾评论与非垃圾评论之间的区分界限（这个界限是由横纵坐标两个指标共同决定的，并且由于两个指标的影响程度不同，界限曲线不一定唯一），之后将待求内容代入模型进行求解，判断其是否属于垃圾评论。



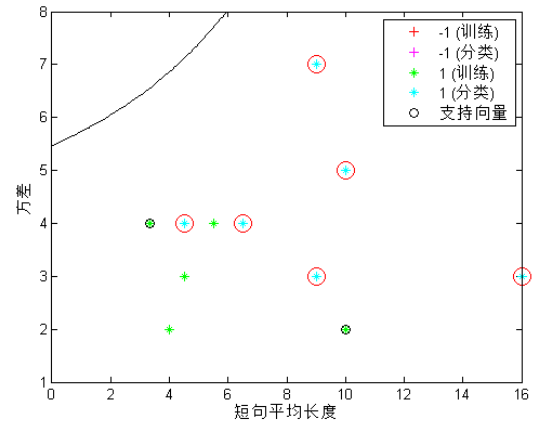
(a)评价语句所占比与有效字符所占比



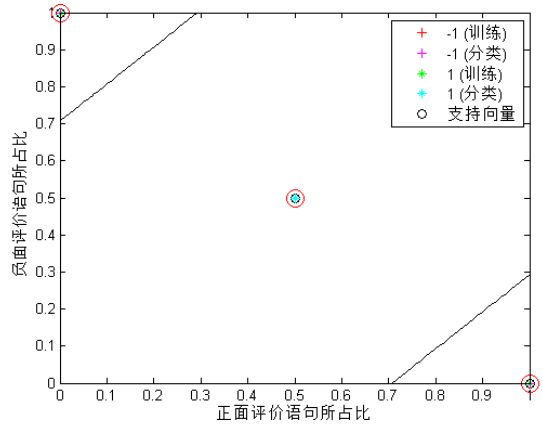
(b)主题词个数与交互信息

图 10: 组合指标对垃圾评论判断的影响

在图10中，空心原点代表支持向量，1表示垃圾信息，−1表示非垃圾信息，图(a)中的黑色曲线即为垃圾评论与非垃圾评论的分界线，在曲线上方的区域对应的评价语句所占比与有效字符所占比组合即为非垃圾评论对应的组合形式，下方对应垃圾评论组合形式。同理，在图(b)中，曲线右下方区域对应非垃圾评论对应的区域，曲线上方对应垃圾评论区域。



(a)短句平均长度与方差



(b)正面与负面评价语句所占比

图 11: 组合指标对垃圾评论判断的影响

与图10相似，图11通过同样的方法给出了短语平均长度与方差组合，正面与负面评价语句所占比组合对垃圾评论判断的结果。段句平均长度越长，且所有短剧的方差越小，评论越有价值，越不可能为垃圾评论。此外，值得注意的话，在综合考虑正面与负面评价语句所占比的结果中有两条分界线，两边的小三角区域对应垃圾评论，这样的结果与我们在之前介绍的过好或者过坏的评论都更有可能为垃圾信息的原则是相一致的。

为了探究不同因素的组合所带来的不同影响，我们将上述的四种组合的前提下的准确率与召回率进行对比。

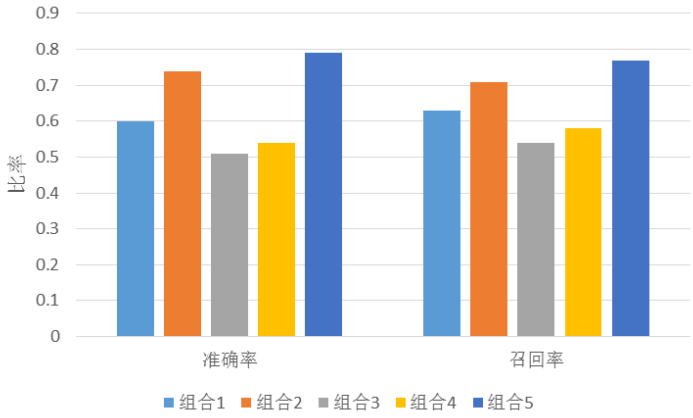


图 12: 双指标组合与多指标组合对比图

在图12中，组合1到组合4为上文介绍的四种指标组合形式，组合为所有指标综合考量的组合形式，可以明显看出，无论是在准确率还是召回率上，综合指标相比较双指标组合都更具准确性。

6 问题二：基于用户信誉度的SVM模型

上述的内容及结果是建立在只关注评论内容的基础上的，但是在实际的交易操作过程中，买家的信誉程度与评论公开度也会对评论内容的好坏造成一定影响。我们随机抽取了天猫某店铺关于iPhone6的500条评价，并且在数据处理过程中考虑了买家的淘宝等级。

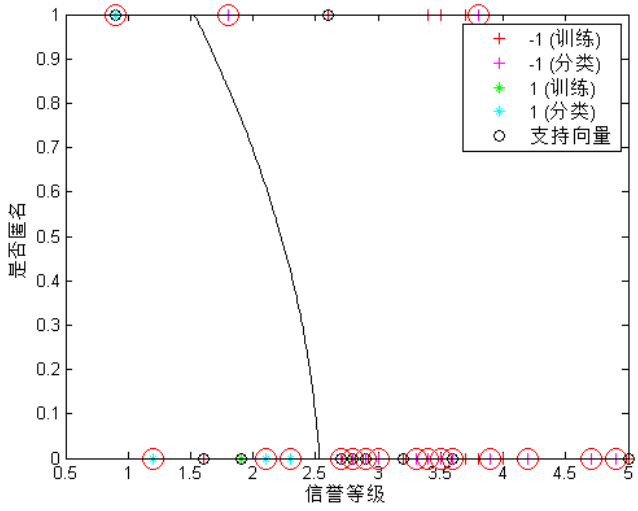
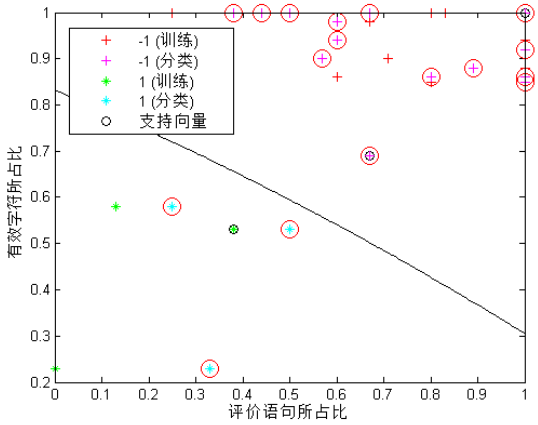
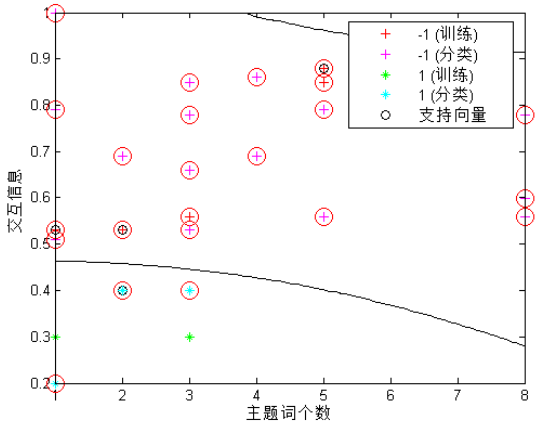


图 13: 信誉等级与是否匿名组合

在图13中，匿名对应的值为0或1，分别代表不匿名和匿名，可以明显看出大多数有效评论都集中在信誉等级较高且匿名的区域，这与实际的评价状况是相符的，因此通过这样模型所得到的评论分类是具有参考价值的。那么，在考虑信誉等级及匿名的基础上，我们又研究了之前讨论过的指标组合对评论垃圾的判断情况（现在相当于是四个指标的组合）。

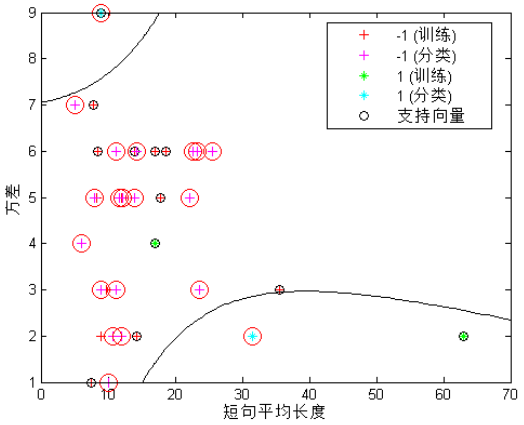


(a)评价语句所占比与有效字符所占比

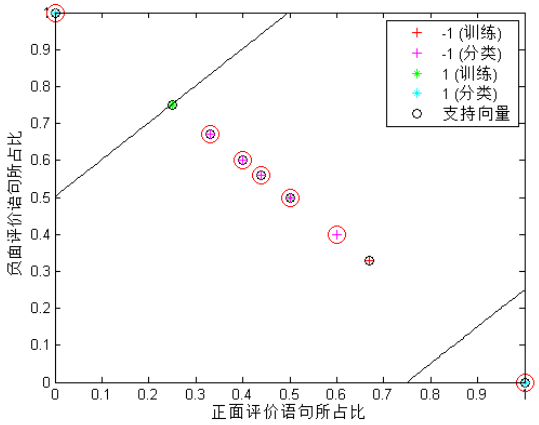


(b)主题词个数与交互信息

图 14: 组合指标对垃圾评论判断的影响（考虑买家信誉等级）



(a)短句平均长度与方差



(b)正面与负面评价语句所占比

图 15: 组合指标对垃圾评论判断的影响（考虑买家信誉等级）

由图14，图15可以看出，垃圾评价与有效评价界限明显变化，而且更多的情况下对应的分界线变为两条，且集中在坐标轴的右侧或上册。从理论上讲，单个条件的单方向变化往往表征着评论向一个极端发展（要么是非常有价值的评价，要么是很无用的评价），因此两条分界线所对应的结果更具科学性。

为了探究不同因素的组合所带来的不同影响，我们将上述的四种组合的前提下的准确率与召回率进行对比。

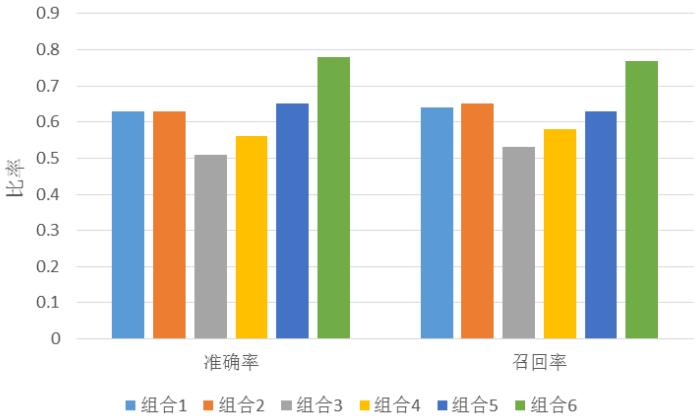


图 16: 双指标组合与多指标组合对比图

在图16中，以考虑评价者属性为前提，组合1到组合4为上文介绍的四种指标组合形式，组合为所有指标综合考量的组合形式，可以明显看出，在考虑评价者属性的条件下，无论是在准确率还是召回率上，综合指标相比较双指标组合都更具准确性。

7 问题三：基于用户行为的SVM模型

根据参考文献 [22]，当考虑到一般产品集合时，由于不用的产品具有不同的特征词，所以在一定程度上评论者的评论的行为能够起到非常长关键的作用，因此在一般的产品集合中提取垃圾评论时，我们在原来的基础上更加注重评论者行为的影响。在某种程度上说，评论者的行为对不同产品中的垃圾评论的提取具有很大的通用性。也就是说在此部分，我们考虑了前面模型中提到的产品评论内容、用户信誉等级，同时考虑到评论者行为对垃圾评论识别和判断的影响。在此部分，我们参考文献 [22]对模型进行了完善。

7.1 模型建立

7.1.1 基于目标产品的垃圾评论者检测模型

针对某具体产品的垃圾评论者检测可以通过观察用户对该产品的评分次数和评论次数实现。若同一用户对同一产品的评分次数越多，且该用户的评分基本上类似，或其评论次数越多，且评论文本相同或者非常相似，则用户可能为垃圾评论者。为检测评论者的以上两种种行为，定义模型如下：

1. 基于用户评分行为的检测模型

$$c_{p,e}\left(u_i\right)=\frac{s_i}{Max_{u_i\in U^{s_i}}} \tag{4}$$

其中，$s_i=\sum_{e_{ij}\in E_{ij},|E_{ij}|>1}|E_{ij}|\cdot sim\left(E_{ij}\right)$，对于相似度函数$sin()$，我们给出如下定义，在一定给定集合下

$$sim\left(E_{ij}\right)=1-Avg_{g_k,e_{k'}}\left|e_k-e_{k'}\right| \tag{5}$$

2. 基于用户评论文本的检测模型 两个评论文本v_k, v'_k之间的相似度可定义为：

$$[sim\left(v_k, v_{k'}\right) = \cos\left(v_k, v_{k'}\right) \tag{6}$$

针对评论文本字数都不多、拼音与汉字共存、简体字与繁体字混用等特点，将评论标题与评论文本合起来作为评论文本对其进行分词，并对得到每一个词提取其拼音串，根据每个文本的拼音串将文本转化成向量。在式（6）中，$\cos\left(v_k, v_{k'}\right)$ 表示评论文本v_k和v'_k拼音串TF-IDF向量的余弦相似度。给 定一个评论集合V_{ij} $(V_{ij} > 1)$，我们可以为评论文本定义一个相似度分数如下：

$$sim\left(V_{ij}\right) = Avg_{v_k, v_{k'} \in V_{ij}} sim\left(v_k, v_{k'}\right) \tag{7}$$

于是，针对目标产品的基于用户评论文本的垃圾评论函数$c_{p,v}\left(u_i\right)$可以定义为：

$$c_{p,v}\left(u_i\right) = \frac{S'_i}{Max_{u'_i \in U^{s'_i}}} \tag{8}$$

其中，$s'_i = \sum_{v_{ij} \in V_{ij}, |V_{ij}|>1} |V_{ij}| \cdot sim\left(V_{ij}\right)$。

3. 基于目标产品的垃圾评论行为得分
根据上面两个模型，将基于目标产品的垃圾评论行为得分定义为：

$$c_p\left(u_i\right) = \frac{1}{2}\left(c_{p,e}\left(u_i\right) + c_{p,v}\left(u_i\right)\right) \tag{9}$$

7.1.2 基于用户评分偏差行为的垃圾评论者检测模型

垃圾评论者为了提升某一个产品或诋毁某一个产品，其与其他人的评分可能会有很大的不同。于是定义了d_{ij}用于表示用户针对某个产品的评分的偏差程度：

$$d_{ij} = e_{ij} - Avg_{e \in E_{*j}} e \tag{10}$$

一个用户基于评分偏差的垃圾评论行为得分可定义为：

$$c_d\left(u_i\right) = Avg_{e_{ij} \in E_{*j}} |d_{ij}| \tag{11}$$

7.1.3 基于用户互动行为的垃圾评论者检测模型

一般来说，如果一个用户发表的垃圾评论较多，其评论的平均回复数会比较少，因此，使用用户的平均回复数作为衡量垃圾评论行为得分的指标。

一个用户基于互动行为的垃圾评论行为得分可定义为：

$$c_r\left(u_i\right) = 1 - \frac{r_i}{Max_{u'_i \in U^{r'_i}}} \tag{12}$$

其中，$r = \frac{\sum R_{ij} \in R_{i*} |R_{ij}|}{|E_{i*}|}$。

7.1.4 基于用户购买行为的垃圾评论者模型

当用户对所评论的产品一次都没有购买过，或者只购买了其中很小一部分的时候，该用户为垃圾评论者的可能性就会很大。于是将基于用户购买行为的垃圾评论行为得分定义如下：

$$c_b(u_i) = 1 - \frac{b_i}{Max_{u'_i \in U^{b'_i}}} \tag{13}$$

其中，$b_i = \frac{|E_b|}{|E|}$ 表示对用户的发表一条评论对产品的平均购买次数。

7.2 对于识别问题的看法

识别评论垃圾的关键是提取表征评论的特征。特别是采用机器学习方法, 对特征选择需进行深入研究。目前, 识别评论垃圾的特征选取主要从评论内容和评论人两个方面考虑。评论内容特征反映评论质量和可信度。从内容和文体的角度分析, 许多研究采用了词性(POS)以及n元文法(n-gram)。一元文法(Unigrams)和二元文法(Bigrams)较常用, 结合基于LIWC 反映与心理学相关的一组语言特征, 在识别欺骗型评论的任务中发挥了效用。对于评论的主观特征, 情感分析被引入, 经验表明, 如果评论的主观表现过于吹捧或者蓄意诋毁, 则极可能是垃圾, 情感分析探测评论内容的主客观度和褒贬性, 利用情感词汇的极性进行测度,但文献 [15]的研究结论表明情感因素对辨识欺骗型评论的作用并不显著, 因为欺骗型评论的识别是真伪的辨别, 如果刻意虚构评论, 则并不容易从单纯的情感词中辨识区分标准。研究指出, 仅从评论内容中提取识别虚假性评论的特征, 辨识效果往往并不十分理想。评论人特征因而被关注。因为评论人特征反映评论撰写者的个人信誉和行为, 通过探测评论人特征,特别其行为表现可以非常准确地预测其发表评论的真伪。评论人的行为特征可表现为其发文量, 发表内容的雷同度, 发表时间以及其评价与大众评价值的偏差等一系列特征因素。文献[41]面向 Yelp 中的真实数据,特别比较了基于内容特征和基于评论人特征的虚假性评论垃圾的识别效果, 发现基于评论人特征的识别效更优。而文献 [15], [16], [17]也指出对虚假性评论的辨别, 评论人的行为特征是评论语言特征的重要补充。通过分析评论人的行为判断其是否为评论垃圾的制造者, 间接识别其发表评论的价值。相关研究 [15], [16], [17]探讨的评论人特征经检验 [18]都具有显著影响作用, 为评论人特征在识别模型中的有效性提供了重要依据。

笔者也认为单纯采用内容或行为特征构建的评论垃圾识别模型会导致辨别信息的丢失和遗漏, 将两方面特征融合可获得更优效果。而另一方面, 对分类问题, 往往分类特征越全面, 整体效果越好。但众多特征间存在约束, 可能导致以特征相互独立为前提的分类器的效率降低，所以, 特征并非越多越好,应增加显著性检验 [18]， 筛选出贡献较大且不相互依 赖的表征特征, 从而提升模型的稳定性和效率。

识别评论垃圾的主流方法之一是利用可指导的机器学习方案 [16], [17], [18]。但对于机器学习方法, 识别的准确率往往取决于用于构建分类器的标注集。然而, 对虚假评论, 人工给出的标注结果带有较大随机性, 很难通过人工阅读来准确判断评论的真实性, 因而研究中最困难的是获取标准的针对虚假评论的标注集。大部分研究工作采用了近似方案来标注虚假评论。如文献 [19]取雷同或近似雷同的评论作为虚假评论, 文献 [21]采用人工标注, 文献 [20]则利用了 AMT(Amazon Mechanical Turk)生成虚假评论, 并结合人工标注生成训练集。尽管采取了一系列的处理和选择的方法, 但标注的数据集都存可靠性问题。如, 文献 [20]的研究在 AMT 模拟的虚假评论数据集上有很好的表现, 但在真实的商业数据集 (Yelp的评论数据)中, 识别效果却并不理想 [21]。因为 AMT 虚构的评论垃圾并不能探测出评论垃圾发布人的真实目的, 无法模拟出其真实的心理状态。可见, 如果标注集的真实性不能得到证, 那么评论垃圾辨识就失去了参照和基准, 结果便不具备信服力。有指导的分类方法在评论垃圾识别上有一定局限, 要想在传统机器学习分类器上有所突破,高质量的标注集的获取至关重要。鉴于标注集的问题, 最近的一组研究尝试采用非指导

的学习方案, 进行评论垃圾识别。文献 [20]采用了频繁项目挖掘实现评论垃圾识别, 频繁项目挖掘在检测个体评论垃圾发布人和评论垃圾发布人群体中得以较好应用, 识别过程的关键在于频繁规则筛选标准的定义。文献 [18]则采用了聚类算法, 通过分析评论人的行为特征, 探测评论垃圾发布人和非评论垃圾发布人分布上的差异, 辨识评论垃圾的撰写者, 识别评论垃圾。研究评论人行为及信誉来识别评论垃圾的方向还包括图论 [4]、 分布规律 [18]- [20]等, 较为新颖, 有待进一步探讨。可见, 为了回避人工标注训练数据集会导致的判断偏差, 很多研究以评论人为突破口。通过分析评论人的行为表现, 迂回地选取评论内容之外的信息征, 来间接判断评论垃圾。相比于直接挖掘评论内容特征的判定方式, 这类方法更能保证判别的效率和稳定性。但这类方法的评判依据毕竟是评论内容之外的间接特征, 是否能够充分地反映信息内容本身的可信度, 值得深入研究, 因而其信服力和准确性仍有待进一步提高。

开放的网络环境资源丰富, 但缺乏监管的现状使网络资讯良莠混杂, 难以被有效利用。真实可信的网络评论能够为用户提供有价值参考, 引导整个行业或产品的改进, 向更符合用户需求的健康方向发展, 但虚假无效的评论垃圾则会误导消费者, 带来更严重的负面效应。辨识资讯真伪, 提高信息质量, 使网络资源可利用价值最大化, 评论垃圾识别是一个值得关注具有社会和应用价值的热点问题。本文从实践研究角度对该领域研究进行了较系统的分析和梳理, 分别从概念辨识、研究范畴、方法以及关键技术等方面对相关研究进行了总结和评述, 得出如下结论:

1. 概念上, 笔者界定“评论垃圾”指故意过分吹捧或过分贬低某种产品的不真实评论以及不包含任何有益成分的非相关在线网络评论。本质上不同于“邮件垃圾”和“网页垃圾”。

2. 研究范畴上, 国内研究中, 对评论可信度与评论有用性的区分不甚明晰。笔者认为评论垃圾识别是评论可信性研究的应用体现。可信性研究关注信息真伪的辨别, 而信息效用价值的研究则是可信性研究的后继。

3. 研究方法上, 辨别评论垃圾的关键是提取表征评论垃圾性的特征。评论内容和评论人特征同样重要, 融合两方面特征的识别模型具有更优的效果。

4. 实现技术上, 由于存在标注偏差, 面向评论内容, 基于机器学习的分类方法存在一定局限。基于评论人及其他外部特征的解决方案有新意和潜力, 但需要深入探索。

8 模型评价

1. 本文从评论内容和用户的行为特征出发，提出一种产品垃圾评论者检测方法。通过找到垃圾评论者发表垃圾评论的行为模式及其文本特征，并将其作为垃圾评论者的检测指标，利用多个指标对垃圾评论者检测。有效地避开了对文本内容深度理解和观点的抽取，避免了单个指标在检测垃圾评论者上的不足。

2. 目前对垃圾评论检测的研究不多，从评论内容和用户行为出发来识别垃圾评论也是一个较新的方法，但是本文方法只是列出垃圾评论者检测的一些可行指标，今后将从消息传播模式角度确定发表垃圾评论的模式，进一步提高识别准确率。

3. 识别评论垃圾利用可指导的机器学习方案。对于机器学习方法, 识别的准确率往往取决于用于构建分类器的标注集。然而, 对虚假评论, 人工给出的标注结果带有较大随机性, 很难通过人工阅读来准确判断评论的真实性，因而在一定程度上造成了结果的不准确性。

参考文献

[1] 聂卉, 王佳佳. 产品评论垃圾识别研究综述[J]. 现代图书情报技术, 2014, (2).

[2] Li F T, Huang M, Yang Y, et al. Learning to Identify Review Spam [C]. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. AAAI Press. 2011:2488-2493.

[3] Ott M, Choi Y J, Cardie C, et al. Finding Deceptive Opinion Spam by Any Stretch of the Imagination [C]. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011: 309-319.

[4] Jindal N, Liu B. Review Spam Detection [C]. In: Proceedings of the 16th International Conference on World Wide Web. New York, NY, USA: ACM, 2007: 1189-1190.

[5] Mukherjee A, Liu B, Wang J, et al. Detecting group review spam[C]//Proceedings of the 20th international conference companion on World wide web. ACM, 2011: 93-94.

[6] Jindal N, Liu B, Lim E P. Finding Unusual Review Patterns Using Unexpected Rules [C]. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2010: 1549-1552.

[7] Lim E P, Nguyen V A, Jindal N, et al. Detecting product review spammers using rating behaviors[C]//Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010: 939-948.

[8] Jindal N, Liu B, Lim E P. Finding typical review patterns for detecting opinion spammers[R]. UIC Tech. Rep, 2010.

[9] 何海江. 一种适应短文本的相关测度及其应用[J]. 计算机工程, 2009, 35(6):88-90. DOI:10.3969/j.issn.1000-3428.2009.06.030.

[10] Bhattarai A, Rus V, Dasgupta D. Characterizing comment spam in the blogosphere through content analysis[C]. //IEEE Symposium on Computational Intelligence in Cyber Security. IEEE, 2009:37 - 44.

[11] 游贵荣, 吴为, 钱沄涛. 电子商务中垃圾评论检测的特征提取方法[J]. 现代图书情报技术, 2014, 30(10):93-100.

[12] Turney P. Mining the web for synonyms: PMI-IR versus LSA on TOEFL[J]. 2001.

[13] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004: 168-177.

[14] NLProcessor – Text Analysis Toolkit. 2000. http://www.infogistics.com/textanalysis.html

[15] Chirita P A, Diederich J, Nejdl W. MailRank: using ranking for spam detection[C]//Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005: 373-380.

[16] Fette I, Sadeh N, Tomasic A. Learning to detect phishing emails[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 649-656.

[17] Cortez P, Correia A, Sousa P, et al. Spam email filtering using network-level properties[M]//Advances in Data Mining. Applications and Theoretical Aspects. Springer Berlin Heidelberg, 2010: 476-489.

[18] Baeza-Yates R A, Castillo C, López V, et al. Pagerank Increase under Different Collusion Topologies[C]//AIRWeb. 2005, 5: 25-32.

[19] Wu B, Goel V, Davison B D. Topical trustrank: Using topicality to combat web spam[C]//Proceedings of the 15th international conference on World Wide Web. ACM, 2006: 63-72.

[20] Ntoulas A,Najork M,Manasse M,et al.Detecting spam web pages through content analysis [C]MProceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland:ACM,2006:83-92

[21] Cortezp P,Correia A,Sousa P,et al.Spam email filtering using network-level properties [C]MProceedings of the 10th industrial conference on Advances in data mining:applications and theoretical aspects.Berlin,Germany:Springer-Verlag,2010:476-489

[22] 邱云飞, 王建坤, 邵良杉等. 基于用户行为的产品垃圾评论者检测研究[J]. 计算机工程, 2012, 38(11):254-257. DOI:10.3969/j.issn.1000-3428.2012.11.077.