文章编号: 1004-4353(2018)03-0266-08

基于 LDA 主题模型的文本语料情感分类改进方法

郭晓慧
(阳光学院 信息工程学院, 福建 福州 350015)

摘要: 针对传统 LDA 主题模型无法体现词与词之间的顺序及关联性这一不足,提出一种改进的加权 W-LDA 情感分类方法. 首先,在该模型的主题采样及其分布期望计算过程中引入平均加权值,以此避免与主题紧密相关词被高频词所淹没,从而提高主题间的区分度;然后,以提取到的高质量文档-主题分布及主题-词向量为基础,引入支持向量机算法(SVM),构建一个集有情感词分析与提取、主题分布计算与情感分类功能的文本语料情感分析方法;最后,利用真实的教学评价数据和公共评论集对本文方法的有效性进行了验证. 结果表明,本文提出的方法在主题区分度、分类准确率以及 F1-Measure 方面均明显优于 SVM 算法和文献[15]中的算法.

关键词: 评论语料; LDA 主题模型; 支持向量机; 情感分类
中图分类号: TP309.3 **文献标识码**: A

The improved method based on LDA topic model for emotion classification of text corpus

GUO Xiaohui
(*Institute of Information Engineering, Yango University, Fuzhou* 350015, *China*)

Abstract: An improved weighted W-LDA emotional classification method is proposed to solve the problem that the traditional LDA topic model can not reflect the order and relevance among words. Firstly, the average weighted value is used in the theme sampling and distribution expectation calculation process of the model, which avoid some important words related to the theme were drowned by high-frequency words. So these measures contribute to improve the degree of descrimination among the subjects. Secondly, based on the extracted high-quality document-subject distribution and theme-word vector, with the support vector machine algorithm (SVM) involved, a emotion classification method on comentary corpus is proposed in this article. Its functions include the analysis and exaction of emotion words, the topic distribution computation and emotion classifiction. Finally, some experiments are perfomed on the real teaching evaluation data and public comment data. The experimental results show that the proposed method has many advantages over the classific SVM and literatur [15] for the degree of descrimination the topics, the classification accuracy and F1-Measure.

Keywords: commentary corpus; LDA topic model; support vector machine; emotion classification

0 引言

文本语料情感分析[1]又称意见挖掘,它通过对网上用户的评论进行分析、处理、归纳和推理来识别

收稿日期: 2018-05-21 **基金项目**: 福建省教育厅科研项目(JA15631)
作者简介: 郭晓慧(1984—),女,讲师,研究方向为个性化推荐算法、数据挖掘.

文本中隐含的情感信息，目前已被广泛应用于商品、新闻评论、教学评价等诸多领域[2]. 文本语料情感分析通常包括语句分词、特征提取与选择、分类模型、识别结果 4 个步骤. 目前，国内外学者对文本语料情感分类方法进行了一些研究，并取得了一些成果. 例如：在特征集提取方面，Hai Zhen 等[3]提出了一种两阶段共现关联规则挖掘方法，该方法采用共生矩阵从观点词挖掘最优关联规则，有效地提高了特征提取精度；Wang Wei 等[4-5]采用混合关联规则挖掘方法和基于片面意见异步挖掘方法有效地挖掘出隐式特征集，提高了有效特征选择的正确性. 在处理语句分词方面，张庆庆等[6-8]使用情感词典对语料集进行分词处理，由于充分考虑到了否定词、大小写字母等词性，大幅增强了情感极性值，得到了较好的中文分词效果. 在分类模型构建方面，刘鸿宇等[9-10]将监督学习方法应用于情感分类，通过采用一元特征、二值特征、形容词打分、位置和特征权重选择等策略，形成了一种集有多种机器算法的情感分类方法，该方法在评论数据集的情感分类上具有良好的效果. 近年来，主题模型越来越受到研究者的重视，其中潜在狄利克雷分配(Lejeune Dirichlet allocation，LDA)模型[11]是一个非常经典的主题模型，可用于识别大型文档集或语料库中的潜在主题信息. 2011 年，夏火松等[12]将 LDA 模型应用于动态自动标注，该模型可引出资源和相关语言聊天的潜在主题，从而可以根据潜在主题对资源进行有效标记. 2016 年，JIN J 等[13]对 LDA 模型做了特征提取和情感极性强度的改进，并应用于社交媒体用户推荐，取得了较好的应用效果. 综上，虽然在情感分类方法及其应用方面取得了一些成果，但是仍然存在诸多问题需要解决，例如词典不够完善、主题区分度不明显等. 为了提高 LDA 的主题分布效果和评论性样本的情感分类质量，本文提出一种基于 LDA 和 SVM 的高精度情感分类方法，并通过实验验证本文方法的有效性.

1　LDA 主题模型的构建

1.1　主题模型思想

LDA 主题模型的主要目标就是从一篇文档或多篇文档计算出其主题-词的概率分布情况. LDA 主题模型是一种典型的词袋模型，也是一种无监督学习算法[14]. 本文讨论的 LDA 主题模型的符号及相关说明如表 1 所示.

表 1　LDA 主题模型中的符号及相关说明

符号	解释说明
K	主题数量
m	文档的数量
n	词数量
M_m	第 m 个文档
N_m	第 m 个文档所包含的词数量
V	所有词总数
$\boldsymbol{\alpha}$	文档集下 Dirichlet 先验超参数，是一个 K 维向量
$\boldsymbol{\beta}$	任一主题下 Dirichlet 先验超参数，是一个 V 维向量
$Z_{m,n}$	第 m 个文档的第 n 个词的主题
$\boldsymbol{\varphi}_{z_{m,n}}$	主题 $Z_{m,n}$ 下的词分布
$\boldsymbol{\theta}_m$	第 m 个文档的主题分布
$W_{m,n}$	第 m 个文档的第 n 个词

假如给定一个主题文档集合 M，它含有 m 个文档，即 $M=\{M_1,M_2,\cdots,M_m\}$，含有 K 个相互独立的主题，每个主题呈现随机多项式概率分布特点. 每个主题下又存在若干个多项式概率分布词汇，这里的多项式概率分布均满足 Dirichlet 分布. 上述文档的表现形式可刻画为 LDA 主题模型，它的生成过程如下：

1) 从 Dirichlet 分布先验知识 $\boldsymbol{\alpha}$ 中取样生成文档 M_m 的主题分布 $\boldsymbol{\theta}_m$.

2) 从 1) 中的主题分布 $\boldsymbol{\theta}_m$ 中取样生成文档 M_m 的第 n 个词汇的主题 $Z_{m,n}$.

3) 从 Dirichlet 分布先验知识 $\boldsymbol{\beta}$ 中重复取样生成与主题 $Z_{m,n}$ 相对应的词汇分布 $\boldsymbol{\varphi}_{z_{m,n}}$.

4) 结合主题 $Z_{m,n}$，从词分布 $\boldsymbol{\varphi}_{z_{m,n}}$ 中采样生成词汇 $W_{m,n}$.

5) 重复执行以上步骤 2) 和 3)，重复执行以上步骤 2)、3)、4)，直至所有词汇被选择.

上述 LDA 生成过程的直观表述如图 1 所示. 图 1 中的矩形表示循环，矩形右下角的数据表示循环次数，空心圆表示隐蔽变量，带底纹的圆表示可观测的变量，带方向的箭头表示各变量间的依赖关系.

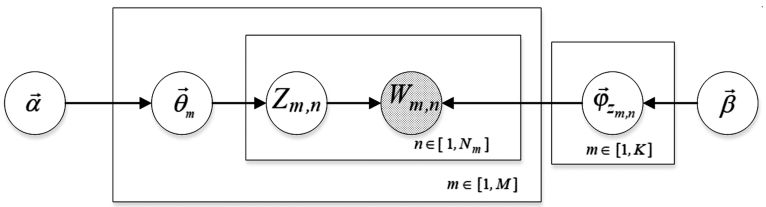


图 1 LDA 主题模型的生成过程

假定现有一个文档集合 M，$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别为文档主题及主题下的词汇下的分布先验参数，$\phi=\{\boldsymbol{\varphi}_{Z_{m,n}}\}$，$K\times V$ matrix，则与图 1 相对应的模型中所有参数的联合分布概率计算如式(1) 所示：

$$P(\boldsymbol{W}_m,\boldsymbol{Z}_m,\boldsymbol{\theta}_m,\phi)=\prod_{n=1}^{N_m}P(W_{m,n}\mid\boldsymbol{\varphi}_{Z_{m,n}})P(Z_{m,n}\mid\boldsymbol{\theta}_m)P(\boldsymbol{\theta}_m\mid\boldsymbol{\alpha})P(\phi\mid\boldsymbol{\beta}). \tag{1}$$

1.2 改进的 W-LDA 主题模型的推导

本文研究的数据样本主要针对的是高校教学评价数据. 文献[12,14-15]的研究结果表明，如果仅使用传统的 LDA 模型，高频词会将代表主题的多数词淹没掉，从而使得主题表达能力受到较大影响. 为解决这一问题，有研究者通过停用词或者设置阈值剔除掉高频词，但是这些方法在一些应用领域中由于缺乏停用词词库或者词库不全，且阈值大小难以把握，因此效果并不理想. 在此背景下，本文使用高斯函数对特征词作加权处理，以此构建一种改进的 W-LDA 主题模型. 具体的实现方法及过程如下：

首先运用式(2) 对文档中的词 a_m 加权：

$$a_m=\exp\left[-\frac{(f_m-f_i)^2}{2\sigma^2}\right], \tag{2}$$

其中方差 $\sigma^2=\dfrac{\sum_{m=1}^{V}(f_m-f_i)^2}{V-1}$，$f_m$ 是第 m 个词的词频数，f_i 是词频数中间的第 i 个词的词频数. 为使数据集被加权前后其总词数大体一致，在式(2) 的基础上作均值化处理，处理公式如式(3) 所示. 式(3) 中的平均加权值 weight_m 将被应用于主题及其词语分布期望中的计算.

$$\text{weight}_m=\frac{n}{\left[\sum_{m=1}^{V}f_m\cdot a_m\right]}\cdot a_m. \tag{3}$$

在 LDA 主题模型推导中，本文采用 Gibbs 抽样算法，该算法快速高效，可生成马尔科夫链，继而可求得一个复杂的多元分布[16]. 从式(1) 和图 1 可知，主题多项式分布 θ 由先验参数 $\boldsymbol{\alpha}$ 生成，$Z_{m,n}$ 主题下的词语多项式分布 ϕ 是由 $\boldsymbol{\beta}$ 生成，则式(1) 可被转化为如式(4) 所示的联合概率分布：

$$P(\boldsymbol{W},\boldsymbol{Z}\mid\boldsymbol{\alpha},\boldsymbol{\beta})=P(\boldsymbol{Z}\mid\boldsymbol{\alpha})P(\boldsymbol{W}\mid\boldsymbol{Z},\boldsymbol{\beta}). \tag{4}$$

求解式(4) 分为两个过程：一是根据先验参数 $\boldsymbol{\alpha}$ 采样主题 $\boldsymbol{Z}$ 的过程，二是根据主题 $\boldsymbol{Z}$ 和先验参数 $\boldsymbol{\beta}$ 采样词语的过程. 式(4) 中的因子 $P(\boldsymbol{Z}\mid\boldsymbol{\alpha})$ 的计算公式为：

$$P(\boldsymbol{Z}\mid\boldsymbol{\alpha})=\int P(\boldsymbol{Z}\mid\theta)P(\theta\mid\boldsymbol{\alpha})\mathrm{d}\theta=\prod_{m=1}^{M}\frac{\Delta(\boldsymbol{n}_m+\boldsymbol{\alpha})}{\Delta(\boldsymbol{\alpha})},\ \boldsymbol{n}_m=\{n_m^{(k_i)}\}_{k=1}^{K}. \tag{5}$$

其中，$n_m^{(k_i)}$ 代表第 k 主题在第 m 文档中出现的次数.

同理，可得到式(5) 中的第 2 个因子的展开公式：

$$P(\boldsymbol{W}\mid\boldsymbol{Z},\boldsymbol{\beta})=\int P(\boldsymbol{W}\mid\boldsymbol{Z},\phi)P(\phi\mid\boldsymbol{\beta})\mathrm{d}\phi=\prod_{Z=1}^{K}\frac{\Delta(\boldsymbol{n}_Z+\boldsymbol{\beta})}{\Delta(\boldsymbol{\beta})},\ \boldsymbol{n}_Z=\{n_Z^{(v)}\}_{v=1}^{V}. \tag{6}$$

其中，$n_Z^{(v)}$ 表示在主题 Z 中词语 v 出现的次数，K 为模型的乘积.

综合式(5) 和式(6)，可得到词语和主题的联合分布公式为

$$P(\boldsymbol{W},\boldsymbol{Z}\mid\boldsymbol{\alpha},\boldsymbol{\beta})=\prod_{Z=1}^{K}\frac{\Delta(\boldsymbol{n}_Z+\boldsymbol{\beta})}{\Delta(\boldsymbol{\beta})}\cdot\prod_{m=1}^{M}\frac{\Delta(\boldsymbol{n}_m+\boldsymbol{\alpha})}{\Delta(\boldsymbol{\alpha})}. \tag{7}$$

在式(7)基础上，依据变量$\boldsymbol{W}$下的隐性主题变量Z的条件分布和狄利克雷分布情况，可求得第m文档的主题及主题下的词语分布期望为：

$$\theta_{m,k}=\frac{n_m^{(k)}+\alpha_k}{\sum_{k=1}^{K}n_m^{(k)}+\alpha_k},\ \varphi_{k,v}=\frac{n_k^{(v)}+\beta_v}{\sum_{v=1}^{V}n_k^{(v)}+\beta_v}. \tag{8}$$

根据式(3)平均加权值的计算方法，当把主题分布到第m文档时，对$n_m^{(k)}$变量不是简单地累加1，而是累加平均加权值.

2　基于评论语料集的情感分析

首先使用 W-LDA 主题模型获得评论语料的主题分布及主题下的词语概率分布表，然后再运用 SVM 对评论做正负二分类，并对分类结果作出科学评价. 基于评论集的情感分类框架如图 2 所示.

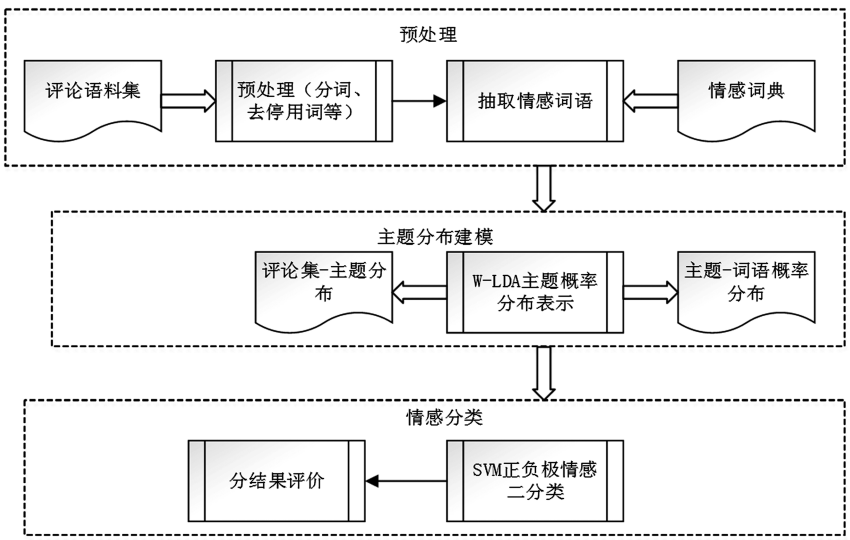


图 2　基于评论语料集的情感分类处理框架

2.1　情感词分析

首先利用情感词典对教学评论语料集进行情感分析. 众所周知，中文情感词典较少，目前被普遍使用的有清华大学李军中文褒贬义词典、知网情感词典和台湾大学中文情感极性词典. 本文在这 3 种情感词典基础上，通过人工处理构建一套适用于教学评价内容的情感词典集，如表 2 所示.

表 2　中文情感词典

类别	个数	分类依据	部分词语
主张词	97	自我定义、感觉	感觉、发现等
正面情感词	1 381	对事物的喜欢	喜欢、真好等
正面评价词	10 841	对事物的肯定、支持	不错、认真等
负面情感词	1 453	对事物的厌恶、不喜欢	讨厌、难受等
负面评价词	12 271	对事物的反对、否定	差、难懂等
程度词	524	形容词或副词的修饰、限定	很、极等

在本文新建的情感词典基础上，以教学评价数据为样本进行实验. 实验发现，从教学评论集上抽取的情感词极性更加明显，非常有利于主题模型的特征表示. 表 3 为测试样本集中的一条教学评价内容的

情感词抽取结果.

表 3 某一教学评价内容的情感词抽取结果

原评价语料	教学手段新颖,引人入胜,深入浅出,肢体语言十分丰富,讲解生动,我真心喜欢,老师讲课的速度降一点就更好了
情感词抽取结果	新颖 引人 入胜 十分 丰富 生动 真心 喜欢 老 更 好 一点

2.2 SVM 分类模型

Corinna Cortes 等提出的支持向量机模型[9]在小样本、非线性和高维模式识别方面优势明显. 依据 RBF 核函数相对稳定,而多项式核函数稳定性较差这一情况,本文选择 RBF 核函数来实现支持向量机模型. 该核函数表现形式为

$$K(X_i, X_j) = \exp(-\gamma \| X - X_e \|^2). \tag{9}$$

公式(9)是空间中任意点 x 到中心 X_e 的欧氏距离的单调函数,该核函数有两个参数:罚因子 C 和核参数 γ. C 越大,误差容限越高;γ 越高,支持向量越少. 本文基于 SCKIT-学习机器学习工具包,采用网格搜索法确定 C 值和 γ 值. 在情感主题和词语提取的基础上,本文再引入 SVM 分类模型对评论数据作情感分类,具体实现过程如下:

1) SVM 的训练过程. 选择 RBF 核函数把 W-LDA 的主题特征分布映射到高维特征空间,然后在高维特征样本空间找出样本的最优分类超平面,进而把它们训练为支持向量和 VC 可信度,从而得到评论语料的分类判别函数.

2) SVM 的分类判断过程. 利用 RBF 核函数将待分类评论集特征向量映射到高维特征空间,并把它作为训练阶段得到的分类判别函数的输入参数,最后由该判别函数输出评论集的情感二分类结果.

上述实现步骤的直观描述如图 3 所示.

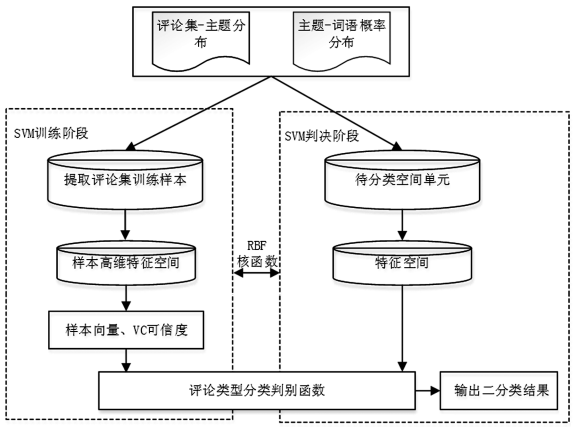


图 3 基于 SVM 的评论情感分类流程

2.3 W-LDA 特征表示与 SVM 分类相结合的情感分析实例

首先在 W-LDA 模型中设置一定主题数 T,把评论文档主题分布范围设置为$[2, T]$,以此获得主题特征(表示向量),然后使用 SVM 对主题特征向量集进行分类. 实验结果表明,经 W-LDA 主题模型处理后的数据集的情感辨识更加精确,具体结果如表 4 所示. 由此可知,在 SVM 算法中融合 W-LDA 模型,其分类质量得到大幅提升,同时也说明使用本文提出的 W-LDA 模型提取的特征更加适用于 SVM 算法分类.

表 4 教学评价数据的情感特征表示

数据	正向词	程度词	负向词
文本评教 1	积极,创造性,激情,感染力	特别	否
文本评教 2	好,清晰,负责任	非常,已经	空
文本评教 3	高,水平	充分,然后	否,少
文本评教 4	激情,严厉,负责	通常	小,不
文本评教 5	鼓励,帮助,大爱	屡次	空
文本评教 6	责备,爱心,友好	经常,充分	也不

常规的教学评价标准通常有 5 个,分别是:教学态度、交流互动、专业技能、教学质量和语言表达. 利用本文改进的 LDA 主题模型对教学评价样本集的特征主题向量的词性进行分析,得到的情感极性分

布结果如图 4 所示. 由图 4 可知，学生在教学质量和交流互动两个主题上对教师的评价较高，而在教学态度、专业技能、语言表达等方面的评价处于中等水平.

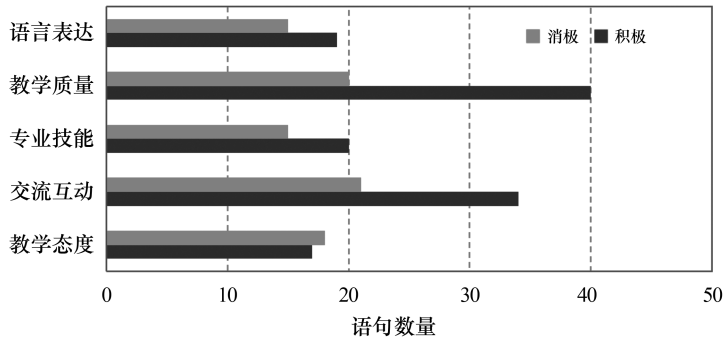


图 4　部分教学评价内容的情感词极性分析结果

为了进一步验证 SVM 分类模型的效果，本文使用 387 条真实教学评价指标得分数据进行验证. 对数据处理后，利用 W-LDA 主题模型计算得到该模型的相关参数分别为：$\alpha=2$，$\beta=0.01$，$K=25$. 实验结果表明，当 $K=25$ 时，模型的精度最高.

为了验证 RBF 核函数的 SVM 分类器效果，本文利用不同的核函数进行了 4 组实验，结果表明基于 RBF 核函数的 SVM 分类器效果较好(图 5).

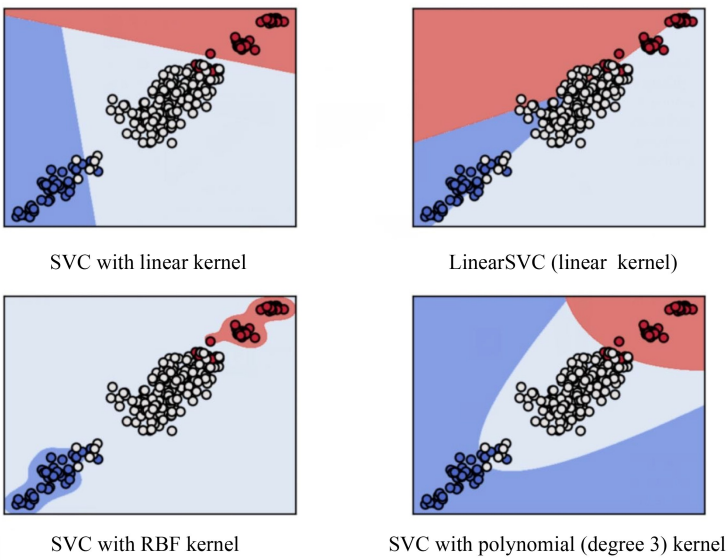


图 5　不同核函数下的 SVM 分类效果

3　实验及结果分析

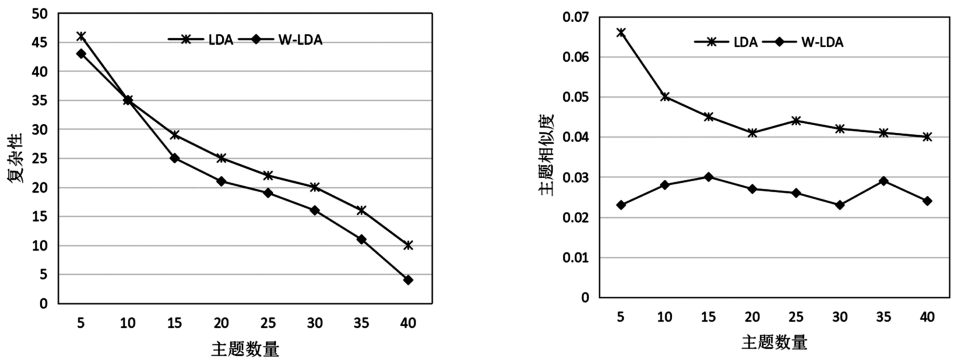
本文实验的硬件环境为：英特尔(R)核心(TM) i7-6700 3.40 GHz，8 G 主存和 1 TB 硬盘的 CPU. 操作系统是 Windows 10，算法的程序代码编写选用 python3 开发工具. 实验选用的分词工具为 Jieba，主题模型采用文中提出的 W-LDA 主题模型算法，分类方法为 RBF 核函数的 SVM 支持向量机算法. 为检验本文算法的有效性，本文同经典 SVM 分类算法及文献[15]中的分类方法进行了对比. 采用的算法评价指标有：准确率(Presicion)、召回率(Recall)和 F1-Measure，其计算公式为：$Presicion$ = 正确分类的记录数 / 分类出的记录总数，$Recall$ = 正确分类的记录数 / 样本中所有已标注的记录数，$F1\text{-}Measure = 2 \times Precisin \times Recall/(Precision + Recall)$. 语料集取于数据堂数据库，其中包括酒店评论集及电子产品评论集两类语料. 这两类语料的极性分布如表 5 所示.

表 5 实验语料集的极性分布情况

语料类别	正向评论数	负向评论数
电子产品评论集	2 730	2 650
酒店评论集	1 080	1 100

3.1 情感主题分布实验

实验采用表 5 所示的通用电子产品评论数据集,使用的情感字典如表 2 所示. 使用 5 折交叉方法验证实验结果,先验参数 $\alpha=50/K$, $\beta=0.01$, 每次训练迭代 1 000 次,实验结果如图 6 所示.



(a) 复杂度对比 (b) 主题独立性对比

图 6 两种主题模型的实验效果对比

由图 6 可知,本文提出的 W-LDA 主题模型比传统的 LDA 主题模型在主题分布计算的复杂值和主题间平均相似度方面有较大优势,这说明 W-LDA 主题模型对评论数据集的主题分布预测能力更强,得到的主题分布质量也更高.

3.2 情感分类实验

在表 5 所示的实验数据(酒店评论集)中抽取 200 条隐含 10 个主题数的评论记录,400 条隐含 20 个主题数的记录,800 条隐含 30 个主题数的记录,1 600 条隐含 40 个主题数的记录. 对上述 4 个不同主题数的数据样本,分别作 4 组实验. 3 种算法的实验结果如表 6 所示.

表 6 情感分类算法实验结果对比

主题数	实验方法	标注记录数	分类结果	正确数	错误数	准确率	召回率	F 值
10	文献[15]方法	200	195	183	12	0.94	0.92	0.93
	SVM 算法	200	182	162	20	0.89	0.81	0.85
	本文方法	200	174	173	1	0.99	0.87	0.93
20	文献[15]方法	400	372	331	41	0.89	0.83	0.86
	SVM 算法	400	384	320	64	0.83	0.80	0.82
	本文方法	400	352	331	21	0.94	0.83	0.88
30	文献[15]方法	800	773	676	97	0.87	0.85	0.86
	SVM 算法	800	720	589	131	0.82	0.74	0.78
	本文方法	800	727	661	66	0.91	0.83	0.87
40	文献[15]方法	1 600	1461	1 255	206	0.86	0.78	0.82
	SVM 算法	1 600	1 433	1 106	327	0.77	0.69	0.73
	本文方法	1 600	1 381	1 235	146	0.89	0.77	0.83

表 6 结果表明:本文方法的正确率和 F1 测量值均优于文献[15]中的算法,但本文方法的召回率略低;同经典 SVM 算法相比,本文方法在正确率、召回率和 F1 测量值方面均具有明显优势.

4　结论

本文研究表明，本文提出的平均加权主题模型（W-LDA 模型）能有效提高主题区分度，且其分类结果的正确率和 F1-Measure 值均优于传统的 SVM 算法和文献[15]中的算法，具有一定的应用价值. 在未来研究中，我们将在数据特征提取阶段充分考虑主题与主题之间的关系，并建立更加完善的情感词库，以期进一步提升算法的分类质量.

参考文献：

[1] 庄丽榕，叶东毅. 基于 CSLSTM 网络的文本情感分类[J]. 计算机系统应用，2018，27(2)：230-235.

[2] 周红庆，吴扬扬. 中文客户评论对象特征的抽取与聚类方法[J]. 微型机与应用，2014，33(15)：69-71.

[3] Hai Zhen, Chang Kuiyu, Kim J. Implicit feature identification via co-occurrence association rule mining[C]//Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing, 2011：393-404.

[4] Wang Wei, Xu Hua, Wan Wei. Implicit feature identification via hybrid association rule mining[J]. Expert Systems with Applications, 2013，40(9)：3518-3531.

[5] Chinsha T C, Joseph S. Asyntactic approach for aspect based opinion mining[C]//2015 IEEE International Conference on Semantic Computing, 2015：24-31.

[6] 张庆庆，刘西林. 基于深度信念网络的文本情感分类研究[J]. 西北工业大学学报(社会科学版)，2016，36(1)：62-66.

[7] Tang D Y, Qin B, Liu T, et al. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015：1422-1432.

[8] 唐晓波，朱娟，杨丰华. 基于情感本体和 KNN 算法的在线评论情感分类研究[J]. 情报理论与实践，2016，39(6)：110-114.

[9] 刘鸿宇，赵妍妍，秦兵，等. 评价对象抽取及其倾向性分析[J]. 中文信息学报，2010，24(1)：84-88.

[10] 尹裴，王洪伟. 面向产品特征的中文在线评论情感分类：以本体建模为方法[J]. 系统管理学报，2016，25(1)：103-114.

[11] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003，3(4)：993-1022.

[12] 夏火松，刘建，朱慧毅. 中文情感分类挖掘预处理关键技术比较研究[J]. 情报杂志，2011，30(9)：160-163.

[13] Jin J, Liu Y, Ji P, et al. Understanding big consumer opinion data for market-driven product design[J]. International Journal of Production Research, 2016，54(10)：3019.

[14] 李实，叶强，李一军，等. 中文网络客户评论的产品特征挖掘方法研究[J]. 管理科学学报，2009，12(2)：185-189.

[15] 李杰，李欢. 基于深度学习的短文本评论产品特征提取及情感分类研究[J]. 情报理论与实践，2018，41(2)：141-146.

[16] 杨丰凯，袁海静. 稳健学生 t 回归模型变点估计的 Gibbs 抽样算法[J]. 统计与决策，2017，22(16)：10-14.

[17] Lecun Y, Bengio Y, Hington G. Deep learning[J]. Nature, 2015，521(7553)：436-444.