

发布日期xxxx 00, 0000, 当前版本日期xxxx 00, 0000。

数字对象标识符10.1109 / ACCESS.2020.Doi号

基于弱标签信息的深度学习情感分类

王传涛^{1,2}, 杨学新^{1,2}, 和林可丁^{1,2}
¹北京建筑工程大学机电与车辆工程学院, 北京102616
²北京市建筑安全监测工程研究中心, 北京102616

通讯作者：杨学新（19801369012@163.com）。

这项工作得到了BUCEA教育科学研究基金（Y1810），北京建筑安全监控工程研究中心开放研究基金计划（202103）和BUCEA 2020年研究生教育和教学质量改善项目（2020013）的部分支持。）。

抽象的 情感分类的目的是解决文本情感倾向自动判断的问题。在线评论的情感分类任务中，传统的深度学习情感分类模型侧重于算法优化以提高模型的分类性能，但是当用于手动标记情感倾向的样本数据不足时，模型的分类性能将很差。一方面，基于弱标签信息的深度学习情感分类模型将弱标签信息引入模型的训练过程中，以减少手动标签数据的使用。另一方面，标记信息薄弱可以在一定程度上代表评论的情绪倾向，但同时也包含杂音，该模型减少了弱标签信息中噪声的负面影响，从而提高了情感分类模型的分类性能。实验结果表明，在酒店在线点评的情感分类任务中，基于弱标签信息的深度学习情感分类模型在不增加人工成本的情况下具有优于传统深度模型的分类性能。

索引条款 深度学习，情感分类，弱标签信息，不平衡分类。

一，引言

互联网的发展为用户提供了一个自由发布在线评论的平台。在线评论包括产品评论，电影评论和服务评论。这些评论包含人们想要表达的情感信息。在线评论中的情感信息可以帮助公司改善产品，帮助政府监控公众舆论并为其他用户提供参考。从评论文本中挖掘情感信息需要情感分析技术，而情感分析技术的基本任务之一就是情感分类。自动区分文本的情感倾向的过程是情感分类。

手动缩放情感趋势样本，这将消耗大量的人力资源。为了清楚起见，以下将人工标记情感倾向的样本数据定义为标记数据。面对大量评论数据，很难手动标记每个评论样本的情绪倾向。许多平台的在线评论不仅包含评论文本，还包含其他一些信息，例如乐谱，表情符号，标签等。这些信息可以视为用户对其自己的评论文本的标记。将此信息引入深度学习模型的模型训练可以减少标记数据的使用。但是，由于某些评论是任意的，因此用户发布评论没有统一的标准。诸如评分之类的信息标记的情感趋势与用户评论文本（例如具有高分的负面评论文本）中表达的情感不一致。如图1所示，用户在电子商务平台上购买的iPhone 11上发布了否定评论文字，但给出了4颗星的积极评分，

传统的深度学习模型专注于算法优化，以提高模型的分类性能并在情感分类任务中表现良好。但是，基于深度学习的传统情感分类模型的训练过程需要

这就是所谓的噪音。由于乐谱，表情符号，标签和其他信息中存在杂音，我们将此信息定义为弱标记信息。包含弱标记信息的样本数据被定义为弱标记数据。将评论文字所表示的弱标记信息与情感倾向不一致的样本定义为噪音样本，并将评论文字所表示的弱标记信息与情绪倾向不一致的样本定义为正确样本。弱标记数据和标记数据之间的区别在于，弱标记数据包含噪音样本，而标记数据不包含噪音样本。

实验表明，在引入弱标签信息参与模型训练的过程中提出的两种方法可以有效减少弱标签信息中包含的噪声对情感分类模型的负面影响，从而提高模型的情感分类性能。

本文的结构如下：第二部分详细介绍了与本文相关的一些工作；第三部分介绍了本文提出的基于弱标记信息的深度学习情感分类模型；第四部分给出了情感分类模型与分类性能对比分析的实验结果。第5节总结了这篇文章，并提出了未来的前景。



图1. 弱标记信息中的噪音示例。

传统研究人员通常将使用弱标记数据等同于使用标记数据，即仅使用弱标记数据来训练情绪分类模型。这解决了标签数据不足的问题，但是由于弱标记数据中存在噪音样本，因此噪音样本将在模型训练期间对模型产生负面影响，并降低模型的分类性能。因此，要减少噪音样本对模型的负面影响，同时将弱标记信息引入模型训练中，是很难实现的。

二。相关工作

A.情感分类

目前，情感分类主要分为两个研究方向：基于字典的情感分类和基于机器学习的情感分类[1-3]。参考文献[4-7]各自提出了一个新的情感词典。实验表明，情感分类任务的分类性能优于传统的情感词典，基于字典的情感分类方法的分类性能优异。但是，情感词典是取决于域的。一旦情感分类任务的应用领域发生变化，基于情感词典的分类模型的分类性能就会下降。参考文献[8]提出使用多域数据创建情感字典来解决情感字典的域依赖问题。尽管情感词典的域依存性有所减弱，但基于词典的情感分类模型最大的问题是情感词典的构建需要大量的人为参与，而且随着网络数据的爆炸性增长，难以解决未知问题。通过手动在情感词典中的单词。

从所有样本中提取少量样本，以手动标记情感倾向，并获得由剩余样本组成的少量标记数据和大量弱标记数据。提出了以下两种方法来训练使用标签数据和弱标记数据的情感分类模型，以减少弱标记数据中的噪音样本对模型的负面影响，从而提高模型的分类性能。

(1) 情感分类模型的训练分为两个阶段。在训练的第一阶段，使用大量的弱标记数据来训练模型，然后在第二阶段使用一些标记数据来继续训练模型以微调模型。

(2) 首先，通过使用一些标记数据和对应于这些标记数据的原始弱标记数据来训练神经网络降噪模型，该原始弱标记数据用于对弱标记数据进行去噪。然后，通过降噪模型对海量弱标记数据进行去噪，并将降噪模型的输出作为情感分类模型的输入，以训练情感分类模型。

参考文献[9]提出使用机器学习技术来完成情感分类的任务。实验表明，基于机器学习的情感分类模型具有很好的分类性能。参考文献[10-12]各自提出了一种新的机器学习算法来完成情感分类任务并实现了较高的分类性能。与基于字典的情感分类模型相比，基于机器学习的情感分类模型避免了单词未知的问题，但是传统的机器学习算法的特征工程仍然需要较高的人工成本。作为机器学习的一个分支，深度学习在以下领域得到了迅速发展。

近年来，在诸如在线评论之类的大型文本的情感分类任务中表现出色。与机器学习技术相比，深度学习技术极大地降低了人工成本。Word2vec, ELMO [13], BERT [14]促进了深度学习在情感分类任务中的应用。参考文献[15-17]为情感分类提出了创新的深度学习模型。实验表明，分类性能优于传统的深度学习模型。但是，这些模型使用的大多数数据集都是用于人工标记情感倾向的标准数据集，而缺乏用于深度学习模型的监督训练的大规模标记数据是深度学习模型的应用瓶颈。

B.弱标记数据

在深度学习模型的训练过程中引入弱标签数据可以解决训练数据不足的问题。目前，关于弱标记数据的研究相对较少。在Twitter挖掘中，参考文献[18]使用半监督分类对症状监测中的相关性进行过滤，并在任务中引入了带有弱标记数据的数据，以完成分类任务。参考文献[19]已经标记了160万个Twitter文本，并建立了表情符号情感词典。实验验证了表情符号（即弱标记信息）在情感分类任务中的有效性。参考文献[20]提出了两个大数据系统来使用表情符号来完成Twitter文本的情感分类任务。实验证明，所提出的两个系统的准确性和鲁棒性都非常好。参考文献[21]在深度学习模型的训练中引入了弱标记数据，并在情感分类任务中实现了出色的分类性能。在情感分类任务中，这些研究人员都使用弱标签数据等效于标签数据，但忽略了弱标记数据中包含的噪声特征。因此，在使用弱标记数据的同时，本文提出了两种方法来减少弱标记数据中包含的噪声样本对情感分类模型的负面影响，从而提高情感分类模型的分类性能。这些研究人员都使用弱标记数据等同于标记数据，但是忽略了弱标记数据中包含的噪声特征。因此，在使用弱标记数据的同时，本文提出了两种方法来减少弱标记数据中包含的噪声样本对情感分类模型的负面影响，从而提高情感分类模型的分类性能。这些研究人员都使用弱标记数据等效于标记数据，但是忽略了弱标记数据中包含的噪声特征。因此，在使用弱标记数据的同时，本文提出了两种方法来减少弱标记数据中包含的噪声样本对情感分类模型的负面影响，从而提高情感分类模型的分类性能。

根据弱标记信息划分所有数据后的情感趋势分布如图2所示：

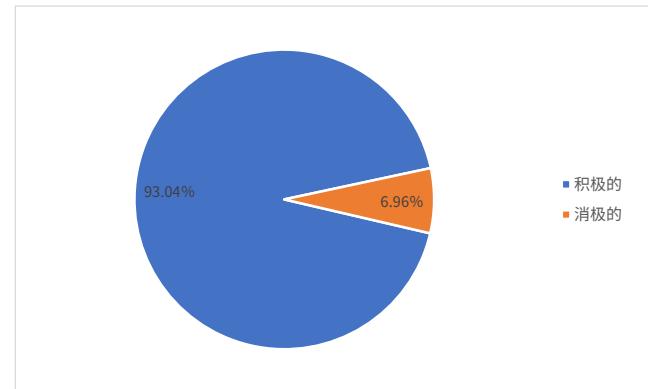


图2. 情绪傾向的数量分布。

评论数据通常包含一些信息，这些信息对后续的情感分类任务没有帮助。使用与数据清理相关的技术可以提高数据质量，从而提高后续情感分类模型的分类性能。具体的数据清理步骤如图3所示：

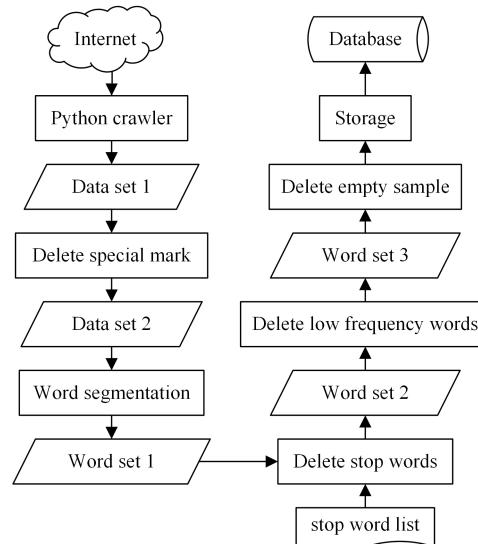


图3. 数据爬网和清理。

三，方法

A.数据清理

本文中使用的示例数据是从酒店网站抓取的北京快捷酒店的评论数据。使用的弱标记信息是用户在发布酒店文字评论时给酒店的分数。根据评分系统，每条评论的总分为5。得分高于2.5分的评论被归为积极情绪倾向，得分低于或等于2.5分的评论被归为负面情绪倾向。总共抓取了983220个在线评论，

删除特殊标记：由于用户在网站上发布自己的评论时并没有统一的规则，因此他们经常在文本评论中包含一些与样本的情感倾向无关的特殊标记。

分词：用户发布的中文在线评论以单词序列的形式出现。分词是指将每个在线评论的单词序列分为多个单独的单词。

删除停用词：停用词是指对整个在线评论没有语义意义的词，例如感叹词，代词等。删除停用词和删除特殊符号具有相同的效果，这有助于

随后的情感分类模型可以更好地捕捉复习句子的主要语义。

删除低频词：由于互联网用户数量众多，每个人都有各种各样的词来表达自己的情感。尽管有些词可以代表一定的情感倾向，但它们的频率太低。对于整个情感分类模型而言，学习低频词的情感倾向弊大于利。

之后，删除空白评论。最后，数据存储在数据库中，以方便后续情感分类模型的使用。

在实验中，首先从所有样本中提取少量样本，以手动标记情感倾向。根据第1节中给出的噪声样本的定义，图4显示了在手动标记样本的情感趋势之前，噪声样本和正确样本的分布。

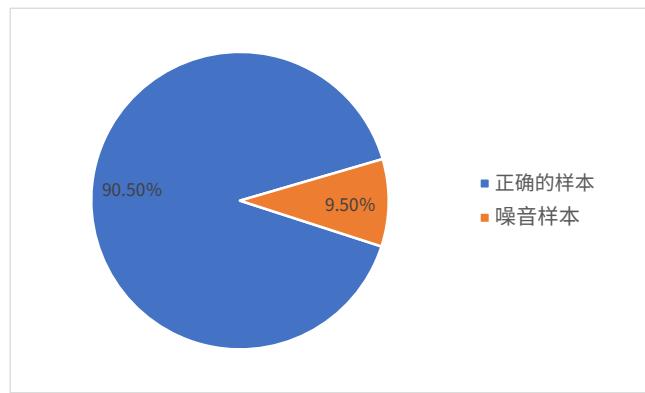


图4。噪声样本和正确样本的数据分布。

实验中总共对19233个样本进行了标记，其中1828个噪声样本占

9.50%。弱标签数据中的噪声样本比例不是很大，因此在大多数情况下，弱标签数据可以代表用户的真实情感倾向。但是，这部分噪声样本仍然会对情感分类模型产生很大的负面影响。因此，如果要进一步提高情感分类模型的分类性能，就不能忽略噪声样本的存在。

B. 基于BiLSTM的情感分类模型

长短期记忆（LSTM）[22]神经网络是一种引入“门”机制的递归神经网络。LSTM神经网络可以捕获更长距离的语义依赖性，避免了由于序列长而导致的传统递归神经网络梯度消失的问题。图5显示了所有评论样本中包含的单词数分布。统计数据显示，大多数评论样本包含较少的单词，并且50个单词以内的样本数量

占总数的98.51%。因此，LSTM适用于酒店评论的情感分类任务。

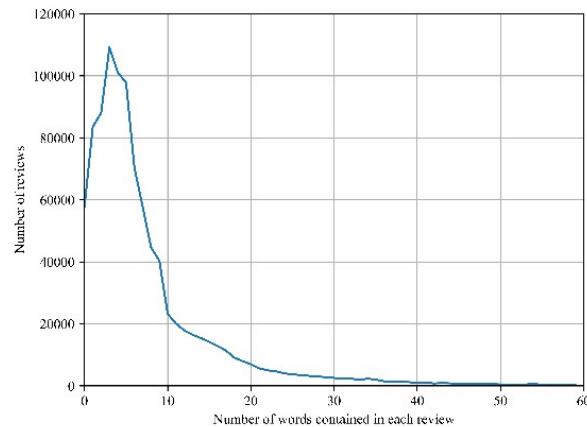


图5。复习样本中包含的单词数的分布。

LSTM的计算过程在（1）～（6）中给出。
指的是Sigmoid激活功能，
• 指权重矩阵之间的点积运算，

$\bullet W, U, V^*$ • 伊夫科。
指的是在
LSTM单位； $x_t, c_t, i_t, f_t, \rho_t, c_t, h_t$ 他们分别
代表当时LSTM单位的输入，单元状态，输入门的值，忘记门的
值，输出门的值，候选单元的状态以及LSTM单元的输出。

$$\tilde{c}_t = \tanh(Wx_t + Uh_t) \quad (1)$$

$$c_t = \tanh(\tilde{c}_t) \quad (2)$$

$$F_t : \text{宽} \times \text{高} \rightarrow \text{英尺 } \rightarrow \text{电压 } \rightarrow \text{英尺 } \quad (3)$$

$$C_t = F_t \cdot C_{t-1} + \tilde{c}_t \quad (4)$$

$$O_t : \text{宽} \times \text{高} \rightarrow \text{ot} \rightarrow \text{电压 } \rightarrow \text{ot} \quad (5)$$

$$h_t = \tanh(c_t) \quad (6)$$

LSTM可以从整体上理解文本语义，并且在情感分类任务中表现良好[23]。但是，LSTM只能捕获单个方向的语义依赖性。双向长期短时记忆（BiLSTM）[24]神经网络是一种改进的LSTM神经网络，可以捕获双向长距离语义依赖性。因此，在实验中，BiLSTM神经网络主要用作深度学习情绪分类模型的基本组成部分，以获得优异的分类性能。图6显示了BiLSTM神经网络的结构。

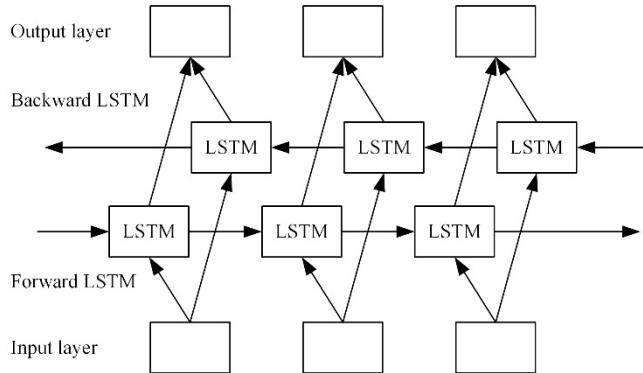


图6. BiLSTM的结构图。

Word2vec用于将审阅数据的文本信息转换为矢量表示，然后将其输入到BiLSTM中。最终的样本情感分类结果是通过将BiLSTM传递到Sigmoid层获得的。具体的情感分类模型如图7所示：

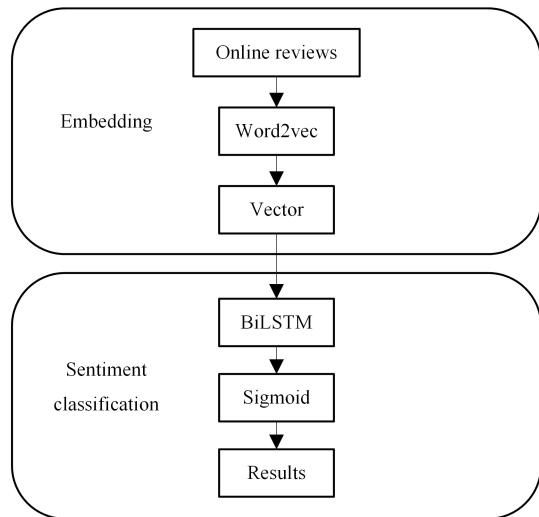


图7. 基于BiLSTM的情感分类模型的结构图。

C. 基于两阶段训练的BiLSTM情感分类模型

基于两阶段训练的两阶段训练的BiLSTM情感分类模型是第一个基于弱标记信息的深度学习情感分类模型。实验中使用的情感分类模型的结构与基于BiLSTM (SC_BiLSTM) 的情感分类模型一致，并且仅在模型训练方法上进行了创新。模型的训练分为两个阶段，两阶段训练方法的示意图如图8所示。

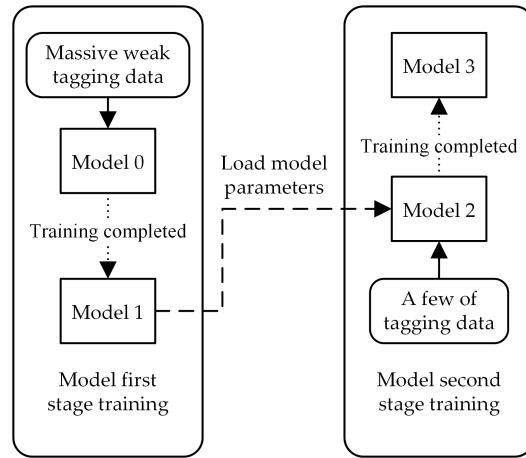


图8. 两阶段模型训练的示意图。

首先，在第一阶段，使用大量的弱标记数据来训练SC_BiLSTM (模型0)。在第3.A节的图4中，可以看出，弱标签数据中的噪声样本所占的比例很小。因此，在使用大量弱标记数据训练模型0之后，我们可以得到情感分类模型（模型1），该模型可以很好地捕获评论语义。之后，将模型1的参数作为第二阶段模型（模型2）训练的初始参数，并将一些标签数据输入模型进行训练，以达到对模型参数的微调目的。减少模型训练第一阶段输入的噪声样本对情感分类模型的负面影响，提高模型的分类性能，并获得最终的情感分类模型（模型3）。表1显示了基于两阶段训练的BiLSTM情感分类模型的训练过程。

表格1。基于两阶段训练的BiLSTM情绪分类模型的训练过程。

算法1：基于两阶段训练的BiLSTM情感分类模型

第一阶段训练的SC_BiLSTM。

输入：数据清理后的大量弱标记数据。

开始：

(1) SC_BiLSTM中的Word2vec将评文本转换为矢量表示形式。

(2) 使用公式 (1-6) 优化SC_BiLSTM模型的参数。

(3) 在SC_BiLSTMmodel收敛后保存模型参数。

输出：保存已完成SC_BiLSTM模型（模型1）的培训的第一阶段。

第二阶段训练的SC_BiLSTM。

输入：数据清理后的一些标记数据。

开始：

(1) 加载模型1的参数以得到模型2。

(2) SC_BiLSTM中的Word2vec将评文本转换为矢量表示形式。

(3) 再次使用公式 (1-6) 优化SC_BiLSTM模型的参数。

(4) 在SC_BiLSTMmodel再次收敛后保存模型参数。

输出：保存已完成SC_BiLSTM模型（模型3）的两阶段培训。

D. 基于弱标记数据降噪的情感分类模型

基于弱标签数据去噪的情感分类模型是基于弱标签信息的第二种深度学习情感分类模型。构建了一个深度学习模型，用于对弱标签数据进行消噪。基于深度学习的弱标签数据的去噪模型使用标签数据的文本信息和标签信息作为输出，并在手动标记情感倾向之前使用相应弱标签数据的文本信息和标签信息作为输入。基于深度学习模型训练的弱标签数据降噪模型 图9显示了基于深度学习的弱标记数据去噪模型的结构。

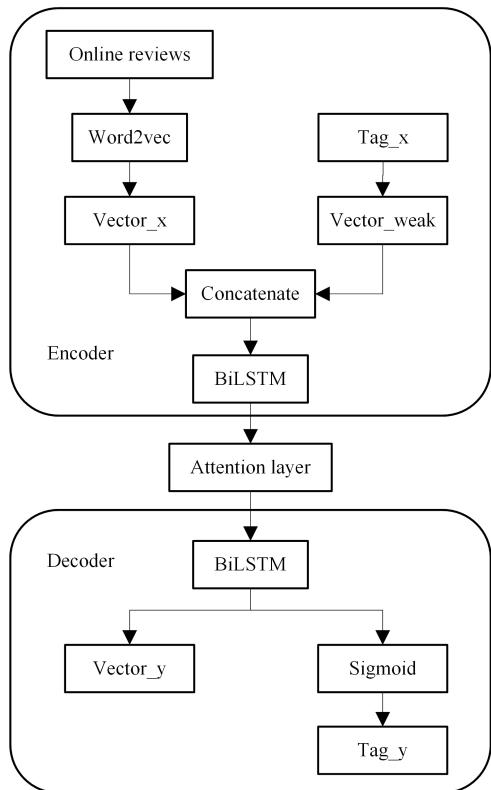


图9. 基于深度学习的弱标记数据降噪模型结构图

我们使用Word2vec将评论文本构造为矢量表示形式（Vector_x），并将相应评论文本的弱标记信息（Tag_x）转换为与Vector_x尺寸相同的矢量（Vector_{weak}）。如果弱标记信息表示正情绪趋势，则Vector_{weak}中的元素全部为1。如果弱标记信息表示负情绪趋势，则Vector_{weak}中的元素全部为0。我们通过连接层将Vector_x和Vector_{weak}联接起来，然后输入然后将它们放入BiLSTM层，以完成基于深度学习的弱标记数据降噪模型的Encoder部分。

通过Encode和Decode之间的注意力层，注意力机制可用于改善深度学习模型的性能。在(7)~(9)中，注意力机制的具体计算步骤为

给定的。在哪里 \tilde{T}_x 是序列长度； C_t 是...的语义向量 t ; \tilde{E}_{ti} 是对齐模型，代表

隐层状态的影响程度 H_{-t} 编码器中BiLSTM的时间一世 在隐藏层状态

S_t 解码器中BiLSTM的时间，由...计算 $-t$ H

和 τ $S_{-t} \cdot \tilde{E}_{ti}$ 是将注意力权重归一化为 t_i 通过softmax。

$$C_t \cdot \frac{\tilde{E}_{ti}}{\tilde{T}_x} \quad (7)$$

$$\cdot \frac{\exp (\tilde{E}_{ti})}{\sum_{k=1}^{\tilde{T}_x} \exp (\tilde{E}_{tk})} \quad (8)$$

$$\tilde{E}_{ti} \cdot a(s_{-t}, h)_{-t} \quad (9)$$

进入深度学习模型的Decoder部分，数据通过BiLSTM层分为两部分，一部分直接作为评论文本的矢量表示（Vector_y）输出，另一部分通过Sigmoid层获得，以获得去噪的标签信息（Tag_y）。

手动标记数据的情感趋势，而无需更改评论样本的文本信息。由于仅修改了噪声样本的标签，因此在训练过程中基于深度学习的弱标签数据降噪模型的输入与输出中的评论文本信息保持相同，即Vector_x与Vector_y相同，均为Word2vec从评论文本转换而来的单词向量。在训练过程中，基于深度学习的弱标记数据去噪模型的输入和输出具有不同的标记信息。输出是标记数据的标记信息，而输出是在手动标记评论样本的情感倾向之前标记数据的弱标记信息。因此，Tag_x和Tag_y不同。

在完成基于深度学习训练的弱标记数据降噪模型后，该模型具有降低弱标记数据噪声的能力。使用该模型时，一旦输入弱标记数据，便获得了噪声降低的输出数据。在训练过程中，文本信息和标签信息相互作用，并且在对弱标记信息进行去噪的过程中对文本信息进行调整。因此，使用模型时，Vector_x和Vector_y不同。在考虑弱标记数据的去噪之后，Vector_x仍然是原始审阅文本的矢量表示，而Vector_y是审阅文本的矢量表示。

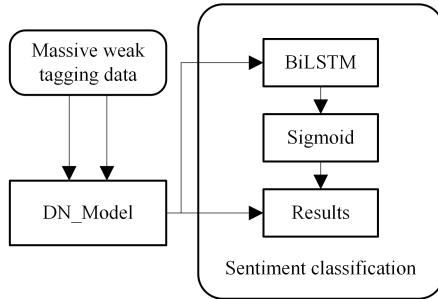


图10. 基于弱标签数据去噪的情感分类模型结构图。

在完成基于深度学习的弱标签数据的去噪模型的训练之后，构造基于BiLSTM的后续情感分类模型（DN_BiLSTM）以完成情感分类任务。图10示出了基于弱标签数据的去噪的整个情感分类模型的结构。图10中的“DN_Model”表示基于深度学习的弱标记数据的经过训练的降噪模型。将大量弱标记数据的文本信息和弱标记信息输入到DN_Model中，并输出经过去噪的文本矢量表示和标记信息。去噪的文本矢量表示用作DN_BiLSTM的输入，去噪的标记信息用作DN_BiLSTM的输出，以完成DN_BiLSTM的训练。因此，

范围	参数值
输入向量长度	50
输出尺寸	128
辍学	0.2
时代	50

表3显示了Text-CNN的参数设置。

表3。 DN_Model中的Text-CNN参数表。

范围	参数值
输入向量长度	50
输出尺寸	128
卷积层数	3
卷积核大小	3; 4; 5
辍学	0.2
时代	50

执行数据清理。数据清理的步骤如3.A中的图3所示。在数据清理中，使用jieba细分对评论文本进行细分，然后使用HIT停用词列表删除停用词。总计

983220酒店在线评论样本已被抓取，并且从其中提取了19,233个样本，以手动标记情感倾向。在19233个手动标记数据中，有30%被选作所有实验组的测试集。

建立了一个实验组（BiLSTM_Tag），该组仅使用一些标记数据来训练基于BiLSTM的情感分类模型。建立了一个实验组（BiLSTM_FullT），它将一些标记数据和大量弱数据混合在一起，以训练基于BiLSTM的情感分类模型。基于两阶段训练用途的BiLSTM情绪分类模型的第一阶段

大量的虚弱的标记数据为了训练（BiLSTM_Weak），第二阶段使用一些标签数据继续基于BiLSTM_Weak（BiLSTM_Continue）进行训练。BiLSTM_Tag，BiLSTM_FullT，BiLSTM_Weak和BiLSTM_Continue使用相同的模型结构，并且仅在模型训练数据上不同。

基于弱标签数据去噪的情感分类模型的训练包括两个模型的训练。首先，我们基于深度学习训练弱标记数据的降噪模型（DN_Model），然后训练情感分类模型（DN_BiLSTM）。DN_Model的输出包含去噪的标记表示，这也是情感分类的结果。因此，也有必要在测试集上测试DN_Model的情感分类性能。

IV. 实验

A. 实验设置

本文进行了两个实验。第一个实验的目的是选择具有最佳降噪性能的降噪模型。第二个实验是选择具有最佳分类性能的情感分类模型。

在第一个实验中，除了第3.D节中图9中的DN_Model模型外，还建立了以该模型为模板的两个对照组。对照组1将图9的DN_Model模型中的所有BiLSTM层更改为BiGRU [25]层，而对照组2将图9的DN_Model模型中的所有BiLSTM层更改为TextCNN [26]层。DN_Model使用的BiLSTM和BiGRU的参数相同。表2给出了BiLSTM和BiGRU的参数。

表2。 DN_Model中BiLSTM和BiGRU的参数表。

表4示出了每种模型的训练集和测试集中的样本的每种情感趋势的数量分布。

表4。 每个数据集中的情感倾向数量分布。

资料集	积极的	消极的
训练套 BiLSTM_Tag	11357	2106
训练套 BiLSTM_FullT	909945	67505
训练套 BiLSTM_Weak	898588	65399
训练套 BiLSTM_Continue	11357	2106
训练套 DN_型号	11357	2106
训练套 DN_BiLSTM	898588	65399
测试集	4838	932

实验中的所有模型都是使用TensorFlow（一个用于深度学习的开源框架）构建的。由于50个字以内的样本数量占样本总数的98.5%，因此Word2vec输出的向量的长度设置为50。用于情感分类模型的BiLSTM神经网络的超参数设置如表5所示：

表5。用于情感分类的BiLSTM神经网络的超参数设置。

范围	参数值
输入向量长度	50
输出尺寸	128
辍学	0.2
时代	20

使用F1，几何平均值（G-mean）[27-29]和准确性作为模型分类性能的评估指标。F1和准确率是常用的分类评估指标。从图2中的部分

从图3.B可以看出，积极情绪倾向和消极情绪倾向的样本数量之间存在较大差距。因此，酒店在线评论的情感分类是一个不平衡的分类任务，而在不平衡的分类任务中，G-mean是古典评价指标。在(10)和(11)中，分别给出了F1和G-mean的计算方法。

$$F1 = \frac{\text{精确召回}}{\text{精确召回} + \text{假负例}} \quad (10)$$

$$G\text{-均值} = \sqrt[n]{\frac{\text{记起}}{\text{一世}}} \quad (11)$$

记起 (12) 中的表示类别的召回率一世。

该实验仅完成了酒店在线评论情感分类的二元分类任务，因此 n 是2。

B. 实验结果

表6显示了三种降噪模型的性能比较，其中粗体数字是每个评估指标的最佳值。

表6。降噪性能比较。

方法	F1	G均值	准确性
比格鲁	0.456	0.0	0.838
文字神经网络	0.830	0.787	0.915
双线性STM	0.833	0.809	0.913

可以得出结论，BiGRU的分类性能较差，因为它受到不平衡数据分布的极大影响。与Text-CNN相比，BiLSTM的F1和准确度基本相同，但在G均值指标上，BiLSTM比Text-CNN高0.022。总体而言，以BiLSTM为组件的DN_Model具有最佳的降噪性能。因此，随后的实验中使用的DN_Model是使用BiLSTM构建的。

表7显示了所有实验组的情感分类表现，其中粗体数字是每个评估指标的最佳值。

表7。比较每个实验组的分类性能。

团体	方法	F1	G均值	准确性
1个	BiLSTM标签	0.803	0.745	0.905
2个	BiLSTM_FullT	0.708	0.568	0.886
3	BiLSTM_Weak	0.708	0.567	0.887
4	BiLSTM_Continue	0.841	0.816	0.918
5	DN_型号	0.833	0.809	0.913
6	DN_BiLSTM	0.859	0.850	0.924

比较BiLSTM_Tag，BiLSTM_FullT和BiLSTM_Weak的分类性能，并制作直方图，如图11所示：

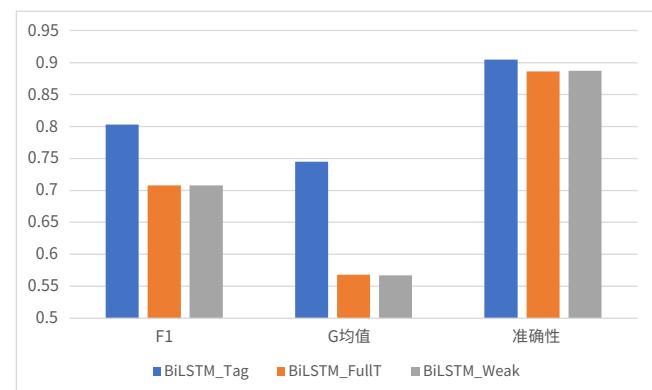


图11。实验组(1)，(2)，(3)的分类性能比较。

从图11可以看出，三个实验组中的所有评价指标BiLSTM-Tag最高。在实验中，BiLSTM_Tag训练仅使用标记数据，而BiLSTM_FullT和BiLSTM_Weak都在训练中使用弱标记数据。BiLSTM_FullT的训练集不仅包含弱标记数据，而且还包含

包含用于BiLSTM-Tag训练的标签数据。实验证明，如果将弱标签数据直接用于深度学习模型训练而不进行处理，则会导致分类模型的性能下降。另外，BiLSTM_FullT和BiLSTM_weak的评估指标值基本相同，这证明在训练过程中将某些标记数据与大量弱标记数据混合不会起到标记数据的作用。BiLSTM_weak的准确度达到0.887，而F1达到0.708，这证明了基于大量弱标记数据训练的情感分类模型已经能够很好地捕获评论文本的情感趋势。

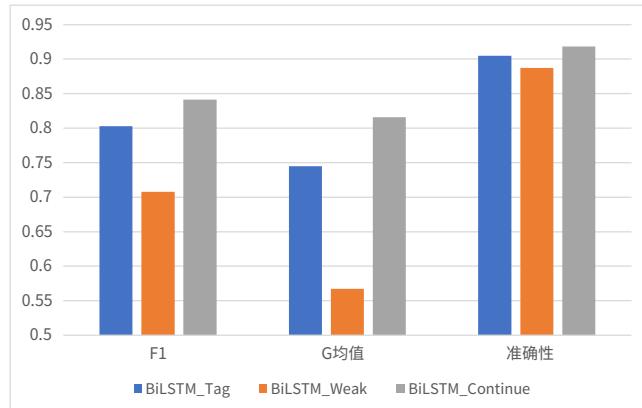


图12。实验组（1），（3），（4）的分类性能比较。

图12显示了BiLSTM_Tag, BiLSTM_Weak和BiLSTM_Continue的分类性能比较。BiLSTM_Continue的分类性能比BiLSTM_Weak更好，这表明基于两阶段训练的BiLSTM情感分类模型可以有效地使用某些标记数据。BiLSTM_Continue在这三个实验组中具有最高的评估指标，这证明基于两阶段训练的BiLSTM情感分类模型使用弱标记数据参与深度学习模型训练，同时减少了噪声样本对弱标记数据的负面影响在模型上，取得了卓越的分类性能。与BiLSTM_Tag相比，BiLSTM_Continue的F1, G平均值和准确性分别提高了0.038、0.071、0.013。

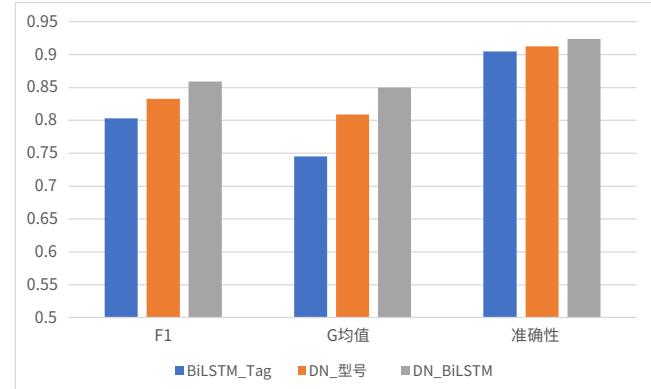


图13。实验组（1），（5），（6）的分类性能比较。

图13显示了BiLSTM_Tag, DN_Model和DN_BiLSTM的分类性能比较。DN_Model还使用一些注释数据进行模型训练，其分类性能优于BiLSTM_Tag。分析此现象的主要原因是DN_Model具有更复杂的神经网络结构。构建基于BiLSTM的情感分类模型的简单原因是，复杂的网络训练需要花费更多时间。如果使用与DN_Model相同的复杂网络结构，则在使用大量的弱标记数据来训练模型时会消耗大量的时间和成本。因此，在实验中，只有具有更复杂网络结构的DN_Model才会使用某些标记数据进行训练。DN_BiLSTM比DN_Model取得了更好的分类效果。分析的原因是，一方面，大量的弱标签数据具有更全面的复查样本特征，另一方面，在用DN_Model对弱标签数据进行去噪后，弱标签数据中的噪声样本对数据的负面影响会更大。模型的分类性能降低。与BiLSTM_Tag相比，DN_BiLSTM增加了

F1, G均值和准确度分别为0.056、0.105、0.019，而DN_BiLSTM在所有评估指标上均达到了所有实验组的最佳值。

BiLSTM_Continue使用标记数据来抵消噪声样本对情感分类模型的分类性能的负面影响。该方法相对简单，而DN_BiLSTM是一种用于学习弱标记数据的降噪方法，以减少噪声样本对分类方法的分类性能的负面影响。因此，DN_BiLSTM具有比BiLSTM_Continue更好的分类性能。

以上结果证明，两种基于弱标记信息的深度学习情感分类模型都比传统的仅基于标记数据进行训练的基于BiLSTM的情感分类模型具有更好的分类性能。

五，结论

本文提出了两种基于弱标签信息的深度学习情感分类模型，分别是基于两阶段训练的BiLSTM情感分类模型和基于弱标签数据去噪的情感分类模型。基于弱标签信息的深度学习情感分类模型使用弱标签数据进行模型训练，同时减少了弱标签数据中噪声样本对情感分类模型分类性能的负面影响，提高了情感分类模型的分类性能。在实验中，将基于弱标记信息的深度学习情绪分类模型与仅使用标记数据进行训练的传统情绪分类模型进行了比较，情感分类模型使用弱标签数据进行训练，但将弱标签数据与标签数据等同起来。实验结果表明，与其他情感分类模型相比，基于弱标签信息的深度学习情感分类模型在酒店在线评论数据的情感分类任务中，在不增加人工成本的前提下，显着提高了分类性能。

下一步，基于弱标签信息的深度学习情感分类模型将应用于电子商务评论和Twitter评论等更多类型的在线评论的情感分类任务，并将尝试使用更多类型的弱标签表情符号，符号等信息，以提高模型的情感分类性能。基于弱标签数据去噪的情感分类模型在所有实验组中获得了最高的分类性能，但它也是时间复杂度最高的模型。因此，在将来，我们应该尝试在保持较高分类性能的同时降低模型的时间复杂度。

参考

- [1] L. Jiang, Y. M. Zhou, M. Zhou, X. Liu and T. Zhao, “目标-依赖的Twitter情绪分类”，程序。第49届Annu。会议副会长。计算语言学。郎技术，卷1，第151-160页，2011年6月。
- [2] S. Kiritchenko, X. Zhu, C. Cherry and S. Mohammad, “NRC-Canada-2014：在客户评论中检测方面和情绪”，程序。8th Int. 车间语义评估。(SemEval)，第437-442页，2014年。
- [3] DT Vo, Y. Zhang, “具有丰富自动功能的目标相关的Twitter情感分类”，程序。IJCAI，第1347-1353页，2015年。
- [4] CS Khoo, SB Johnkhan, “基于词典的情感分析：六个情感词典的比较评估”，情报科学杂志，卷44号4，第491-511页，2018。
- [5] N. Rizun, Y. Taranenko and W. Waloszek, “通过对潜在的语义关系建模，提高了情感分类的准确性”，信息，卷9，没有十二页，307，2018。
- [6] MZ。Asghar, A。Khan, S。Ahmad, M。Qasim, 爱荷华州。Khan “使用基于规则的分类方案的词典增强的情感分析框架”，一等奖卷12号2，pp.e0171649,2017年。
- [7] A. Al-Saffar, S. Awang, H. Tao, N. Omar, W. Al-Saiagh, M. Al-bared, “基于组合分类方法和Senti-lexicon算法的马来语情感分析”，一等奖卷13号4，第pp.e0194852,2018。
- [8] Jha V, Savitha R, Shenoy PD等。关键词：情感感知词典，多领域情感分类
计算机与电气工程 2018, 69: 585-597。
- [9] B. Pang, L. Lee和S. Vaithyanathan, “竖起大拇指？分类 使用 机器 学习 技术” 自然语言处理中的经验方法，卷10，第79-86页，2002年。
- [10] Y. Wang, “基于迭代的朴素贝叶斯情绪考虑表情符号属性的微博多媒体帖子分类”，多媒体工具和应用程序，2020年，第1-16页。
- [11] SN Alyami和Olatunji SO, “支持的应用 使用基于Twitter的数据集进行阿拉伯语情感分类的矢量机，” 信息与知识管理杂志，卷19号2020年第1期，第2040018页。
- [12] 徐F, 潘Z.夏和R.夏, “电子商务产品评论 基于朴素贝叶斯连续学习的情感分类 框架，” 信息 加工 和 管理，2020年，第102221页。
- [13] 我是Peters, M.Neumann, M.Iyyer, M.Gardner, C.Clark, K. Lee和L. Zettlemoyer, “深度上下文文化的词表示”，arXiv预印本arXiv。pp.1802.05365, 2018。[14] J.Devlin, MW. Chang, K. Lee和K. Toutanova, “伯特：培训深层双向变压器以理解语言，” arXiv预印本arXiv，pp. 1810.04805, 2018。[15] R. Fei, Q. Yao和Y. Zhu, “深度学习结构 基于改进的交叉熵和权重的跨域情感分类，” 科学编程，2020年。
- [16] M. Wang, ZH Ning和T. Li, “信息几何学” 用于情感分类的增强型模糊深度信念网络，“国际机器学习与控制论杂志 卷10号11，页3031-3042，2019。
- [17] W. Li, P. Liu和Q. Zhang, “一种改进的方法 基于深度神经网络的情感注意机制的文本情感分类，” 未来的互联网，卷11号4，第pp.96,2019。
- [18] O. Edo-Osagie, G. Smith, I. Lake, O. Edeghere, La. De, B. 伊格莱西亚(Iglesia)，“使用半监督分类进行Twitter挖掘以在症状监测中进行相关性过滤”，一号第14卷，第7，页.e0210689，2019。
- [19] PK Novak, J. Smailović和B. Sluban, “表情符号，” 一，卷10号，第12页，e0144296, 2015年。
- [20] A. Kanavos, N. Nodarakis和S. Sioutas, “大比例尺Twitter情感分类的实现，” 算法 卷10号1, pp.33, 2017。
- [21] Z. Jianqiang, G. Xiaolin和Z.Xuejun, “深度卷积神经网络进行Twitter情绪分析，” IEEE Access, 第6卷, 23253-23260页, 2018。
- [22] S. Hochreiter和J. Schmidhuber, “长期短期记忆，” 神经计算 卷9, 没有8, 第1735-1780页, 1997年。
- [23] X. Zhu, P. Sobihani和H. Guo, “长期短期记忆 在递归结构上，” PMLR国际机器学习大会，2015年第1604-1612页。
- [24] M. Schuster和KK Paliwal, “双向复发 神经网络，” 中 IEEE信号处理事务,

- 卷 45 号 11, 第 2673-2681 页, 1997, doi: 10.1109 / 78.650093。
- [25] 赵。K., 范·梅里恩布尔。B 和 Gulcehre。C, “学习使用 RNN 编码器/解码器进行统计机器翻译的短语表示,” 计算机科学, 2014。
- [26] 金。Y., “用于语句的卷积神经网络分类,” Eprint Arxiv, 2014。
- [27] B. Krawczyk, 英国电信。McInnes 和 A. Cano, “情感混合人工智能系统国际会议使用二值化技术从多类不平衡 Twitter 数据中进行分类,” 史宾格湛, 2017 年第 26-37 页。
- [28] S. Li, G. Zhou, Z. Wang, SYM。李和王 R “情感分类失衡,” 第 20 届 ACM 信息和知识管理国际会议论文集, 第 2469-2472 页, 2011 年。
- [29] R. Xu, T. Chen, X. Y. Xia, Q. Lu, B. Liu, 和 X. Wang, “嵌入情绪和情感分类中数据失衡的成分,” 认知计算 第 7 卷, 第 2, 第 226-240 页, 2015 年。



林可丁 出生于中国山西
1996. 获得北京建筑工程大学工业工程学士学位
。自 2018 年以来, 他一直在北京土木建筑大学攻读工业工程硕士学位。他目前的研究兴趣包括深度学习和自然语言处理。



王传涛 1981 年生于中国湖北。他先后获得齐齐哈尔大学数学和应用数学理学学士学位, 2004 年, 2007 年重庆大学计算数学理学硕士学位和博士学位。2011 年获得北京交通大学系统工程专业博士学位。

从 2011 年到 2016 年, 他担任助理教授
在 这 工业的 工程
北京民政大学系

工程与建筑。自 2017 年以来, 他一直是副教授, 至今仍在这所大学的同一系工作。他是两本书和三十多篇文章的作者。他的研究兴趣包括数据挖掘, 深度学习和自然语言处理。



杨学新 出生于中国山东,
1997 年。于 2018 年在中国山东工业大学获得工业工程学士学位。他目前正在北京建筑工程大学获得工业工程硕士学位。他目前的研究兴趣包括数据挖掘和深度学习以及自然语言处理。