

文章编号:1001-9081(2016)10-2767-05

DOI:10.11772/j.issn.1001-9081.2016.10.2767

基于用户回复内容观点支持度的评论有用性计算

李学明, 张朝阳*, 余维军

(重庆大学 计算机学院, 重庆 400030)

(*通信作者电子邮箱 nwpuzhangcy@163.com)

摘要:针对有监督评论有用性预测方法中的训练数据集难以构造,以及无监督方法缺乏对情感信息支撑的问题,提出基于语义和情感信息构建一种无监督模型,用于对评论有用性进行预测,同时考虑了评论和评论下回复内容对观点的支持度用来计算观点的有用性得分,进而得到评论的有用性。同时,提出结合句法分析和改进潜在狄利克雷分配(LDA)模型的评论摘要方法用于评论有用性预测模型中的观点提取,基于句法分析结果构建 must-link 和 cannot-link 两种约束条件指导主题模型学习,在保证召回率的同时提高模型准确率。该方法在实验数据集上能取得 70% 左右的 F_1 值和 90% 左右的排序准确率,且实例应用也表明该方法对结果具有较好的解释性。

关键词:评论有用性;观点支持度;情感分析;观点摘要;用户回复

中图分类号:TP301.6 **文献标志码:**A

Review helpfulness based on opinion support of user discussion

LI Xueming, ZHANG Chaoyang*, SHE Weijun

(School of Computer, Chongqing University, Chongqing 400030, China)

Abstract: Focusing on the issues in review helpfulness prediction methods that training datasets are difficult to construct in supervised models and unsupervised methods do not take sentiment information into account, an unsupervised model combining semantics and sentiment information was proposed. Firstly, opinion helpfulness score was calculated based on opinion support score of reviews and replies, and then review helpfulness score was calculated. In addition, a review summary method combining syntactic analysis and improved Latent Dirichlet Allocation (LDA) model was proposed to extract opinions for review helpfulness prediction, and two kinds of constraint conditions named must-link and cannot-link were constructed to guide topic learning based on the result of syntactic analysis, which can improve the accuracy of the model with ensuring the recall rate. The F_1 value of the proposed model is 70% and the sorting accuracy is nearly 90% in the experimental data set, and the instance also shows that the proposed model has good explanatory ability.

Key words: review helpfulness; opinion support; sentiment analysis; opinion summarization; user discussion

0 引言

评论信息是用户发表的针对某种商品的反馈意见,从某种程度上反映出商品的质量和体验效果等,这就为用户购买商品提供了丰富的决策意见。许多网站已经采取了相应的措施用以提高用户发表评论的质量,但网上仍然充斥着大量的垃圾和虚假评论,用户难以从海量的评论数据中辨别评论信息的可信度,也难以便捷地找到自己需要的评论信息。尽管大部分网站都提供评论排序显示的功能,这些排序方法看起来简单、直观,但效果往往不够理想。例如:按时间排序可能会导致用户错失很多高质量的评论,用户容易受到虚假评论的误导;按评论有用性得票率排序会存在得票不公平性的问题等。因此,一种科学合理的评论排序方法就显得尤为重要^[1]。近年来针对这方面的研究工作主要集中在三个方面:垃圾评论、评论质量、观点摘要(评论总结),Liu^[2]对此作了很多的梳理。

Jindal 等^[3]将垃圾评论分为三类:虚假评论、未针对特定产品和服务的评论以及非评价性质的评论,并指出第二、三类垃圾评论较少,而且使用监督学习的方法容易识别出,而虚假评论则复杂许多,因为即使是人工也很难识别出故意伪造的

看似真实的评论。虚假评论常用分类、回归等相关机器学习方法来识别,并使用近似重复的特征来度量虚假评论^[3]。基于用户的非典型行为特征^[4]、挖掘非预期的分类关联规则^[5]和基于网络图^[6]等非监督学习方法被用来发现垃圾评论者或垃圾评论,有研究者进一步对垃圾评论群组的识别进行了相关研究^[7-8],黄婷婷等^[9]从商品特征关联度的角度来评测评论的可信度。评论质量的研究主要围绕评论的有用性展开,常用的解决方法是将其当作回归问题^[10-11]或分类问题^[12],有研究者基于评论的有用性提出了相关的评论排序模型如 SVR (Support Vector Regression)^[13]、Ranking SVM (Support Vector Machine)^[14]。观点摘要可以简单看作多文档文本摘要,且更侧重于针对产品的各方面特征挖掘其情感倾向。文献[15-16]通过词性标注以及关联算法对观点摘要进行研究,后续相关研究大多从中延伸,如扈中凯等^[17]通过词性路径模板的方式提取产品特征词。

在生成观点摘要的基础上,本文尝试从情感倾向分析的角度探索评论有用性的问题。以往文献都是通过对评论内容投票研究评论有用性,但评论内容投票只能反映粗粒度的统计信息,且存在有偏性,而针对评论内容的回复(即评论回复)则能反映其他用户对评论内容的观点倾向。本文通过挖

收稿日期:2016-03-21;修回日期:2016-06-17。 基金项目:国家自然科学基金资助项目(90818028)。

作者简介:李学明(1967—),男,重庆人,教授,博士,主要研究方向:数据挖掘、大数据处理; 张朝阳(1991—),女,河南洛阳人,硕士研究生,主要研究方向:数据挖掘、自然语言处理; 余维军(1991—),男,四川南充人,硕士研究生,主要研究方向:数据挖掘、自然语言处理。

掘评论回复的情感倾向,结合相同观点具有一致有用性的特点,提出基于回复内容的观点支持度(Opinion Support based on User Discussion, OSUD)模型,计算观点支持度,探索评论内容的有用性。虽然 Ryang 等^[18~19]已经进行了相关的工作,但其只是简单地利用评论回复的统计信息,结合关键词的 TF-IDF(Term Frequency-Inverse Document Frequency)值来得到关键词搜索排序模型,并没有深入挖掘评论回复蕴含的深层观点和情感信息。

1 本文 OSUD 模型

本文模型 OSUD 参考了文献[20]中通过情感分析对微博进行可信度评估的方法,该方法基于两个观察结果:1)一个信息如果被更多的人认可则可能是真实的;2)更多的人产生争论则有可能是谣言。同理,在本文中,评论的有用性也是基于这两个假设产生。如果支持一条评论的用户得分相对于反对的用户得分更高,则这条评论被认为可能更具有用性,对评论的支持或反对则通过评论和评论回复中所涉及的观点的语义情感信息表达出。

关于一个产品的评论,其有用性计算是基于评论中所持观点在评论和评论回复中的支持度得到的,其简易形式可以表示为支持和反对该评论观点的用户得分差值的比例,用户被定义为评论下直接或间接的响应者。每个用户有不同的排名值,且可以根据某些特征计算得到用户的权重,进而每个用户都是可以被量化并归一处理的。

最后,由于一条评论可能含有一个或者多个观点(针对多个特征进行的评价),对某一特征持有相同态度的观点具有相同的有用性,而一条评论的有用性度量则通过评论中所持观点的有用性加权得到。

1.1 模型描述

1.1.1 评论回复用户的多样性比率

考虑某种特殊情况,评论下的所有回复均为评论者自己所发表,这为计算观点的有用性带来噪声,且本文认为,当一条评论下回复的不同用户比率越高,更能真实反映出评论内容的有用性,将这一表征回复用户比率的特征定义为评论回复用户的多样性比率 dsr ,定义如式(1)所示:

$$dsr(r) = \frac{u_0(r)}{rep_T(r) - rep_R(r)} \quad (1)$$

其中: $rep_T(r)$ 表示评论 r 下的回复总数; $rep_R(r)$ 表示评论作者本身发表的回复数; $u_0(r)$ 表示评论 r 下所有回复用户的数目。当 $rep_T(r) = 0$ 或者 $rep_T(r) = rep_R(r)$ 时, $dsr(r) = 0$ 。

1.1.2 观点支持度

评论是用户针对产品某方面或多方特征的观点表述,基于此本文将评论有用性问题转化为观点有用性问题,针对一条评论 r ,为了计算评论 r 的有用性,需要首先计算该评论所包含的特征及观点 $Opinion(op_1, op_2, \dots, op_n)$,然后针对每一个观点 op ,计算其有用性 $Helpfulness(op)$ 。对于每一个产品特征,有正面和负面两种观点表述,为了计算观点的有用性,需要首先计算用户分别对正面和负面两种观点的支持得分,将与观点 op 相同的观点支持得分记为支持度 $Support(op)$,与 op 相反的观点支持得分记为否定度 $Oppose(op)$ 。

假定 $Agree(op)$ 表示与评论 r 所持观点 op 相同的所有评论的集合(包括 r), $Confirm(m)$ 和 $Deny(m)$ 分别表示评论 m 下所有支持和反对观点 op 的回复用户的集合,其中 $m \in$

$Agree(op)$, $R_i = [0,1]$ 为用户 i 的排名值。则相同观点评论下用户回复对观点的支持度 $Support(op)$ 如式(2)所示, $Score(m)$ 表示评论下其他用户的回复信息对该评论的支持得分如式(3)所示。

$$Support(op) = \sum_{m \in Agree(op)} (dsr(m) \times Score(m)) \quad (2)$$

$$Score(m) = 1 + \sum_{i \in Confirm(m)} R(i) - \sum_{i \in Deny(m)} R(i) \quad (3)$$

1.1.3 观点否定度

同样地,令 $Disagree(op)$ 表示与评论 r 所持观点 op 相反的所有评论的集合, $Confirm(n)$ 和 $Deny(n)$ 表示评论 n 下所有支持和反对观点 op 的回复用户集合,其中 $n \in Disagree(op)$ 。则持有相反观点的评论下用户回复对该观点的否定度表示如式(4)所示, $Score(n)$ 表示评论下其他用户的回复信息对该观点的支持得分如式(5)所示。

$$Oppose(op) = \sum_{n \in Disagree(op)} (dsr(n) \times Score(n)) \quad (4)$$

$$Score(n) = 1 + \sum_{i \in Confirm(n)} R(i) - \sum_{i \in Deny(n)} R(i) \quad (5)$$

1.1.4 观点有用性

进而,观点 op 的有用性表示如式(6)所示:

$$Helpfulness(op) = \frac{Support(op) - Oppose(op)}{Support(op) + Oppose(op)} \quad (6)$$

分母使得表达式取值范围为 $[-1, 1]$ 。正数取值表示该评论有一定的有用性,值越大,则有用性越高,取值为 1 表示没有反对该观点的用户;相反,如果取值为负,则表示该评论可能并不具有有用性,或者说该评论不被用户所信任,其不被信任的程度通过 $|Helpfulness(op)|$ 度量,取值为 -1 表示没有用户支持该观点。

1.1.5 观点多样性

不同观点对用户感知力度有所区别,同样对评论有用性影响程度也有所差别,被用户提及更多的观点说明更加受用户关注,对评论的有用性影响越大。为了表征不同观点对评论有用性影响的权重,定义观点多样性 $dsr(op)$ 如式(7)所示:

$$dsr(op) = \frac{rev_T(op)}{\sum_{o \in Opinion} rev_T(o)} \quad (7)$$

其中: $rev_T(op)$ 表示持有观点 op 的评论数。

1.1.6 评论有用性

进而,可以得到评论有用性的计算公式如式(8):

$$Helpfulness(r) = \sum_{o \in Opinion(r)} dsr(o) \times Helpfulness(o) \quad (8)$$

同观点有用性一样,取值为负数表示不具有某种程度的有用性,正数表示具有一定程度的有用性。同时,求和公式使得含有有用观点越多的评论有用性越大;相反,持有无用观点越多的评论无用性越大。

1.2 观点摘要

观点摘要是计算以观点支持度为基础的评论有用性的关键步骤,观点摘要提取的准确度直接影响后续计算的准确率。

句法分析的方法可以保证较好的准确率但召回率较差,而无监督的潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)主题模型方法虽然能提高召回率,但难以保证准确率,本文基于此提出句法分析和主题模型结合的 SA-LDA 方法对产品特征进行抽取,首先基于句法分析的方法对产品所在类目的所有产品评论进行特征观点对提取,自动构建语料库用于生成两种约束条件,约束条件的构建考虑了对同一特征类

别的不同表述和对特征类有决策作用的词,借此辅助 LDA 模型进行训练。针对某个产品的评论数据,对观点句进行提取,基于语料库通过改进的 SA-LDA 模型得到特征聚类结果。

1.2.1 预处理

预处理主要是对评论文本进行分词、POS (Part-Of-Speech) 标注和合并、分句以及并列句的拆分和合并。分词和 POS 标注使用 NLPIR 提供的接口,对连续的 POS 标签进行合并,如“机器手感还不错”,得到的 POS 标签序列为“nnda”,合并后为“nda”,有助于提取到短语结构和后续的特征模板匹配。分句将评论文本拆分为以句子为单位,并对并列的词语和表述进行拆分和合并。

1.2.2 基于句法分析的语料库构建

评论中的产品特征属性词通常以名词、名词短语或动词为主,评价观点情感词则常以形容词、动词或名词为主,且特征词与情感词之间的搭配通常有类似于“名词 + 形容词”的模式,如“反应速度快”“电池容量小”。本文基于文献[2]中的词性路径模板提取特征词的方法提取出所有的候选产品特征词和评价词,并对其进行聚类。

聚类方法使用传统 K-means 算法,其初始聚类中心是随机指定的,后续每个类别的聚类中心通过计算每个特征词与其他所有词之间的距离,取距离之和最小的词作为聚类中心,结合语义相似度和观点相似度对特征词之间的距离进行度量,其相似度计算公式如式(9)所示:

$$\text{sim}(f_1, f_2) = a \times \text{ssim}(f_1, f_2) + b \times \text{osim}(f_1, f_2) \quad (9)$$

其中: $\text{ssim}(f_1, f_2)$ 表示语义相似度; $\text{osim}(f_1, f_2)$ 表示观点相似度。语义相似度融合 HanLP 基于同义词的语义距离、基于知网 HowNet 词语语义相似度的 WordSimilarity 以及字符串的编辑距离三种方式进行度量;观点相似度则通过计算用观点词表示的空间向量模型 $\text{Feature}(o_1, o_2, \dots, o_n)$ 之间的相似度,其中: o_1 到 o_n 表示特征所对应的观点表述。

基于句法分析聚类的结果,只保留名词、形容词和动词,对于评论中显式出现的特征,将其特征所在分句中出现的词语加入该类特征预料中。语料库格式表示为 {特征, {特征词}, {〈词项, 词频〉}}, 其中:{特征词} 表示特征类别下所有出现的特征词表达形式;{〈词项, 词频〉} 表示特征类别下出现的词项以及出现的次数,用以描述特征关联词。

1.2.3 SA-LDA 主题聚类

基于特征词与观点词共现矩阵,对评论中的主观句进行识别,用于 SA-LDA 主题聚类,并基于前文构建的语料知识库,构建约束条件用于约束和指导模型进行主题学习。

LDA 模型倾向于学习得到宽泛的主题,且很多时候受频繁词的影响会使得对某些词的主题学习产生歧义,先验知识可以在 LDA 模型主题学习时对其进行约束和修正,进而得到更为有效的聚类结果。根据上一节所得到的语料库等先验知识,本文自动构建两种约束条件:Must-link 和 Cannot-link。

Must-link 约束指定了哪些词对应该划分在同一个主题下,本文考虑两种类型的 Must-link:

① 特征词。对同一特征的不同特征词表述,如“价格”和“价钱”,必定从属于同一个主题下,每个特征类下的特征词两两构成 Must-link。

② 贡献词。通过广义的词共现矩阵计算得到,若某个词经常出现在某个特征表述语句中,且基本不在其他特征表述语句中出现,则该词可以隐式地表达某个特征或主题,如“便

宜”可以隐含表达特征“价格”。特征和词语的共现通过 $\text{frequency} * \text{PMI}(f, w)$ 得到,如式(10)所示:

$$\text{frequency} * \text{PMI}(f, w) = P_{f\&w} * \ln \frac{P_{f\&w}}{P_f P_w} \quad (10)$$

其中: $P_{f\&w}$ 是特征 f 和语料库中指示词 w 的联合出现概率; P_f 是特征 f 在语料库语句中出现的概率; P_w 是指示词 w 在语句中出现的概率。

考虑通用表达词如“不错”“还好”等可能经常与很多特征搭配出现,本文通过计算某个词关联的特征数与总特征数的比值,将超过某个阈值的词从词表中去除,因为该词不具有指示作用。一般地,本文考虑每个特征类目下 $\text{frequency} * \text{PMI}(f, w)$ 值前 5 的词语,将该特征类目下的所有特征词与每个词语匹配构成 Must-link。

Cannot-link 约束定义了哪些词对不应划分在同一主题下,同样,本文也考虑两种类型的 Cannot-link:

① 特征词。不同特征类别下的特征词,如“价格”和“电池”,必定不属于同一个主题,不同特征类别下的特征词两两构成 Cannot-link。

② 非共现词。若一个词在语料库中跟某个特征类从来不产生交集,本文将其定义为非共现词,本文中每个主题只对应一个特征类,所以有理由认为该词不应划分到该特征类目所在主题下。

为了更准确地描述 Must-link 和 Cannot-link,本文引入同义词的概念,对每个 Must-link 和 Cannot-link,通过同义词对约束条件进行扩展。

基于 Gibbs-LDA 和两种先验知识的约束条件,在计算词项 w 属于某个主题 z 的概率分布时,利用先验知识产生的约束条件对其进行修正,使得属于与特征类联系紧密的词项被划分到特征类所在主题,即:在主题更新过程中,加入一个约束条件影响因子 $\delta(w_i, z_j)$ 作为先验知识对模型进行调和,如式(11)所示:

$$P(z_i = j | z_{-i}, w, \alpha, \beta) = \delta(w_i, z_j) \left(\frac{n_{-i,j}^{(w)} + \beta}{\sum_{w'} n_{-i,j}^{(w')} + W\beta} \right) \left(\frac{n_{-i,j}^{(d_i)} + \alpha}{\sum_j n_{-i,j}^{(d_i)} + T\alpha} \right) \quad (11)$$

对于文档中每个词语的主题更新,假设当前所计算的词是 w_i ,针对当前主题 z_j , $\delta(w_i, z_j)$ 计算方法如下:

① 如果 w_i 受 Must-link 约束,且约束特征为当前主题所表达的特征,则 $\delta(w_i, z_j) = 1$ 且 $\delta(w_i, z_{-j}) = 0$,其中 z_{-j} 表示除 z_j 外的其他主题。

② 如果 w_i 受 Cannot-link 约束,且约束特征为当前主题所表达的特征,则 $\delta(w_i, z_j) = 0$ 且 $\delta(w_i, z_{-j}) = 1$ 。

③ 其他情况下, $\delta(w_i, z_j | j = 1, 2, \dots, T) = 1$ 。

1.3 回复内容的支持度计算

用户回复对评论的支持度通过回复内容的情感值得到,而回复内容的情感值与评论内容的情感值也有关联,如表 1 所示。

表 1 不同情况下正负面回复对评论的支持或反对示例

| 评论 | 正面/负面 | 回复内容 | Confirm/Deny |
|-------|-------|-------------|--------------|
| 拍照很差劲 | 正面 | 确实是这样 | Confirm |
| | | 我看还不错呀 | Deny |
| | 负面 | 这拍照效果确实不怎么样 | Confirm |
| | | 是你技术不行吧 | Deny |

所以,对于负面的评价,需要对回复内容提取特征词,如果提取到与计算观点相似的特征词,则分析情感表述是否一致;否则,直接以回复内容的情感值表示是否是支持评论的观点。本文的情感分析工具使用 SnowNLP 的情感分析组件,通过对获取到的标记为 good 和 bad 的评论进行模型训练,可以很好地适用于本文的应用场景。

2 实验结果分析

本文从“中关村在线”网站获取数据,抓取相关产品的评论和回复信息,去除没有作者信息和评分等无效评论,过滤掉投票数低于 50 的评论,得到最终的数据集如表 2 所示。以有用投票率(有用投票数与总投票数的比值)作为评论有用性的评价标准,采用平均有用投票率作为区分是否有用评论的阈值。

表 2 评论数据集

| 产品 | 评论数 | 有用评论数 | 无用评论数 | 有用投票率阈值 |
|--------------|-----|-------|-------|---------|
| 苹果 iPhone 6 | 154 | 88 | 66 | 0.71 |
| 小米红米 Note | 103 | 51 | 52 | 0.80 |
| 三星 Galaxy S4 | 270 | 171 | 99 | 0.72 |
| 尼康 D7000 | 94 | 56 | 38 | 0.62 |

2.1 评价指标

由于评论投票数的有偏性,本文首先将评论的有用性当作二分类问题,通过分类问题常用的准确率、召回率和 F 值,将本文的 OSUD 模型与 LR (Linear Regression) 模型、SVR 模型^[14] 和 REDQ (Ranking based on Estimated Discussion Quality)^[19] 模型效果进行对比;接着将其当作排序问题,利用常用的排序准确率 $P@10$ 进行比较;最后通过实例分析本文模型的实验效果。

令 R 是所有评论的集合, H 表示真正有用的评论集合, U 是真实无用的评论,而 H^* 表示实验预测为有用的评论集合, U^* 表示实验预测为无用的评论集合。准确率、召回率和 F 值分别定义如下:

$$Precision = \frac{|H|}{|R|} \times \frac{|H \cap H^*|}{|H^*|} + \frac{|U|}{|R|} \times \frac{|U \cap U^*|}{|U^*|} \quad (12)$$

$$Recall = \frac{|H|}{|R|} \times \frac{|H \cap H^*|}{|H|} + \frac{|U|}{|R|} \times \frac{|U \cap U^*|}{|U|} \quad (13)$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

2.2 实验结果分析

2.2.1 准确率、召回率与 F 值比较

使用三种模型对评论数据集进行有用性预测,得到准确率、召回率、 $F1$ 值和排序准确率 $P@10$ 的实验结果如表 3 所示。

按平均准确率从高到低排名为:OSUD、SVR、LR、REDQ。

按平均召回率从高到低排名为:SVR、OSUD、LR、REDQ。

按平均 $F1$ 值从高到低排名为:SVR、OSUD、LR、REDQ。

按平均 $P@10$ 值从高到低排名为:OSUD、REDQ、SVR、LR。

对于调和平均 $F1$ 值:REDQ 取得的效果最差,OSUD 的

效果与 SVR 相当且整体上优于 LR 模型,且对于评论数据集存在有偏的情况下效果优于 SVR 模型,即 OSUD 更偏向于预测有用性的评论,这是由于模型计算过程中更倾向于正面情感导致的。REDQ 效果较差的原因是本文实验比较的时候只使用了其回复的统计信息,而原文中是将其当作搜索排序问题,加入了关键词的影响,将统计信息作为辅助才能达到较好的实验效果,所以可看出 REDQ 本身对评论有用性的影响意义不大。

通过 $P@10$ 衡量前 10 条评论的排序准确率:OSUD 整体上的排序准确率高于其他三种方法,特别是对于数据有偏的情况下,可以取得更好的效果;SVR 和 LR 对于数据分布均匀的情况下可以训练出较好的模型,但由于数据集较小,难以学习得到更好的效果。

实验结果验证了本文 OSUD 模型的有效性。从实验结果可看出,本文 OSUD 模型基于观点支持度的思想,从另一个层面度量了评论的有用投票率,而本文基于有用投票率作为验证指标,SVR 与 LR 模型也是对评论有用投票率进行拟合,这也很好说明了本文 OSUD 模型通过对评论和回复观点的合理建模,可以达到甚至优于通过机器学习训练得到的效果。

表 3 实验结果比较

| 比较项目 | 方法 | 实验数据 | | | | |
|--------|------|--------------|--------------|--------------|--------------|--------------|
| | | 苹果 iPhone6 | 小米红米 Note | 三星 Galaxy S4 | 尼康 D7000 | |
| 准确率 | OSUD | 0.764 | 0.722 | 0.802 | 0.755 | 0.761 |
| | REDQ | 0.536 | 0.618 | 0.479 | 0.520 | 0.538 |
| | SVR | 0.732 | 0.794 | 0.758 | 0.747 | 0.758 |
| | LR | 0.683 | 0.788 | 0.712 | 0.741 | 0.731 |
| 召回率 | OSUD | 0.641 | 0.600 | 0.654 | 0.630 | 0.631 |
| | REDQ | 0.364 | 0.327 | 0.393 | 0.289 | 0.343 |
| | SVR | 0.652 | 0.670 | 0.624 | 0.639 | 0.646 |
| | LR | 0.627 | 0.640 | 0.576 | 0.619 | 0.616 |
| $F1$ 值 | OSUD | 0.697 | 0.655 | 0.720 | 0.687 | 0.690 |
| | REDQ | 0.434 | 0.428 | 0.432 | 0.372 | 0.417 |
| | SVR | 0.690 | 0.727 | 0.685 | 0.689 | 0.698 |
| | LR | 0.654 | 0.706 | 0.637 | 0.675 | 0.668 |
| $P@10$ | OSUD | 0.900 | 0.800 | 0.900 | 1.000 | 0.900 |
| | REDQ | 0.600 | 0.500 | 0.700 | 0.500 | 0.575 |
| | SVR | 0.800 | 0.930 | 0.840 | 0.750 | 0.830 |
| | LR | 0.770 | 0.910 | 0.810 | 0.780 | 0.818 |

2.2.2 实例分析

表 4 为评论实例,其中有用投票数为 70,总投票数 79,得到有用投票率为 0.886。

表 4 评论实例

| 分类 | 评价 |
|------|--|
| 优点 | 网速快,内存大、手机薄、屏幕大 |
| 缺点 | 没有 4G 网络、像素可不可以不到 1300 万啊、电池毫安太少了,要是电池在大点就更好了 |
| 整体评价 | 苹果 6S 是可不可以 4G 网络啊、像素可不可以再高一点、要是网速再快点那就更好了,价钱太贵了能不能再便宜点而且屏幕还是 IPS 技术、总之不是很好、不过比以前的手机好了很多 |

得到特征观点对并对其观点有用性进行分析如表 5 所示。

表 5 实例评论的特征观点对及有用性分析

| 特征 | 观点 | 正面评论数 | 负面评论数 | 有用性 |
|----|----|-------|-------|-------|
| 网速 | 正面 | 2 | 4 | -0.13 |
| 内存 | 正面 | 4 | 17 | -0.31 |
| 屏幕 | 正面 | 47 | 15 | 0.46 |
| 像素 | 负面 | 7 | 28 | 0.12 |
| 电池 | 负面 | 14 | 35 | 0.38 |
| 价格 | 负面 | 3 | 58 | 0.72 |

针对每个观点, 计算其观点的有用性, 下面实例展示对“网速”持正面观点的有用性计算, 其统计信息如表 6 所示。

表 6 对“网速”持正面观点所对应的统计信息

| 观点 | 评论数 | Confirming Replies | Denying Replies |
|----------|-----|--------------------|-----------------|
| Agree | 2 | 15 | 7 |
| Disagree | 4 | 23 | 5 |

为了直观地展示评论有用性的计算, 此处不考虑评论的权重 $dvr(v)$, 则可得到针对“网速”正面评价的支持度为 10:

$$\text{Suppose(网速)} = \sum_{i=1}^n \left(1 + \left(\sum_{j=1}^{m_i} 1 - \sum_{k=1}^{n_i} 1 \right) \right) = 2 + 15 - 7 = 10$$

同样, 得到其否定度为 22:

$$\text{Oppose(网速)} = \sum_{i=1}^n \left(1 + \left(\sum_{j=1}^{m_i} 1 - \sum_{k=1}^{n_i} 1 \right) \right) = 4 + 23 - 5 = 22$$

进而, 针对“网速”正面评价的观点有用性为 -0.375, 表示用户对该观点持负面态度, 即不具有有用性。

$$\text{Helpfulness(网速)} = \frac{10 - 22}{10 + 22} = -0.375$$

同样, 不考虑观点的权重 $dvr(op)$, 直接对该评论所含观点的有用性进行求和得到评论有用性为 1.24。

从结果可看出, 该评论含有两条相对无用的观点, 4 条有用的观点, 求和得到最终评论的有用性, 使得含有有用观点越多, 最终评论有用性越高, 相反, 无用观点越多, 最终评论有用性越低; 而且中间计算结果可以很直观地反映出用户对每个观点的支持度, 相对于简单的观点摘要, 观点的有用性囊括了其他回复用户对该观点的支持度, 更加真实可信。

3 结语

本文基于观点支持度度量的方法从评论回复的角度分析评论有用性问题, 验证了用户回复信息所表达的语义对评论有用性的影响, 且通过本文模型, 基于观点的有用性可以很好地解释评论有用性问题。本文模型也可作为预测模型的一个特征综合其他统计语义等特征训练得到更好的实验效果。

作者排名衡量每个用户发表意见的影响值, 排名值的合理性可以提高算法的准确率, 本文由于缺乏作者排名的数据, 将所有作者均衡处理, 有一定的缺陷。评论的回复有很大的随意性, 对于无关句的处理需要进一步的实验和验证。

参考文献:

- [1] 林煜明, 王晓玲, 朱涛, 等. 用户评论的质量检测与控制研究综述[J]. 软件学报, 2014, 25(3): 506–527. (LIN Y M, WANG X L, ZHU T, et al. Survey on quality evaluation and control of online reviews[J]. Journal of Software, 2014, 25(3): 506–527.)
- [2] LIU B. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1–167.
- [3] JINDAL N, LIU B. Opinion spam and analysis[C]// Proceedings of the 2008 International Conference on Web Search and Data Mining. New York: ACM, 2008: 219–230.
- [4] 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. New York: ACM, 2010: 939–948.
- [5] JINDAL N, LIU B, LIM E P. Finding unusual review patterns using unexpected rules[C]// Proceedings of the 19th ACM International Conference on Information and Knowledge Management. New York: ACM, 2010: 1549–1552.
- [6] WANG G, XIE S, LIU B, et al. Identify online store review spammers via social review graph[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(4): 61.
- [7] MUKHERJEE A, LIU B, WANG J, et al. Detecting group review spam[C]// Proceedings of the 20th International Conference Companion on World Wide Web. New York: ACM, 2011: 93–94.
- [8] MUKHERJEE A, LIU B, BLANCE N. Spotting fake reviewer groups in consumer reviews[C]// Proceedings of the 21st International Conference on World Wide Web. New York: ACM, 2012: 191–200.
- [9] 黄婷婷, 曾国荪, 熊焕亮. 基于商品特征关联度的购物客户评论可信排序方法[J]. 计算机应用, 2014, 34(8): 2322–2327. (HUANG T T, ZENG G S, XIONG H L. Trustworthy sort method for shopping customer reviews based on correlation degree with product features[J]. Journal of Computer Applications, 2014, 34(8): 2322–2327.)
- [10] CHOSE A, IPEIROTIS P G. Designing novel review ranking systems: predicting the usefulness and impact of reviews[C]// Proceedings of the Ninth International Conference on Electronic commerce. New York: ACM, 2007: 303–310.
- [11] LIU Y, HUANG X, AN A, et al. Modeling and predicting the helpfulness of online reviews[C]// Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. Piscataway, NJ: IEEE, 2008: 443–452.
- [12] ZENG Y-C, KU T, WU S-H, et al. Modeling the helpful opinion mining of online consumer reviews as a classification problem[J]. Computational Linguistics and Chinese Language Processing, 2014, 19(2): 17–32.
- [13] HSIEH H Y, WU S H. Ranking online customer reviews with the SVR model[C]// Proceedings of the 2015 IEEE International Conference on Information Reuse and Integration. Piscataway, NJ: IEEE, 2015: 550–555.
- [14] XU B, ZHAO T-J, WU J-W, et al. Automatically ranking reviews based on the ordinal regression model[C]// AICI 2011: Proceedings of the Third International Conference on Artificial Intelligence and Computational Intelligence. Berlin: Springer, 2011: 126–134.
- [15] HU M, LIU B. Mining and summarizing customer reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2004: 168–177.
- [16] LIU B, HU M, CHENG J. Opinion observer: analyzing and comparing opinions on the Web[C]// Proceedings of the 14th International Conference on World Wide Web. New York: ACM, 2005: 342–351.

(下转第 2776 页)

4 结语

在自然语言处理过程中,仅仅依靠统计或者是规则方法来发现新词难以满足要求,所以本文采用统计与规则相结合的方法来发现新词。由于微博数据的文本一般较短,对其进行新词发现会存在数据稀疏的问题,而 N-Gram 算法能够较好地解决微博数据稀疏的问题,所以本文在 N-Gram 算法基础上,融合词语的互信息和邻接熵的统计量方法,以及词频和停用字等过滤规则相结合进行微博新词发现。实验结果表明,本文方法对微博新词的发现效果有所提高。但是,由于微博数据量的庞大,采用 N-Gram 算法对其进行分词会产生大量的词串,在实验的过程中发现其耗费的时间较长。今后将针对这个问题对其进行相应的改进,使其效率提高。

参考文献:

- [1] 张海军, 史树敏, 朱朝勇, 等. 中文新词识别技术综述[J]. 计算机科学, 2010, 37(3): 6–12. (ZHANG H J, SHI S M, ZHU C Y, et al. Survey of Chinese new words identification[J]. Computer Science, 2010, 37(3): 6–12.)
- [2] PAVEL P, PAVEL S. Combining association measures for collocation extraction [C]// COLING-ACL 2006: Proceedings of the COLING/ACL on Main Conference Poster Sessions. Stroudsburg: Association for Computational Linguistics, 2006: 651–658.
- [3] 丁溪源. 基于大规模语料的中文新词抽取算法的设计与实现[D]. 南京: 南京理工大学, 2011: 44–45. (DING X Y. Design and implementation of Chinese new words extraction algorithm based on large scale corpus[D]. Nanjing: Nanjing University of Science and Technology, 2011: 44–45.)
- [4] 陈飞, 刘奕群, 魏超, 等. 基于条件随机场方法的开放领域新词发现[J]. 软件学报, 2013, 24(5): 1051–1060. (CHEN F, LIU Y Q, WEI C, et al. Open domain new word detection using condition random field method[J]. Journal of Software, 2013, 24(5): 1051–1060.)
- [5] 李文坤, 张仰森, 陈若愚. 基于词内部结合度和边界自由度的新词发现[J]. 计算机应用研究, 2015, 32(8): 2302–2304. (LI W K, ZHANG Y S, CHEN R Y. New word detection based on inner combination degree and boundary freedom degree of word[J]. Application Research of Computers, 2015, 32(8): 2302–2304.)
- [6] 李明. 针对特定领域的中文新词发现技术研究[D]. 南京: 南京航空航天大学, 2012: 35–40. (LI M. New words discovery re-

(上接第 2771 页)

- [17] 庾中凯, 郑小林, 吴亚峰, 等. 基于用户评论挖掘的产品推荐算法[J]. 浙江大学学报(工学版), 2013, 47(8): 1475–1485. (HU Z K, ZHENG X L, WU Y F, et al. Product recommendation algorithm based on users' reviews mining[J]. Journal of Zhejiang University (Engineering Science), 2013, 47(8): 1475–1485.)
- [18] RYANG H, YUN U. Ranking book reviews based on user discussion[M]// PARK J J, ADELI H, PARK N, et al. Mobile, Ubiquitous, and Intelligent Computing. Berlin: Springer, 2014: 7–11.
- [19] RYANG H, YUN U. Ranking method for book reviews based on estimated discussion quality[M]// PARK J J, STOJMENOVIC I, JEONG H Y, et al. Computer Science and Its Applications. Berlin: Springer, 2015: 171–177.

search for specific areas[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 35–40.)

- [7] 林自芳, 蒋秀凤. 基于词内部模式的新词识别[J]. 计算机与现代化, 2010(11): 162–164. (LIN Z F, JIANG X F. A new method for Chinese new word identification based on inner pattern of word [J]. Computer and Modernization, 2010(11): 162–164.)
- [8] 霍帅, 张敏, 刘奕群, 等. 基于微博内容的新词发现算法[J]. 模式识别与人工智能, 2014, 27(2): 141–145. (HUO S, ZHANG M, LIU Y Q. New words discovery in microblog content[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(2): 141–145.)
- [9] 周超, 严馨, 余正涛, 等. 融合词频特性及邻接变化数的微博新词识别[J]. 山东大学学报(理学版), 2015, 50(3): 6–10. (ZHOU C, YAN X, YU Z T, et al. Weibo new word recognition combining frequency characteristic and accessor variety[J]. Journal of Shandong University (Natural Science), 2015, 50(3): 6–10.)
- [10] 崔世起. 中文新词检测与分析[D]. 北京: 中国科学院研究生院, 2006: 15–22. (CUI S Q. Chinese new words detection and analysis [D]. Beijing: Graduate University of Chinese Academy of Sciences, 2006: 15–22.)
- [11] YE Y M, WU Q Y, LI Y, et al. Unknown Chinese word extraction based on variety of overlapping strings[J]. Information Processing & Management, 2013, 49(2): 497–512.
- [12] HUANG J H, POWERS D. Chinese word segmentation based on contextual entropy[C]// Proceedings of the 17th Asian Pacific Conference on Language, Information and Computation. Singapore: [s. n.], 2003: 152–158.

Background

This work is partially supported by the National Science and Technology Support Program of China (2013BAB06B04), the Natural Science Foundation of Jiangsu Province (BK20130852), the Science and Technology Project of Jiangsu Water Resources Department (2013025), the Huaneng Group Company Headquarters Technology Project of China (HNKJ13-H17-04).

YAO Rongpeng, born in 1989, M. S. candidate. His research interests include big data, data management.

XU Guoyan, born in 1971, Ph. D. professor. Her research interests include big data, data management.

SONG Jian, born in 1991, M. S. candidate. His research interests include big data, data management.

- [20] WIDYANTORO D, WIBISONO Y. Modeling credibility assessment and explanation for tweets based on sentiment analysis [J]. Journal of Theoretical and Applied Information Technology, 2014, 70(3): 540–548.

Background

This work is partially supported by the National Natural Science Foundation of China (90818028).

LI Xueming, born in 1967, Ph. D., professor. His research interests include data mining, big data processing.

ZHANG Chaoyang, born in 1991, M. S. candidate. Her research interests include data mining, natural language processing.

SHE Weijun, born in 1991, M. S. candidate. His research interests include data mining, natural language processing.