面向产品评论分析的短文本情感主题模型

熊蜀峰 [1, 2] 姬东鸿 [1]

摘 要 情感主题联合生成模型已经成功应用于网络评论分析. 然而, 随着智能终端设备的广泛应用, 由于屏幕及输入限制, 用户书写的评论越来越短, 我们不得不面对短评论中的文本稀疏问题. 本文提出了一个针对短文本的联合情感–主题模型 SSTM (Short-text sentiment-topic model) 来解决稀疏性问题. 不同于一般主题模型中通常采用的基于文档产生过程的建模方法, 我们直接对整个语料集合的产生过程建模. 在产生文档集的过程中, 我们每次采样一个词对, 同一个词对中的词有相同的情感极性和主题. 我们将 SSTM 模型应用于两个真实网络评论数据集. 在三个实验任务中, 通过定性分析验证了主题发现的有效性, 并与经典方法进行定量对比, SSTM 模型的文档级情感分类性能也有较大提升.

关键词 情感分类, 情感主题模型, 主题模型, 短文本主题模型, 文本稀疏

引用格式 熊蜀峰, 姬东鸿. 面向产品评论分析的短文本情感主题模型. 自动化学报, 2016, **42**(8): 1227−1237

DOI 10.16383/j.aas.2016.c150591

A Short Text Sentiment-topic Model for Product Review Analysis

XIONG Shu-Feng[1, 2] JI Dong-Hong[1]

Abstract Topic and sentiment joint modelling has been successfully used in sentiment analysis for opinion text. However, we have to face the text sparse problem in opinion text when the length of text becomes shorter and shorter with popularity of smart devices. In this paper, we propose a joint sentiment-topic model SSTM (short-text sentiment-topic model) for short text. Unlike the topic model which models the generative process of each document, we directly model the generation of the whole review set. In the generation process of corpus, we sample a word-pair each time, in which the two words have the same sentiment label and topic. We apply SSTM to two real life social media datasets with three tasks. In the experiment, we demonstrate the effectiveness of the model on topic discovery by qualitative analysis. On the quantitative analysis of document level sentiment classification, SSTM model achieves better performance compared with the existing approaches.

Key words Sentiment classification, sentiment topic model, topic model, short text topic mode, text sparse

Citation Xiong Shu-Feng, Ji Dong-Hong. A short text sentiment-topic model for product review analysis. *Acta Automatica Sinica*, 2016, **42**(8): 1227−1237

产品评论挖掘技术是辅助分析海量评论信息的一种有效手段, 其目标是检测出文本中所表达的对某一话题的情感 (观点) 信息, 根据分析的粒度可以分为文档级、句子级和元素级[1−5]. 对于评论文本而言, 其包含的观点信息中两个最重要内容分别是评价目标 (在产品评论中称为 aspect) 和情感极性.

情感极性通常是由情感词汇来表达, 有情感表达词的地方通常都有评价目标词, 然而, 同一个情感表达短语在修饰不同的评价目标时可能会表示不同的极性. 如图 1 所示, 当 "小" 用来修饰不同的评价目标 "耗电" 和 "音量" 时, 它的情感极性正好相反. 虽然 "小" 和 "大" 是一对反义词, 但在句子 $R1$ 中, 分别修饰 "耗电" 和 "内存" 时, 都表示正面极性. 为了利用情感词与评价目标之间的相互依存关系, 一些研究工作提出采用无监督 (弱监督) 主题模型 (Topic model) 来处理此问题[6−8].

$R1$: *外观时尚, 屏幕清晰, 铃声悦耳, 性能稳定, 耗电小, 功能齐全, 内存大.*

$R2$: *上网慢! 音量小!*

图 1 两条评论文本信息

Fig. 1 Two opinion texts

学者们提出的无监督的主题模型解决了情感词与评价目标相互依存的问题. 然而, 随着互联网终端的广泛使用, 又需要面临一个新的问题 — 文本稀疏.

收稿日期 2015-09-15 录用日期 2015-12-28
Manuscript received September 15, 2015; accepted December 28, 2015
国家自然科学基金 (61373108, 61173062, 61133012) 和国家社会科学重大招标计划项目 (11&ZD189) 资助
Supported by National Natural Science Foundation of China (61373108, 61173062, 61133012), and The Major Program of the National Social Science Foundation of China (11&ZD189)
本文责任编委 张民
Recommended by Associate Editor ZHANG Min
1. 武汉大学计算机学院 武汉 430072 2. 平顶山学院 平顶山 467099
1. Computer School of Wuhan University, Wuhan 430072 2. Pingdingshan University, Pingdingshan 467099

随着移动互联网终端的广泛使用, 为了适应较小的屏幕以及受限的输入设备, 人们提交的产品评论文本的长度也变得越来越短[9]. 如图 1 所示, 在评论 $R1$ 中, 26 个字表达了用户对 7 个主题 (评价目标) 的观点. 目前大部分购物网站和产品查询网站的用户评论文本都具有此类简洁的表达和鲜明的观点. 随着文本内容的变短, 数据稀疏性问题也越来越成为亟需解决的问题. 本文的研究目的正是要为数据稀疏问题提供一种解决方案.

针对主题建模中的文本稀疏问题, 一些研究工作将短文本组合成较长的伪文档后再进行训练学习[10−11]; 另一种建模方式是基于这样的假设: 一段短文本只对应于一个单一的主题[12−13]. 最近, 文献 [14] 提出直接对词对共现过程进行建模. 上述的所有工作都仅仅对短文本中的主题建模, 而没有考虑情感极性信息. 在我们的方法中, 通过建模全局的词对生成过程, 联合检测情感极性和主题.

尽管很多有监督学习方法也取得了较好的效果[15−18], 但是有监督学习方法要依赖于高成本的人工标注语料. 为了减小人工成本, 我们提出一个弱监督的短文本情感主题模型 (Short-text sentiment-topic model, SSTM), 该模型是一个概率混合模型, 通过直接对全局范围内的词对 (Word-pair) 生成过程建模来学习短文本中的情感和主题信息. 模型中的 "词对" 是指在特定的上下文中的两个无序的共现词. 具体而言, 我们首先将整个语料看成是一个共现词对集合 (A bag of co-occurred word-pairs). 然后对词对集合的生成过程进行建模, 即通过一个混合模型依次采样语料中的每一个词对, 这个混合模型包括一组主题语言模型和一组情感语言模型. 通过学习 SSTM 模型, 我们得到语料级别的情感−主题组成信息和全局的情感主题分布信息. 并可以进一步推导出每个文档的情感分布和主题分布. 我们在两个评论文本数据集上对提出的方法进行了评估. 实验结果表明 SSTM 能够准确地发现文本中的主题并进一步检测出情感极性, 检测准确率明显高于经典的同类方法.

本文的主要贡献概括为以下几点:

1) 针对评论文本短小的特性, 提出了一个词对情感主题模型来同时检测评论文本中的主题和情感 (第 2.2 节).

2) 对模型的 Gibbs 采样方法进行了推导, 解决了模型的参数估计问题 (第 2.3 节).

3) 提供了一种有效的方法来估计模型学习过程无法获得的文档级别的主题和情感极性 (第 2.4 节).

4) 通过在两个真实数据集上的实验, 我们证明了提出的 SSTM 模型同时检测评论文本的主题与情感极性的有效性 (第 4 节).

1 相关工作

通过前面的讨论, 我们知道主题 (评价目标) 和情感极性是观点文本中用户所关心的两项重要信息. 因此采用概率混合模型对主题和情感极性联合建模是一种很自然的解决方案. 已经有很多学者提出基于 LDA (Latent Dirichlet allocation) 的模型来解决这个问题[19−29]. 在文献 [30] 中, 作者根据一些特性将此类方法分成了若干类别. 用于方法分类的这些特性如下:

1) 用一个隐含变量建模词/用不同的变量分别建模目标词和星级.

2) 建模文本中的所有词/只建模观点表达词语.

3) 建模目标词与星级间的依存有关系/不考虑依存关系.

4) 只使用评论语料/额外使用附加数据.

由于前两项特性属于模型内在特性, 两个不同的模型的内存特性通常不同, 用于划分大类时粒度过细, 而后两项涉及外部知识和外部数据, 需要人工干预, 因此我们根据后两项特性进行区分, 将 SSTM 划分为不考虑依存关系且不使用附加输入数据一类.

根据这两项特性的划分, 与 SSTM 模型同类的相关的方法主要包括以下几个代表性的工作:

1) JST (Joint sentiment-topic model). 此模型总体框架比 LDA 多了一层, 也就是在文档层与主题层之间加入了一个附加的情感层[8], 形成词、文档、主题和情感四层结构. 在此结构中, 情感极性与文档相关, 主题与情感极性相关, 而词同时与情感极性和主题相关. 在 JST 模型中, 一个句子中的所有词的情感标签和主题各自独立, 采样时完全依赖单个词在文档中的统计信息, 其生成过程没有考虑文档内容的文本稀疏问题, 而在 SSTM 模型中, 一个词对有着共同的主题和情感标签, 并且采样时统计词对在整个语料中的共现信息, 在更大范围的统计学习下解决了稀疏问题.

2) ASUM (Aspect and sentiment unification model). 由文献 [6] 提出, 和 JST 一样由四层结构组成. 二者的不同之处在于 ASUM 模型中, 同一个句子中的词都来自于同一个语言模型, 而 JST 模型中句子中的词可以来自不同的语言模型. 在 SSTM 模型中, 我们沿用 ASUM 中部分假设, 即约束同一个句子生成的词对来自于同一个情感模型, 但是只要求单个词对中的两个词来自于同一个主题模型. 因为 ASUM 中的假设过强, 特别是对于评论信息, 因为用户在一句话中经常评论某实体的相关方面, 如手机的屏幕与其电池的待机能力, 这两方面相互影响, 用户也经常在一句话中同时评论.

3) STDP (Senti-topic model with decom-

posed prior). 文献 [31] 提出先验分解的情感主题模型, 作者将情感极性的生成过程分解为两个阶段. 在第一阶段中, 先检测一个词是情感词还是主题词, 如果是情感词则进入第二阶段. 第二阶段主要是识别词的极性标签. STDP 将极性标签与情感词检测分别独立进行, 割裂了二者之间的天然联系和影响. 在 SSTM 模型当中, 极性标签是由情感词和主题词共同决定的, 而不是分开的两个阶段. 其次, STDP 需要人工构造先验知识来检测一个词是情感词还是主题词. 并且这样生成的先验规则并不一定适合于所有领域和不同语言 (比如中文与英语). SSTM 模型无需构造先规则, 具有更好的领域适应性. 由于人工指导行为需要消耗人力成本, 为了尽量自动化地完成观点分析任务, 我们的模型试图最小化人工指导. 因此, 我们除了采用一个公共可用的情感词典用于对齐人类情感与机器识别的情感之外, 模型中不再使用任何规则.

上面提到的三个模型主要针对足够长的传统媒体文本, 比如电影评论、餐馆评论等 (电影评论文档平均长度 668, 餐馆评论文档平均长度 153. 详细评测数据统计信息可查阅三个模型对应的文献)[1]. 而我们采集的中文购物网站的评论数据, 文档平均长度分别为 32 和 20. 如果不考虑短文本稀疏问题, 传统模型的单个文档生成过程建模时没有足够数量的词统计信息来发现词之间的主题相关性. 这个问题会进一步影响情感极性的识别. 为了克服建模单文档生成过程会遇到的文本稀疏问题, 我们采用类似于 BTM 模型[14] 中的方法, 即对整个语料级别的词对生成过程建模. 不同之处在于, 我们的混合模型联合检测情感与主题, 而 BTM 仅考虑主题信息.

近年来的一些其他主题建模工作也考虑到了短文本中的词稀疏问题[32−34]. 但所有这些工作都只是建模文本中的主题信息 (不考虑情感信息), 并且大部分方法都是应用于其他任务和不同的领域数据.

监督学习方法在评论分析任务上也取得了很好的效果. 文献 [18] 提出利用机器学习方法对文档的情感进行分类, 通过实验说明传统的方法 (朴素贝叶斯、最大熵和 SVM) 在情感分类任务上并没有达到基于主题的文本分类任务上那么好的性能, 进一步说明了情感分类的任务面临的巨大挑战. 文献 [15] 提出一种包容层级 (Subsumption hierarchy) 结构来形式化词汇特征信息, 提高了观点分类任务性能. 文献 [16] 提出基于度量标签的元算法来对评论进行评级, 该算法可以保证类似的元素可以获得相似的评级标签. 文献 [17] 利用句子中词语间的句法关系作为特征, 采用 SVM 算法对文档进行情感分类, 取得了很好的效果.

2 情感主题模型

前文的分析表明, 情感词与评价目标是两个相互依存的部分, 无监督的主题建模方法通过隐藏变量来定义文档的主题信息, 通过在主题模型中加入情感层来定义隐藏的情感极性信息, 从而在模型中自然地考虑情感词与评价目标 (主题) 间的依存关系. 因此主题模型很适合本任务, 并且能够将传统的评价目标发现与情感分类两个独立任务集成到一个统一的模型中, 同时完成评价目标发现与情感分类任务. 因此, 本文主要研究基于主题模型的主题学习和文档级的情感极性分类. 在本节, 我们将描述所提出的 SSTM 模型, SSTM 通过建模整个语料的生成过程来同时学习短文本中的主题与情感极性标签. 在 SSTM 中, 我们采用 Gibbs 采样方法来进行参数推断. 随后, 我们对生成过程无法直接推断出的单个文档的主题与情感极性标签进行了估计.

2.1 观点文本的表示

直观上而言, 一个词的情感极性标签是由情感词和其上下文所决定的. 如图 1 所示, 当出现在词对 ⟨耗电, 小⟩ 中时, “小” 的情感极性是正, 而出现在词对 ⟨音量, 小⟩ 中时, 其情感极性为负. 因此上下文中抽取的词对往往包含着用来检测情感极性的重要信息. 然而, 当文本比较短时, 上下文信息有限, 主题−情感模型将遇到稀疏性问题. 如文献 [14] 指出的那样, 一个有效的方法是使用全局的词对共现模式来学习文本主题. 所以我们将主题学习与情感极性识别融合到一个统一的框架中, 即从词对生成过程中同时学习短文本中的主题与情感极性标签. 换句话说, SSTM 使用全局的词对生成过程来代替传统的单个文档的生成过程来建立模型.

在 SSTM 中, 第一步就是要将观点文本表示成词对集合, 其中词对是从每篇文档中抽取出来的. 一个词对 b 是由两个无序的词组成, 其含义是这两个词在一定的窗口大小范围内同时出现在文档中. 在我们的实验中, 窗口大小为 10. 例如, 在文档 $R2$ “上网慢! 音量小!”, 可以抽取出 5 个词对 ⟨上网, 慢⟩、⟨上网, 音量⟩、⟨上网, 小⟩、⟨慢, 音量⟩、⟨慢, 小⟩ 和 ⟨音量, 小⟩. 在生成过程中, 每个词对中的词将被指派相同的主题和情感极性标签. 我们从语料中抽取所有词对形成一个集合用来表示评论文本.

我们在实验部分证明了词对 采样方法 对文本稀

[1] 尽管文献 [31] 在短文本数据集上进行了实验, 但 STDP 模型本身并没有考虑文本稀疏问题.

表 1 论文中符号的含义
Table 1 Meanings of the notations

符号	描述	符号	描述
D	文档数量	β	ϕ 的非对称 Dirichlet 先验参数, $\beta = \{\{\{\beta_{z,l,i}\}_{k=1}^{\mathrm{T}}\}_{l=1}^{S}\}_{i=1}^{V}$
M	词对数量		
T	主题数目	α	θ 的 Dirichlet 先验参数
S	情感极性数	γ	π 的 Dirichlet 先验参数
V	词汇表大小	Θ	主题的多项式分布
b	词对, $b = (w_i, w_j)$	z_t	第 t 个词的主题
w	词	l_t	第 t 个词的情感极性标签
z	主题	$\boldsymbol{B}$	词对集合
l	情感极性标签	$\boldsymbol{z_{-t}}$	除第 t 个词以外的其他所有词的主题分布
$\pi_{k,l}$	主题 k 和情感极性 l 上的分布	$\boldsymbol{l_{-t}}$	除第 t 个词以外的其他所有词的情感极性
Π	情感极性标签的多项式分布	$N_{k,l,i}$	词 w_i 指派为主题 k 和情感极性 l 的次数
$\phi_{k,l,w}$	词 w 基于主题 k 和情感极性 l 的分布	$N_{k,l}$	指派为主题 k 和情感极性 l 的词的数量
Φ	词的多项式分布	$N'_{(\cdot)}$	句子计数
θ_k	主题 k 的分布	N_k	主题 k 中的词的数量

疏问题进行了有效的解决, 尽管模型可能会引入噪音数据 (这是基于窗口的方法的通用问题). 实际上, 这些噪音数据对模型的学习影响很小. 比如前面提到的一些词对, 如 ⟨慢, 音量⟩ 和 ⟨慢, 小⟩ 都是噪音数据, 可观察出这些词对不是描述同一主题 (评价目标). 对于这种情况, 我们有一个合理的解释: 在对整个词对集合进行统计学习时, 这些噪音词对的出现频率会非常小. 也就是说, 通过全局性的共现规律的统计能够降低噪音数据对模型的负面影响. 这也是统计学习方法的基于数据统计的优势体现. 事实上, 除了文献 [14] 以外, ASUM[6] 也使用了类似的基于窗口的方法. 这两个方法都取得了较好的实验效果. 特别是 ASUM 模型, 其假设一个句子中的所有词具有相同的主题和情感极性标签. 在 SSTM 模型中, 我们放松了 ASUM 模型中的假设, 将约束从整个句子范围内缩小到固定窗口内的两个词.

2.2 SSTM 模型

SSTM 模拟所有用户发布的整个文档集的生成过程. 按照用户的习惯, 书写评论的时候, 通常会对相关产品的某些方面进行评价, 如图 1 中的例子, 评论 $R1$ 中用户对手机的 7 个方面进行了评价, 并且对各方面的描述大概采用了相同数量的文字. 在某些情况下, 用户可能会使用更多的文字来描述自己更关注的方面, 如, 长期出差的用户可能更在意手机的电池续航能力, 在发表评论时也就会花更多的文字来描述对电池的观点. 在生成模型中, 可以用概率分布来建模对各方面, 用户使用更多文字描述的方面具有更高的概率分布.

从一个用户的角度来讲, 文档的生成过程大致如下:

1) 用户要对某款笔记本电脑发表自己的观点, 首先他用一个主题分布来确定评价目标的比例, 比如, 所有内容中 30 % 关于内存, 30 % 关于外观, 20 % 关于运行速度, 最后 10 % 关于电池.

2) 对于要评价的每一个主题, 他确定要表达的情感极性比例, 如 80 % 正面和 20 % 负面.

3) 随后他选择相应的词对来表达前面确定好的主题和情感极性下的观点.

假设我们有一个语料包含 D 篇文档, $C = \{d_1, d_2, d_3, \cdots, d_D\}$; 采用第 2.1 节所描述的方法从 C 中抽取得到的词对集合 $B = \{b_1, b_2, \cdots, b_M\}$. SSTM 首先建模集合 B 的生成过程, 然后采用一种近似方法来估计每篇文档 d 的情感–主题分布. 假设总共有 T 个主题, 按索引 $\{1, 2, \cdots, T\}$ 排列. 对于一个主题 T, 有 l 个情感极性标签与之相关. 给定 α、β 和 γ 为 Dirichlet 先验. 论文中符号含义如表 1 所示.

我们形式化地定义如图 2 所示的 SSTM 的生成过程如下:

1) 采样一个主题分布 $\theta \sim Dir(\alpha)$.

2) 对于每个主题 $k \in \{1, 2, 3, \cdots, T\}$

a) 采样一个情感分布 $\pi_k \sim Dir(\gamma_k)$;

b) 对于每一个情感标签 $l \in \{1, 2, 3, \cdots, S\}$;

c) 采样一个词分布 $\phi_{kl} \sim Dir(\beta)$.

3) 对于每个词对 $b \in B$

a) 选择一个主题 $z \sim Mult(\theta_k)$;

b) 根据主题 z, 选择一个情感标签 $l \sim Mult(\pi_{k,l})$;

c) 根据情感标签 l 和主题 z, 选择两个词 $w_i, w_j \sim Mult(\phi_{l,z})$.

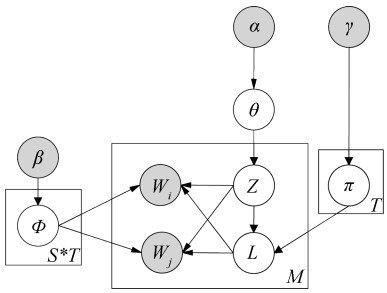


图 2　SSTM 模型的图表示

Fig. 2　SSTM model

我们同时给出了 JST 和 ASUM 模型的图表示, 如图 3 和图 4 所示, 其中 R 表示同一句子中的词数量. 我们的模型与二者不同时之处在于: 1) SSTM 模型不考虑文档的生成过程; 2) SSTM 在同一个主题和情感标签下生成词对 (w_i, w_j), 而 ASUM 在相同主题和情感标签下生成同一个句子的所有词, JST 中各个词的主题和情感标签均相互独立生成.

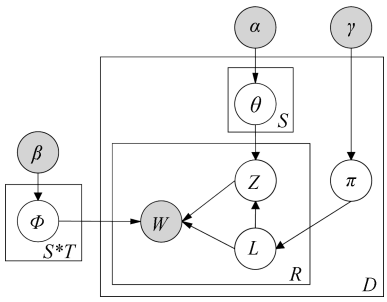


图 3　JST 模型的图表示

Fig. 3　JST model

在情感模型中, 需要提供合适的先验知识, 使用模型能够建立正确的情感标签到真实的人类情感间的关系. 换言之, 模型学习到的每个情感标签类别需要与人类情感一一对应. 在 SSTM 模型中, 我们使用情感词典来引入人类情感信息. 具体而言, 情感信息的引入是通过非对称先验参数 β 来实现的. 假设一个词 w_i 来自于情感词典, 则它的情感分布的 Dirichlet 先验通过以下公式计算:

$$\beta_{w_i,l} = P_{w_i}(l) * \beta_0 \tag{1}$$

其中, $P_{w_i}(l)$ 是预先定义的词典中的词 w_i 关于情感标 l 的概率, β_0 是一个基本因子 (如 0.05). 假设有三种情感极性: 中性、正面和负面, 且 w_i 在情感词典中的极性为正面, 我们预先定义中性的概率为 $P_{w_i}(neutral) = 0.009$, 正面的概率 $P_{w_i}(positive) = 0.99$ 和负面的概率 $P_{w_i}(negative) = 0.001$. 这样定义的含义是 w_i 有 99 % 概率在语料中表示正面极性, 0.9 % 概率表示中性和 0.1 % 概率表示负面.

2.3　模型推断

为了估计 SSTM 模型中的参数 θ、ϕ 和 π, 需要计算后验分布 $P(z, l|\boldsymbol{B})$, 也就是给出集合 $\boldsymbol{B}$ 的条件下的, 主题为 z 和情感标签为 l 的条件概率. 此概率很难直接计算, 所以我们采用 Gibbs 采样方法进行近似推断. Gibbs 的全条件概率分布如式 (2):

$$P(z_t = k, l_t = l|\boldsymbol{B}, \boldsymbol{z_{-t}}, \boldsymbol{l_{-t}}) \propto \frac{P(\boldsymbol{B}|\boldsymbol{z}, \boldsymbol{l})}{P(\boldsymbol{B_{-t}}|\boldsymbol{z_{-t}}, \boldsymbol{l_{-t}})} \cdot \frac{P(\boldsymbol{l}|\boldsymbol{z})}{P(\boldsymbol{l_{-t}}|\boldsymbol{z_{-t}})} \cdot \frac{P(\boldsymbol{z})}{P(\boldsymbol{z_{-t}})} \tag{2}$$

式 (2) 中第一部分的分子, 通过对 ϕ 积分, 可以得到:

$$P(\boldsymbol{B}|\boldsymbol{z}, \boldsymbol{l}) = \prod_{k=1}^{T}\prod_{l=1}^{S} \frac{\Gamma(V\beta_{k,l})}{\Gamma(\beta_{k,l})^V} \frac{\prod_{i=1}^{V}\Gamma(N_{k,l,i} + \beta_{k,l,i})}{\Gamma(\sum_{i=1}^{V} N_{k,l,i} + V\beta_{k,l})} \tag{3}$$

第二部分的分子, 通过对 π 积分, 可以得到:

$$P(\boldsymbol{l}|\boldsymbol{z}) = \prod_{l=1}^{S} \frac{\Gamma(T\gamma_l)}{\Gamma(\gamma_l)^T} \frac{\prod_{k=1}^{T}\Gamma(N_{k,l} + \gamma_{k,l})}{\Gamma(\sum_{k=1}^{T} N_{k,l} + T\gamma_l)} \tag{4}$$

第三部分的分子, 通过对 θ 积分, 可以得到:

$$P(\boldsymbol{z}) = \frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T} \frac{\prod_{k=1}^{T}\Gamma(N_k + \alpha)}{\Gamma(\sum_{k=1}^{T} N_k + T\alpha)} \tag{5}$$

通过类似的处理方式, 三个分母也可以变换后得到:

$$P(\boldsymbol{B_{-t}}|\boldsymbol{z_{-t}}, \boldsymbol{l_{-t}}) = \prod_{k=1}^{T}\prod_{l=1}^{S} \frac{\Gamma(V\beta_{k,l})}{\Gamma(\beta_{k,l})^V} \frac{\prod_{i=1}^{V}\Gamma(\{N_{k,l,i}\}_{-t} + \beta_{k,l,i})}{\Gamma(\sum_{i=1}^{V}\{N_{k,l,i}\}_{-t} + V\beta_{k,l})} \tag{6}$$

$$P(\boldsymbol{l_{-t}}|\boldsymbol{z_{-t}}) = \prod_{l=1}^{S} \frac{\Gamma(T\gamma_l)}{\Gamma(\gamma_k)^T} \frac{\prod_{k=1}^{T}\Gamma(\{N_{k,l}\}_{-t} + \gamma_{k,l})}{\Gamma(\sum_{k=1}^{T}\{N_{k,l}\}_{-t} + T\gamma_l)} \tag{7}$$

$$P(\boldsymbol{z_{-t}}) = \frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T} \frac{\prod_{k=1}^{T}\Gamma(\{N_k\}_{-t} + \alpha)}{\Gamma(\sum_{k=1}^{T}\{N_k\}_{-t} + T\alpha)} \tag{8}$$

通过式 (3) ∼ (8) 替换式 (2) 中对应的部分, 并利用 Gamma 函数性质, 可以推导出 Gibbs 采样每次迭代中的条件分布概率:

$$P(z_t = k, l_t = l|\boldsymbol{B}, \boldsymbol{z}_{-\boldsymbol{t}}, \boldsymbol{l}_{-\boldsymbol{t}}) \propto \frac{(\{N_{k,l,w_{i,1}}\}_{-t} + \beta)(\{N_{k,l,w_{i,2}}\}_{-t} + \beta)}{(\{N_{k,l}\}_{-t} + V\beta + 1)(\{N_{k,l}\}_{-t} + V\beta)} \frac{(\{N_{k,l}\}_{-t} + \gamma_{k,l})}{(\{N_k\}_{-t} + T\gamma_l)} \frac{(\{N_k\}_{-t} + \alpha)}{(M_{-t} + S\alpha)} \tag{9}$$

其中, M 是词对的总数而不是词汇表大小.

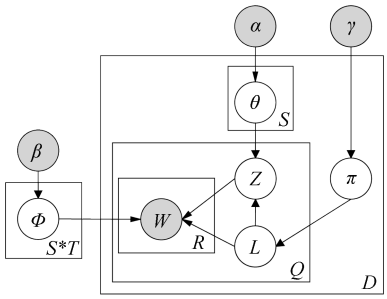


图 4 ASUM 模型的图表示

Fig. 4 ASUM model

给定超参 α、β 和 γ, 词对集合 B 和其对应的主题 z, 情感标签 l, 我们可以利用贝叶斯规则和 Dirichlet 共轭特性推断出参数 θ、ϕ 和 π:

$$\theta_k = \frac{N_k + \alpha}{M + S\alpha} \tag{10}$$

$$\phi_{k,l,w} = \frac{N_{k,l,w} + \beta}{N_{k,l} + V\beta} \tag{11}$$

$$\pi_{k,l} = \frac{N_{k,l} + \gamma_{k,l}}{N_k + T\gamma_l} \tag{12}$$

2.4 推断文档的情感极性和主题

为了解决文本稀疏问题, SSTM 没有对文档的生成过程建模. 因此, 我们需要提供一个必要的步骤来近似估计文档的情感和主题分布. 我们用如下公式近似文档 d 的情感极性:

$$L_d = \arg\max_{l \in L} N_d^{(l)} \tag{13}$$

其中, $N_d^{(l)}$ 是文档 d 中情感标签为 l 的词数量. 由于生成过程是基于词对, 我们需要估计每一个词的情感标签, 通过如下公式:

$$L_w = \arg\max_{l \in L} P(l|w) \tag{14}$$

其中, 词 w 的情感标签为 l 的概率 $P(l|w)$ 可以通过贝叶斯公式推导:

$$P(l|w) = \frac{\sum\limits_z P(z)P(l|z)P(w_i|l,z)}{\sum\limits_l (\sum\limits_z P(z)P(l|z)P(w_i|l,z)} \tag{15}$$

其中, $P(z) = \theta_k$, $P(l|z) = \pi_{k,l}$ 且 $P(w_i|l,z) = \phi_{k,l,i}$.

同样可以近似估计词 w 的主题:

$$P(z|w) = \frac{P(z)\sum\limits_l P(l|z)P(w_i|l,z)}{\sum\limits_z (P(z)\sum\limits_l P(l|z)P(w_i|l,z))} \tag{16}$$

尽管我们采用这样一种频率计算的方式来近似文档的情感和主题, 但其实验效果良好. 更复杂的处理方式可以进一步研究.

3 实验设置

3.1 数据集

我们使用两个在线购物网站的评论文本数据集来验证我们的方法. 一个数据集是来自于京东[2] 的笔记本电脑产品评论, 另一个是 IT168[3] 网站的手机产品评论数据集. 在进行中文分词之后, 我们的预处理工作还包括移除标点、数字和停用词. 经过预处理后的数据集的统计信息如表 2 所示. 在实验中, 我们随机选择其中 50 % 作为验证集用于调试参数, 另外 50 % 用作测试集.

表 2 语料统计信息

Table 2 Statistics of the text corpus

	笔记本	手机
文档平均词数	20	32
评论数	3 988	2 289
词汇表大小	7 964	8 787
正面评论数	1 993	1 146
负面评论数	1 995	1 943

3.2 情感词典

因为情感词典能够提供必要的知识用于识别情感极性, 我们使用知网 (HowNet)[4] 情感词典为 SSTM 模型提供先验信息. 知网情感词典包括大约 5 000 正面和 5 000 负面词汇. 如第 2.2 节所介绍, 我们使用情感词典来影响先验参数 β.

3.3 对比方法和参数设置

在主题识别任务上, 我们与标准 LDA 和 BTM 进行了对比分析, 结果见第 4 节. 对于情感分类任

[2] http://www.jd.com
[3] http://product.it168.com
[4] http://www.keenage.com/html/c_index.html

务, 虽然我们的方法是无监督学习 (词典信息仅用于情感对齐), 但是为了全面衡量 SSTM 的性能, 除了选择三个代表性的方法进行定量对比外, 我们还设计了与有监督的分类算法 SVM 的对比实验.

基线方法直接对文档中的词汇极性数进行统计, 其中词汇的极性直接由情感词典获取, 统计数量多的极性作为文档的极性. LDA 是经典的主题模型方法, BTM 类似于去掉情感层的 SSTM. BTM、JST 和 ASUM 已经在文章前面部分介绍过.

在实验中, 对于其他主题模型均使用各自原始论文中相同的超参设置. 对于 SSTM 模型, 我们使用对称参数 α 的值为 0.03、γ 的值为 0.02. 参数 β 为非对称, 如第 3.2 节所介绍, 我们设置基本因子 β_0 的值为 0.05.

4　实验结果与分析

为了评估 SSTM 模型的性能, 我们设计了三个任务: 主题发现、情感相关的主题发现和文档级的情感极性分类.

4.1　主题发现

对于主题模型, 发现主题词是一个主要任务. 因为我们的模型是设计用来进行观点文本分析, 评价目标自然地就作为主题来看待. 我们发现主题数目设置为 25 时, 可以获得较好的主题发现与情感极性识别效果. 因此, 本节中所有的实验都是基于此设置. 表 3 和表 4 分别列出了 SSTM 发现的笔记本和手机数据集的主题词. 其中加粗的词与当前的主题无关, 也就是主题识别错误的词. 我们可以看到 SSTM 和 BTM 的主题词错误小于 LDA 模型 (定量对比见表 5 和表 6).

为了便于理解, 我们给模型识别出的每一个主题人工指定一个标签. 我们只列举出每一个数据集中的 3 个样例 主题, 每个主题 取前 10 个词 (按概率

表 3　笔记本数据集中发现的部分主题词列表

Table 3　Example topics discovered from LAPTOP dataset

SSTM			BTM			LDA		
外观	电池	散热性	外观	电池	散热性	外观	电池	散热性
指纹	电池	散热	太	电池	散热	容易	电池	好
钢琴	小时	热	容易	时间	好	指纹	小时	散热
漂亮	长	温度	指纹	小时	不错	外壳	时间	声音
烤漆	比较	好	**键盘**	**键盘**	电池	钢琴	长	风扇
好	时间	烫	烤漆	比较	度	烤漆	续航	小
模具	续航	CPU	比较	长	热	表面	比较	温度
屏幕	使用	硬盘	不错	好	温度	**亮点**	使用	热
外壳	上网	风扇	外壳	不错	声音	**感觉**	**键盘**	运行
文字	**小**	机器	钢琴	使用	使用	**说**	**小巧**	**轻**
呵呵	芯	比较	**屏幕**	续航	CPU	**屏幕**	芯	时

表 4　手机数据集中发现的部分主题词列表

Table 4　Example topics discovered from MOBILE dataset

SSTM			BTM			LDA		
拍照	媒体播放	屏幕	拍照	媒体播放	屏幕	拍照	媒体播放	屏幕
拍摄	播放	屏幕	像素	MP3	屏幕	效果	支持	屏幕
功能	速度	好	摄像头	播放	色	摄像头	MP3	显示
支持	不错	显示	拍摄	耳机	显示	像素	播放	比较
屏幕	影音	色	数码	效果	TFT	拍照	内存	色彩
像素	手机	效果	手机	好	效果	照片	**蓝牙**	色
材质	处理器	彩色	支持	音乐	色彩	拍摄	卡	清晰
照片	格式	设计	倍	听	手机	拍	格式	**高**
摄像头	MP3	TFT	效果	**功能**	好	数码	**扩展**	**铃声**
拍照	流畅	机子	相机	不错	26 万	相机	文件	**方便**
数码	文件	**人**	拍照	比较	像素	倍	视频	TFT

逆序排列). 可以看到: 1) 每一个主题列表下的词很好的与产品的某一个属性 (Aspect) 相关联; 2) 这些词有较好的主题内部连贯性. 例如, 表 3 中的第 2 列大概是关于电池, 通过列表中的词, 我们可以直接推测出该款笔记本的 "电池" 可使用较 "长" "时间", "续航" 能力 "不错".

然而, 在一些主题下有一些无关的 "噪音词". 例如, 在第 1 列中的词 "指纹"、"钢琴" 好像与主题 "外观" 无关, 但是通过语料分析, 我们发现了其中的原因: 1) 分词器错误地将 "钢琴烤漆" 分成了两个词 "钢琴" 和 "烤漆"; 2) 在一些评论中会提到 "钢琴烤漆面成了指纹采集器, 很容易留下指纹". 而且, 还有一些情绪词出现在列表中, 如第 1 列的 "呵呵". 这也不难解释, 人们通常会利用情绪词来强调和描述情感. 另外还有一些词同时出现在多个主题中, 如表 4 第 2 列、第 4 列及第 6 列中的 "手机". 这些词可以视为来自于一些全局主题, 这些全局主题也可以看做是其他主题的公共子主题. 如果要用于细粒度的主题发现, 就需要过滤这些公共主题, 我们留作将来的工作.

表 5 笔记本数据集上的 CM 值 (%)

Table 5 CM (%) on laptop dataset

方法	标注员 1	标注员 2	标注员 3	标注员 4	平均值
LDA	58	50	60	56	56
BTM	70	66	75	72	70.75
SSTM	69	64	72	67	68

表 6 手机数据集上的 CM 值 (%)

Table 6 CM (%) on mobile dataset

方法	标注员 1	标注员 2	标注员 3	标注员 4	平均值
LDA	69	65	71	74	69.75
BTM	76	74	81	81	78
SSTM	75	72	79	78	76

我们通过以上的定性分析证明了 SSTM 用于主题词发现的可行性, 下面我们对发现的主题词内部的相关性进行定量评估. 对于主题模型的量化评估仍然是一个有待研究课题[35], 各种评价方法都有一定的局限性. 最近, 文献 [36] 提出了一个合理的基于人工判断的评估方法 CM (Coherence measure). CM 方法能够克服传统的基于困惑度 (Perplexity) 评价方法的缺点, 评价方法的优缺点分析可参考文献 [35]. 我们沿用文献 [36] 的 CM 方法来评估我们的实验结果. 由 4 个标注员对每个主题中的前 10 个候选词进行评判. 首先, 标注员判断能否从一个主题中的大部分 (或全部) 词中抽象出一个可理解的话题. 如果主题中的一组词不可以抽象为一个话题, 那么 10 个词都被标记为不相关. 否则, 标注员依次判断每一个词是否和抽象出的话题内容相关. CM 则定义为主题内的相关词数目与候选词总数的比率. 对每个数据集, 我们随机选择 10 个主题进行评估. 评估结果如表 5 和表 6 所示. SSTM 模型的主题相关性结果与 BTM 相当, 它们的结果均高于 LDA 模型. 由于 SSTM 在 BTM 基础上增加了一个情感层, 在对情感进行估计时不可避免地会产生误差, 对一下层主题发现的效果产生了负面影响, 因此 SSTM 在本任务上的性能略低于 BTM. 在手机数据集上的结果总体要好于笔记本数据集. 主题发现任务的实验结果说明 SSTM 模型中的词对采样方法达到了与 BTM 模型中相当的水平, 从一定程度上解决了文本稀疏性问题. 下一节我们进一步测试 SSTM 模型的情感极性识别效果, 这是 LDA 模型和 BTM 模型没有考虑的, 因此我们只进行定性评估.

4.2 情感相关的主题发现

第二个实验是发现情感相关的主题. 本实验是同时根据词的主题和情感极性进行分类. 识别出来的主题可以用来提供与情感极性相关的信息, 比如哪个主题是正面或者负面? 为什么某个主题是负面评价?

如表 7 所示, 我们列举了每个数据集中 5 个情感相关的主题 (3 个正面和 2 个负面). 以笔记本数据集为例, 对于一个问题: 为什么做工得到负面评价? 列表中的词回答了这个问题: 触摸板不好用、盖子小而且产品有瑕疵.

4.3 文档级的情感极性分类

在本节, 我们讨论 SSTM 模型情感极性识别的定量评估结果. 对每一个数据集, 其中的每条评论都有一个二元情感标签 (正面和负面). 我们的实验中采用不同主题数目进行了对比, 主题数从 5 到 30, 步长为 5. SSTM 在每个主题数设置下的结果均优于其他两个经典的情感主题模型. 然而, 只有主题数目接近实际评价目标时, 主题情感模型的情感识别结果才能更精确, 在第 4.1 节我们已经发现主题数设置为 25 时能取得较好的主题词发现结果. 因此在本实验中, 我们统一设置所有情感主题模型的主题数目为 25. 各种方法识别出来的情感极性准确率对比结果如表 8 所示. 其中 SVM (Uni) 表示采用一元词特征训练分类器, SVM (Bi) 表示采用二元词特征训练分类器. 对于 SVM 分类器, 由于二元词特征考虑了前后词间的关系, 因此其效果要好于一元特征, SSTM 达到了接近算法 SVM 的性能. 当然, 由于 SVM 是监督学习算法, 其总体性能还是要好于无监督方法. SSTM 作为一种弱监督模型, 达到了

接近监督学习算法 SVM 的性能, 这充分说明了其有效性. 由于主题数目对于情感分类会产生影响, 我们绘制了折线图来进行分析, 如图 5 所示. 随着主题数目的增加, 识别性能有一些波动, 但 SSTM 模型的总体性能都要高于其他两个主题模型. 由于不知道语料所讨论的真实话题数目, 我们在开发集上对不同的主题数进行调试. 最初设定主题数目为 5, 相当于粗粒度的主题划分, 随着粒度的细化, SSTM 性能逐渐提升. 因为 SSTM 考虑语料全局范围内关联的词对之间的情感与主题, 而 JST 和 ASUM 模型不考虑, 因此变化曲线不如 SSTM 明显. 特别是 ASUM 由于假设较强 (同一个句子具有相同主题和情感), 因此主题数增多可能造成句子主题和情感的离散化, 对其整体性能造成了负面影响, 特别是手机数据集上的分类性能呈单调下降.

对于两个数据集, SSTM 模型的情感极性识别性能均好于其他方法. 我们相信这主要得益于良好的主题发现, 因为情感极性往往与其所描述的主题相关. 正如文献 [6] 所讨论的那样, JST 在其原始论文[8] 提供的电影评论数据集上取得较好的效果, 但是在我们的数据集和文献 [6] 的数据集上的性能有较大的下降. ASUM 基于这样一个假设: 同一个句子中的词来自于同一个情感–主题模型. 对于短文本来说, 这是一个很强的假设, 采样过程中没有足够

表 7　SSTM 发现的部分情感相关的主题词列表

Table 7　Example sentiment-specific topics discovered by SSTM

笔记本					手机				
正面			负面		正面			负面	
快递	性价比	外观	做工	售后	铃声	外观	按键	输入法	信号
速度	不错	小	有点	电话	铃声	设计	按键	短信	信号
东西	价格	漂亮	禁用	服务	不错	外观	手感	输入法	网络
京东	机器	喜欢	触摸板	差	耳机	不错	感觉	键	差
质量	便宜	买	需要	客服	听	好	好	切换	无
好	款	外观	外壳	送货	声音	感觉	操作	拼音	检测
发货	性能	本本	盖子	快递	放	喜欢	不错	数字	移动
问题	好	不错	小	货	音乐	漂亮	容易	麻烦	关机
比较	电脑	好	老版	无	好	时尚	使用	选	质量
很快	超值	键盘	掉	态度	耳朵	手感	摇杆	手	故障
送货	降价	适合	瑕疵	前台	效果	机身	舒服	标点符号	通话

表 8　情感极性识别结果 (主题数目设置为 25)

Table 8　Sentiment identification results (The number of topics is 25.)

	基线	JST	ASUM	SSTM	SVM (Uni)	SVM (Bi)
笔记本	0.637645	0.50677	0.57754	0.65503	0.66047	0.70021
手机	0.602188	0.53698	0.43694	0.64201	0.64476	0.68953

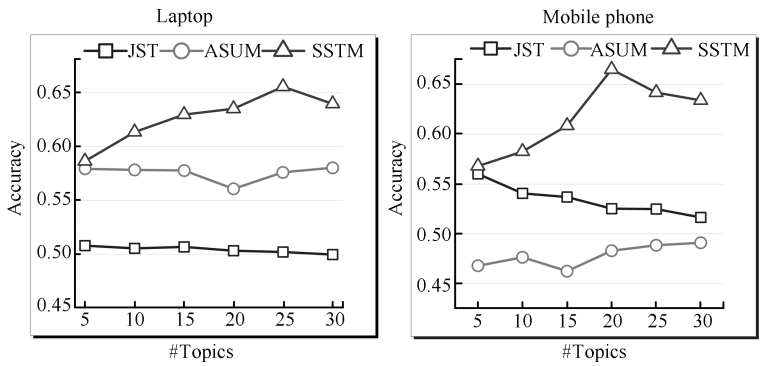


图 5　主题数目对三个主题模型情感识别性能的影响

Fig. 5　The impact of topic numbers in three topic models

的句子来估计模型参数. 我们放宽假设, 使得一个句子中的词可以来源于多个主题, 并能够进一步提高情感识别性能. 在我们的实验中, JST 和 ASUM 的效果都要低于基线系统. 值得注意的是, JST 与 ASUM 模型在两个数据集上的情感识别效果正好相反. 根据文献 [6] 中报告的长文本数据集的结果和我们短文本数据集上的结果的差异, 我们分析造成这种不稳定的原因是由于文本的稀疏所引起的. SSTM 在两个数据集都取得了较好的效果, 这也充分证明了我们模型的有效性.

5 结语

在本文的工作中, 我们提出了一个弱监督的短文本情感主题模型 SSTM. 此模型采用一个联合的情感主题识别方法, 通过此方法可以减小评论文本稀疏性对识别效果的负面影响. 在 SSTM 模型中, 我们将整个语料表示为一个词对集合. 然后通过模拟此集合的生成过程, SSTM 发现了隐含于词共现模式中的信息, 从而有效地识别了主题和情感极性, 在两个真实数据集上的实验结果证明了 SSTM 模型不仅能学习出高质量的主题, 还能准确地识别出文档级别的情感极性.

尽管 SSTM 模型取得了较好的效果, 仍然还有一些方面可以进一步研究. 比如, 对于第 4.1 节提到的公共主题的过滤问题, 可以尝试通过在模型中增加一个全局主题来解决. 另外, 虽然我们目前的简单方法能有效估计文档级别的情感极性, 探索更复杂的方法也是一项将来的工作. 因为 SSTM 模型设计时考虑了商品评论文本的特殊性, 所以可以更进一步修改模型以适用于其他社会媒体数据, 如微博、微信等数据.

References

1 Fang L, Huang M L, Zhu X Y. Exploring weakly supervised latent sentiment explanations for aspect-level review analysis. In: Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management. New York, NY, USA: ACM, 2013. 1057−1066

2 Xu Bing, Zhao Tie-Jun, Wang Shan-Yu, Zheng De-Quan. Extraction of opinion targets based on shallow parsing features. *Acta Automatica Sinica*, 2011, **37**(10): 1241−1247
(徐冰, 赵铁军, 王山雨, 郑德权. 基于浅层句法特征的评价对象抽取研究. 自动化学报, 2011, **37**(10): 1241−1247)

3 Zhao Yan-Yan, Qin Bing, Liu Ting. Integrating intra- and inter-document evidences for improving sentence sentiment classification. *Acta Automatica Sinica*, 2010, **36**(10): 1417−1425
(赵妍妍, 秦兵, 刘挺. 基于图的篇章内外特征相融合的评价句极性识别. 自动化学报, 2010, **36**(10): 1417−1425)

4 Liu B. *Sentiment Analysis and Opinion Mining*. San Rafael, CA: Morgan Claypool Publishers, 2012.

5 Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008, **2**(1−2): 1−135

6 Jo Y, Oh A H. Aspect and sentiment unification model for online review analysis. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2011. 815−824

7 He Y L, Lin C H, Alani H. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies—Volume 1. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011. 123−131

8 Lin C H, He Y L. Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2009. 375−384

9 Zhang Lin, Qian Guan-Qun, Fan Wei-Guo, Hua Kun, Zhang Li. Sentiment analysis based on light reviews. *Journal of Software*, 2014, **25**(12): 2790−2807
(张林, 钱冠群, 樊卫国, 华琨, 张莉. 轻型评论的情感分析研究. 软件学报, 2014, **25**(12): 2790−2807)

10 Weng J S, Lim E P, Jiang J, He Q. TwitterRank: finding topic-sensitive influential twitterers. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2010. 261−270

11 Hong L J, Davison B D. Empirical study of topic modeling in twitter. In: Proceedings of the 1st Workshop on Social Media Analytics. New York, NY, USA: ACM, 2010. 80−88

12 Zhao W X, Jiang J, Weng J S, He J, Lim E P, Yan H F, Li X M. Comparing twitter and traditional media using topic models. *Advances in Information Retrieval*. Heidelberg, Berlin, Germany: Springer, 2011. 338−349

13 Gruber A, Weiss Y, Rosen-Zvi M. Hidden topic Markov models. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. San Juan, Puerto Rico: Omnipress, 2007. 163−170

14 Yan X H, Guo J F, Lan Y Y, Cheng X Q. A biterm topic model for short texts. In: Proceedings of the 22nd International Conference on World Wide Web. New York, NY, USA: ACM, 2013. 1445−1456

15 Riloff E, Patwardhan S, Wiebe J. Feature subsumption for opinion analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006. 440−448

16 Pang B, Lee L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005. 115−124

17 Matsumoto S, Takamura H, Okumura M. Sentiment classification using word sub-sequences and dependency sub-trees. *Advances in Knowledge Discovery and Data Mining*. Heidelberg, Berlin, Germany: Springer, 2005: 301−311

18 Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—Volume 10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002. 79−86

19 Titov I, McDonald R. Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th International Conference on World Wide Web. New York, NY, USA: ACM, 2008. 111−120

20 Titov I, McDonald R T. A joint model of text and aspect ratings for sentiment summarization. In: Proceedings of ACL-08: HLT. Columbus, Ohio, USA: Association for Computational Linguistics, 2008. 308−316

21 Li F T, Huang M L, Zhu X Y. Sentiment analysis with global topics and local dependency. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence. Carol Hamilton, USA: Association for the Advancement of Artificial Intelligence, 2010. 1371−1376

22 Wang H N, Lu Y, Zhai C X. Latent aspect rating analysis without aspect keyword supervision. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2011. 618−626

23 Moghaddam S, Ester M. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: ACM, 2011. 665−674

24 Mukherjee S, Basu G, Joshi S. Joint author sentiment topic model. In: Proceedings of the 2014 SIAM International Conference on Data Mining. Philadelphia, PA, USA: SIAM, 2014. 370−378

25 Zhao W X, Jiang J, Yan H F, Li X M. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. 56−65

26 Li F T, Wang S, Liu S H, Zhang M. Suit: a supervised user-item based topic model for sentiment analysis. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. Carol Hamilton, USA: Association for the Advancement of Artificial Intelligence, 2014. 1636−1642

27 Moghaddam S, Ester M. The FLDA model for aspect-based opinion mining: addressing the cold start problem. In: Proceedings of the 22nd International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013. 909−918

28 Zhang Y, Ji D H, Su Y, Wu H M. Joint naïve Bayes and LDA for unsupervised sentiment analysis. *Advances in Knowledge Discovery and Data Mining*. Heidelberg, Berlin, Germany: Springer, 2013. 402−413

29 Zhang Y, Ji D H, Su Y, Sun C. Sentiment analysis for online reviews using an author-review-object model. *Information Retrieval Technology*. Heidelberg, Berlin, Germany: Springer, 2011. 362−371

30 Moghaddam S, Ester M. On the design of LDA models for aspect-based opinion mining. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2012. 803−812

31 Li C T, Zhang J W, Sun J T, Chen Z. Sentiment topic model with decomposed prior. In: Proceedings of the 2013 SIAM International Conference on Data Mining. Philadelphia, PA: SIAM, 2013. 767−775

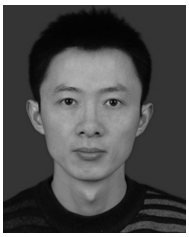
32 Wang X R, McCallum A. Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2006. 424−433

33 Phan X H, Nguyen L M, Horiguchi S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: Proceedings of the 17th International Conference on World Wide Web. New York, NY, USA: ACM, 2008. 91−100

34 Lim K W, Buntine W. Twitter opinion topic model: extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2014. 1319−1328

35 Chang J, Boyd-Graber J L, Gerrish S, Wang C, Blei D M. Reading tea leaves: how humans interpret topic models. In: Proceedings of the 2009 Advances in Neural Information Processing Systems. San Diego, CA, USA: NIPS Foundation, Inc., 2009. 288−296

36 Xie P T, Xing E P. Integrating document clustering and topic modeling. In: Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence. Cambridge, MA, USA: Association for Uncertainty in Artificial Intelligence, 2013.



熊蜀峰 武汉大学计算机学院博士研究生, 平顶山学院讲师. 主要研究方向为自然语言处理, 机器学习和观点挖掘.
E-mail: xsf@whu.edu.cn
(**XIONG Shu-Feng** Ph. D. candidate at the Computer School of Wuhan University and lecturer at PingDingShan University. His research interest covers natural language processing, machine learning, and opinion mining.)

姬东鸿 武汉大学计算机学院教授. 主要研究方向为自然语言处理, 数据挖掘和生物信息处理. 本文通信作者.
E-mail: dhji@whu.edu.cn
(**JI Dong-Hong** Professor at the Computer School of Wuhan University. His research interest covers natural language processing, data mining, and biological information processing. Corresponding author of this paper.)