计算机科学
COMPUTER SCIENCE

基于不平衡数据与集成学习的属性级情感分类

林夕，陈孜卓，王中卿

引用本文

林夕，陈孜卓，王中卿. 基于不平衡数据与集成学习的属性级情感分类[J]. 计算机科学, 2022, 49(6A): 144-149.

LIN Xi, CHEN Zi-zhuo, WANG Zhong-qing. Aspect-level Sentiment Classification Based on Imbalanced Data and Ensemble Learning[J]. Computer Science, 2022, 49(6A): 144-149.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210500205

基于不平衡数据与集成学习的属性级情感分类

林 夕 陈孜卓 王中卿
苏州大学计算机科学与技术学院 江苏 苏州 215006
(linxi350904583@foxmail.com)

摘 要 情感分类一直是自然语言处理领域的重要研究部分。该任务一般是将带有情感色彩的样本分类成正类和负类两种类别。在很多理论模型中,都假设正负类数据样本是平衡的,而在现实中正负类样本一般是不平衡的。提出一种基于属性级的LSTM集成学习的方法,针对不平衡样本数据进行属性级情感分类。首先,对数据集进行欠采样处理,将其分成多组;其次,为每组数据分配一种分类算法进行训练;最后,将多组模型融合,得到最终分类结果。一系列的实验结果显示,基于属性级的LSTM集成学习的方法明显提高了分类的准确性,其性能优于传统的LSTM模型分类方法。

关键词:不平衡数据;LSTM;集成学习;情感分类;属性词
中图法分类号 TP391

Aspect-level Sentiment Classification Based on Imbalanced Data and Ensemble Learning

LIN Xi,CHEN Zi-zhuo and WANG Zhong-qing
School of Computer Science and Technology,Soochow University,Suzhou,Jiangsu 215006,China

Abstract Sentiment classification remains an important part of the field of natural language processing. The general task is to classify the emotional data into two categories,which is positive and negative. In many models,it is assumed that the positive and negative data are balanced. Contrarily,the two class of data are always imbalanced in reality. This paper proposes an ensemble learning model based on aspect-levelLSTM to process aspect-level problem. Firstly,the data sets are under-sampled and divided into multiple groups. Secondly,a classification algorithm is assigned to each group of data for training. Finally,it yields the classification result through joining all models. The experimental results show that the ensemble learning model based on aspect-level LSTM significantly improves the accuracy of classification,and its performance is better than the traditional LSTM model.

Keywords Imbalanced data,LSTM,Ensemble learning,Sentiment classification,Aspect word

1 引言

情感分类是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程[1]。随着互联网技术的普及,越来越多的网络用户会在网络平台上发表带有主观情感的评论。对这些评论信息进行管理分析,对用户情感以及需求的分类成为网络平台重点关注的技术之一。因此,情感分类在自然语言处理中占有重要地位。

近些年来,情感分类是国内外研究的热点之一,主要分析文本所表达的含义和情感信息并且将文本划分成正面或反面倾向两种或几种类型。情感分类是对文本作者倾向性观点、态度的划分,所以又被称为观点分析、倾向性分析等。一般的情感分类理论研究中,普遍都会假设正类样本和负类样本是平衡的,也就是两者数量相同。但是在实际收集到的数据中,正负类样本往往是不平衡的,这就产生了进一步的研究。如果使用传统的机器学习模型对不平衡样本进行情感分类,会使最终的分类结果倾向于样本数量大的类别,导致模型失真,分类性能下降[2]。因此,对不平衡样本情感分类进行研究是很有必要的。

不平衡分类问题是机器学习领域中经常存在的问题。一般提升不平衡分类问题准确率的方法有三大种:重采样技术、特征选择和代价敏感学习。本文采用最广泛使用的欠采样方法来处理不平衡分类问题,达到提高分类器性能的目的。不平衡样本中存在一种类别数据远大于另一种类别数据的现象,而欠采样是对大类样本进行采样减少数据样本的个数,使其与小类样本数据数量接近,然后进行学习。欠采样因为随机丢弃了一些样本,所以会丢失重要的信息。因此欠采样之后一般采用集成学习的方法,将大类样本划分为若干组分别和小类样本组合以供不同学习器进行机器学习。这样既达到了对于每个学习器都是样本平衡的,又使得全局中没有丢失信息。

本文采用基于长短记忆网络(LongShort-Term Memory,LSTM)[3]集成学习并结合属性词的方法进行不平衡样本的情感分类。我们对淘宝商品用户评论信息进行采集,发现情感倾向不平衡的现象。因此我们先欠采样评论,将其分成多组,再每组分别基于LSTM,TD-LSTM[4](Target-Dependent LSTM)以及拓展的神经网络模型循环进行学习,然后使用Bagging方法对各组训练好的模型进行集成,最后预测给定数据集得到本文模型的准确率。本文采用的方法结果与传统分类算法分类结果对比显示,集成后的分类器性能得到提高,效果更优,较好地解决了不平衡数据上的属性级情感分类问题。

本文第2节介绍情感分类、不平衡分类及LSTM神经网络的相关工作;第3节介绍基于属性级LSTM集成学习不平衡

通信作者:王中卿(wangzq@suda.edu.cn)

情感分类方法；第 4 节进行实验并对结果进行分析；最后总结全文并展望未来。

2　相关工作

2.1　情感分类

情感分类是指根据文本所表达的含义和情感信息将文本划分成褒扬的或贬义的两种或几种类型，是对文本作者倾向性和观点、态度的划分，因此有时也称倾向性分析（Opinion Analysis）。在情感分类中研究最多的就是划分正负两类，因此本文只研究二分类问题。纵观目前主观性文本情感倾向性分析的研究工作，主要研究思路分为基于语义的情感词典方法和基于机器学习的方法，而基于有监督的机器学习方法是目前的主流方法。

情感分类是按照文本持有的情感、态度进行判断分类，因此现有机器学习的分类方法基本上都可以运用到情感分类中。基于机器学习的情感分类大致流程一般如下：首先人工标注文本倾向性作为训练集，然后提取文本情感特征，通过机器学习的方法构造分类器，最后用分类器对样本进行极性分类。常用的分类方法有贝叶斯分类器[5]、支持向量机[6]、最大熵分类器[7]等。

最早从事情感分析研究的 Pang 等[8]使用词袋（Bag-of-Feature）框架选定文本的 N 元语法（N-Gram）和词性（POS）等作为情感特征，分别使用朴素贝叶斯、最大熵模型和支持向量机的方法将电影评论分类。在 Pang 等的研究基础上，后续研究主要是把情感分类作为一个特征优化任务[9]。

近年来，深度学习在自然语言处理领域取得广泛运用。情感分类渐渐被划分为粗粒度的句子级情感分类和细粒度的属性级情感分类。Hochreiter[3]于 1997 年提出 LSTM，之后在句子级情感分析领域取得进展[10-11]。句子级别情感分类的准确率一般难以达到普通文本分类的水平，主要是情感文本中复杂的情感表达和大量的情感歧义造成的。因此，更加细粒度的属性级别情感分类总体性能表现得更为出色，成为目前的研究热门。Wang 等[12]将 LSTM 和注意力机制结合，相较于句子级的 LSTM 模型分类效果有所提升。Tang 等[4]提出 TD-LSTM 模型，创新点是对属性词与其上下文分别建模整合。Wu 等[13]通过利用 BERT 模型，在注意力模型中添加上下文信息来提升分类性能。Jiang 等[14]利用 Bi-LSTM 获得句子与属性词的语义依赖性，构造出一种 MAN（Mutual Attention Neural networks）模型，进一步深化了属性级情感分类研究。我们将以上属性级模型选择并利用到本文的方法中。

随着训练语料库的发展以及分类算法的进步，情感分类任务的准确度得到提高，基于机器学习、深度学习的分类将会有广阔的发展前景。

2.2　不平衡分类

现有的情感分类研究基本上是基于平衡分布的样本，不平衡分类问题在已有的研究中还是很稀少。如果数据存在严重的不平衡，预测得出的结论往往有偏差，即分类结果会偏向于较多观测的类。因此，不平衡分类问题非常值得研究。

处理不平衡分类的方法一般从 3 个角度入手，即数据层面、特征层面和算法层面。其中主流的方法是重采样技术、特征选择和代价敏感学习。

特征选择是利用特征层面的信息，从特征集合中选取合适的特征子集，来增加不平衡样本中多类和少类的区分度，有效处理不平衡分类问题。而代价敏感学习关注错误代价较高类别的样本，以分类错误总代价最低为诊断算法的优化目标。Wang 等[15]将特征选择方法引入到不平衡情感分类的研究中，在提升分类效果的同时降低了特征向量的维度。

重采样方法是运用最为广泛的，而在采样中又分为过采样和欠采样。过采样基本思想是增加少数样本使数据平衡，最简单的方法是随机复制少类样本。欠采样与过采样相反，通过减少多类样本使样本平衡。将重采样技术应用到不平衡分类中，可以取得优异的效果。采样技术在近几年不断被改进，取得突破。Ye 等[16]提出的基于聚类融合欠采样的改进欠采样方法，有效提高了模型中的样本质量，增强了欠采样算法的抗噪声能力。Lin 等[17]将基于聚类技术的欠采样方法引入到情感分类问题中，取得了很好的效果。

3　基于 LSTM 集成学习的不平衡情感分类

3.1　欠采样集成学习

集成学习是通过组合多个分类器进行学习的模式机制，能够吸收每个分类器的优点，提高分类性能。欠采样方法能避免样本不平衡带来的分类结果的偏移[18]。两者结合，能大大提高分类性。欠采样集成学习步骤一般为：1）多类欠采样成多组，和少类样本组成多个训练样本的输入；2）每个样本集合使用不同或相同的分类器模型训练；3）使用集成技术集成结果输出。集成技术有 Bagging，Boosting 和 Stacking 等方法。本文采用 Bagging 方法集成，算法步骤如下。

（1）从原始样本集中抽取训练集。每轮从原始样本集中使用欠采样方法抽取 n 个训练样本。共进行 k 轮抽取，得到 k 个训练集。

（2）每次使用一个训练集得到一个模型，k 个训练集共得到 k 个模型。

（3）根据上步得到的 k 个模型，采用投票的方式得到分类结果。

我们提出一种基于欠采样和多分类算法的集成方法，具体框架如图 1 所示，图中的基模型都是基于 LSTM 神经网络的分类算法循环训练，直至训练出 k 个模型再进行情感分类以及 Bagging 集成。

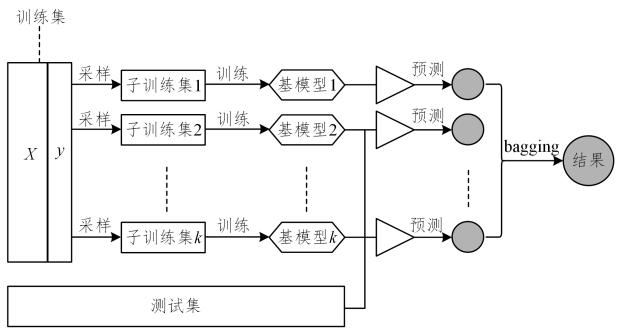

图 1　基于 LSTM 集成学习分类器的框架

Fig. 1　Framework of ensemble learning classifier based on LSTM

3.2　基于属性级 LSTM 集成学习算法

在本文的方法框架中，欠采样后的每组样本都要分配一种分类算法，因此我们提出了多种供样本分配的分类算法。一般情况下，分类发布方法差异较大时，集成后性能更好[19]。通过这些分类算法集成学习，分类性能比单个分类器性能

更高[18]。正如3.1节所提到，我们提出的分类算法都是基于LSTM神经网络，在LSTM的基础上扩展与属性词相关，这样能够使集成学习时分类方法多样化，提高分类准确率并且能够解决属性级情感分类问题。

具体分类算法如下。

(1)LSTM：LSTM是一种特殊RNN类型，能够学习长期依赖关系。Hochreiter和Schmidhuber[3]于1997年提出LSTM，之后AlexGraves[10]进行推广，将其应用于其他领域。

LSTM是情感分析领域中语义组合性能表现最为出色的神经网络模型[20]，能够有效避免梯度消失、梯度爆炸问题，且能够从具有多层次抽象的子表达式计算较长表达式的表示。句子特征表示可以作为预测句子情感极性的特征[4]。

如图2所示，在LSTM中每个词表示成向量，所有词向量都属于预训练好的词嵌入矩阵。LSTM能使当前词向量w_t与之前的输出h_{t-1}递归变换，将可变长度的词向量映射到固定长度的向量。

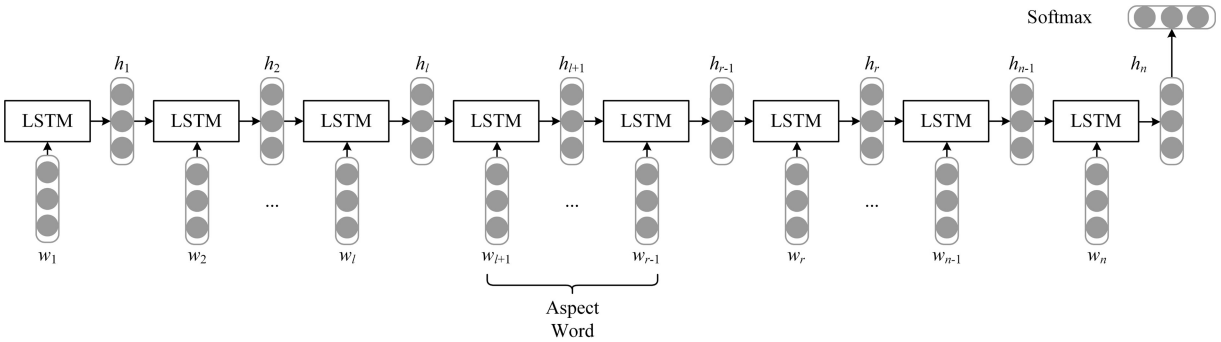

图2 LSTM分类器框架

Fig. 2 Framework of LSTM

传统RNN变换方程如下：

$$h_t=\tanh(W\cdot[h_{t-1};w_t]+b) \tag{1}$$

其中，$W\in R^{d\times 2d}$，$b\in R^d$。与RNN相比，LSTM单元增加了3个神经门来控制更新历史信息：输入门、遗忘门和输出门。LSTM单元计算如下：

$$i_t=\sigma(W_i\cdot[h_{t-1};w_t]+b_i) \tag{2}$$

$$f_t=\sigma(W_f\cdot[h_{t-1};w_t]+b_f) \tag{3}$$

$$o_t=\sigma(W_o\cdot[h_{t-1};w_t]+b_o) \tag{4}$$

$$g_t=\tanh(W_r\cdot[h_{t-1};w_t]+b_r) \tag{5}$$

$$c_t=i_t\odot g_t+f_t\odot c_{t-1} \tag{6}$$

$$h_t=o_t\odot\tanh(c_t) \tag{7}$$

其中，$\odot$代表向量间点乘；σ是sigmoid函数；W_i，b_i，W_f，b_f，W_o，b_o是输入门、遗忘门和输出门的权重矩阵。计算每一层的值，最后一个隐藏向量作为输出，再利用Softmax层分类为正负类。Softmax层函数如下：

$$Softmax_i=\frac{\exp(x_i)}{\sum_{i'=1}^{2}\exp(x_{i'})} \tag{8}$$

以上的LSTM能够计算整个句子的情感表示，但是并没有考虑属性词，无法解决属性级情感分类问题，也就是同一句话属性词不同但最后情感分类的特征表示相同。例如，“手机壳用料很不错，就是有些色差。”这句话中用料和色差两个属性词的情感极性明显是不同的，但在LSTM只能是相同的特征表示。因此我们做了些许改进，原本的模型中输入一整句话时，我们现在将属性词与整句话拼接作为输入。也就是说，输入由$[w_1;w_2;\cdots;w_n]$变为了$[w_1;w_2;\cdots;w_n;v_a]$，v_a代表属性词的词向量。这样，LSTM模型就能做到输出与属性词相关。

(2)TD-LSTM[4]：改进后的LSTM能进行情感分类，也考虑了属性词，但无法非常有效并且有说服力地解决属性级情感分类问题。因此，TD-LSTM被提出，它能够与属性词相关。其实现方法是根据属性词分别进行前后文建模，这样上下文信息都被考虑到，作为情感分类的特征表示，并且属性词不同，其特征表示也不同。图3为TD-LSTM分类器框架，模型中使用两个LSTM神经网络$LSTM_L$和$LSTM_R$，将一句话关于属性词分为两段，属性词左端文本及属性词本身作为输入从左至右运行$LSTM_L$，属性词右端及属性词本身作为输入从右至左运行$LSTM_R$。最后，利用Softmax层组合两段，进行情感极性分类。这样，属性词不同，两端输入大不相同，情感极性能够被细化得更为清楚。

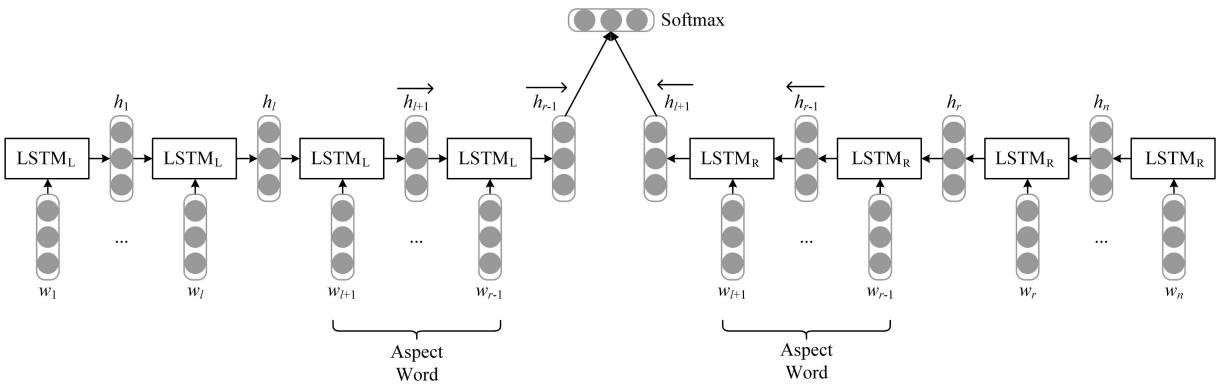

图3 TD-LSTM分类器框架

Fig. 3 Framework of TD-LSTM

(3)AT-LSTM[12]：扩展 LSTM 算法，在隐藏向量之上加入注意力机制[21]。AT-LSTM 基于注意力机制，捕捉句子关键部分。

图 4 为 AT-LSTM 分类器框架。

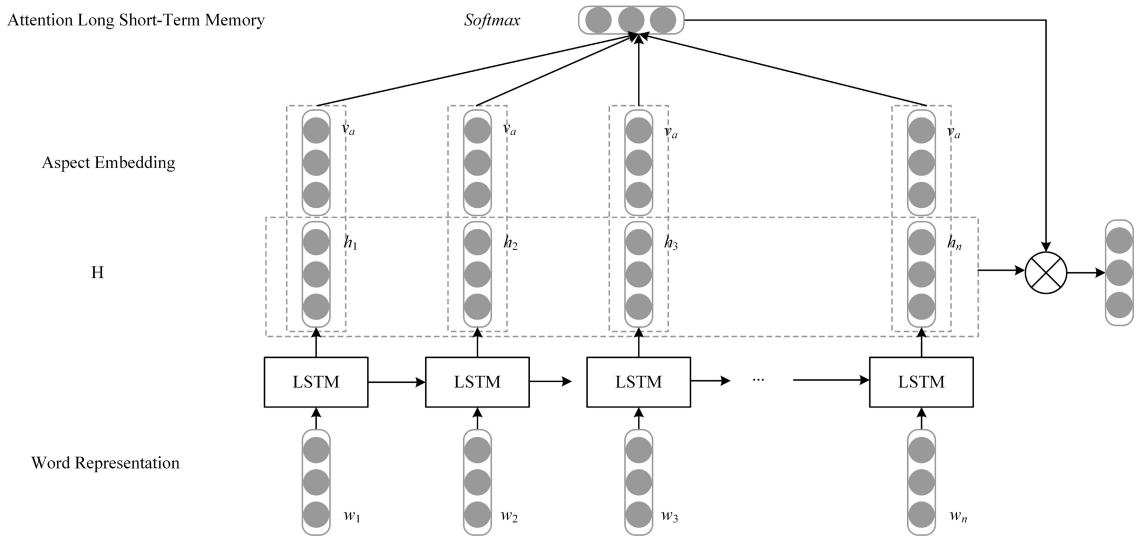

图 4　AT-LSTM 分类器框架

Fig. 4　Framework of AT-LSTM

设 LSTM 中每一层隐藏向量集合是 H，H 为$[h_1,\cdots,h_N]$($H\in R^{d\times N}$)，其中 d 是隐藏层的大小，N 是句子的长度。新设置v_a代表属性词的词向量，$e_N\in R^N$是全 1 列向量。注意力机制产生了新式：

$$M=\tanh\left(\begin{bmatrix}W_hH\\W_vv_a\otimes e_N\end{bmatrix}\right) \tag{9}$$

$$\alpha=Softmax(w^{\mathrm{T}}M) \tag{10}$$

$$r=H\alpha^{\mathrm{T}} \tag{11}$$

其中，$M\in R^{(d+d_a)\times N}$，$\alpha\in R^N$，$r\in R^d$，$W_h\in R^{d\times d}$，$W_v\in R^{d_a\times d_a}$，$w\in R^{d+d_a}$。$v_a\otimes e_N$中运算符是指将v_a重复 N 次后再拼接，结果为$[v_a;v_a;\cdots;v_a]$。通过以上式子，得出最终 AT-LSTM 中句子表示和 Softmax 分类函数：

$$h^*=\tanh(W_pr+W_xh_N) \tag{12}$$

$$y=\mathrm{Softmax}(W_sh^*+b_s) \tag{13}$$

其中，$h^*\in R^d$，W_p 和W_x 都是待训练的参数，W_s 和b_s 是 Softmax 层的参数。从 AT-LSTM 神经网络中看出，该算法考虑到属性词，因此模型分类效果比无属性词的 LSTM 分类效果更好。

(4)Change-Part：在 TD-LSTM 分类算法中稍作修改，将训练集中原句左端或右端改为另外一句的左端或右端，目的是使模型适用性更广，放大同一情感极性中的共性。具体算法过程是将测试集中正负类数据先分开，同一类样本中每一句话关于属性词左端文本与随机另一句话的右端文本匹配，再组合成输入，利用 TD-LSTM 进行训练。

在模型训练过程中，我们选择交叉熵作为损失函数：

$$J=-\frac{1}{m}\sum_{i=1}^{m}[a_i\ln y_i+(1-a_i)\ln(1-y_i)] \tag{14}$$

其中，m 是样本数量，a_i 是实际的标签，y_i 是预测的输出。考虑到本文中的情感分类都是二分类问题，a_i取 1 或 0。

综合以上 4 种分类算法，我们得到一种基于欠采样和多分类算法的集成方法，具体步骤如下：1)多类样本进行 n 次欠采样和少类样本组合成 n 组样本(n 为不平衡比，多类/少类)；2)每组样本集合循环分配分类算法训练；3)使用 Bagging 方法集成得到集成学习模型并输出结果：

$$y=\frac{1}{K}\sum_{i=1}^{K}y_i \tag{15}$$

其中，K 是分类器个数，y_i就是每个分类器预测的结果，y 为最终预测的结果。方法步骤中的基模型都是在上文的 4 种分类算法中循环选取，直至训练出 k 个模型再进行情感分类以及 Bagging 集成。

4　实验与分析

4.1　实验样本不平衡情况

本文利用淘宝 3 种商品类别的中文评论语料作为实验数据，对其进行统计并分析后，我们发现评论中普遍存在样本不平衡情况。我们在语料中标注了属性词和情感词及其情感倾向。例如，“电池寿命长”，属性词为“寿命”，“长”是情感词，而其情感极性就为正。语料中 3 种商品类别是：家具、百货和厨房用具，其相关评论情感极性分布情况如表 1 所列。

表 1　淘宝商品评论情感极性分布情况

Table 1　Emotional polarity distribution of Taobao reviews

类别	正类	负类	正类/负类
家具	34059	1677	20.31
百货	19663	2057	9.56
厨房	21310	2274	9.37

从表 1 中可以看出每类数据都是正类样本远大于负类样本，各类别不平衡比(正类/负类)都大于 9，甚至在家具类样本中不平衡比达到 20 以上。我们可以分析出在淘宝评论中情感极性分布不平衡是正常现象，而其实在大多数领域中主观性文本的情感极性分布是普遍不平衡的。

4.2　实验设置

我们使用淘宝 3 种商品类别的中文评论数据作为实验语料。语料中 3 种商品的评论都预先划分为两种情感类别——正类和负类，详细分布情况见表 1。

实验中我们将数据划分成两部分，其中 80% 为训练集，20% 为测试集。用到的分类算法有 LSTM，TD-LSTM 和

AT-LSTM。LSTM 神经网络基于 Keras 搭建[1)]。考虑到实验性能，训练时间不能过长，本实验中词向量维度设置需适量。

本文采用词嵌入方式作为文本的输入，词向量维度设置成 300。在集成学习时，正类样本划分成 10 组与负类样本组成 10 组样本(评论数据正类与负类的不平衡比都大约等于 10 或者大于 10)作为每组分类器的分类样本。

在上述的实验中我们均以正确率(Accuracy)作为评价模型的标准，这样对比比较直观。正面评论文本和负面评论文本的预测情况可具体分为 TP(实际正面，预测正面)，FP(实际正面，预测负面)，FN(实际负面，预测正面)，TN(实际负面，预测负面)4 种。Accuracy 的计算公式为：

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

4.3 实验结果与分析

本文提出的基于 LSTM 的集成学习中的分类算法都考虑到了属性词。其中使用的 LSTM 是 3.2 节中提出的改进方法，即输入是句子和属性词的结合，其余的 TD-LSTM，AT-LSTM 和 Change-Part 都是按上文所述进行训练。

利用如下方法比较分类效果，找出最合理的处理不平衡分类的方法。

(1)完全训练＋TD-LSTM：各类别训练集全部作为训练样本，分类算法为 TD-LSTM。

(2)欠采样＋TD-LSTM：各类别训练集中多类样本(正类)欠采样，选择部分数量以达到与少类样本(负类)平衡的效果，再使用 TD-LSTM 训练。

(3)过采样＋TD-LSTM：各类别训练集中少类样本(负类)过采样，复制样本数量以达到与多类样本(正类)平衡的效果，再使用 TD-LSTM 训练。

(4)欠采样＋LSTM：与(2)相同，区别是分类算法为 LSTM。

(5)欠采样＋Change-Part：欠采样后，在训练集中相同类里关于属性词随机匹配左右段样本，重新组合后形成新的训练集，再使用 TD-LSTM 训练。

(6)欠采样＋AT-LSTM：欠采样后，使用 LSTM 训练，加入注意力机制。

(7)欠采样＋单分类器集成学习：欠采样出 10 组样本，全部使用 TD-LSTM 算法训练，得到集成的模型。

(8)欠采样＋三分类器集成学习：欠采样出 10 组样本，循环使用 LSTM，TD-LSTM 和 AT-LSTM 算法训练，得到集成的模型。

(9)欠采样＋四分类器集成学习：欠采样出 10 组样本，循环使用 LSTM，TD-LSTM，Change-Part 和 AT-LSTM 算法训练，得到集成的模型。

第一组实验是讨论欠采样的有效性(利用方法(1)－方法(3))。分别通过欠采样、过采样和完全采样 3 种方法处理数据，再同样使用 TD-LSTM 算法进行情感分类，目的是对比 3 种方法下对同样数据进行分类的准确率。观察表 2 可以看出，欠采样方法的分类效果相较于其他两种有大幅度的优异性。欠采样相较于过采样有着平均 17.7%的提升，相较于完全采样有平均 12.8%的提升。其原因可能在于过采样和完全训练的样本偏向于多类，预测结果也偏向多类，导致准确率下降。

表 2 不同采样方式的分类结果

Table 2 Results of different sampling methods

类别	欠采样	过采样	完全采样
家具	0.84	0.69	0.72
百货	0.81	0.69	0.72
厨房	0.82	0.72	0.75

第二组实验是 LSTM，TD-LSTM，Change-Part 和 AT-LSTM 4 种分类算法效果的对比(利用方法(4)－方法(6))。对于样本先进行欠采样后，将样本数据运用到上文所提的 4 种算法中分别训练得到单个分类算法的效果，如表 3 所列。可以看出，TD-LSTM 在单分类器中分类性能表现最好，相较于 LSTM 性能提升大约 3%，相较于 AT-LSTM 提升大约 5%，与 Change-Part 结果相近。这样的结果是符合预期的，因为 LSTM 虽然考虑到属性词，但输入差别小；而 TD-LSTM 基于属性词建模，属性词不同，输入情况大不相同，情感极性自然不同，TD-LSTM 情感分析显然更为准确。AT-LSTM 并没有带来性能提升，我们分析认为是词向量中混合不同情感导致注意力机制起了少许负面作用。Change-Part 原本就是在 TD-LSTM 基础上修改而成，因此效果也相似。

表 3 不同分类算法分类结果

Table 3 Results of different classification algorithms

类别	LSTM	TD-LSTM	Change-Part	AT-LSTM
家具	0.80	0.84	0.83	0.79
百货	0.79	0.81	0.79	0.78
厨房	0.80	0.82	0.81	0.78

第三组实验利用效果最好的 TD-LSTM 单分类器集成与 3 分类器集成(LSTM＋TD-LSTM＋AT-LSTM)和 4 分类器集成(LSTM＋TD-LSTM＋Change-Part＋AT-LSTM)比较(利用方法(7)－方法(9))。集成学习时，正类样本划分成 10 组与负类样本组成 10 组样本(评论数据正类与负类的不平衡比都大约等于 10 或者大于 10)，然后循环使用分类算法形成 10 个分类器集成。表 4 中显示，4 分类器集成相较于单分类器集成性能提升 4.4%左右，可见多分类器的集成学习方法是有效的。我们可以得出结论：欠采样的基于 LSTM 类分类器集成学习方法确实在处理不平衡样本问题时非常有效。除此之外，4 分类器集成相比 3 分类器集成，情感分类效果也有小幅度提升。分类器种类越多，差异越大，集成后分类效果会有提升。

表 4 各分类器集成分类结果

Table 4 Results of different ensemble classification algorithms

类别	TD-LSTM	3-classifiers	4-classifiers
家具	0.83	0.84	0.86
百货	0.83	0.87	0.87
厨房	0.84	0.88	0.88

通过上面多组实验比较，我们得出结论：欠采样后多分类器集成学习方法的性能比传统单分类器方法性能高得多。本文方法是能够较好地处理不平衡数据上的属性级情感分类问题的。

1) http://keras.io/

结束语　本文提出了一种基于 LSTM 的集成学习情感分类方法，来处理常见的不平衡数据问题。方法中先采用欠采样方法来得到多组训练语料，然后利用 LSTM 和 TD-LSTM 等属性词相关的算法对每组训练集训练，最后集成学习，使用该模型对测试集分类。最终的实验结果表明，文中方法确实有效，能够比较优秀地分类不平衡数据，处理属性级情感分类问题。该方法相比于非欠采样方法或者单传统算法分类的性能都要高得多。

本文只是提出一种处理不平衡数据情感分类问题的思路，方法还有很大的改进空间。未来工作中，我们会考虑其他分类算法的集成学习，并且拓广应用范围，选择英文语料或者其他类别的主观性评论。除此之外，欠采样方法的改进也是值得进一步研究的问题。我们会在下一步研究这些问题，力求提升情感分类性能。

参 考 文 献

[1] ZHAO Y Y, QIN B, LIU T. Text sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834-1848.

[2] BARANDELA R, SANCHEZ B J S, GARCIA V, et al. Strategies for learning in class imbalance problems[J]. Pattern Recognition, 2003, 36(3): 849-851.

[3] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[4] TANG D, QIN B, FENG X, et al. Effective LSTMs for Target-Dependent Sentiment Classification [J]. arXiv: 1512. 01100, 2015.

[5] XU F, PAN Z, XIA R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework[J]. Information Processing & Management, 2020, 57(5): 102221.

[6] MULLEN T, COLLIER N. Sentiment analysis using support vector machines with diverseinformation sources[C]// Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. 2004: 412-418.

[7] XIE X, GE S, HU F, et al. An improved algorithm for sentiment analysis based on maximum entropy[J]. Soft Computing, 2019, 23(2): 599-611.

[8] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment Classification using Machine Learning Techniques[C]// 2002 Conference on Empirical Methods in Natural Language Processing. 2002: 79-86.

[9] JAYANAG B, VINEELA K, VASAVI S. Feature Subsumption for Sentiment Classification of Dynamic Data in Social Networks using SCDDF[J]. International Journal of Advanced Computer Science and Applications, 2012, 3(9): 1575-1605.

[10] GRAVES A. Supervised sequence labelling with recurrent neural networks [M]. Berlin: Springer, 2012.

[11] LONG F, ZHOU K, OU W. Sentiment analysis of text based on bidirectional LSTM with multi-head attention[J]. IEEE Access, 2019, 7: 141960-141969

[12] WANG Y, HUANG M, ZHU X, et al. Attention-based LSTM for Aspect-level Sentiment Classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.

[13] WU Z, ONG D C. Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis[J]. arXiv: 2010. 07523, 2020.

[14] JIANG N, TIAN F, LI J, et al. MAN: Mutual Attention Neural Networks Model for Aspect-Level Sentiment Classification in SIoT[J]. IEEE Internet of Things Journal, 2020, 7(4): 2901-2913.

[15] WANG Z H, WANG Z Q, LI S S, et al. Feature Selection for Imbalanced Sentiment Classification[J]. Journal of Chinese Information Processing, 2013, 27(4): 113-119.

[16] YE F, JIANG Y S. Unbalanced classification method based on clustering and under-sampling[J]. Computer Application and Software, 2020, 37(1): 298-303.

[17] LIN W C. Clustering-based undersampling in class-imbalanced data[J]. Information Sciences, 2017, 409-410: 17-26.

[18] LIU X Y, WU J, ZHOU Z H. Exploratory Undersampling for Class-Imbalance Learning[J]. IEEE Transactions on Systems Man & Cybernetics Part B, 2009, 39(2): 539-550.

[19] KITTLER J, HATEF M. On combining classifiers [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(3): 226-239.

[20] LI J, LUONG M T, JURAFSKY D, et al. When Are Tree Structures Necessary for Deep Learning of Representations? [C]// The 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 2304-2314.

[21] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. arXiv: 1409. 0473, 2014.

LIN Xi, born in 2000. His main research interests include natural language processing and so on.

WANG Zhong-qing, born in 1987, Ph.D, is a member of China Computer Federation. His main research interests include natural language processing and sentiment analysis.