

基于 TSCM 模型的网络短文本情感挖掘

黄发良¹, 李超雄¹, 元昌安², 汪焱¹, 姚志强¹

(1. 福建师范大学软件学院, 福建福州 350007; 2. 广西师范学院计信学院, 广西南宁 530023)

摘要: 针对网络短文本情感挖掘问题, 提出一种新的基于 LDA 和互联网短评行为理论的主题情感混合模型 TSCM, TSCM 模型中的整篇评论中每个句子的主题分布是不同的, TSCM 产生词的流程是先确定词的情感极性, 再确定词的主题, TSCM 考虑了词与词之间的联系. 真实数据集 Movie 与 Amazon 上的大量实验表明, 与代表性算法 JST、S-LDA、D-PLDA 和 SAS 相比较, TSCM 模型能对用户真实情感与讨论主题进行更加有效的分析建模.

关键词: 情感分析; 主题情感混合模型; LDA

中图分类号: TP273

文献标识码: A

文章编号: 0372-2112 (2016)08-1887-05

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2016.08.017

Mining Sentiment for Web Short Texts Based on TSCM Model

HUANG Fa-liang¹, LI Chao-xiong¹, YUAN Chang-an², WANG Yan¹, YAO Zhi-qiang¹

(1. Faculty of Software, Fujian Normal University, Fuzhou, Fujian 350007, China;

2. School of Computer and Information Engineering, Guangxi Teachers Education University, Nanning, Guangxi 530023, China)

Abstract: For sentiment analysis of web short texts, a topic sentiment combining model (TSCM) is proposed based on LDA and web review behavioral theory, which is founded on the assumption that topic distribution of each sentence in a review is unique and different from that of other sentences. Generative process of TSCM is to first determine sentiment orientation of each word and then topic of each sentence in a review while taking word relation into consideration. Extensive experiments on real-world datasets (Movie and Amazon) show that TSCM significantly outperforms JST, S-LDA, D-PLDA and SAS in terms of the accuracy of sentiment classification and topic detection.

Key words: sentiment analysis; topic sentiment mixture; latent dirichlet allocation (LDA)

1 引言

以自由开放共享为核心精神的 Web 2.0 使得用户成为互联网的主角, 诸如社交网站、微博和 BBS 论坛之类的平台为网民发表意见和交流情感提供了经济便捷的渠道. 一般来说, 用户在这些平台上发表的言论比较简短却又饱含着丰富的个人情感. 研究如何高效挖掘隐藏于这些鱼目混杂的社交媒体网络言论中的观点与情感有助于政府机构、企业组织与理性个体的管理决策.

网络短文本情感挖掘正在吸引着来自人工智能、数据挖掘、自然语言处理等不同领域研究者的广泛关注^[1-3], 涌现出的各种算法大致可归纳为三类: 有监督情感挖掘、无监督情感挖掘与半监督情感挖掘. 有(半)监督情感挖掘方法不同程度地利用训练语料来训练生成文本情感分类器, 一般具有较高的分类准确率, 但获

取训练样本的昂贵代价极大地限制此类方法应用. 因此, 以 JST^[4]、S-LDA^[5] 与 ASUM^[6] 等为代表的无监督情感分类方法近年来备受青睐, 此类方法能有效地避免传统无监督情感分类方法具有的情感词典依赖性缺点, 能达到较好的情感识别效果. 然而, 现有的这些 LDA 情感主题模型还不能很好地捕获网络短评用户的真实情感.

互联网社会学相关研究表明^[7]: 互联网用户在对商品、服务、新闻等对象进行评价时, 往往会先确定评论的情感极性, 然后再对评价对象的各个方面进行评价, 即先确定情感极性, 再确定各个句子主题.

基于上述观测, 本文提出了一个基于 LDA 和互联网短评行为理论的主题情感模型混合 TSCM, 该模型在综合考虑互联网用户短评行为习惯的基础上利用吉布斯采样技术实现情感与主题挖掘, 真实短评数据集上的实验结果表明该模型能较好地互联网短评进行情

感与主题挖掘.

2 相关工作

基于主题模型的无监督情感挖掘主要是通过应用 LDA 主题建模技术^[8]对主观性文本进行学习来实现隐含情感知识的发现.

Mei 等^[9]提出一个主题情感模型 TSM 进行主题及其相关情感的演化分析. Titov 等^[10]应用 MG-LDA 提取评论对象中的各个被评价,然后提出 MAS 模型对情感进行总结, MAS 模型要求评论对象的每个方面至少在部分评论中被评价过,然而,这对真实评论文本数据集来说是不实际的. Dasgupta 等^[11]提出一种基于用户反馈的谱聚类技术进行网络文本的无监督情感分类,聚类分析过程涉及数据特征都是具有情感倾向的主题,然而,在该分析过程中需要人为指定最重要的特征维. Lin 等提出一种基于 LDA 模型的 JST 模型^[4],该模型将文本情感标签加入 LDA,形成一个包含词、主题、情感和文档的四层贝叶斯概率模型. 电影评论数据集上的实验表明, JST 模型的分类效果要优于 Pang 等的有监督分类. 观测到 JST 模型中的 Gibbs 采样推理过程中出现大量“1”的现象, He^[12]对 LDA 模型的目标函数进行修改,即:在建立情感先验分布时,应用广义期望标准来表达情感词的情感期望. Jo 等^[6]提出一个与 JST 类似的情感分类主题模型 ASUM,将 JST 中的主题替换为方面. 为了克服 JST 的不足, Li 等^[5]提出与 JST 类似的四层贝叶斯概率模型 Dependency-Sentiment-LDA,引入一个转移变量来刻画单词之间的情感关联性. Lin 等^[13]提出 JST 的变体 Reverse-JST,该变体在没有层次先验知识时与 JST 是等价的,但在加入情感先验知识时, JST 具有更强的主题情感分类能力. Brody 等^[14]对主题词进行了情感识别,然而没有建立文档或句子的情感模型. 基于产品评分是与产品某个方面质量的优劣是相互依赖的, Moghadda 等^[15,16]提出 ILDA 模型,通过增加相关参数来改进 LDA,依据产品的文本评论同时实现产品属性方面的提取与评分,同时还提出运用 bag-of-phrases 模型对文本提取主题词与情感词的 D-PLDA 模型. Mukherjee 等^[17]提出 SAS 模型,假设我们已有待建模语料的种子词集,然后利用这些种子词集对 aspect 词语进行簇分析,进而得到文本的 aspect 词语与情感词语.

3 网络短文本情感挖掘

为了方便说明 TSCM 模型及其运用,对相关符号进行如下约定: α, β 与 γ 分别是(文档, 句子, 情感)-主题分布、(情感, 主题)-词语分布与文档-情感分布的 Dir 参数, θ, μ 与 π 分别是(文档, 句子, 情感)-主题分布、(情感, 主题)-词语分布与文档-情感分布, z, l 与 w 分别

是主题、情感与词语变量, D, S, W, K, L 与 V 分别是指文档数、句子数、单篇文档的词语数、主题数、情感数与文档词库的词语数.

3.1 主题情感混合模型

“文档-主题-单词”三层贝叶斯模型 LDA(图 1(a))是通过概率推导来寻找数据集的语义结构,从而得到文本的主题. 该模型假设文档是由不同主题组成的且一个主题是单词集合的概率分布,在此假设下,文档单词的产生可分为两个阶段,首先从文档-主题分布中选择一个主题,然后根据随机选择的主题从主题-单词分布中选择一个单词.

为了弥补 LDA 情感层的缺失,我们通过在 LDA 中嵌入情感层构造主题情感混合模型 TSCM(图 1(b)),在 TSCM 中,情感标签与文档关联,主题标签与句子关联,在此基础上,主题又与情感标签关联,词语同时与主题和情感标签关联.

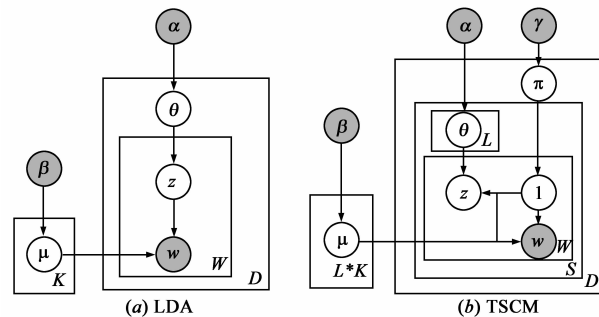


图1 LDA 与 TSCM

对于文档集 $C = \{d_1, d_2, \dots, d_D\}$, 与文档集 C 对应的词典的大小为 V , 文档 d_i 是由 W_i 个单词组成的序列. TSCM 产生文档集 C 的过程可简单归结为如下两个步骤:(1) 初始化 TSCM 模型的先验分布参数 $\Theta = \{\theta, \mu, \pi\}$, 具体地, μ, π 与 θ 分别服从狄利克雷分布 $\text{Dir}(\beta)$ 、 $\text{Dir}(\gamma)$ 与 $\text{Dir}(\alpha)$, 其中 β 是指单词在文档集 C 中出现的先验次数, γ 是指情感极性标签在文档 d 中出现的先验次数, θ 服从 $\text{Dir}(\alpha)$ 分布, α 是指主题在文档 d 的句子 s 中出现的先验次数;(2) 概率生成文档集 C 中的单词, 此生成过程可简单描述如下: 首先从文档-情感分布 π_d 中选出一个情感标签 l, l 服从 $\text{Mul}(\pi_d)$ 分布 ($\text{Mul}(\ast)$ 表示多项分布); 接着根据产生的情感标签 l , 从(文档, 句子, 情感)-主题分布 θ_{dsl} 中选出一个主题 z, z 服从 $\text{Mul}(\theta_{dsl})$ 分布, 这里 θ 与 LDA 的 θ 不同, LDA 中一篇文档只有一个 θ , 而 TSCM 里一篇文档里 θ 的个数是文档句子数与情感极性种类数的乘积; 最后根据选出的情感 l 与主题 z , 从(情感, 主题)-词语分布 μ_{lz} 中选择一个单词 w, w 服从 $\text{Mul}(\mu_{lz})$ 分布.

3.2 模型推导

TSCM 模型利用吉布斯采样技术对概率分布进行

推导. 为了得到参数 θ, μ 与 π 的分布, 我们需要计算联合分布 $p(z_i = z, l_i = l | z_{-i}, l_{-i}, w)$, 其中 z_{-i} 与 l_{-i} 分别是指除了文档 d 中第 i 个词以外的其他词的主题与情感极性. 对联合分布 $P(w, z, l) = P(w | z, l) P(z, l) = P(w | z, l) P(z | l) P(l)$ 进行欧拉展开可得:

$$P(w | z, l) = \left(\frac{\Gamma(V\beta)}{[\Gamma(\beta)]^V} \right)^{L \times K} * \prod_{l=1}^L \prod_{k=1}^K \frac{\prod_{v=1}^V \Gamma(n_{l,k,v} + \beta)}{\Gamma(n_{l,k} + V\beta)} \quad (1)$$

$$P(z | l) = \left(\frac{\Gamma(K\alpha)}{[\Gamma(\alpha)]^K} \right)^{D \times S \times L} * \prod_{d=1}^D \prod_{s=1}^S \prod_{l=1}^L \frac{\prod_{k=1}^K \Gamma(n_{d,s,l,k} + \alpha)}{\Gamma(n_{d,s,l} + V\alpha)} \quad (2)$$

$$P(l) = \left(\frac{\Gamma(L\gamma)}{[\Gamma(\gamma)]^L} \right)^D * \prod_{d=1}^D \frac{\prod_{l=1}^L \Gamma(n_{d,l} + \gamma)}{\Gamma(n_d + L\gamma)} \quad (3)$$

其中 $n_{l,k,v}$ 表示单词 v 同时属于主题 k 和情感极性 l 的频数, $n_{l,k}$ 表示所有同时属于主题 k 和情感极性 l 的单词总的频数, $n_{d,s,l,k}$ 表示在短评 d 中的第 s 个句子中主题 k 属于情感极性 l 的频数, $n_{d,s,l}$ 表示属于情感极性 l 的主题出现在短评 d 中的第 s 个句子中的总频数, $n_{d,l}$ 表示在短评 d 中情感极性 l 出现的频数, n_d 表示短评 d 中情感极性标签的总频数, $\Gamma(*)$ 表示伽马函数.

根据上述公式, 吉布斯采样联合概率可表示为:

$$p(z_i = z, l_i = l | z_{-i}, l_{-i}, w) \propto \frac{\{n_{d,s,l}^k\}_{-i} + \alpha}{\{n_{d,s,l}\}_{-i} + K\alpha} * \frac{\{n_{l,k}^v\}_{-i} + \beta}{\{n_{l,k}\}_{-i} + V\beta} * \frac{\{n_d^l\}_{-i} + \gamma_l}{\{n_d\}_{-i} + \sum_{l=1}^L \gamma_l} \quad (4)$$

其中, $\{n_{d,s,l}^k\}_{-i}$ 表示短评 d 中的第 s 个句子里, 除了当前单词 i 外, 主题 k 属于情感极性 l 的频数, $\{n_{d,s,l}\}_{-i}$ 表示短评 d 中的第 s 个句子里, 除了当前单词 i 外, 属于情感极性 l 的所有主题的总频数. $\{n_d^l\}_{-i}$ 表示除了当前单词外, 单词 v 同时属于主题 k 和情感极性 l 的频数, $\{n_{l,k}\}_{-i}$ 表示除了当前单词外, 所有属于主题 k 和情感极性 l 的单词总频数. $\{n_{l,k}^v\}_{-i}$ 表示除了当前单词外, 短评 d 中情感极性 l 的频数, $\{n_d^l\}_{-i}$ 表示除了当前单词外, 短评 d 中情感极性标签的总频数, 即短评 d 中的单词总数.

参数 $\Theta = \{\theta, \mu, \pi\}$ 的最大似然估计如下:

$$\theta_{dsl}^k = \frac{n_{d,s,l}^k + \alpha}{n_{d,s,l} + K\alpha} \quad (5)$$

$$\mu_{lk}^v = \frac{n_{l,k}^v + \beta}{n_{l,k} + V\beta} \quad (6)$$

$$\pi_d^l = \frac{n_d^l + \gamma_l}{n_d + \sum_{l=1}^L \gamma_l} \quad (7)$$

其中, θ_{dsl}^k 表示短评 d 中的第 s 个句子中, 主题 k 属于情感极性 l 的概率, μ_{lk}^v 表示词语 v 同时属于主题 k 和情感极性 l 的概率, π_d^l 表示短评 d 中情感极性 l 出现的概率. 对于一篇需要进行情感分析的文档, 我们可以根据 $P(l)$ 来算出该短评属于 l 情感极性的概率 (l 为积极情感或消极情感), 若短评属于积极情感的概率大于短评属于消极情感的概率, 则判断该短评的情感极性为积极, 反之, 则为消极.

3.3 情感挖掘

运用 TSCM 模型进行文档情感极性判定的过程可描述如下:

对于短评集 $C = \{d_1, d_2, \dots, d_{|C|}\}$, 首先对每篇短评 $d = \{w_1, w_2, \dots, w_n\}$ 中的每个单词 w 随机分配情感极性 l 与主题 k , 并且更新参数向量 $Temp\theta = (n_{d,s,l}^k, n_{d,s,l}, n_{l,k}^v, n_{l,k}, n_d^l, n_d)$, 直到短评集 C 中每个短评 d 的所有单词都被分配情感极性 l 与主题 k . 然后对每篇短评 d 中的每个单词 w , 计算 $p(z_i = z, l_i = l | z_{-i}, l_{-i}, w)$, 并且更新向量 $Temp\theta$, 循环这个过程 X 次 (X 是指定的循环控制参数). 接着从第 $X+1$ 次开始, 对每篇短评 d 中的每个单词 w , 计算 $p(z_i = z, l_i = l | z_{-i}, l_{-i}, w)$, 更新向量 $Temp\theta$, 并且每 Y 次 (Y 是人为规定的参数) 更新 $\theta_{dsl}^k, \mu_{lk}^v$ 和 π_d^l 的值, 循环这个过程直到总的循环次数 (包括前面的 X 次) 达到 Z 次 (Z 是人为规定的总循环次数). 最后通过 π_d^l 我们就可以判断文档的情感极性, 对于短评 d , 若 $\pi_{d_1}^l > \pi_{d_2}^l$ (其中 l_1 为积极情感, l_2 为消极情感) 则短评 d 的情感极性为积极情感, 反之短评 d 的情感极性则为消极情感.

4 实验与分析

为了定量分析 TSCM 的性能, 我们选择代表性情感混合模型 JST、S-LDA、D-PLDA、SAS 作为比较对象, 利用基准评论数据集 (Movie, Amazon 中的 Books, Music, Electronics) 分别从情感分类准确率、主题数对准确率的影响与主题提取三方面进行分析.

4.1 准确率分析

各种模型在不同数据集上分类准确率的实验结果见图 2. 由图 2 可以看出: (1) 四种模型都在长评论集 Movie 具有相对较高的分类准确率 (考虑到 SAS 需要种子词, 而我们的实验语料只有 Movie 数据集有种子词, Amazon 数据集并没有种子词, 所以本文只用 SAS 对 Movie 数据集进行实验), 而在短评论集 Amazon 的分类准确率相对较低, 这说明四种模型在情感分析时都存在着信息量丰富度更大的评论的偏好; (2) 和 JST、S-LDA、D-PLDA、SAS 相比较, 除了在数据集 Music 上, TSCM 的分类正确率在其他所有的数据集上都要明显高于其他四者, 尤其在长评论集上 TSCM 的优势表现特别突出, 即使是数据集 Music, TSCM 在分类正确性上也

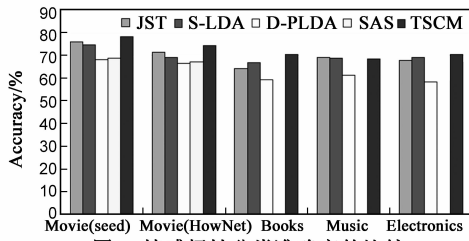


图2 情感极性分类准确率的比较

与模型 JST、S-LDA 表现相当并高于 PLDA; (3) 对比 Movie 数据集在不同词典下的准确率发现, HowNet 先验处理方法导致的情感分类准确率要低于 seed 先验处理方法, 这可能是某些情感词的 HowNet 预定情感极性值与其在 Movie 中的实际情感极性值存在差异, 从而影响了模型对情感极性的判断. 通过上述分析, 不难得出如下结论: (1) TSCM 可以有效地提高网络短文本情感极性分类的准确率; (2) 和 JST、S-LDA、D-PLDA 与 SAS 相比较, TSCM 更适合对较长评论进行情感极性分类, 同时 TSCM 也存在种子情感词的域相关问题.

4.2 主题数对准确率的影响

考虑到 TSCM 模型的主题情感混合特性, 在此对主题数对情感极性分类准确率产生的影响展开实验分析. 实验结果如图 3 所示. 由图 3 可知, 在 4 个实验数据集中, TSCM 的准确率几乎在所有不同主题数下都比 JST 和 S-LDA 高, 只有极少数情形 (主题数为 5、25 的 Movie, 主题数为 10 的 Music). 这表明了 TSCM 具有比 JST 和 S-LDA 更好的性能.

4.3 主题提取

TSCM 能同步进行评论的主题和情感分析, 本实验利用 TSCM 模型从 4 个数据集中分别提取积极情感与消极情感主题词, 并以此来评估提取出的单词对于判断情感极性是否有用. 实验结果如表 1 所示, 在此仅列出出现概率最高的 20 个单词.

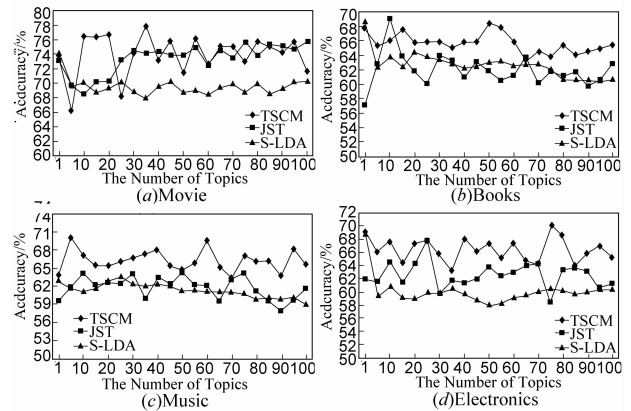


图3 主题数鲁棒性的比较

从表 1 可以看出, Movie 数据集中提取的主题词存在较多的情感词, 比如, 积极情感词 good、amaze、pretty 等与消极情感词 bad、bore、worst 等, 积极情感词中 good 出现的频率最高, 而消极情感词中 bad 与 bore 出现频率最高. 此外在 Movie 数据集下也提取出与电影有关的主题词 (如: 积极情感下的 director、actor 应是电影的导演与演员有关). 而从 Books、Music 和 Electronics 数据集中提取的主题词则较少出现情感词, 具体地说, 在 Books 中提取的主题词主要与书名或书的内容有关 (如: Fair 是描述书里经常出现的集市. Whitomb 是书中出现的惠特科姆酒店). 在 Music 数据集中提取的主题词主要与乐器、歌名或歌的内容有关 (如: Piano、Rain 可能是描述韩国钢琴家李闰珉弹的钢琴曲 kiss the rain. life、woman 可能指一些描述女人生活的歌曲). 在 Electronics 数据集中提取的主题词主要是电子产品的品牌、属性等 (HONDA、Nissan 都是汽车品牌, 而消极情感下的 sound、quality 则可能是描述一款耳机的声音质量). 由此可见, TSCM 提取主题词对我们理解短评内容与短评情感极性都有重要的作用, 因而其有效实现了主题发现与情感分析的完美结合.

表 1 不同数据集中提取的主题词

数据集	积极情感词	消极情感词
Movie	film movie make character scene time good play stori end director work thing plot amaze pretty show actor feel perform	bad bore worst half stupid wast terrible clich aw slow disappoint pointless dozen player lame unfunny danger offenses wise preview
Books	contest contract valid predecessor Fair Pope confront revise induce Clark DOE rescue tender enjoy Anyway autom glow verbatim rock compete	Dru Whitcomb Bank surprise disappoint event opportune shame uninspire detach connect grain today follow disturb area property Mosley eesource Arsenal-Chelsea
Music	Famili Muldowney dimens opend Gounod diction LHCB improvise Happen Lone Steven clever even Curti medit Parti Atmosphere Wright Piano Rain	entertain by Ludacri tell life women And speak met h-o produce put decent studio please record disc flat-out atrocious download
Electronics	Quality Deserv Power Key Come Honda Civic Nissan Altima c340 lost sell real cigarette Write reflect buyer leak upper confide	bought return unit Each final gave The player constantly Bad error skip slightest smudg sound quality nice price work essential

5 结束语

随着 Web2.0 的快速发展,人们可以在网上针对现实事件进行评论,挖掘隐藏在这些个性化评论中的情感与观点能有效辅助用户个体、企业组织等的决策行为. 针对传统主题情感混合模型的不足,本文提出了一个新的基于 LDA 与互联网短评行为理论的主题情感混合模型 TSCM. 实验表明,TSCM 具有良好的网络短文本情感分析性能.

参考文献

- [1] Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and trends in information retrieval, 2008, 2(1-2): 1-135.
- [2] Tang H, Tan S, Cheng X. A survey on sentiment detection of reviews [J]. Expert Syst Appl, 2009, 36(7): 10760-10773
- [3] 吕品, 钟璐, 唐琨皓. 在线产品评论用户满意度综合评价研究[J]. 电子学报, 2014, 42(4): 740-746.
Lv P, Zhong L, Tang K. Customer satisfaction degree evaluation of online product review[J]. Acta Electronica Sinica, 2014, 42(4): 740-746 (in Chinese)
- [4] Lin C, He Y. Joint sentiment/topic model for sentiment analysis[A]. Proceedings of CIKM[C]. New York: ACM, 2009. 375-384.
- [5] Li F, Huang M, Zhu X. Sentiment analysis with global topics and local dependency[A]. Proceedings of AAAI[C]. Atlanta: AAAI, 2010. 1371-1376.
- [6] Jo Y, Oh A H. Aspect and sentiment unification model for online review analysis[A]. Proceedings of WSDM[C]. NY: ACM, 2011. 815-824.
- [7] 淘宝评价流程[EB/OL]. <http://wenku.baidu.com/view/1ea83bd751e79b8969022629.html>, 2014-08-15.
- [8] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [9] Mei QZ, Ling X, Wondra M, et al. Topic sentiment mixture: Modeling facets and opinions in weblogs[A]. Proceedings of WWW[C]. New York: ACM, 2007. 171-180
- [10] Titov I, McDonald R. Modeling online reviews with multi-grain topic models[A]. Proceedings of WWW[C]. NY: ACM, 2008. 111-120
- [11] Dasgupta S, Ng V. Topic-wise, sentiment-wise, or otherwise? Identifying the hidden dimension for unsupervised text classification[A]. Proceedings of EMNLP[C]. Singapore: ACL 2009. 580-589
- [12] He Y. Latent Sentiment Model for Weakly-Supervised Cross-lingual Sentiment Classification[M]. Advances in Information Retrieval. Berlin: Springer, 2011. 214-225.
- [13] Lin C, He Y, 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.
- [14] Brody S, Elhadad N. An unsupervised aspect-sentiment model for online reviews[A]. Proceedings of ACL[C]. Stroudsburg: ACL, 2010. 804-812.
- [15] Moghaddam S, Ester M. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews[A]. Proceedings of SIGIR[C]. New York: ACM, 2011. 665-674.
- [16] Moghaddam S, Ester M. On the design of LDA models for aspect-based opinion mining[A]. Proceedings of CIKM[C]. New York: ACM, 2012. 803-812.
- [17] Mukherjee A, Liu B. Aspect extraction through semi-supervised modeling[A]. Proceedings of ACL[C]. Stroudsburg: ACL, 2012. 339-348.

作者简介



黄发良 男, 1975 年生于湖南永州. 福建师范大学软件学院副教授. 研究方向为数据挖掘、智能信息系统.
E-mail: huangfl@fjnu.edu.cn



李超雄 男, 1991 年生于福建莆田. 硕士研究生, 研究方向为数据挖掘与知识发现.