

张倩,瞿有利.用于网络评论分析的主题-对立情感挖掘模型[J].计算机科学与探索,2013,7(7):620-629.

ISSN 1673-9418 CODEN JKTYA8  
Journal of Frontiers of Computer Science and Technology  
1673-9418/2013/07(07)-0620-10  
doi: 10.3778/j.issn.1673-9418.1305015

E-mail: fcst@vip.163.com  
<http://www.ceaj.org>  
Tel: +86-10-89056056

## 用于网络评论分析的主题-对立情感挖掘模型\*

张 倩<sup>+</sup>,瞿有利

北京交通大学 计算机与信息技术学院,北京 100044

### Topic-Opposite Sentiment Mining Model for Online Review Analysis\*

ZHANG Qian<sup>+</sup>, QU Youli

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

+ Corresponding author: E-mail: carol.zhangqian@gmail.com

ZHANG Qian, QU Youli. Topic-opposite sentiment mining model for online review analysis. *Journal of Frontiers of Computer Science and Technology*, 2013, 7(7): 620-629.

**Abstract:** At mining product topics and opposite sentiment information of the topics from online reviews, in order to help manufactures and service providers improve their products and services, and help customers make decisions, this paper proposes a topic model called topic-opposite sentiment mining model (TOSM) based on latent Dirichlet allocation (LDA), which assumes that all words in a single sentence are generated from one topic and one sentiment. This paper extends LDA to TOSM with adding the sentiment layer, so that TOSM can detect topics and opposite sentiment of topic simultaneously from reviews. Moreover, this paper uses sentiment lexicon in TOSM to make the opposite sentiment represented clearly. This paper does three experiments with the reviews of electronic devices from Amazon and the reviews of restaurants from Yelp. The experimental results show that the topic-sentiment found by TOSM matches evaluative details of the reviews, and TOSM outperforms other generative models.

**Key words:** topic model; latent Dirichlet allocation (LDA); sentiment; review mining; topic-opposite sentiment mining model (TOSM)

**摘 要:**为了挖掘网络评论中的产品主题和主题的对立情感信息,以帮助生产商和服务商改进产品和服务质量,帮助消费者做出购买决策,基于LDA(latent Dirichlet allocation)提出了一个用于网络评论分析的主题-对

---

\* The Fundamental Research Funds for the Central Universities of China under Grant No. 2011JBM231 (中央高校基本科研业务费专项资金).

Received 2013-03, Accepted 2013-05.

CNKI网络优先出版:2013-05-29, <http://www.cnki.net/kcms/detail/11.5602.TP.20130529.1347.001.html>.

立情感挖掘模型(topic-opposite sentiment mining model, TOSM), 模型中假设句子为分配主题和情感的最小单位。该模型在LDA的基础上增加情感层, 将LDA的三层结构拓展为四层, 能同时得到主题以及主题的对立情感信息。为了使对立情感的描述更准确, 在情感层中融入了情感词典先验信息。在Amazon网站的电子产品评论和Yelp网站的饭店评论数据集上进行了三组实验, 实验表明, TOSM挖掘到的观点主题与评论中有价值的细节描述相匹配, TOSM模型的情感分类结果优于其他模型。

**关键词:** 主题模型; LDA; 情感; 评论挖掘; 主题-对立情感挖掘模型(TOSM)

**文献标志码:** A    **中图分类号:** TP391.1

## 1 引言

Royal Pingdom公司《2011年全球互联网产业发展状况报告》指出, 全球互联网用户已达21亿。随着互联网的广泛应用, 越来越多的人在网上发表自己对某种产品或服务的评论。有研究<sup>[1]</sup>表明, 丰富的网络评论比传统媒体更有经济价值。对网络上的评论进行<主题, 情感>挖掘, 不仅能够帮助其他消费者了解产品, 并做出购买决策; 还可以帮助企业对用户偏好进行有效分析, 从而为用户提供更好的服务, 提高自身企业的竞争力。因此, 如何从大量网络评论中迅速有效地提取出评论的主题和情感尤其重要。

机器学习技术被广泛应用于情感分类。Turney<sup>[2]</sup>提出了一种利用搜索引擎估计短语褒贬倾向的方法, 进而使用篇章中短语极性的平均值代表整体的情感倾向, 但分类精度低于有监督学习方法<sup>[3]</sup>。Pang等人<sup>[4]</sup>用传统的机器学习方法进行情感倾向的挖掘工作, 以互联网上的电影评论文本作为语料, 采用了不同的特征选择方法, 应用朴素贝叶斯、最大熵、支持向量机(support vector machine, SVM)对电影评论进行分类。他们在之后的工作<sup>[5]</sup>中采用主观检测器, 并只对评论的主观部分进行分类, 使得情感分类精确度得以提升。Pang等人<sup>[4-5]</sup>使用的电影评论后来成为许多研究<sup>[6-7]</sup>的基准。之后的情感分类方法的研究<sup>[8-10]</sup>着重于同时应用于多领域。

以上这些方法存在两个缺陷:

(1) 只关注情感分类, 而不考虑评论中的主题分布, 使得挖掘结果的有效性受限。

(2) 大多数方法<sup>[4-5, 8, 10]</sup>是监督学习方法, 需要对手工标注的评论进行模型训练, 而手工标注的评论难以获取。

无监督的主题模型TAM(topic-aspect model)<sup>[11]</sup>、TSM(topic sentiment mixture)<sup>[12]</sup>、ASUM(aspect and sentiment unification model)<sup>[13]</sup>和JST(joint sentiment/topic model)<sup>[14]</sup>能够同时抽取评论中的主题和情感。TAM和TSM认为主题分布和情感分布是独立的, 这样的设置不能反映主题和情感间的相互影响。ASUM和JST考虑了主题和情感的相关性, 对<情感, 主题>对进行建模, 这种方法能同时获取评论的主题和该主题的某个情感信息, 但是不能保证获取到每个主题的两个对立情感信息。

本文提出了一个新的概率主题模型: 主题-对立情感挖掘模型(topic-opposite sentiment mining model, TOSM)。该模型采用文档-主题-情感-词四层产生式结构, 对于文档中每一个词的产生, 先从文档-主题分布中选择主题, 再从主题-情感分布中选择情感, 主题和情感确定后选择词。上述词的产生过程中, 选定主题后, 在主题-情感分布中选择情感时, 可以选择到正面情感和负面情感, 因此该模型能同时获得评论的主题及主题的两个对立情感信息。通过模型计算将每一篇评论表示为主题的概率分布, 每个主题又表示为情感的概率分布。

## 2 主题-对立情感挖掘模型

主题-对立情感挖掘模型是一个产生式主题模型。评论中描述同一主题的词趋向于邻近的词, 因此模型中假设句子为分配主题和情感的最小单位<sup>[13]</sup>。该模型在LDA(latent Dirichlet allocation)<sup>[15]</sup>的基础上增加情感层, 将LDA的三层结构拓展为四层, 即文档层、主题层、情感层和词层。为了使对立情感的描述更准确, 在情感层中融入了情感词典先验信息。

2.1 主题-对立情感挖掘模型描述

将TOSM产生一篇评论的过程形象地描述为评论者写评论的过程。写一篇有关饭店的评论,评论者首先确定要描述的主题的分布,如:该评论30%关于服务水平,50%关于食物,20%关于价格。然后再决定每一个主题的情感分布,如:对于服务水平50%满意,50%不满意;对于食物70%满意,30%不满意;对于价格60%满意,40%不满意。最后,在每个句子表示同一个主题和同一个情感的规则下完成整篇评论。

TOSM模型的图形表示如图1所示。其中空心点表示隐含变量,实心点表示可观察值,虚线代表情感词典的引入。表1为模型变量的解释。

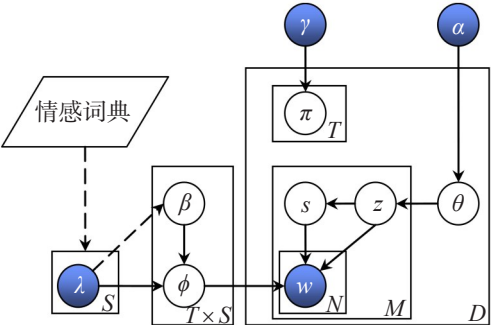


Fig.1 TOSM model  
图1 TOSM模型

该模型生成一篇评论的过程如下:

- (1)对于每一个主题  $z$  和情感  $s$ ,抽取词分布  $\phi_{zs} \sim \text{Dirichlet}(\beta_s)$ , 其中  $z \in \{1, 2, \dots, T\}, s \in \{1, 2, \dots, S\}$ 。
- (2)对每一篇文档  $d, d \in \{1, 2, \dots, D\}$ :
  - (2.1)抽取文档的主题分布  $\theta_d \sim \text{Dirichlet}(\alpha)$ 。
  - (2.2)对每一个主题,抽取情感分布  $\pi_{dz} \sim \text{Dirichlet}(\gamma)$ 。
  - (2.3)对每一个句子:
    - ①选取主题  $k \sim \text{Multinomial}(\theta_d)$ 。
    - ②主题为  $k$  时,选取情感  $j \sim \text{Multinomial}(\pi_{dk})$ 。
    - ③选取单词  $w \sim \text{Multinomial}(\phi_{kj})$ 。

2.2 情感词典

按情感进行分类,词可分为两类:情感词和普通词,其中情感词又分为正面情感词和负面情感词。在不同的主题中能表达不同情感的词属于普通词;在不同的主题中只能表达一个情感的词称为情感词。如:普通词“camera, notebook, look”,情感词

Table 1 The variables of TOSM and their explanations

表1 TOSM模型变量及其含义

变量	含义
$D$	语料库中的文档数,即评论数
$M$	文档中的句子数
$N$	句子中的词数
$S$	情感的个数
$T$	主题的个数
$V$	词汇表的大小
$z$	主题
$s$	情感
$w$	词
$W$	词汇表里词的编号最大值
$\phi$	词的多项式分布
$\theta$	主题的多项式分布
$\pi$	情感的多项式分布
$\alpha(k)$	$\theta$ 的 Dirichlet 先验参数
$\beta_s(w)$	$\phi$ 的 Dirichlet 先验参数
$\gamma_s(j)$	$\pi$ 的 Dirichlet 先验参数
$\lambda_{sw}(w)$	引入情感词典后选择 $\beta$ 的转换函数

“good, bad, annoying, excellent”,并且“good, excellent”不可能出现在负面情感类别中,“bad, annoying”不会出现在正面情感类别中。

本文使用的情感词典 PARADIGM 来自 Turkey<sup>[16]</sup>。词典分为两部分:正面情感词典和负面情感词典,其中每一部分包括7个强情感词,其他为普通情感词,强情感词表达的情感比普通情感词更强烈。词典如表2所示,表中用粗体标注了强情感词,并且词典中词的先后顺序无影响。

2.3 参数λ和β的设置

设计参数λ和β是为了将词的情感先验信息融入到模型中,通过先验信息对词的产生概率进行奖惩。参数λ检测对某个词是否奖惩,参数β决定对该词的奖惩强度。

当前情感为  $s$  时,对于每一个词  $w$ ,根据以下公式计算  $\lambda_{sw}(w, s)$ :

$$\lambda_{sw}(w, s) = \begin{cases} 1, & S(w) \text{ 与当前情感 } s \text{ 相同, 对该词奖励} \\ -1, & S(w) \text{ 与当前情感 } s \text{ 对立, 对该词惩罚} \\ 0, & \text{普通词, 不进行奖惩} \end{cases}$$

Table 2 Sentiment lexicons PARADIGM

表2 情感词典 PARADIGM

分类	情感词
正面情感词典	<b>good, nice, excellent, positive, fortunate, correct, superior,</b> amazing, best, fun, awesome, comfortable, enjoy, fantastic, favorite, glad, great, happy, impressive, love, perfect, thank, attractive, recommend, satisfy, worth
负面情感词典	<b>bad, nasty, poor, negative, unfortunate, wrong, inferior,</b> annoying, complain, disappointed, hate, junk, mess, not like, not recommend, not worth, problem, regret, sorry, terrible, upset, trouble, unacceptable, not good, waste, worst, worthless

其中, 函数  $S(w)$  是单词  $w$  所在的情感词典的情感。

$\beta_{sw}$  是情感为  $s$  时, 词  $w$  的分布先验参数。根据以下公式计算  $\beta_{sw}(\lambda_{sw}(w, s))$ :

$$\beta_{sw}(\lambda_{sw}(w, s)) = \begin{cases} \beta_1, \lambda_{sw}(w, s) = 1, \beta_1 \text{ 为奖励参数} \\ \beta_2, \lambda_{sw}(w, s) = -1, \beta_2 \text{ 为惩罚参数} \\ \beta_3, \lambda_{sw}(w, s) = 0, \beta_3 \text{ 为普通参数} \end{cases}$$

以  $w = \text{excellent}$  为例,  $S(w) = \text{正面情感}$ 。当前情感  $s$  为正面情感时,  $\lambda_{sw}(w, s) = 1, \beta_{sw}(\lambda_{sw}(w, s)) = \beta_1$ , 对  $\text{excellent}$  进行奖励; 当前情感  $s$  为负面情感时,  $\lambda_{sw}(w, s) = -1, \beta_{sw}(\lambda_{sw}(w, s)) = \beta_2$ , 对  $\text{excellent}$  进行惩罚。因此, 词  $\text{excellent}$  在正面情感下的产生概率最大。

如果  $w = \text{book}$  等普通词, 当前情感  $s$  为正面情感或负面情感,  $\lambda_{sw}(w, s)$  都为 0, 此时  $\beta_{sw}(\lambda_{sw}(w, s)) = \beta_3$ 。

## 2.4 模型推理

本文使用 Gibbs 抽样算法<sup>[17]</sup>对 TOSM 进行参数估计和统计推理。为了获得  $\theta, \pi$  和  $\phi$  三个分布, 首先估算  $z$  和  $s$  的后验分布。第  $t$  个句子在给定主题和情感时的抽样分布为  $P(z_t = k, s_t = j | w, z_{-t}, s_{-t}, \alpha, \beta, \gamma)$ , 其中  $z_{-t}$  和  $s_{-t}$  分别表示文档  $d$  中除了第  $t$  个句子以外的所有句子的主题和情感的取值向量。

词、主题、情感的联合概率可以分解为以下三项:

$$P(w, s, z) = P(w | s, z) P(s, z) = P(w | s, z) P(s | z) P(z)$$

(1)  $P(w | s, z)$  引入分布  $\phi$  后, 得到:

$$P(w | s, z) = \left( \frac{\Gamma(V\beta)}{\prod_{v=1}^V \Gamma(\beta)} \right)^{S \times T} \frac{\Gamma(\sum_{w=1}^W C_{k j w}^{TSW} + \beta_{jw})}{\Gamma(\sum_{w=1}^W (C_{k j w}^{TSW} + \beta_{jw}) + N_i)} \times \prod_{w=1}^W \frac{\Gamma(C_{k j w}^{TSW} + \beta_{jw} + N_{iw})}{\Gamma(C_{k j w}^{TSW} + \beta_{jw})}$$

其中,  $C_{k j w}^{TSW}$  表示主题为  $k$  并且情感为  $j$  的词  $w$  的个数;  $N_i$  表示句子  $i$  中所有词的个数;  $N_{iw}$  表示句子  $i$  中词  $w$  的个数;  $\beta_{jw}$  表示词  $w$  的分布先验参数。

(2)  $P(s | z)$  引入分布  $\pi$  后, 得到:

$$P(s | z) = \left( \frac{\Gamma(\sum_{j=1}^S \gamma_{kj})}{\prod_{j=1}^S \Gamma(\gamma_{kj})} \right)^{D \times T} \prod_{d=1}^D \prod_{k=1}^T \frac{\Gamma(C_{dkj}^{DTS} + \gamma_{kj})}{\Gamma(C_{dk}^{DT} + \sum_{j=1}^S \gamma_{kj})}$$

其中,  $C_{dkj}^{DTS}$  表示在文档  $d$  中主题为  $k$  并且情感为  $j$  的句子的个数;  $C_{dk}^{DT}$  表示文档  $d$  中主题为  $k$  的句子的个数。

(3)  $P(z)$  引入分布  $\theta$  后, 得到:

$$P(z) = \left( \frac{\Gamma(\sum_{k=1}^T \alpha_k)}{\prod_{k=1}^T \Gamma(\alpha_k)} \right)^D \prod_{d=1}^D \frac{\Gamma(C_{dk}^{DT} + \alpha_k)}{\Gamma(C_d^D + \sum_{k=1}^T \alpha_k)}$$

其中,  $C_d^D$  表示文档  $d$  中的所有句子数。

将  $\theta, \pi, \phi$  边缘化后, 得到每次迭代抽样时文档中第  $t$  个句子的主题、情感抽样分布, 即主题、情感的条件后验概率:

$$P(z_t = k, s_t = j | w, z_{-t}, s_{-t}, \alpha, \beta, \gamma) \propto \frac{C_{dk}^{DT} + \alpha_k}{\sum_{k'=1}^T (C_{dk'}^{DT} + \alpha_{k'})} \times \frac{C_{dkj}^{DTS} + \gamma_{kj}}{\sum_{j'=1}^S (C_{dkj'}^{DTS} + \gamma_{kj'})} \times \frac{\Gamma(\sum_{w=1}^W C_{k j w}^{TSW} + \beta_{jw})}{\Gamma(\sum_{w=1}^W (C_{k j w}^{TSW} + \beta_{jw}) + N_i)} \times \prod_{w=1}^W \frac{\Gamma(C_{k j w}^{TSW} + \beta_{jw} + N_{iw})}{\Gamma(C_{k j w}^{TSW} + \beta_{jw})} \quad (1)$$

上述公式中,  $C_{dkj}^{DTS}, C_{k j w}^{TSW}, C_{dk}^{DT}$  计数时都要排除第  $t$  个句子。

文档  $d$  中主题为  $k$  的近似概率为:

$$\theta_{dk} = \frac{C_{dk}^{DT} + \alpha_k}{\sum_{k'=1}^T (C_{dk'}^{DT} + \alpha_{k'})} \quad (2)$$



文档 $d$ 中主题为 $k$ 时情感为 $j$ 的近似概率为:

$$\pi_{dkj} = \frac{C_{dkj}^{DTS} + \gamma_{kj}}{\sum_{j'=1}^S (C_{dkj'}^{DTS} + \gamma_{kj'})} \quad (3)$$

词 $w$ 的主题为 $k$ 并且情感为 $j$ 的近似概率为:

$$\phi_{kjw} = \frac{C_{kjw}^{TSW} + \beta_{jw}}{\sum_{w'=1}^V (C_{kjw'}^{TSW} + \beta_{jw'})} \quad (4)$$

### 3 实验及实验结果分析

为了评价 TOSM 模型,进行三组实验:

(1)分析情感分类精度和主题个数的关系。

(2)分析 TOSM 模型挖掘到的观点主题,所谓的观点主题即为带情感信息的主题。

(3)评估并比较 TOSM 模型与 ASUM、JST 和 baseline 的情感分类精度。

#### 3.1 实验数据集及预处理

实验中用到两个数据集:一个是 Amazon(<http://www.amazon.com>)网站有关电子产品的评论数据集,称为 Electronics;另一个是 Yelp(<http://www.yelp.com>)网站的饭店评论数据集,称为 Restaurants。

##### 3.1.1 数据集介绍

(1)Electronics 共有 7 种类别: Air Conditioner, Canister Vacuums, Coffee Machine, Digital SLRs, Laptops, MP3Player, Space Heater。Electronics 数据集共有 24 261 条评论,7 个类别的评论数均等。

(2)Restaurants 数据集通过随机选取 72 家饭店的评论,组成 25 461 条评论。

##### 3.1.2 数据集预处理步骤

(1)移除评论中的网址及非英文词。

(2)根据“.”、“!”、“?”以及换行符来分割句子,删除句中的标点符号,并删除超过 50 个词的长句。

(3)用 Porter Stemmer (<http://tartarus.org/~martin/PorterStemmer/>)算法<sup>[18]</sup>对数据集进行词根处理来减少词汇表的大小,并一定程度上预防数据稀疏问题,删除数据集中出现频率小于 2 的单词。

预处理后 Electronics 数据集的词汇表大小为 14 124, Restaurants 数据集的词汇表大小为 17 308。

#### 3.2 模型参数取值

TOSM 模型需要 3 个超参数:  $\alpha$ 、 $\beta$  和  $\gamma$ 。根据经验,设定实验中 3 个超参数取值为:  $\alpha=0.1$ ;  $\gamma=1$ ;  $\beta_1=0.1$ ,  $\beta_2=0$ ,  $\beta_3=0.001$ 。

有研究表明 LDA 模型使用非对称先验时的结果远远超过使用对称先验得到的结果<sup>[19]</sup>。模型 TOSM 中,  $\beta$  是观点主题的词分布先验参数,是  $T \times S \times V$  的向量,采用非对称形式,向量中的每一个  $\beta_{zsw} \in \{\beta_1, \beta_2, \beta_3\}$ 。 $\beta$  是将情感词典融入模型的一个关键因素,另外一个关键因素是模型初始化,初始化时需要给语料库中出现在情感词典中的情感词赋相应的情感,如果使用随机初始化,则达不到融入情感先验信息的效果。在正面情感下,出现在正面情感词典中的词使用奖励参数 0.1,出现在负面情感词典中的词使用惩罚参数 0,不出现在任何情感词典中的词使用普通先验参数 0.001,反之同理。这样就设定正面情感词典中的词不出现在负面情感类别下,负面情感词典中的词也不出现在正面情感类别下。

#### 3.3 情感分类精度计算方法

计算模型的情感分类精度需要评论的真实情感分类和模型计算出的评论情感分类。

##### 3.3.1 评论的真实情感分类

Electronics 和 Restaurants 数据集都采用 5 星打分制,打分为 1 星或 2 星的评论为负面评论,4 星或 5 星的评论为正面评论,打分为 3 星的为中立评论。由此得到每篇评论的真实情感分类。

##### 3.3.2 模型对评论的情感分类

(1)ASUM 和 JST 模型:通过分布  $\pi$  得到评论的情感分类,分布  $\pi$  中某评论的正面情感概率如果大于负面情感概率,则被认为是正面评论,反之同理。如果两个情感类别概率相等,该评论为中立评论。

(2)TOSM 模型:  $\pi$  分布是每个主题的对立情感分布,需要通过计算才能得到评论的情感分布。假设一个句子中的词属于同一情感类别,因此先计算评论中每个句子的情感分布,再得到整篇评论的情感分布来确定该评论的情感分类。

(3)baseline:每篇评论的情感分类是根据评论中包含正面情感词的句子数和包含负面情感词的句子

数来决定的。

得到数据集中评论的真实情感分类和模型计算的情感分类后,根据以下公式计算情感分类精度:

$$Accuracy = \frac{N_{\text{same}}}{N_{\text{total}}} \quad (5)$$

其中,  $Accuracy$  为情感分类精度;  $N_{\text{same}}$  为真实情感分类和模型计算的情感分类相同的评论数;  $N_{\text{total}}$  为数据集的评论总数。

### 3.4 实验1 情感分类精度和主题个数的关系

采用TOSM模型同时对主题和主题的对立情感进行建模,模型设定不同的主题个数会得到不一样的情感分类精度。实验1分析了在先验信息一样的情况下,主题个数取值对模型情感分类精度的影响。

#### 3.4.1 实验设定

用模型ASUM、JST和TOSM在Electronics和Restaurants数据集上进行实验,模型参数取值参见3.2节。每个模型主题个数取值为  $T \in \{10, 30, 50, 70, 90\}$ 。实验结果表明,每个模型在数据集上迭代5 000次后收敛。

#### 3.4.2 实验步骤

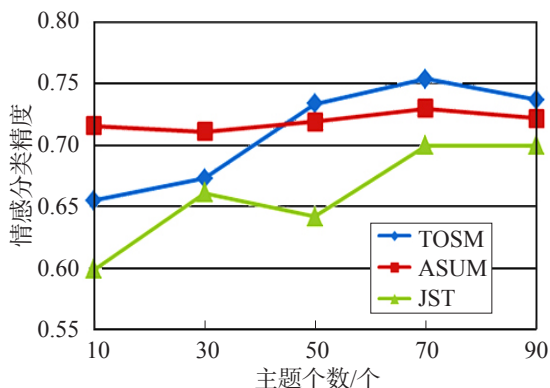
(1)根据上述设定,每个模型分别取不同的主题个数在两个数据集上进行实验,模型在数据集上收敛后保存模型计算得到的分布。

(2)参照式(5)计算模型的情感分类精度。

#### 3.4.3 实验结果与分析

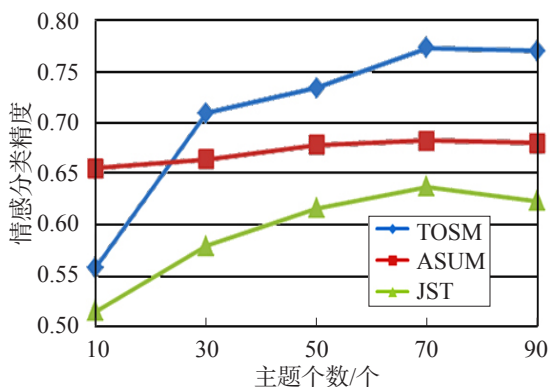
实验1的结果如图2所示。由图2可以看出,ASUM和TOSM在两个数据集上的表现都比JST模型好得多,在Restaurants中,TOSM和JST的情感分类精度最大差距达到15%。ASUM模型在不同的主题个数下,精度都维持在一定范围内,差距甚微,主题个数为70时精度达最高值。

TOSM在两个数据集中,当主题个数为10时,情感分类精度都低于ASUM模型,随着主题个数增加,精度提升并超过ASUM,在主题个数为70时精度最高。从图2(b)中可以看到主题个数为70时,TOSM的情感分类精度比ASUM高10%。3个模型的精度为最高值时主题个数为70。



(a) Sentiment classification accuracy on Electronics dataset

(a) Electronics数据集上的情感分类精度



(b) Sentiment classification accuracy on Restaurants dataset

(b) Restaurants数据集上的情感分类精度

Fig.2 Sentiment classification accuracy vs. different topic number setting

图2 模型在不同主题个数下的情感分类精度

图2中TOSM模型曲线表明,主题个数的选取对情感分类精度的影响很大,模型用于不同数据集时需要考虑数据集最适合的主题个数。

### 3.5 实验2 观点主题的发现

实验2的目的是对TOSM模型挖掘到的观点主题进行评价分析。TOSM模型每个主题下有两个对立的情感分类,本文定义结合了其中一个对立情感的分类的主题为观点主题。表3中Laptop-positive和Camera-negative类别都是一个观点主题,表中主题名称为手动标注。

#### 3.5.1 观点主题下词的贡献度

每一个观点主题下包含了许多词,有的词对观点主题的主题描述或情感表达贡献很大,而有的词

对观点主题几乎没有贡献,如代词、冠词等。本文借用文献[20]中的词项得分(term score)原理来计算观点主题下词的贡献度。包含词 $w$ 的观点主题越多,词 $w$ 对包含它的观点主题的贡献度越低。

根据式(6)计算观点主题 $n$ 下的每一个词 $w$ 的贡献度  $Contribution_{n,w}$  :

$$Contribution_{n,w} = P_{n,w} \text{lb} \frac{P_{n,w}}{\left(\prod_{n'=1}^{T \times S} P_{n',w}\right)^{\frac{1}{T \times S}}} \quad (6)$$

其中,  $n \in \{1, 2, \dots, T \times S\}$ ,  $w \in \{1, 2, \dots, W\}$ ,  $P_{n,w}$  为模型计算得到的观点主题 $n$ 下产生词 $w$ 的概率。

3.5.2 实验设定

用模型TOSM和ASUM在Electronics和Restaurants数据集上进行实验,模型参数取值参见3.2节。主题个数设定为70,模型在数据集上迭代5 000次后收敛。

3.5.3 实验步骤

- (1)TOSM(ASUM)模型在数据集上收敛后,计算每个观点主题(主题)下产生每个词的概率。
- (2)根据式(6)计算每个词的贡献度,并保存每个观点主题(主题)下贡献度前100的词。
- (3)每个观点主题(主题)下的词以贡献度由高到低排序,手动标注观点主题(主题)的名称。

3.5.4 实验结果与分析

表3列出了TOSM模型在Electronics数据集上发现的观点主题的部分结果。可以看到,每一个主

题分为两个有对立情感的观点主题。这样的结果不仅能使人们了解到评论中大家讨论了什么,还能对被讨论主题的优缺点进行对比,使得挖掘到的信息更全面。表4列出了ASUM模型在Electronics数据集上发现的主题的部分结果。可以看到,不是每一个主题都存在两个对立的情感信息,从而无法对评论中用户表述的情感进行全面的分析。

TOSM发现的观点主题与被评论对象有很高的相关性,同时其情感表达明确。如表3中列出的Laptop主题,出现的词都描述了笔记本的部件或性能。在正面情感下阐述了其电池待机时间长和运行速度快的优点,以及因为屏幕容易留下指印和灰尘,触控板使用不舒服而产生了负面情感。

TOSM挖掘到了更多的信息,导致主题的分类更细致。ASUM挖掘到综合的“screen”主题,包括“MacBook, MP3Player”等产品,如表4中的多个“screen”主题。但是TOSM却挖掘到了不同产品的“screen”主题,表3中列出了两个screen主题,由于用户对不同产品关注的特性不一样,表中结果可以看出两个主题用词的差别。

对于同样的产品,TOSM能挖掘到更多产品相关的信息。如表3中的Camera主题,TOSM挖掘到了“easi, amateur, learn, price, sunlight, color”等表述词,可以推断评论描述的是一个入门单反,评论认为价格合理并易于上手,适合初学者,但是在色彩表现方

Table 3 An example of discovering sentiment-topic by TOSM on Electronics dataset

表3 TOSM在数据集Electronics上发现的观点主题例子

Laptop		Screen(mac)		Screen(camera)		Camera	
positive	negative	positive	negative	positive	negative	positive	negative
batteri	dust	screen	notebook	imag	lcd	camera	pixel
hour	fingerprint	macbook	mac	color	view	photo	screen
power	scratch	color	inch	camera	screen	amateur	bright
life	glossy	widscreen	dust	superb	shoot	easi	color
bright	slipperi	stylish	glare	clear	photo	Nikon	glossi
fast	uncomfort	backlight	reflect	vivid	displai	price	sunlight
wifi	sharp	feel	screen	sharp	frame	novice	matt
laptop	glare	bright	matt	detail	zoom	overall	dim
game	trackpad	better	displai	good	mirror	upgrad	monitor
procesor	dust	stylish	fingerprint	pixel	disappoint	learn	finger

Table 4 An example of discovering topic by ASUM on Electronics dataset  
表4 ASUM在数据集Electronics上发现的主题例子

Positive				Negative			
screen	screen	screen	price	price	screen	service	ipod
macbook	screen	photo	worth	monei	screen	return	music
screen	bright	qualiti	monei	save	fingerprint	order	Ipod
brighter	resolut	superb	well	yourself	dust	back	download
gorgeou	color	screen	save	wast	glare	amazon	softwar
displai	sharp	vivid	cost	horribl	reflect	wait	itune
better	pictur	camera	cheaper	garbag	smaller	today	difficult
beauti	pixel	vivid	paid	spend	notebook	bought	without
glossi	glossi	brilliant	valu	suck	light	ipod	lose
color	crisp	view	extra	dollar	color	week	manual
movi	size	lcd	spend	heater	ssd	refund	playlist

Table 5 The sentiment classification accuracy of 4 models  
表5 4个模型的情感分类精度 (%)

模型	Restaurants	Electronics				
		Electronics overall	Canister Vacuums	Digital SLRs	Laptops	Space Heaters
TOSM	77.3	75.4	79.9	85.2	83.2	80.4
JST	68.2	70.0	71.2	67.4	71.0	70.2
ASUM	63.7	73.0	70.7	87.0	74.6	73.2
baseline	68.5	72.0	70.0	85.0	73.6	67.4

面有所欠缺。类似这样的考虑了消费者需求的观点主题,使得挖掘到的信息更有经济价值。

3.6 实验3 情感分类精度

实验3通过情感分类精度来评价 TOSM 模型发现观点主题的质量,对 TOSM、JST、ASUM 和 baseline 的情感分类精度进行对比分析。

3.6.1 实验设定

用模型TOSM、ASUM、JST和baseline在Electronics、Restaurants数据集和Electronics的7个子数据集上进行实验,模型参数取值参见3.2节。Electronics、Restaurants数据集的主题个数设为70,Electronics子数据集的主题个数设为20,每个模型在数据集上迭代5 000次后收敛。

3.6.2 实验步骤

(1)ASUM、JST对每个数据集收敛后,根据分布 $\pi$ 计算模型对评论的情感分类,再根据式(5)计算模

型的情感分类精度。

(2)TOSM对每个数据集收敛后,先计算模型对评论的情感分类,再根据式(5)计算模型的情感分类精度。

(3)baseline根据情感词典信息计算对评论的情感分类,再根据式(5)计算情感分类精度。

3.6.3 实验结果和分析

表5列出了4个模型的情感分类精度,以及Electronics的4个子类Canister Vacuums、Digital SLRs、Laptops和Space Heaters的结果,每个数据集中的最高精度用黑体标注。可以看出,JST在数据集上表现平平。TOSM模型虽然不是每个数据集上都有最高的精度,但是总体表现很出色,超过了其他几个模型。对比TOSM模型在Electronics上的表现,发现在子类数据集上的表现要比在整个数据集上好,Digital SLRs、Laptops和Space Heaters数据集上的精度超过了80%。



这是由于子类数据集相当于对原数据集上的主题进行了细分,该模型获得了更多信息,使得结果更优。

TOSM在Electronics、Restaurants数据集上的情感分类精度优于ASUM、JST和baseline,在Electronics的子数据集上的表现比整个数据集要好。

#### 4 结束语

本文提出了一个同时挖掘评论中的主题及主题的对立情感的产生式模型TOSM。模型假设句子为分配主题和情感的最小单位,在LDA的基础上增加情感层,将LDA的三层结构拓展为四层。为了使对立情感的描述更准确,在情感层中融入了情感词典先验信息。在情感分类实验中,TOSM的性能超越了其他模型,得到了较好的结果。

对将来的工作有三方面规划:

(1)研究TOSM模型应用到不同数据集时不同参数取值的影响,将模型应用到不同的领域和不同的数据集上。

(2)TOSM模型融入对否定的检测能优化情感分析结果。例如“the taste is not good”,“not good”表达的是负面情感,但是如果没有考虑到“not”和“good”的连接关系,很难得到正确的结果。在之后的工作中希望把否定检测方法<sup>[20-21]</sup>融入到本文的模型中。

(3)TOSM模型利用情感词典作为先验信息来区分评论中的主题词和情感词,将来可以将一些监督学习的组建引入到模型中,如通过结合最大熵组建和主题模型<sup>[22]</sup>,让TOSM模型同时具有监督学习和主题模型的特点,从而得到更好的结果。

#### References:

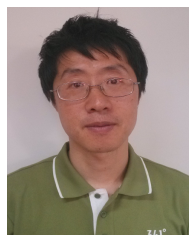
- [1] Pang Bo, Lee L. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135.
- [2] Turney P D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Stroudsburg, PA, USA: Association for Computational Linguistics, 2002: 417-424.
- [3] Chaovalit P, Zhou Lina. Movie review mining: a comparison between supervised and unsupervised classification approaches[C]//Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05). Washington, DC, USA: IEEE Computer Society, 2005: 112c.
- [4] Pang Bo, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP '02). Stroudsburg, PA, USA: Association for Computational Linguistics, 2002: 79-86.
- [5] Pang Bo, Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts[C]//Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04). Stroudsburg, PA, USA: Association for Computational Linguistics, 2004: 271.
- [6] Whitelaw C, Garg N, Argamon S. Using appraisal groups for sentiment analysis[C]//Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05). New York, NY, USA: ACM, 2005: 625-631.
- [7] Kennedy A, Inkpen D. Sentiment classification of movie reviews using contextual valence shifters[J]. Computational Intelligence, 2006, 22(2): 110-125.
- [8] Blitzer J, Dredze M, Pereira F. Biographies, Bollywood, boom-boxes and blenders: domain adaptation for sentiment classification[C]//Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL '07). Stroudsburg, PA, USA: Association for Computational Linguistics, 2007: 440-447.
- [9] Aue A, Gamon M. Customizing sentiment classifiers to new domains: a case study[C]//Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP '05), 2005.
- [10] Li Shoushan, Zong Chengqing. Multi-domain sentiment classification[C]//Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (HLT-Short '08). Stroudsburg, PA, USA: Association for Computational Linguistics, 2008: 257-260.
- [11] Paul M, Girju R. A two-dimensional topic-aspect model for discovering multi-faceted topics[C]//Proceedings of the 24th AAAI Conference on Artificial Intelligence, Atlanta,

- USA, Jul 11-15, 2010.
- [12] Mei Qiaozhu, Ling Xu, Wondra M, et al. Topic sentiment mixture: modeling facets and opinions in weblogs[C]//Proceedings of the 16th International Conference on World Wide Web (WWW '07). New York, NY, USA: ACM, 2007: 171-180.
- [13] Jo Y, Oh A H. Aspect and sentiment unification model for online review analysis[C]//Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM '11). New York, NY, USA: ACM, 2011: 815-824.
- [14] Lin Chenghua, He Yulan, Everson R, et al. Weakly supervised joint sentiment-topic detection from text[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(6): 1134-1145.
- [15] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation[J]. The Journal of Machine Learning Research, 2003, 3: 993-1022.
- [16] Turney P D, Littman M L. Measuring praise and criticism: inference of semantic orientation from association[J]. ACM Transactions on Information Systems, 2003, 21(4): 315-346.
- [17] Walsh B. Markov chain Monte Carlo and Gibbs sampling. 2004.
- [18] Porter M F. An algorithm for suffix stripping[J]. Program: Electronic Library and Information Systems, 1980, 14(3): 130-137.
- [19] Wallach H, Mimno D, McCallum A. Rethinking LDA: why priors matter[J]. Advances in Neural Information Processing Systems, 2009, 22: 1973-1981.
- [20] Das S, Chen M. Yahoo! for Amazon: sentiment parsing from small talk on the Web[C]//Proceedings of the 2001 European Finance Association Annual Conference (EFA 2001), Barcelona, 2001.
- [21] Eguchi K, Lavrenko V. Sentiment retrieval using generative models[C]//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP '06). Stroudsburg, PA, USA: Association for Computational Linguistics, 2006: 345-354.
- [22] Zhao W X, Jiang Jing, Yan Hongfei, et al. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP '10). Stroudsburg, PA, USA: Association for Computational Linguistics, 2010: 56-65.



ZHANG Qian was born in 1990. She is a master candidate at School of Computer and Information Technology, Beijing Jiaotong University. Her research interests include data mining and natural language processing.

张倩(1990—),女,云南大理人,北京交通大学计算机与信息技术学院硕士研究生,主要研究领域为数据挖掘,自然语言处理。



QU Youli was born in 1974. He received his Ph.D. degree in computer science from Beijing Institute of Technology in 2000. Now he is a senior engineer and master supervisor at Beijing Jiaotong University. His research interests include artificial intelligence, machine learning and natural language processing, etc.

瞿有利(1974—),男,吉林长春人,2000年于北京理工大学计算机专业获得博士学位,现为北京交通大学高级工程师、硕士生导师,主要研究领域为人工智能,机器学习,自然语言处理等。