

旅游场景下的基于深度学习的文本方面级细粒度情感分类^①

刘文远^② 郭智存 郭丁丁^③

(燕山大学信息科学与工程学院网络感知与大数据智能研究室 秦皇岛 066004)

摘 要 方面级细粒度情感分类是指针对文本数据,分析其在指定方面的情感极性。由于获取到的评论样本往往涉及不同的方面,导致各个方面的情感极性不平衡。为了减少不平衡数据对模型训练的影响,本文提出了一种新的数据平衡方法——批处理平衡方法(BB),用来平衡多标签多类别数据。同时,由于评论文本蕴含多个方面,传统模型结构往往每次只能预测一个方面的情感。为了提高情感挖掘效率,本文提出了自动关注不同方面的情感注意力网络——双向循环卷积注意力网络(Attn-Bi-LCNN)模型。模型会同时关注不同方面的不同情感信息形成情感语义矩阵,根据情感矩阵进行情感预测。对比实验表明,模型取得了更好的预测结果和更快的运算速度。

关键词 情感分类;深度学习;神经网络;注意力机制

0 引言

随着社会的发展和人们经济水平的提高,人们往往从互联网上获取到各种旅游景点的信息。互联网的出现不仅使得信息的传播更为便捷,更激发起人们分享信息的欲望,尤其是当今的年轻人更愿意在各种旅游网站上发表自己的言论以此表达自己的情感。挖掘这种大量的含有个人情感的评论,不仅能够全面了解景点各个属性,同时也能够为其他文本应用领域带来辅助作用,因此,进行文本的方面级情感分析是很有必要的。

在文本的方面级细粒度情感分类研究中,最常见的问题有两个:一个是在数据处理过程中,获取到的数据类别不平衡的问题;另一个是在文本情感预测过程中,设计的模型架构的问题。在解决数据不平衡的问题中,常见的采样方式并不能对方面级文本进行类别平衡^[1]。针对文本的情感分类问题所采用的模型,目前最为有效的办法是搭建深度学习

网络。深度学习模型的建立降低了特征工程,无需人为提取特征,同时海量的数据也能够支持大规模网络中权重的训练。在文本情感分类领域,文献[2]在卷积神经网络(convolution neural network, CNN)的基础上引入词上下文信息来进行文本情感分类。文献[3]采用多任务卷积神经网络解决文本中情感分布问题。文献[4]提出循环卷积神经网络(recurrent convolution neural network, RCNN)对文本进行分类。文献[5]利用多尺度注意力机制关注文本不同部分来解决文本情感分类问题。文献[6]采用关注不同粒度词语进行方面级情感分类。文献[7]采用在高维空间生成样本来解决样本类别不平衡的问题。文献[8]采用循环神经网络来关注文本不同方面的情感倾向。文献[9]采用分层使用注意力机制的方式计算文本情感倾向。尽管每个模型都采取不同的方式来提取句子中的情感特征,但模型的最终结果仍是单一情感向量,只能预测某一方面的感情,不能同时预测所有方面的情感,导致模型在应用时耗费大量计算时间。

① 国家自然科学基金(61672448)和秦皇岛市科学技术研究与发展计划(202003B043)资助项目。

② 男,1968年生,博士,教授;研究方向:无线传感器网络和移动网络;E-mail: wylu@ysu.edu.cn。

③ 通信作者, E-mail: guodingding@ysu.edu.cn。

(收稿日期:2020-11-11)

针对方面级细粒度文本的数据不平衡问题,本文设计了一种全新的平衡方式——批处理平衡(batch balance, BB),对数据进行平衡化处理。针对传统模型预测效率过低问题,本文在批处理平衡数据的基础上,提出了一种新型的深度模型结构——双向循环卷积注意力网络(bidirectional convolutional recurrent attention network, Attn-Bi-LCNN)模型。Attn-Bi-LCNN 模型利用注意力机制能够同时关注句子中涉及到不同情感,生成情感语义矩阵,利用情感语义矩阵达到同时预测所有方面情感极性、提高模型实用性的目的。

本文主要贡献如下。

(1) 针对获取到的方面级细粒度文本数据类别不平衡问题设计了一种新的批处理平衡算法,用以解决上采样和下采样两种传统平衡算法不能平衡方面级细粒度文本数据问题。

(2) 探讨了神经网络权重初始化对模型学习能力的影 响,将正态分布与 Kaiming 分布结合在一起共同作用于神经网络权重的初始化,避免了由于神经网络过深而出现的输出结果消失或爆炸等问题。

(3) 针对传统神经网络生成单一语义向量,导致模型在方面级情感分类问题上预测效率较低的问题,设计了一种新的双向循环卷积注意力网络模型,用来提高预测效率。

本文将在第 1 节介绍情感分类领域内的相关工作;在第 2 节介绍针对多方面文本数据不平衡问题,采用全新的文本数据处理流程;在第 3 节对模型进行具体介绍;在第 4 节中介绍模型的实际学习能力;在第 5 节中作出结论。

1 相关工作

随着数据量的激增和硬件计算水平的大幅度提高,深度学习逐渐成为解决问题的主要的手段,同样在文本的方面级细粒度情感分类任务中,如何利用神经网络提取文本中各个方面的情感信息,成为研究人员研究的重点。

文献[10]基于循环神经网络提出了目标依赖长短期记忆网络(target-dependent long short term

memory, TD-LSTM)模型,分别将文本的正序和倒序输入到模型中,取得了很好的效果。文献[11]基于长短期记忆网络(long short term memory, LSTM)的基础上引入了词语的位置信息,利用词语的不同位置赋予不同的权重来关注句子中不同方面词语对情感的影响。文献[12]将文本的方面级情感分类问题由直接分类问题转换为文本的序列标注问题,利用循环神经网络时序性对预测结果进行时序性输出,取得了较好的效果。文献[13]利用深度记忆网络的记忆功能提高模型的长期记忆能力,从而提高模型的学习能力。

文献[14]将卷积神经网络与注意力机制结合起来,提出了一种卷积记忆网络,用来捕捉句子中词与词之间的关系,取得较好的效果。文献[15]在采用注意机制的基础上,对文本中各个词语的位置信息进行了重新编码,将每个词语独特的位置特征进行编码表示,并用于文本的方面级情感分类。

文献[16]在初始记忆网络的基础上构建了一种情感记忆网络,利用情感记忆网络强化模型对情感极性的敏感度,对于文本的情感分类起到了较好的效果。文献[17]将方面词与文本中不同的方面对应的情感标签联系起来进行联合学习。

尽管以上模型都取得了较好的效果,但是其特定的模型结构只会针对特定方面输出单一语义向量。在实际应用时,由于文本内蕴含多个方面,单一语义向量会耗费大量时间,为解决这一问题,本文提出了 Attn-Bi-LCNN 模型,能够进一步提高模型预测效率。

2 动态平衡数据处理

利用深度学习解决自然语言处理领域内的问题时,首先针对于文本的特点,需要进行文本的预处理。传统文本的处理流程包括文本的分词、去除停用词、文本可变长度处理和词语的向量化。针对数据不平衡的问题,本文在数据处理阶段加入了批处理平衡算法进行数据平衡。针对神经网络过深可能导致前向传播输出消失或爆炸问题,本文加入了神经网络权重预处理。数据处理流程如图 1 所示。

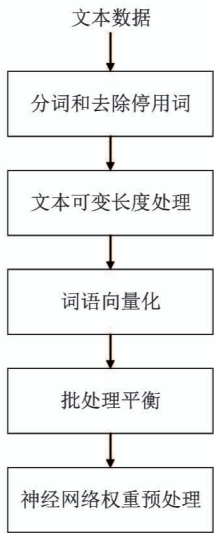


图 1 数据处理流程图

2.1 分词和去除停用词

文本的分词是指将完整的中文文本语句分割成词组。例如对于“这个酒店卫生不错,就是价格贵了一些”这一评论文本进行分词得到“这个 酒店 卫生 不错, 就是 价格 贵 了 一些”分词结果,词与词之间用空格隔开。随后,为了进一步降低数据噪声,本文将去除对于情感分类无意义的词语。

文本的去除停用词是指将分好词之后的语句去除不必要的词语。在去除停用词阶段,根据数据特点,本文在采用哈工大停用词典的同时加入了自己收集的停用词集合。哈工大停用词典是针对中文词的,但是在评论中可能会涉及到许多英文单词,但并不是所有的英文单词都具有意义,例如:“烤鸭味道很好,不比北京的烤鸭差,ps:面皮非常多也好吃,还有葱花,都搭配得很 good”评论中,单词“ps”并不能代表实际意义。可以通过去除英文无用词来达到削减噪声的作用,同时也节约了计算成本。

2.2 文本可变长度处理

在自然语言处理领域,与图像处理不同的是,图像处理问题往往可以获得大小相同的图像矩阵,但是由于文本数据的特殊性,获取到的文本往往是长短不一的,这时为了方便模型训练,需要进行文本的可变长度处理。

由于循环神经网络结构的特殊性,网络权重是在时间上共享的,所以循环神经网络能够处理任意长度的文本数据。

在批处理训练方式中,每次只向模型输入原始数据集的一小部分,但是每批数据里面评论文本的长度是长短不一的,需对较短文本进行填充处理,即对较短评论数据进行填 0 操作,使填充之后的长度与同批数据内最大长度相同。填充结果如表 1 所示。

表 1 文本填充示例

评论文本								原始长度
8	9	50	60	2	3	0	0	6
2	3	3	88	4	5	5	45	8
5	13	25	69	78	21	14	0	7

2.3 词语向量化

词语向量化是指将中文词语转化为稠密词向量。稠密向量化表示有两点:一是克服了文本的维度灾难,可以人为规定词向量维度大小。另一点是稠密词向量可以表示词与词之间的联系。本文采用 Word2vec 进行词语的向量化。

为了探索不同词向量维度对模型学习结果的影响,本文针对不同词向量维度进行了对比实验。在对比实验中,主要采取了 4 个不同大小的维度,分别是 50、100、150、200。词向量不同维度的大小代表了同义词在不同空间维度的向量表示,即同一词语在线性空间中的位置,用于以后的线性映射。

从图 2 中可以看出,词向量大小为 50 表达最好,但是各个维度之间的差异并不是很大。Word2vec 等词向量训练工具训练出的词向量,与 one-hot 形式词向量相比,不仅能够克服词向量维度灾难,能够灵活规定词向量维度大小,同时具有更深层次的语义联系。但是从图中可以看出,词向量维度过大,可能并不表示蕴含的信息越丰富,这是因

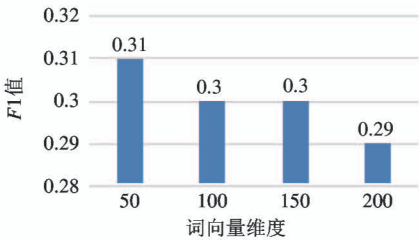


图 2 词向量维度对比

为,词向量的维度越大,可能包含的噪音就越多。同时由于词向量的维度过大,也会导致后续计算量过大的问题。稠密词向量相比于 one-hot 型词向量而言,克服了向量的稀疏性,并且稠密词向量能够蕴含更多的语义信息。但是,并不是稠密词向量的维度越大蕴含信息就越丰富,需要结合特定语境,来确定词向量维度大小。

2.4 批处理平衡

本文所采用的数据为评论文本数据,训练数据共计 105 000 条,共包括“交通是否便利”、“服务人员态度”、“上菜速度”等 20 个小方面,而对于每条评论而言,可能会涉及到 20 个方面中的其中几种,对于涉及到的方面又存在“好”、“中”、“差”3种情感倾向。

由于文本数据本身涉及了多个方面,每个方面涉及到不同的情感,部分方面数量统计数据如表2

表 2 部分方面数量统计表

	未提及	差	中	好
交通是否便利	81 382	1318	1046	21 254
距离商圈远近	83 680	586	533	20 201
是否容易寻找	80 605	3976	2472	17 947
排队等待时间	92 763	3034	4382	4821
服务人员态度	42 410	8684	12 534	41 372

所示。

根据表 2 可以看出,训练数据的各个类别数量严重不平衡,例如在“交通是否便利”方面,存在 81 382条训练文本没有提及该方面,同时在提及到此方面的文本中,“差”类别和“中”类别的情感倾向又远小于情感倾向为“好”的类别数量。

为解决数据不平衡问题,本文提出了独特的平衡数据算法——BB 算法。BB 算法流程如算法 1 所示。

算法 1 批处理平衡算法

```
Input: data  $x$ ,  $x$  labels
Output: data  $y$ ,  $y$  labels
1. import Counter
2. class num = dict( Counter( $x$  labels) ) #统计标签内各类别个数
3. if len(class num) == 1:
4.      $y$  labels = [2...]
5. elif len(class num) == 2:
6.      $y$  labels = reverse(  $x$  labels ) #将  $x$  的标签按位取相反值,即涉及的两个类别
7. elif len(class num) == 3:
8.      $y$  labels = reverse( max( $x$  labels), min( $x$  labels) ) #将  $x$  标签中最多最少的互换位置
9. else:
10.     $y$  labels = reverse( max( $x$  labels), min( $x$  labels) ) && reverse( other 2 class )
11. for label in  $y$  labs:
12.    generate ( data  $y$  aspect + sentiment word )
13. return data  $y$ ,  $y$  labels
```

在 BB 算法中,算法会循环遍历每一次进行训练的样本,针对每一条数据生成一条对应的平衡数据,所以算法的时间复杂度是 $O(n)$ 。由于每次训练都会生成等量的平衡数据,需要额外的内存空间,所以空间复杂度也是 $O(n)$ 。

为了探索批处理平衡是否能够有效避免由于数据不平衡导致的学习器具有倾向性的问题,本文设置了对比实验进行验证。

从图 3 可以看出,进行数据平衡对比能够有效促进神经网络学习,获得更高的 F1 值,模型具有更好的学习能力。

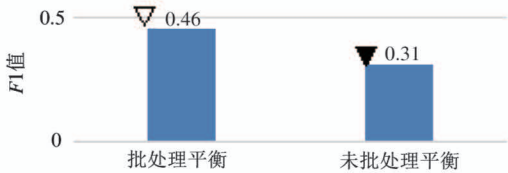


图 3 数据是否平衡对比

2.5 神经网络权重预处理

在神经网络计算过程中,其前向传播的本质即为矩阵乘法。在神经网络的前向计算中,如果初始化的初始权重随意设置,可能会引起前向传播计算爆炸或计算消失等问题。例如在图 4 中,如果初始化矩阵数值过大,并且神经网络层数较深,那么数据每经过一层神经网络,输出值就会快速增长。输出结果过大则会导致神经网络在反向传播过程中更新缓慢,从而导致学习缓慢,甚至使神经网络学习器无法学习。同理,如果神经网络初始权重过小,则会导致最终输出结果过小,从而导致输出结果计算的损失收敛缓慢或无法收敛,导致神经网络无法进行学习。

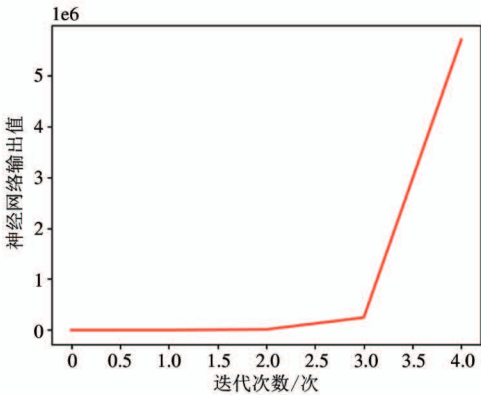


图 4 未进行权重预处理

为避免前向传播中出现结果爆炸或者消失现象,本文采用正态分布和 Kaiming 初始化相结合的方式对神经网络的隐藏层进行权重初始化,其中全连接层采用正态分布初始化,模型中的卷积层采用 Kaiming 初始化,通过两种初始化结合的方式对网络模型各个代表层进行权重初始化。

根据图 5 可以看出,模型的输出值围绕 1 上下

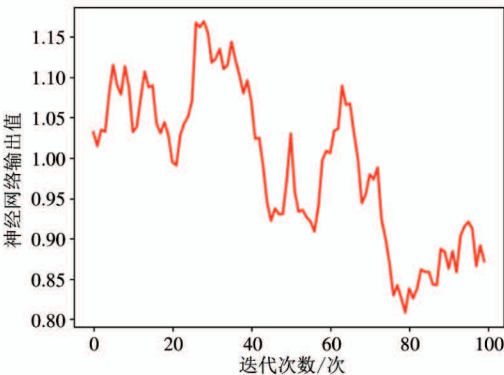


图 5 进行权重预处理

波动,尽管经过较深的神经网络前向计算,模型的输出值也没有出现爆炸或者消失的情况,所以进行神经网络的权重预处理能够有效避免前向计算中的输出爆炸或消失的问题,有效促进神经网络的收敛并进行学习。

3 Attn-Bi-LCNN 模型

3.1 模型输入

假定训练句子 $S = (w_1, \cdots w_i, \cdots w_n)$, 其中 w_i 代表句子中第 i 个词,句子共有 n 个词。那么,模型的输入即为训练句子 s 的词向量矩阵 V ,其中 $V \in R^{n \times d}$, d 为词语向量化维度。句子矩阵的第 i 行即为句子中 w_i 的词向量表示。

方面级别细粒度情感分类就是将训练数据输入到模型中,并由模型判断各方面情感极性。例如句子“这个酒店环境很好,就是有些贵”,需要将句子分词之后去除停用词得到的结果进行向量化,并将向量化之后的句子矩阵输入到模型中。当判断“环境”方面时,模型输出应为积极;当判断“价格”方面时,模型输入结果为消极。

本文采用批处理随机梯度下降,即每次输入模型中的数据为若干条待训练文本。因为循环神经网络能够处理任意长度的文本序列,所以在输入数据中需要对数据进行可变长度处理——对较短长度的句子进行填充。最终输入到模型中的数据大小为 $R^{B \times L \times d}$, 其中 B 为批处理数据量, L 为批处理数据中句子最大长度, d 为词向量维度。

为了避免模型在训练过程中出现过拟合的现象,本文采用正则化(regularization)与 Dropout 相结合的方式避免模型的过拟合。正则化是在机器学习或者深度学习中用来优化模型参数的一种重要手段,其基本表达形式为在损失函数之后添加额外约束项,如式(1)所示。

$$error = LossFunction + \beta \parallel w \parallel \tag{1}$$

在使用 Dropout 的神经网络中,神经网络会在特定层随机丢弃掉一些神经元来降低参数的数量,与此同时,也降低了模型过拟合的风险。应用 Dropout 另一个好处就是增加了模型的泛化能力。由于

在应用 Dropout 的过程中,神经网络每一次迭代都会随机丢弃掉一些神经元,而这些神经元并不是被固定的,而是随机丢弃的,这就造成每一次训练都是在原有网络基础上进行微改的神经网络,相当于进行了集成训练。

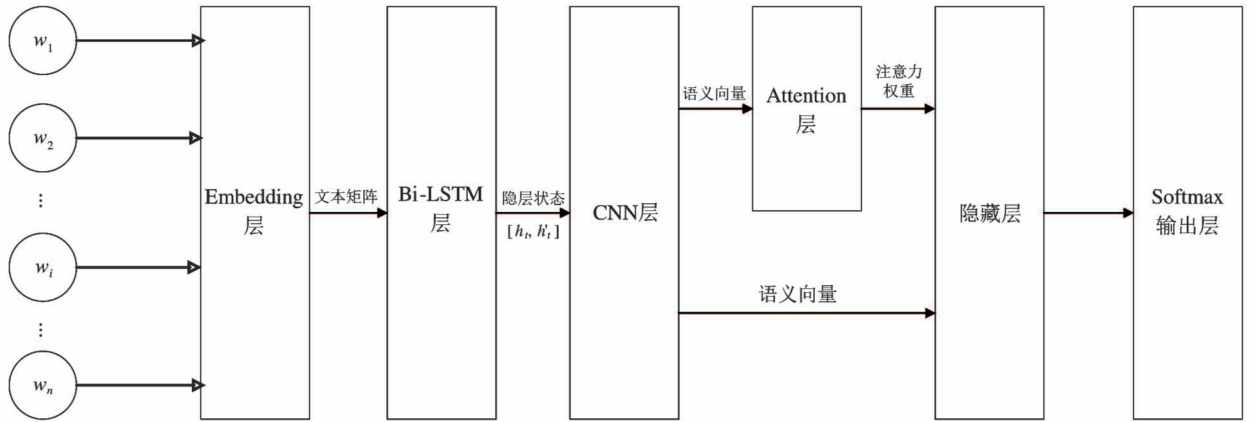


图6 Attn-Bi-LCNN 模型结构图

3.2.1 Embedding 层

模型中的 Embedding 层的作用是将词语映射为词向量,Embedding 层的本质即为一层全连接层,其内部权重即为词向量。通过 Embedding 层之后,将训练文本转换为文本矩阵,矩阵的第 i 行即为句子中第 i 个词的词向量。

$$S = R^{L \times d} \quad (2)$$

如式(2)所示,文本矩阵 S 大小为 $L \times d$, 其中 L 为矩阵行数,同时也是句子词语数, d 为词语向量化维度。

3.2.2 Bi-LSTM 层

Bi-LSTM 层的作用是为了提取句子中单词之间的顺序特征,并对句子表示进行记忆存储。Bi-LSTM 表示双向 LSTM,当文本矩阵中的词语按照时间顺序输入到双向 LSTM 中,会获取当前词的前向语义表示 h_t 和反向语义表示 h'_t ,将二者拼凑起来得到当前词的完整语义表示 $[h_t, h'_t]$ 。

同样,文本向量矩阵经过 Bi-LSTM 层的输出为文本的高维语义表示,输出结果为

$$S_{LSTM} = Bi-LSTM(S) = R^{l \times d} \quad (3)$$

其中, S_{LSTM} 代表文本高维语义表示矩阵, l 代表句子长度, d_{LSTM} 代表 Bi-LSTM 隐藏层维度大小,矩阵的

3.2 模型结构

Attn-Bi-LCNN 主要由 Embedding 层、Bi-LSTM 层、CNN 层、Attention 层及输出层组成。Attn-Bi-LCNN 模型示意图如图 6 所示。

每一行即为对应词的完整语义表示 $[h_t, h'_t]$ 。

3.2.3 CNN 层

CNN 层的作用是为了提取高维语义表示中的词与词之间的关系。循环神经网络的第 i 个词会受到前 $i-1$ 个词的记忆影响。与循环神经网络不同,此处 CNN 的作用是为了提取相邻若干个词之间的关系,对窗口之外的词并无任何影响。

本处 CNN 层采用的是前文介绍的 CNN 的一维卷积,其卷积结果由卷积核大小和卷积核步数决定。

$$S_{CNN} = C(S_{LSTM}) = R^{1 \times d} \quad (4)$$

其中, S_{CNN} 为卷积之后的句子向量, C 代表卷积操作, d_{CNN} 代表 S_{CNN} 的维度大小,具体值由设置的超参数决定。

3.2.4 隐藏层

隐藏层等同于前面介绍的全连接神经网络,起到特征融合和降维的作用。对于特征向量,可以通过设置隐藏层大小,进行升维或降维。

3.2.5 输出层

本文研究的问题为分类问题,所以选择 Softmax 函数作为输出层激活函数,其计算结果即为模型预测结果。Softmax 层计算公式如式(5)所示。

$$Softmax = \frac{e_i}{\sum_j e_j} \quad (5)$$

其中, e_i 代表向量中第 i 个元素的值, $\sum_j e_j$ 代表向量中所有元素的和。

3.3 Attention 层设计实现

模型中计算得出的情感语义矩阵的每一行即为每一方面的情感向量。不同方面的情感极性不同,同时也意味着不同方面对于句子的关注点是不同的,本模型采用注意力机制解决这一问题。矩阵的每一行情感语义向量都是通过 Attention 层与 CNN 层得出的语义向量进行计算,并赋予不同的关注权重,实现对不同方面的情感更好的预测。

本文采用的是自我注意力机制,其注意力权重计算公式如式(6)所示。Attention 层网络结构图如图 7 所示。

$$A = softmax(V_a \tanh(W_m D^T)) \tag{6}$$

其中, A 为注意力权重矩阵,矩阵的每一列代表一个方面的注意力权重; D^T 代表卷积神经网络卷积核得出的语义向量所组成的矩阵,矩阵的每一行为图 7 中的 d_{CNN}^i ,代表不同卷积核卷积得到的特征向量; W_m 代表 D^T 的权重矩阵,相当于一层全连接神经网络,其神经元数目为 m 个,为超参数; V_a 与伪随机初始向量不同,此处的 V_a 为方面词词向量矩阵,矩阵的每一行代表训练数据的各个方面词向量化得到的词向量,此处实验中可以进行微调更新或者不参与更新。最后通过 Softmax 激活函数对结果进行激活得到注意力权重矩阵 A 。

$$S = A \times D \tag{7}$$

得到不同方面的注意力权重之后,将权重矩阵 A 与卷积层得到的语义向量进行相乘,得到不同方面的情感语义矩阵 S 。

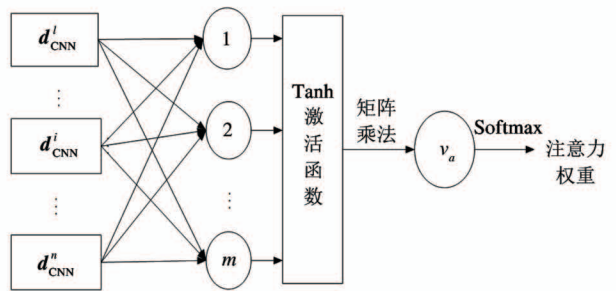


图 7 Attention 结构示意图

4 性能分析

本文针对评论文本的方面级细粒度情感问题在同一数据集上对不同模型进行了评估。

4.1 数据集

本文采用的数据集为 2018 年 AI Challenger 全球 AI 挑战赛中细粒度用户评论情感分析赛道数据。该数据集是专门用来判断文本各方面情感倾向的数据集,主要内容为用户对用餐地点的各个方面的评价,包括距离商圈远近、服务人员态度、价格水平、食物口感等 20 个方面的评价,共有 105 000 条用于训练的带标注数据,15 000 条验证数据。数据中的每条文本并不会涉及到所有的 20 个方面,所以模型的任务是判断方面是否被提及同时对提及到的方面进行情感预测。

对数据的预处理主要为前文所介绍的文本分词和去除停用词。本文采用 jieba 分词作为分词工具,因为用户在评价菜品或表明位置时可能会涉及到很多专有名词,例如“五香杏仁”、“紫菜蛋花汤”、“老吴酒馆”等,如果直接进行分词可能会出现把菜名或者地名分开等情况,造成分词不准确。所以,本文加入了在网上搜集的地名及菜名的词语作为自定义词典加入到 jieba 分词中,使其分词更加准确,避免造成歧义。

在去除停用词阶段,本文采用了哈工大停用词词表。同时,因为数据文本为用户评论,所以还包括很多人为噪声,例如“#”、“:)”、“^”等人为构建的网络表情符号,本文采用正则表达式过滤符号噪声。经过分词及去除停用词处理之后的文本,短则几十词,长则上百词,所以在进行模型训练时还需进行数据的可变长度处理。数据案例及标注如表 3 所示。

4.2 模型对比

本文提出的 Attn-Bi-LCNN 模型与其他传统情感分类模型相比,在解决多标签细粒度情感分类问题上更具有实用性,能够在验证数据集上更快地得出结果。这是因为传统的情感分类神经网络往往是将一句待预测样本转化为单一语义向量,并将此向

量用于情感分类。但是由于在多标签犀利去情感分类问题上,一句样本往往蕴含着不同方面的不同情感,通过单一语义情感并无法准确判断各方面的情感倾向。同时,由于其模型只会生成单一语义向量,

在进行预测时,每次运算只会预测一个方面。如果需要将样本内所涉及的 20 个方面全部进行预测,则需要将样本循环输入神经网络进行 20 次运算,这样会导致模型在实际应用时耗费大量计算时间。

表 3 数据实例及标注

评论文本	卫生	服务	价格
酒店很不错,很干净,服务员也非常的友善。	正面	正面	未提及
酒店位置很好,有停车场。房间干净,整洁。窗帘是电动的,很赞! 有小冰箱里面的饮料都是免费的。早餐品种丰富而且味道好。尤其感谢酒店礼宾部员工,我走的时候已经把车玻璃帮我擦干净! 非常感谢!	正面	正面	未提及
酒店环境一般吧,但是价格太贵了,下次不来了。	中性	未提及	负面

为了解决上述问题,本文提出的 Attn-Bi-LCNN 模型创新性地提出了在原有神经网络层上加入注意力机制并生成情感语义矩阵。情感语义矩阵与传统单一情感语义向量不同的是,情感矩阵的每一行向量都代表了不同方面的情感倾向。而情感矩阵内部的各个向量之间,通过引入注意力机制关注句子中不同部位,得出关注不同类别的情感倾向。为了验证 Attn-Bi-LCNN 模型的有效性,本节将从模型在验证集上的运算时间和预测准确程度两方面来对比 Attn-Bi-LCNN 模型的有效性。

4.2.1 模型准确度对比

为了探索 Attn-Bi-LCNN 模型准确度,涉及对比的传统神经网络模型有 RCNN、CNN、CNN + Attention、Bi-LSTM + Attention。实验验证集数据为15 000 条点评数据,涉及 20 个方面,每个方面分为 4 个类别:积极、中性、消极、未提及。评价指标依然采取 F1 值,F1 值越大,模型表现越好。对比模型如表 4 所示。

表 4 对比模型示例

模型名称	输出形式
Attn-Bi-LCNN	情感语义矩阵
RCNN	情感语义向量
CNN	情感语义矩阵
CNN + Attention	情感语义矩阵
Bi-LSTM + Attention	情感语义矩阵

RCNN 模型是应用在文本领域内的经典模型,由文献[4]在 2015 年 AAAI 会议上提出,在文本分类领域具有很好的适用性。RCNN 创新性的将卷积神经网络中的最大池化引入到循环神经网络中,解决了样本不等长的问题,实现了文本对齐。CNN 采用的是前文介绍的一维卷积,通过卷积核依次扫描样本,来模拟实现 N-gram 提取特征的过程,并通过反向传播来实现端到端训练。CNN + Attention 代表的是将注意力机制引入卷积神经网络,通过应用 CNN 提取相邻词语特征进行预测,是文本领域内经典的神经网络算法。Bi-LSTM + Attention 在原有双向循环神经网络的基础上引入注意力机制,通过注意力机制来实现对句子各个部位之间的关注。

根据图 8 可以看出,本文提出的 Attn-Bi-LCNN 模型取得了最好预测效果,其次是 RCNN 模型,表现最差的是 CNN + Attention 模型。出乎意料的是,

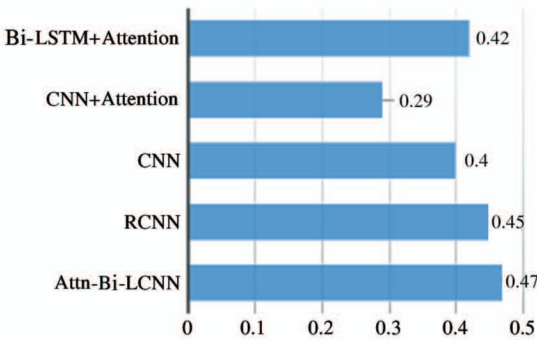


图 8 模型准确度对比

采用一维卷积的 CNN 模型取得了不错效果,而其相对应加上注意力机制模型却表现不佳。原因是单纯的 CNN 模型结构相对简单,而且一维卷积所提取的特征是相邻词组之间的特征,所以无需增添注意力机制进行补充。

Bi-LSTM + Attention 模型也取得了不错的结果。LSTM 能够有效地提取文本的时序性特征,使得前面的词语能够对后续的文本产生影响,同时由于 LSTM 具有长期记忆,在一定程度上避免了梯度消失现象。加入的注意力机制能够对句子中不同的部分产生关注,增添了对特征部位的选择,提高了模型的学习能力。RCNN 与其他模型相比虽然没有引入注意力机制,但是模型创新性地引入了卷积神经网络中的最大池化,这不仅解决了文本对齐问题,同时最大池化也起到了提取重要特征的作用,提升了模型的学习能力。

相比而言,Attn-Bi-LCNN 模型取得了最好的预测效果。首先 Attn-Bi-LCNN 模型主要由 CNN、LSTM、注意力层组成,其中 CNN 用以提取文本数据的相邻词组之间的抽象联系,将相邻词语联系起来。随后采用 LSTM 提取词与词之间的时序联系,将词语转换为高维语义向量。最后采用注意力机制关注句子中不同部位对情感判断的影响,将重要部位赋予相对大的权重。Attn-Bi-LCNN 模型结合了 CNN、LSTM、注意力层等网络层优点,取得了较好的学习能力,同时也证明了本文提出的 Attn-Bi-LCNN 模型的有效性。

4.2.2 模型预测时间对比

本文另一创新点是将于预测的情感语义向量提升为情感语义矩阵,使得情感语义矩阵能够同时预测不同方面的情感,提高了模型的应用性。为了验证 Attn-Bi-LCNN 模型在实际应用中的高效性,本文将与 RCNN 模型进行对比。RCNN 模型结构的特点是将训练文本转换为单一语义向量,利用语义向量进行情感预测。为此,本文将统计 RCNN 模型预测单一方面时间,同时和 Attn-Bi-LCNN 模型预测的 20 个方面计算时间进行对比。

根据图 9 可以看出,Attn-Bi-LCNN 模型运算时间要明显低于 RCNN 模型运算时间。同时 RCNN

模型还只是预测一个方面,其模型预测时间就要比 Attn-Bi-LCNN 模型预测时间多近 3 倍。本文提出的 Attn-Bi-LCNN 模型运算时间略高于前面 3 个模型,这是因为相对于前面 3 个模型,Attn-Bi-LCNN 模型结构更为复杂,参数较多。所以,Attn-Bi-LCNN 模型其情感语义矩阵结构是完全高效的,具有较高的实用价值。

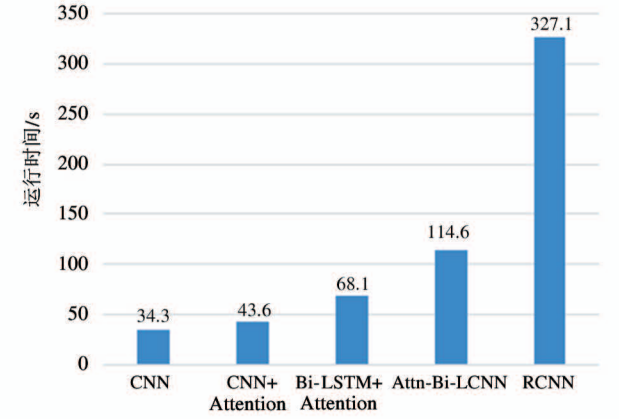


图 9 模型预测时间对比

5 结 论

针对 CNN 与 LSTM 提取特征的特点,将二者结合起来进行特征提取。利用 LSTM 能够提取序列数据的时序特征,将文本数据提取成高维语义特征向量。再利用 CNN 能够提取局部特征的特点,提取高维语义向量之间的局部联系,形成情感向量。

为了关注文本中的不同方面的不同情感,将注意力机制引入模型,利用注意力机制关注句子不同部位的特点生成情感语义矩阵。与传统模型生成情感向量不同,语义矩阵包含的情感更为丰富,能够同时预测不同方面的不同情感。

针对不同参数对算法的影响进行了测试,选取最优参数集。对不同算法模型进行了实验对比,验证了模型学习能力。同时对算法运算时间进行了对比,验证了模型的高效性。

综上所述,本文对评论文本的方面级细粒度情感分类进行了研究,研究成果还可以进行扩展。本文所采用的词向量预训练工具所处语境较为单一,后续可能采用更为优秀的词向量预训练工具进行训练。此外,在该数据集中,所有的方面是事先人为标

注好的,但是当今每天会产生海量数据,人为标注不现实。为此,对自动提取方面模型的探索仍是有必要的。

参考文献

- [1] ZHU M, XU C, WU Y F B. IFME: information filtering by multiple examples with under-sampling in a digital library environment[C] // Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, Indianapolis, USA, 2013: 107-110
- [2] KIM Y. Convolutional neural networks for sentence classification[C] // Empirical Methods in Natural Language Processing, Doha, Qatar, 2014: 1746-1751
- [3] ZHANG Y, FU J, SHE D, et al. Text emotion distribution learning via multi-task convolutional neural network[C] // The 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 4595-4601
- [4] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification[C] // The 29th AAAI Conference on Artificial Intelligence, Austin, USA, 2015: 2267-2273
- [5] FAN F, FENG Y, ZHAO D. Multi-grained attention network for aspect-level sentiment classification[C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018: 3433-3442
- [6] WANG J, LI J, LI S, et al. Aspect sentiment classification with both word-level and clause-level attention networks[C] // The 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 4439-4445
- [7] FERNÁNDEZ A, GARCIA S, HERRERA F, et al. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary[J]. *Journal of Artificial Intelligence Research*, 2018, 61: 863-905
- [8] LIU F, COHN T, BALDWIN T. Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis[C] // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, USA, 2018: 278-283
- [9] ISLAM J, MERCER R E, XIAO L. Multi-channel convolutional neural network for Twitter emotion and sentiment recognition[C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, USA, 2019: 1355-1365
- [10] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[C] // Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 2016: 3298-3307
- [11] BAO H, LI S T. Object-level sentiment analysis combining Bi-LSTM and positional relationship[J]. *Information System Engineering*, 2018(3): 149-151
- [12] WANG W, PAN S J, DAHIMEIER D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis[C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, USA, 2016: 616-626
- [13] TANG D, QIN B, LIU T. Aspect level sentiment classification with deep memory network[C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, USA, 2016: 214-224
- [14] FAN C, GAO Q, DU J, et al. Convolution-based memory network for aspect-based sentiment analysis[C] // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, USA, 2018: 1161-1164
- [15] LI L, LIU Y, ZHOU A. Hierarchical attention based position-aware network for aspect-level sentiment analysis[C] // Proceedings of the 22nd Conference on Computational Natural Language Learning, Brussels, Belgium, 2018: 181-189
- [16] WANG S, MAZUMDER S, LIU B, et al. Target-sensitive memory networks for aspect sentiment classification[C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018: 957-967
- [17] SCHMITT M, STEINHEBER S, SCHREIBER K, et al. Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks[C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018: 1109-1114

Text aspect-level fine-grained sentiment classification based on deep learning in travel scenarios

LIU Wenyuan, GUO Zhicun, GUO Dingding

(The Key Laboratory of Software Engineering of Hebei Province, School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

Aspect-level fine-grained sentiment classification is to analyze the emotional polarity of the text data in a given aspect. As the sample of comments often involves different aspects, the emotional polarity of each aspect is unbalanced. In order to reduce the impact of unbalanced data on model training, this paper proposes a new data balance method——batch balance (BB), which is used to balance multi-label and multi-category data. At the same time, because the commentary text itself contains multiple aspects, the traditional model structure can only predict one aspect of emotion at a time. In order to improve the efficiency of emotion mining, this paper proposes an attention network that automatically focuses on different aspects of emotions——a bidirectional circular convolutional attention network model. The model will pay attention to different affective information from different aspects at the same time to form the affective semantic matrix, and make affective prediction according to the affective matrix. The comparative experiments show that the proposed model achieves better prediction results and faster computation speed.

Key words: sentiment classification, deep learning, neural network, attention