

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2014.028

多策略中文微博细粒度情绪分析研究

欧阳纯萍[†] 阳小华 雷龙艳 徐强 余颖 刘志明

南华大学计算机科学与技术学院, 衡阳 421001; [†] E-mail: ouyangcp@126.com

摘要 针对中文微博的用户情绪分析问题, 提出一种基于多策略融合的细粒度情绪分析方法。首先采用朴素贝叶斯算法对微博的有无情绪分类问题进行研究, 然后构建有情绪微博的 21 维特征向量, 最后采用 SVM 和 KNN 算法对微博进行细粒度情绪分析。以新浪微博作为实验对象, 结果表明多策略集成方法要好于单一分类算法。而在多策略集成方法中, “NB+SVM” 方法又略优于 “NB+KNN” 方法。

关键词 细粒度情绪分析; 中文微博; 朴素贝叶斯; SVM; KNN

中图分类号 TP391

Multi-strategy Approach for Fine-grained Sentiment Analysis of Chinese Micro-blog

OUYANG Chunping[†], YANG Xiaohua, LEI Longyan, XU Qiang, YU Ying, LIU Zhiming

School of Computer Science and Technology, University of South China, Hengyang 421001; [†] E-mail: ouyangcp@126.com

Abstract Fine-grained sentiment analysis of Chinese micro-blog is investigated and a method of multi-strategy fusion is proposed. Firstly, the authors apply Naive Bayesian to identify sentiment or non-sentiment about micro-blog. Secondly, based on emotion ontology, a method for how to form 21 sentiment features vectors of micro-blog is presented. At last, fine-grained sentiment of micro-blog is classified based on SVM and KNN respectively. Experiment results show that multi-strategy fusion is better than simplex, in addition, “NB+SVM” strategy is better than “NB+KNN” strategy.

Key words fine-grained sentiment analysis; Chinese micro-blog; naive Bayesian; support vector machine (SVM); K Nearest Neighbor (KNN)

微博是 Web2.0 时代人们分享与传播信息的重要平台, 使人们能够更加方便地评论热点事件和抒发自己的情绪。因此, 国内外学者围绕微博展开了一系列研究, 其中微博情绪分析是研究热点之一^[1]。国外学者采用了基于 SVM 的距离监督学习、基于语义的关联分析、基于 KNN 的语料强化学习、基于情感词的语义标注等方法对 Twitter 进行了情感分类研究^[2-5]。由于中文微博比英文微博具有更丰富的语义信息, 上述方法无法直接应用到中文微博情绪分析中^[6]。所以国内学者们从中文微博的特点出发, 致力于中文微博的情感分类和情绪分析研究。谢丽星等^[7]提出一种基于 SVM 的层次结构多策略

中文微博情感分类方法, 引入主题相关的特征后, 基于层次结构的多策略方法在正负情感分类上准确率达到 67.283%。刘志明等^[8]分别使用 3 种机器学习算法、3 种特征选取算法以及 3 种特征项权重计算方法对微博正负情感分类进行了研究, 实验证明采用 SVM、信息增益和 IF-IDF 三者结合的方法对微博情感分类效果较好。韩忠明等^[9]以 HowNet 情感词典为基础, 构建了一个计算短文本情感倾向性的自动机, 在短文本的正负情感分类上较 SVM 方法有非常大的优势。文献[10-12]均采用朴素贝叶斯分类器(NB)对微博的正负情感和主客观分类问题进行研究。

湖南省自然科学基金项目(13JJ4076, 11JJ6047)和湖南省教育厅优秀青年项目(13B101)资助
收稿日期: 2013-07-09; 修回日期: 2013-10-11; 网络出版时间: 2013-11-08 11:27

从上述研究可以发现,粗粒度的中文微博情感分类是目前学者们普遍关注的热点。所谓的粗粒度情感分类是指从 positive(积极)和 negative(消极)两类对中文微博的情感进行分析。但在实际应用中,微博的细粒度情绪往往更能展示用户对于事物的态度。细粒度情绪实际上是对粗粒度情绪的进一步分类,把“积极”和“消极”两类再划分为“欢乐”、“高兴”、“哀伤”、“厌恶”等更贴近人们真实情绪的分类。例如“话说大约十年前,十块钱炸肉岗尖一盘,加一瓶啤酒,就是我一顿饱饱的午餐。现在呢,图中这寥寥十数根,就是十块钱的肉,算开胃菜么?通胀的猪肉伤不起啊”这条微博很难判别是属于积极还是消极的情感,原因是它包含的情绪太多。其中,“话说大约十年前,十块钱炸肉岗尖一盘,加一瓶啤酒,就是我一顿饱饱的午餐。”表达了作者的欢喜情绪。“现在呢,图中这寥寥十数根,就是十块钱的肉,算开胃菜么?”表达了作者的惊讶情绪。“通胀的猪肉伤不起啊”则表达了作者悲伤的情绪。从这条微博的细粒度情绪中我们可以发现作者对于十年前的物价表现出积极正面的情感,而对于现在的物价则是展示出消极负面的情感。所以对微博的细粒度情绪进行分析,可以避免单纯从正负情绪两类对微博进行划分的笼统性,更细致的分类也能有效提高微博情绪识别的准确度。

本文主要针对中文微博的细粒度情绪分析进行研究,提出一种基于情感词汇本体的多策略集成分析法。首先采用朴素贝叶斯分类器对微博进行有无情绪分类,为微博的细粒度情绪分析提供预分类基础。然后利用情感词汇本体对有情绪微博进行特征选择及特征值权重计算,最后分别采用 SVM 和 KNN 分类器对其进行 7 类(anger 愤怒、disgust 厌恶、fear 恐惧、happiness 高兴、like 喜好、sadness 悲伤、surprise 惊讶)细粒度的情绪分析。研究表明,基于多策略融合的分析方法要好于单一分类算法。而在多策略融合分析方法中,“NB+SVM”方法又略优于“NB+KNN”方法。

1 中文微博的有无情绪分类

对中文微博的细粒度情绪分析是建立在此微博是一条含有情绪的文本基础上,例如“《变形金刚 3》港澳 6 月 29 日上映,大陆 REAL.D.3D 及 IMAX 3D 上映时间尚不明确。”这条微博只是客观阐述了一个事实,并没有包含任何情绪,所以不需要进行细粒

度的情绪分析。因此,在进行微博的细粒度情绪分析之前,需要先对微博进行粗粒度划分,即把微博分为有情绪和无情绪两类。本文采用朴素贝叶斯算法进行有无情绪分类。

1.1 词特征和词性特征选择

特征选择是实现朴素贝叶斯分类器的前提,特征选择的好坏在很大程度上决定该算法的性能^[13]。微博的特征是以字词来表示,因此首先需对微博进行分词,然后从分词结果中进行特征选择,选择出那些能代表有情感和无情感类别的微博特征词。我们选用张华平博士的分词工具 NLPPIR 对中文微博进行分词,NLPPIR 相对于老版本新增了微博分词功能,但是对于一些常用在在微博中表示情绪的表情词和缩略词并未收录在词典里,所以在分词之前,需将这些词汇加入自定义词典中。通过对 NLP&CC2013 评测组提供的 4000 条新浪微博进行分析,共提取了如 NND,MLGB,TMD,[大笑],[泪],[可怜]等 365 个词汇构建了自定义情绪词库,同时也把这个词库作为了分词工具 NLPPIR 的自定义词典。接着我们针对微博的分词结果,从词和词性两个方面考虑中文微博的特征选择。

1) 词特征选择:分别统计每个词在 4000 条训练集的有无情绪两类中出现的频率,然后依据式(1)筛选出特征词:

$$\frac{Wordfreq_{1,i}}{Wordfreq_i} > k \& \& Wordfreq_{1,i} > n, i = 0,1, \quad (1)$$

$i=0$ 时,表示词在有情绪分类中的频率与在无情绪分类中的频率比值大于 k ,而且词在有情绪分类中出现的频率同时大于 n ; $i=1$ 时,表示词在无情绪分类中的频率与在有情绪分类中的频率比值大于 k ,而且词在无情绪分类中出现的频率同时大于 n 。经过试验测试得到 k 和 n 的初值分别为 4 和 20 时,按这种方式筛选出的词在有情绪微博中出现的频率有明显差别,因此它们对有无情绪的文本分类具有表征作用,可以作为微博文本的特征。

2) 词性特征选择:考察剩余词的词性特征,发现某些词性在主观表达和客观表达中出现的频率也有很大差别的。统计每个一类词性(NLPPIR 分词器定义的,如[t], [n]等)和个别表征人称作用强的二类词性(如[rr], [ne]等)在有情绪和无情绪两种情况下出现的频率,特别地,将所有的表情符也作为一类词性进行有无情绪词频统计。

最终经过上述两个步骤的选择,我们得到 796

个特征词和 19 种词性, 组成 815 个供朴素贝叶斯分类器使用的特征集合, 详细的特征选择算法描述如下。

```

    输入: D, 已分词语料;  $k=4, n=20$ ;
    输出: 特征集合;
    FeatureSet={ } //特征初始集合为空;
    在 D 中计算出每个词在有情绪两类中出现的频率集合 WordFreq={ } //集合中的每个元素都是一个三元组 <word(词), freq(频率), pos(词性)>;
    RemainWordFreq={ } //经过词特征选择后剩余的词频集合, 初始为空;
    for 每个词 in WordFreq { // 词特征选择;
        if(词在有情绪中的频率/词在无情绪中的频率> $k$  and 词在有情绪中的频率> $n$ )or (词在无情绪中的频率/词在有情绪中的频率> $k$  and 词在无情绪中的频率> $n$ );
        FeatureSet = FeatureSet  $\cup$  WordFreq;
    }
    Else;
    RemainWordFreq = RemainWordFreq  $\cup$  WordFreq;
    }
    在 RemainWordFreq 中计算出每个词性在有情绪和无情绪两类中出现的频率集合 PosFreq={ } ; //集合中的每个元素都是一个二元组<pos(词性), freq(频率)>, 初始每个词性频率都为 0;
    for WordFreq in RemainWordFreq { //词性特征的选择;
        if (当前的 wordfreq.pos == pos) // pos 为在 RemainWordFreq 集合中所有出现词性的其中之一;
        posfreqi.freq= posfreqi.freq+ wordfreq.freq //计算每种词性在有情绪和无情绪两类中出现的频率;
    }
    FeatureSet = FeatureSet  $\cup$  PosFreq ;
    return FeatureSet。
    
```

1.2 基于朴素贝叶斯的有无情绪分类

朴素贝叶斯分类算法是基于贝叶斯定理的一种基于假设的先验概率的学习算法。它的基本思想是: 已知类别集合, 求解在待分类项出现的条件下集合中各个类别出现的概率, 哪个值最大, 就认为待分类项属于哪个类别。其中, 各类别的条件概率计算如下所示:

$$\begin{aligned}
 P(b_i|A) &= P(A|b_i)P(b_i) / P(A) \\
 &= P(a_1|b_i)P(a_2|b_i)\dots P(a_m|b_i)P(b_i) \quad (2) \\
 &= P(b_i) \prod_{j=1}^m P(a_j|b_i),
 \end{aligned}$$

$P(b_i)$ 是类别 b_i 的先验概率, $P(a_j|b_i)$ 是特征属性 a_j 在类别 b_i 中的后验概率。由于需要分类的类别只有有情绪和无情绪两类, 所以 b_0 表示无情绪类, b_1 表

示有情绪类。则定义先验概率和后验概率的计算方法如下:

$$P(b_i) = \frac{\sum(a|b_i)}{A}, i = 0, 1, \quad (3)$$

$$P(a_j|b_i) = \frac{\sum F(a_j)|b_i + 1}{\sum(a|b_i) + \sum C}, \quad (4)$$

其中, $\sum a|b_i$ 表示类 b_i 下特征词的总数; A 表示训练集中所有特征词的数目; $\sum F(a_j)|b_i$ 表示类 b_i 下特征词 a_j 出现的次数; $\sum C$ 表示不重复的特征词总数。

本文采用多项式贝叶斯分类模型对微博进行有无情绪分类, 具体步骤如下:

- 1) 对训练集中的每条微博分词后, 构建其特征集合;
- 2) 计算有无情绪两个类别的微博在训练集中的出现概率 $P(b_i)$ ($i=0$ 表示无情感, $i=1$ 表示有情感);
- 3) 计算特征集合中的每个特征属性分别对有无情绪两个类别的条件概率 $P(a_j|b_i)$ (a_j 表示第 j 个特征词);
- 4) 输出分类器;
- 5) 构建待分类微博的特征集合;
- 6) 分别计算每一个特征属性在有情绪和无情绪两个类别下的概率乘积:

$$P1 = P(\text{有情绪}) \prod_{j=1}^m P(a_j|\text{有情绪}),$$

$$P2 = P(\text{无情感}) \prod_{j=1}^m P(a_j|\text{无情感});$$

- 7) 比较 $P1$ 和 $P2$ 的值, 值较大者即为待分类微博所属的情绪类别。

2 微博细粒度情绪分析

在已得到的有无情绪分类基础上, 需要对有情绪的微博再进行细粒度情绪分析。

2.1 细粒度情绪特征选择及权重计算

情感词是微博情绪分析中较为重要的表征, 本文所使用的情感词主要来自于大连理工情感词本体, 共收录 27476 个词, 这些情感词被分为 7 大类 21 小类。具体的 7 类为乐、好、怒、哀、惧、恶、惊; 21 类为快乐(PA)、安心(PE)、尊敬(PD)、赞扬(PH)、相信(PG)、喜爱(PB)、祝愿(PK)、愤怒(NA)、悲伤(NB)、失望(NJ)、疚(NH)、思(PF)、慌(NI)、恐惧(NC)、羞(NG)、烦闷(NE)、憎恶(ND)、贬责(NN)、

妒忌(NK)、怀疑(NL)、惊奇(PC)^[14]。其中7大类情绪作为微博的细粒度情绪最终输出结果,而21小类情绪则作为微博特征向量中的特征属性。不过由于中文微博存在大量变形词、新词和潜在情感词,仅依靠情感词汇本体库远不能覆盖微博中表达情绪的所有词汇。因此我们对4000条测试集微博进行卡方检验,以获得对各类细粒度情绪有表征作用的词,并给这些词人工赋予情感强度、极性属性,从而扩展情感词汇本体库。

由于微博文本较短,特征关键词少,如果直接根据情感本体中的情感词,将每条微博表示为由情感词构成的特征向量,则将会得到一个非常稀疏的微博语料矩阵。在计算时,不仅会增加程序计算的时间复杂度,而且会影响最后的分类性能。因此,将情感本体库中确定的21类细粒度情感分类作为每条微博的最终特征,即每条微博将表示为21维的情感特征向量,实现对语料的特征矩阵降维的目的。

特征权重被用于衡量特征项在文本表示中的重要程度或者区分能力的强弱。将微博表示为21维的情感特征向量之后,每一维的权重将按式(5)计算:

$$\text{Weight}_m = \sum_i^w \alpha \times \text{intension}_{(i)} + \beta \times \text{polarity}_{(i)}, \quad (5)$$

$$(m = 1, 2, \dots, 21, \alpha, \beta \in [0, 1], \alpha + \beta = 1)$$

它表示第 m 维的权重等于属于这一维的所有情感词的强度和极性加权之和的累加。式中 m 表示维度标识,取值为1到21之间的整数; w 表示属于第 m 维情感类的总词数; intension 表示情感强度; polarity 表示情感极性; α, β 是取值在0到1之间的可调参数,设定为 $\alpha=0.9, \beta=0.1$,最终形成一个包含所有微博特征的21维向量矩阵。

2.2 基于SVM的细粒度情绪分析

SVM(支持向量机)是一种监督式的新颖的小样本机器学习算法。它避开传统的归纳到演绎的过程,实现高效的从训练样本到预测样本的“转导推理”。本文使用台湾大学林智仁教授等开发的 Libsvm^[15]用于SVM分类模型的训练及测试。由于最终要输出的是7类细粒度情绪中的一类,因此属于基于SVM的多标签分类问题。在处理SVM多类问题时,往往将其分解成一系列的二分类问题加以解决。常见的处理方法包括一对多法(1-v-r)、一对一法(1-v-1)以及有向无环图法(DAGSVM)。我们采用1-v-1方法构建SVM多类分类器,此方法在计算效率和分类效果上有着更好的性能。

1-v-1方法分别选取2个不同类别构成一个SVM子分类器,共有 $k(k-1)/2$ 个子分类器,本文为7分类问题,所以最终构造了21个SVM子分类器。选取训练集4000条微博中的2172条标注为有情绪的博客作为SVM分类模型的训练集,训练参数采用 Libsvm提供的默认参数,核函数选择径向基函数高斯核,以文本文件形式输入训练集微博的21维特征向量。训练得到一个七分类的SVM分类器后,再对50000条测试集(NLP&CC2013评测组提供)中经朴素贝叶斯分类为有情绪的30393条微博进行细粒度情绪分析。

2.3 基于KNN的细粒度情绪分析

KNN(K Nearest Neighbors)算法又叫K最近邻算法。它的基本思想是:假设样本空间里每个样本都有唯一的类标记表示这些样本是属于某一个分类。在给定待分类项后,计算样本空间中与待分类项距离最近的 K 个样本,根据这 K 个样本中哪个类别的样本最多,则判定待分类项就属于该类别。KNN算法需要人为确定 K 的取值。如果 K 值选择过小,则得到的近邻数较少,会降低分类精度,同时也会放大噪声数据的干扰。而如果 K 值选择过大,则有可能把距离最近的,但实际上并不相似的文本包含进来,造成噪声增加而导致分类效果的降低。关于 K 值的选择,到目前为止还没有一种非常行之有效的方法,关于此方面的论文也较少,所以大部分研究仍然采用的是重复实验,不断调整的方法来确定 K 值。本文通过采用不同的 K 值来计算训练集中微博分类的准确度,发现 K 值为21时分类的准确度较高。

将测试集中的30393条有情绪微博和训练集中的2172条有情绪微博均表示为21维的特征向量,然后执行KNN算法进行细粒度情绪分析。算法过程描述如下。

输入:待分类微博的特征向量, K 值为21;

输出:7类情感中的一类;

求向量相似度:用余弦公式 $\text{CosSim}(\text{测试集}, \text{训练集})$,计算待分类微博与训练集中的每一条微博的向量相似度;

对计算结果,按照相似度值进行降序排序;

取前21个相似度计算结果;

统计对应的21条训练集微博所属的情感类别,把属于相同类别的向量相似度值进行累加;

累加和最大的相似度值所对应的情感类别即为最终待分类微博所属的情感类别;

返回 7 类细粒度情感中的一种。

3 实验结果与分析

我们参加了 2013 年 CCF 自然语言处理与中文计算会议的中文微博情绪分析评测任务, 实验采用的训练集和测试集来自 NLP&CC2013 评测组提供的新浪微博数据集, 包括 4000 条已标注的微博和 50000 条待分析的微博, 其中 4000 条已标注微博中有情绪微博为 2172 条, 无情绪微博为 1828 条。除了实现“NB+SVM”和“NB+KNN”两种多策略方法以外, 本文还对使用单一分类算法 SVM 或 KNN 进行比较试验, 即使用一种分类算法一次性将无情绪和 7 类细粒度情绪微博分析。实验使用的训练和测试数据情况如表 1 所示。

微博情绪判别及分类评测的目标: 对于输入的整条微博, 首先判断出该微博是否包含情绪, 其次对包含情绪的微博, 判别其情绪分类为 anger 愤怒、disgust 厌恶、fear 恐惧、happiness 高兴、like 喜好、sadness 悲伤、surprise 惊讶 7 类中的一种。情绪判别任务使用正确率(Precision)、召回率(Recall)和 F 值(F -measure)来评价, 微博情绪分类任务中, 使用宏平均(Macro)和微平均(Micro)的准确率(Precision)、召回率(Recall)、 F 值(F -measure)作为评价指标的评价标准。根据上述目标和评价标准, 共有 19 支队伍提交了情绪判别和分类任务的评测结果, 我们提出的“NB+KNN”方法在情绪判别任务中 F 值为 0.7271, 在情绪分类任务中 F 值为 0.3412, 均居 19 支参赛队伍之首。表 2, 3 和 4 为分别采用 4 种分类方法完成情绪判别和分类任务的实验结果(MAX 为评测任务中所有提交结果中的最高值)。

表 1 实验数据简表

Table 1 Short form of experimental data

分类算法	训练集	测试集
SVM	3500 条已标注的微博	500 条已标注的微博
KNN	3500 条已标注的微博	500 条已标注的微博
NB+SVM	NB: 4000 条已标注的微博	NB: 50000 条未标注微博
NB+KNN	SVM 和 KNN: 2172 条已标注为有情绪的微博	SVM 和 KNN: 50000 条中使用 NB 算法分出的 30393 条有情绪的微博

表 2 情绪判别实验结果

Table 2 Experimental result of sentiment identification

	Precision	Recall	F -measure
SVM	0.6212	0.7274	0.6701
KNN	0.6495	0.7220	0.6838
NB+SVM	0.6626	0.8012	0.7254
NB+KNN	0.6676	0.7982	0.7271
MAX	0.7494	0.9517	0.7271

表 3 情绪分类宏平均评价结果

Table 3 Macro average of sentiment classification

	Macro Precision	Macro Recall	Macro F -measure
SVM	0.1922	0.1805	0.1862
KNN	0.2088	0.2308	0.2192
NB+SVM	0.2109	0.1996	0.2051
NB+KNN	0.2704	0.3064	0.2873
MAX	0.2844	0.3064	0.2873

表 4 情绪分类微平均评价结果

Table 4 Micro average of sentiment classification

	Micro Precision	Micro Recall	Micro F -measure
SVM	0.2283	0.2509	0.2391
KNN	0.2381	0.2773	0.2562
NB+SVM	0.2539	0.3070	0.2779
NB+KNN	0.3133	0.3746	0.3412
MAX	0.3834	0.3976	0.3412

从整体的实验结果来看, 中文微博的有无情绪判别效果比细粒度情绪分类好。在单项任务上, “NB+SVM”和“NB+KNN”在有无情绪判别中比单一的 SVM 和 KNN 分类算法效果好将近 5%, 而在细粒度情绪分类中, “NB+SVM”, “NB+KNN”比单一算法的 F 值高将近 11%, 说明多策略的方法更适合细粒度的情绪分析。从整个评测结果来比较, “NB+KNN”在有无情绪判别中的准确率略低于最高值, 召回率低了将近 20%, 但是 F 值为整个评测结果中的最高值。在细粒度情绪分类中, “NB+KNN”的宏平均准确率、微平均准确率和微平均召回率略低于最高值, 而宏平均的召回率、 F 值, 以及微平均的 F 值均在整个评测结果中位居首位。说明“NB+KNN”在情绪的判别和分类上略优于“NB+SVM”方法, 虽然“NB+KNN”在准确率和召回率上略逊色于其它方法, 但是就 F 值而言, “NB+KNN”在情绪的判别和分类上都取得了较好的效果。虽然我们没有对算法进行改进, 但是我们分阶段使用了不同的分类算法, 从实验结果可见这种多策略组合方案优于直接进行微博细粒度情绪分析的方法。

4 结论

本文把朴素贝叶斯(NB)算法分别与支持向量机(SVM)算法和 K 最近邻(KNN)算法进行了融合,用于处理中文微博的情绪判别及分类问题。其中“NB+KNN”方法在情绪判别及分类中均取得了较好的效果。通过实验发现: 1) 在进行细粒度情绪分析前先完成粗粒度的情绪分类,能较好地提高分析效果; 2) 同时基于词和词性的特征选择,可以提高朴素贝叶斯算法在粗粒度情绪分类中的分类性能; 3) 在进行有情绪微博的特征选择时使用卡方检验和情感词汇本体,对后续使用 KNN 算法的分类效果也有所提升。

我们所提出的多策略方法在准确率和召回率上还有很大的改进空间,下一步工作主要从以下几方面入手: 1) 尝试从语义依存角度去扩展现有的情感词典; 2) 对微博的特征选择方法进行比较和改进; 3) 考虑在多策略方法中引入集成学习、深度学习,在分类算法的创新性上做进一步研究。

参考文献

- [1] Ramage D, Dumais S, Liebling D. Characterizing microblogs with topic models // Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Washington, DC: The AAAI Press, 2010: 130-137
- [2] Jansen B J, Zhang Mimi. Twitter power: tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 2009: 2169-2188
- [3] Barbosa L, Feng J. Robust sentiment detection on twitter from biased and noisy data//Proceedings of the 23rd International Conference on Computational Linguistics. Beijing: Tsinghua University Press, 2010: 36-44
- [4] Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using Twitter Hashtags and smileys // Proceedings of the 23rd International Conference on Computational Linguistics. Beijing: Tsinghua University Press, 2010: 241-249
- [5] Si Jianfeng, Mukherjee A, Liu B, et al. Exploiting topic based Twitter sentiment for stock // Proceedings of the 50st Annual Meeting of the Association for Computational Linguistics. Jeju Island, Korea, 2013: 115-120
- [6] 周胜臣, 瞿文婷, 石英子, 等. 中文微博情感分析研究综述. 计算机应用与软件, 2013, 30(3): 161-164,181
- [7] 谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取. 中文信息学报, 2012, 26(1): 73-83
- [8] 刘志明, 刘鲁. 基于机器学习的中文微博情感分类实证研究. 计算机工程与应用, 2012,48(1): 1-4
- [9] 韩忠明, 张玉沙, 张慧, 等. 有效的中文微博短文本倾向性分类算法. 计算机应用与软件, 2012, 29(10): 89-93
- [10] 林江豪, 阳爱民, 周咏梅, 等. 一种基于朴素贝叶斯的微博情感分类. 计算机工程与科学, 2012, 34(9): 160-165
- [11] 杨武, 宋静静, 唐继强. 中文微博情感分析中主客观句分类方法. 重庆理工大学学报: 自然科学版, 2013, 27(1): 51-56
- [12] 张珊, 于留宝, 胡长军. 基于表情图片与情感词的中文微博情感分析. 计算机科学, 2012, 39(11A): 146-148, 176
- [13] 周茜, 赵明生, 扈旻. 中文文本分类中的特征选择研究. 中文信息学报, 2004,18(3): 17-23
- [14] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造. 情报学报, 2008, 27(2): 180-185
- [15] Chang C C, Lin C J. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27