

一种循环卷积注意力模型的文本情感分类方法



陈千^{1,2} 车苗苗¹ 郭鑫¹ 王素格^{1,2}

1 山西大学计算机与信息技术学院 太原 030006

2 山西大学计算智能与中文信息处理教育部重点实验室 太原 030006

(chenqian@sxu.edu.cn)

摘要 情感分类对推荐系统、自动问答、阅读理解等下游应用具有重要应用价值,是自然语言处理领域的重要研究方向。情感分类任务直接依赖于上下文,包括全局和局部信息,而现有的神经网络模型无法同时捕获上下文局部信息和全局信息。文中针对单标记和多标记情感分类任务,提出一种循环卷积注意力模型(LSTM-CNN-ATT,LCA)。该模型利用注意力机制融合卷积神经网络(Convolutional Neural Network,CNN)的局部信息提取能力和循环神经网络(Recurrent Neural Network,RNN)的全局信息提取能力,包括词嵌入层、上下文表示层、卷积层和注意力层。对于多标记情感分类任务,在注意力层上附加主题信息,进一步指导多标记情感倾向的精确提取。在两个单标记数据集上的F1指标达到82.1%,与前沿单标记模型相当;在两个多标记数据集上,小数据集实验结果接近基准模型,大数据集上的F1指标达到78.38%,超过前沿模型,表明LCA模型具有较高的稳定性和较强的通用性。

关键词:循环神经网络;卷积神经网络;注意力机制;情感分类

中图分类号 TP391

Recurrent Convolution Attention Model for Sentiment Classification

CHEN Qian^{1,2}, CHE Miao-miao¹, GUO Xin¹ and WANG Su-ge^{1,2}

1 School of Computer & Information Technology, Shanxi University, Taiyuan 030006, China

2 Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Taiyuan 030006, China

Abstract Sentiment classification has important application value for downstream applications, including recommendation system, automatic question answering and reading comprehension. It is an important research direction in the field of natural language processing. The task of sentiment classification depends on global and local information hidden in context. However, existing neural network models can not capture the local and global information of context at the same time. In this paper, a recurrent convolutional attention model (LSTM-CNN-ATT, LCA) is proposed for single label and multi-label sentiment classification tasks. It uses attention mechanism to fuse the local information extraction ability of convolutional neural network and the global information extraction ability of recurrent neural network, including word embedding layer, context representation layer, convolution layer and attention layer. For the multi-label sentiment classification task, the topic information is added to the attention layer to further guide the accurate extraction of multi-label emotion tendency. The F1 index on two single label datasets reaches 82.1%, which is equivalent to the frontier single label model. On two multi-label datasets, the experimental results on small datasets are close to the benchmark model, and the F1 index on large datasets reaches 78.38%, which is higher than the state-of-the-art model. It indicates that LCA model has high stability and strong universality.

Keywords Recurrent neural network, Convolutional neural network, Attention mechanism, Sentiment classification

1 引言

情感分析是自然语言处理领域中的重要研究任务之一。随着各种社交网络平台的快速发展,越来越多的人喜欢在网络上发表意见和看法。人们发布的内容不仅包括对热点问题

的讨论,还包括对特定商品的评论。从如此庞大的短文本数据集中提取有价值的信息,分析用户的情感倾向,对下游任务包括推荐系统、问答系统、阅读理解等具有重要应用价值。在推荐系统中,情感分析不仅可以方便用户更快、更好地搜索到自己所需的信息,还能方便商家及时了解用户的真实需求,推

收到日期:2020-01-13 返修日期:2020-06-19 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61502288,61403238);山西省基础研究计划项目(201901D111032,201701D221101);山西省重点研发计划项目(201803D421024)

This work was supported by the National Natural Science Foundation of China(61502288,61403238), Natural Science Foundation of Shanxi Province(201901D111032,201701D221101) and Key Research and Development Project of Shanxi Province(201803D421024).

通信作者:郭鑫(guoxinjsj@sxu.edu.cn)

荐更符合用户喜爱的商品。在问答系统中,通常会对问题进行分类,或对答案进行匹配,很多问题都可以利用分类的思想进行处理,如观点分析和序列标注等问题。

深度学习作为机器学习中的一个重要分支,目前已成功地应用于计算机视觉、语音识别和自然语言处理(Natural Language Processing, NLP)领域,并取得了重大突破进展。RNN 和 CNN 是两种常用的文本处理表示和建模方法。RNN 可以对整个序列建模并捕获长期依赖关系^[1]。然而,对整个序列进行整体建模有时会成为一种负担,并且可能会忽略文本中的关键部分^[2]。相比之下, CNN 能够很好地提取局部特征^[3]并具有位置不变特征^[4]。为了实现同时对全局语义信息和文本局部特征进行抽取,并抽取局部信息对类别标签的重要性权重信息,本文提出一种注意力机制融合 RNN 和 CNN 的神经网络模型,在单标记和多标记数据集上对方法模型进行验证,实验结果表明,本文提出的模型具有一定的可行性和有效性。

2 相关工作

文本情感分类的研究方法大致可以分为两类:传统的机器学习方法和神经网络学习方法。特征工程长期以来主导着传统浅层情感分析任务的性能。各种基于规则和统计的特征,如词特征和 N-grams 特征,被用来对文本进行表示,并输入到线性回归模型^[5]和支持向量机^[6]等机器学习模型中进行判断。传统的解决方案存在两大缺陷:1)提取人工特征所需劳动量大;2)表示稀疏性问题。

深度学习在自然语言处理领域得到快速发展,其中多层感知器(Multilayer Perceptron, MLP)^[7]、CNN^[8]、RNN^[9]和注意力机制^[10]等多种神经网络结构被广泛应用于文本情感分析任务,并获得了更好的句子语义表示。Kim^[8]首次利用预训练的词向量结合 CNN 抽取文本表示,并将该文本表示应用于情感分析任务。Tang 等提出利用 LSTM^[11]对句子之间的情感关系进行建模,解决了梯度消失和梯度爆炸问题。Xu 等^[12]提出一种具有缓存机制的 LSTM,用于捕获长期情绪信息。但针对特定目标的情感,其局部特征体现在句中的不同地方, LSTM 无法捕获序列特征的权重区别。为了解决该问题,机器翻译领域广泛使用的注意力机制^[10]被引入情感分析任务中。Tian 等^[13]使用双向 GRU(Gated Recurrent Unit)结合注意力机制,在短文本情感分类任务中取得了较好的效果,证明注意力机制通过权重计算可以识别出文中与该情感类别最相关的词语,进而提升模型的可解释性。Devlin 等^[14]提出的基于深度双向 Transformer 的预训练语言模型 BERT,其模型内部结构是完全基于 Self-Attention 的,证明引入注意力机制更容易捕获句子中远距离的相互依赖的特征。Wang 等结合 LSTM 和注意力机制提出一种 ATAE 方面的情感分析模型,该模型在输入时融合方面编码的信息,在 LSTM 输出的隐藏状态后添加了注意力机制^[15]。Cheng 等提出 HEAT 模型,进一步利用层次注意力机制来捕捉方面信息,以完成句子特定方面的情感分析工作,从而提升方面级情感分析的准确率^[16]。Xue 等提出一种门控的卷积神经网络 GCAE,在卷积过程中加入方面信息,用来针对不同的方面进

行情感分类^[17]。Jiang 等^[18]提出一种细粒度的 LSTM-CNN 注意力分类模型,但 LCA 在以下几点上不同于该模型:首先选择双向 LSTM,以更充分地挖掘了上下文信息;其次, LCA 中 CNN 层在池化之前进行了 attention 操作,能有效保留池化操作丢失的局部信息;最后, LCA 不仅能用于方面情感分析,也适用于一般文本分类任务,从而能够较好地利用 attention 融合 LSTM 远程依赖关系和 CNN 局部特征,有效地提高情感分类的准确性。

尽管不同网络结构相互融合的工作已经开展,但是将 CNN, RNN, Attention 3 种模型架构相结合的模型较少。因此本文针对文本情感分析任务,设计通过 CNN 和 RNN 分别提取局部和全局特征,使用注意力机制根据主题特征重要性将特征进行融合,并在单标记和多标记数据集上验证了模型的性能。

3 循环卷积注意力模型(LCA)

RNN 在整个序列建模过程中能捕获长期依赖关系,得到一个具有全局上下文信息的词表示,其中包含了文本的语义信息以及上下文信息。CNN 具有良好的局部特征抽取能力,可以对上下文表示信息中重要的局部信息进行抽取。但 CNN 中的池化操作会丢失大量有价值的信息,忽略了局部与整体之间的关联性,为了提取全局和局部特征,有效地对局部特征与类别间的关联程度进行建模,本文提出一个循环卷积注意力模型,其中 RNN 采用的是长短期记忆模型(LSTM)。

3.1 模型结构

LCA 模型由 4 层构成,分别是表示层、全局编码层、局部编码层和注意力层。表示层采用词嵌入的方式对文本信息进行预表示,文本 $S = \{w_1, w_2, \dots, w_n\}$, n 为文本中词的个数。将每个词映射到一个低维连续的向量空间 $\mathbf{E}^w \in \mathbb{R}^{v \times d}$, 其中 v 为词表大小, d 为词嵌入的维度。使用预训练的词向量权重初始化 Embedding 层,得到该短文本对应的词向量矩阵 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, $\mathbf{x}_i \in \mathbf{E}^w$ 。常见的预训练词嵌入有 Word2Vec^[19]和 Glove^[20]。本文采用 Glove 作为模型的输入表示。

全局编码层采用 LSTM,能有效捕获上下文语义信息,并将其融合到词嵌入中。因此,我们使用第一层 Bi-LSTM 得到的隐藏层状态作为上下文嵌入的序列编码表示。将前向 LSTM 单元在 w_i 上的隐藏状态表示为 $\overrightarrow{\text{LSTM}}(w_i)$, 将后向 LSTM 的隐藏状态 $\overleftarrow{\text{LSTM}}(w_i)$ 表示拼接作为单词表示,得到上下文信息嵌入的单词表示 $h_i^{(1)} \in \mathbb{R}^{2d}$, 如式(1)所示:

$$h_i^{(1)} = [\overrightarrow{\text{LSTM}}(w_i); \overleftarrow{\text{LSTM}}(w_i)], i \in [1, n] \quad (1)$$

局部编码层采用 CNN,将得到的 LSTM 层表示输入到卷积层进行卷积操作,对每个大小为 k 的窗口进行卷积,计算结果如式(2)所示:

$$c_i = g(\mathbf{W} \times \mathbf{x}_{i:i+k-1} + b) \quad (2)$$

其中, c_i 表示特征映射中第 i 个特征值; g 表示非线性激活函数,本文采用双曲正切函数 \tanh ; b 表示偏置量; \mathbf{W} 表示滤波器中权重矩阵; $\mathbf{x}_{i:i+k-1}$ 是文本句子中第 i 个词到 $i+k-1$ 个词的长度。本文模型使用多个滤波器进行特征学习,通过卷积操作后得到的句子特征矩阵 $\mathbf{C} \in \mathbb{R}^{n-k+1}$, 如式(3)所示:

$$\mathbf{C} = [c_1, c_2, \dots, c_{n-k+1}] \quad (3)$$

标准的 Bi-LSTM 或 CNN 模型并不能识别出文本不同部分对于情感分类的不同贡献。引入注意力层用于捕获句子中不同位置对特定类别的情感贡献程度,其输入数据来源于式(3),对于单标记分类任务,使用一个随机的向量 \mathbf{h} 作为注意力机制的 Query,用来计算每个卷积窗口的权重 α 。其权重指在推断文本情感极性时,卷积窗口被正确注意到的概率。 α_i 通过一个注意力模型计算得到,其核心在于通过一个得分函数将特征表示 c_i 和注意力向量表示 Query 联系起来,通过 softmax 函数将其归一化得到概率分布。注意力模型结构如式(4)和式(5)所示:

$$\alpha_i = \frac{\exp(\text{score}(\mathbf{h}, \mathbf{c}_i))}{\sum_{i=1}^n \exp(\text{score}(\mathbf{h}, \mathbf{c}_i))} \quad (4)$$

$$\mathbf{z}_s = \sum_{i=1}^n \alpha_i \mathbf{c}_i \quad (5)$$

常用的 score 函数计算方法有 3 种:点积、二次型和拼接^[21]。本文采用点积方式,即 $\text{score}(\mathbf{h}, \mathbf{c}_i) = \langle \mathbf{h}, \mathbf{c}_i \rangle$ 。接着 CNN 层的池化特征和注意力的融合特征进行拼接,将拼接后的句子表示送入 softmax 分类器中。LCA 模型的整体结构如图 1 所示。图 1 的左部分用于处理多标记分类任务,右部分用于处理单标记分类任务。

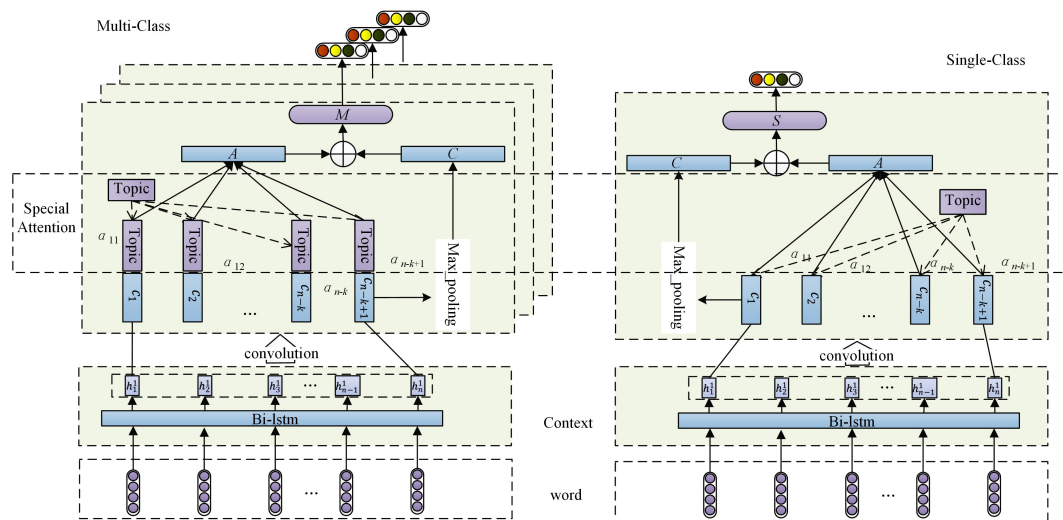


图 1 LCA 模型结构示例图

Fig. 1 LCA model structure diagram

对于多标记分类任务,文本的主题类别一般与特定词语有着很强的关联性。例如,我们借助主题类别来判断文本在某个主题类别下的情感,或判断这个主题类别是否存在。为了利用这个主题信息,我们在每个卷积的特征中加入主题向量 T_i ,并使用主题向量作为查询,使用注意力机制对与主题相关的信息进行抽取。本文将此多标记问题转化为多个单标记分类问题,使用多个分类器分别对每个主题进行预测。具体操作如式(6)、式(7)所示:

$$\alpha_i = \frac{\exp(\text{score}(T_i, (c_j; T_i)))}{\sum_{i=1}^n \exp(\text{score}(T_i, (c_j; T_i)))} \quad (6)$$

$$\mathbf{z}_s = \sum_{i=1}^n \alpha_i (c_i; T_i) \quad (7)$$

其中,“;”为拼接操作。最后,针对每个主题 T_i 会得到一个表示 \mathbf{z}_s ,分别用于判断每个标签。

3.2 模型训练

对于单标记数据集,模型训练采用 Adam 优化器。通过最小化交叉熵损失函数来优化网络模型,完成情感分类任务,交叉熵损失函数为:

$$L = - \sum_{i=1}^N y_i \log(y_i) \quad (8)$$

其中, N 为训练样本个数, y 为预测类别, \hat{y} 为实际类别。

对于多标记数据集,我们将注意力机制中的随机向量替换成与类别相关的主题向量,生成的注意力表示即为针对类别主题生成的表示,分别与卷积层的 Max-pooling 特征进行拼接,为每个类别生成一个类别表示,再输入全连接层进行分

类。最小化交叉熵损失函数对模型进行优化训练,交叉熵损失函数如下:

$$L = - \sum_{j=1}^N \sum_{i=1}^M \hat{y}_{ij} \log(y_{ij}) \quad (9)$$

其中, \hat{y}_{ij} 为样本 j 在类别 i 下的真实标签, y_{ij} 为预测标签。

4 实验设计与结论

4.1 数据集

本文在两个不同领域的单标记公开数据集和两个多标记数据集中进行实验,以验证所提模型的效果。

(1)Movie Review(MR):MR 是一个电影情感分析的数据集评论。它仅包含单个标记,分为积极和消极两类。

(2)Stanford Sentiment Treebank with Five Labels (SST-5):SST-5 是包含单标记的斯坦福情绪树库,其中包含 5 类情绪:非常消极的、消极的、中性的、积极的和非常积极的。

(3)SemEval 2014 Restaurant:将 SemEval Restaurant 2014 数据集进行简单处理,使得其转化为一个多标记的数据集。文本中涉及的几个主题方面包含 food,price,service,ambience 和 misc,每个主题具有 4 种类标签:-2,-1,0,1,分别代表未提及、负面、中性和正面情感。

(4)SemEval 2014 Restaurant Large:通过合并 2014-2016 年 3 年的 Restaurant 评论数据集。使用与数据集(3)相同的处理方式得到了一个更大的数据集,称为 Restaurant-Large,包含 restaurant,food,drinks,ambience,service,price,misc 和 location 8 个主题方面。

SemEval 数据集为方面级分析数据集,为了对本文提出多标记模型进行验证,我们对 SemEval Restaurant 2014 数据集进行合并,每条数据会涉及多个方面类别,且每个方面类别涉及 3 种情感:-1(负向)、0(中立)、1(正向)。并不是每一个样本中都包含所有主题,但每个样本都至少涉及一个主题输入,因此引入一个主题向量来预测某一主题类别下的情感。以上 4 种数据集的统计信息如表 1 所列。其中, *No. label* 代表标签类别的个数, *Max len.* 代表最大句长, *Train* 代表训练样本个数, *Test* 代表测试集个数, *CV* 代表采用交叉验证方式切分标注数据集。此外,对 4 个数据集进行正负例统计,发现 MR 数据集正负例比例为 1:1, SST-5 的 5 级情绪样本比例也较均衡;对于多标记数据集,Restaurant 样本的方面主题比例极不均衡,表现得尤为突出。

表 1 4 种数据集统计信息

Table 1 Statistics information of four datasets

Dataset	No. label	Max len.	Train	Test	Proportions
MR	1	56	10662	CV	1:1(n:p)
SST-5	1	53	11855	2210	13:26:19:27:15 (-2:-1:0:1:2)
Restaurant 2014	5	69	3518	973	33:9:16:11:31
Restaurant-Large	8	69	4656	2426	11.5:0.8:24.1:5.3: 0.3:17.2:6.9

4.2 实验设计

对于二分类问题,采用标准 F1 作为评测指标;对于多分类问题,由于 Macro-F1 受样本数量少的影响较大,因此采用 Micro-F1 值作为本文实验结果的评测指标。具体计算公式如式(10)和式(11)所示:

$$P_{micro} = \frac{\bar{TP}}{\bar{TP} + \bar{FP}}, R_{micro} = \frac{\bar{TP}}{\bar{TP} + \bar{FN}} \quad (10)$$

$$F_{1-micro} = \frac{2 \cdot P_{micro} \cdot R_{micro}}{P_{micro} + R_{micro}} \quad (11)$$

其中, \bar{TP} , \bar{FP} , \bar{FN} 分别代表全部标签的 True Positive 平均实例个数、False Positive 平均实例个数和 False Negative 平均实例个数。

4.3 实验结果分析

词向量选用 300 维 GloVe 词向量;由于数据量较小, LSTM 的隐层维度设置为 100 维,以便更好地优化词表示;优化器为常用的 Adam 优化器;dropout 设为 0.5。实验结果采用十折交叉验证,为确保模型的可重复性,设置随机数种子为 10,学习率为 0.001,批量大小为 128。为了验证所提方法在单标记文本分类任务上的有效性,在 MR 和 SST-5 数据集上我们将 LCA 模型同以下模型进行了对比: Standard-RNN^[22], Bi-LSTM^[21], CNN^[8] 和 Tree-LSTM^[21]。此外,还重现了以下前沿模型。

DSCNN^[23]: DSCNN 通过 LSTM 对预先训练好的词嵌入进行处理,然后利用卷积算子提取特征,分层构建文本表示。与 Tree-LSTM 相比, DSCNN 不依赖于解析器。

ID-LSTM(Information Distilled LSTM)^[24]: 其使用的基础模型是基于 LSTM 的,针对不同时刻会有 Retain 和 Delete 两种不同的动作。

HS-LSTM^[24]: 其使用的模型是 2 个层次化的 LSTM,一个用来将 word 转化成 phrase 向量,另一个将生成的 phrase

向量转化成 sentence 向量。

MVCNN^[25]: 其是 Yin 提出的一个多通道卷积神经网络。它结合了不同版本的预先训练的词嵌入,使用可变大小的卷积滤波器进行不同粒度的短语特征提取。

由表 2 可以看出, HS-LSTM 模型在 MR 数据集上取得了较好的结果,但是在 SST-5 数据中表现不佳。而本文模型在两个数据集上都取得了较高的准确率,并且在 MR 上的性能与 HS-LSTM 相当,具有较强的泛化能力。表明 LCA 能够捕获全局的语义和局部的结构表示,以及其包含的注意力机制具有良好的特征融合能力,自适应地调整多个特征表示的权重,使得模型拥有较好的性能。

表 2 MR 和 SST-5 数据集上的比较实验

Table 2 Comparative experiments on MR and SST-5 datasets

Model	MR	SST-5
Standard-RNN	77.4	46.4
bi-LSTM	79.7	49.1
CNN	81.5	48
MVCNN	—	49.6
Tree-LSTM	80.7	50.1
DSCNN	81.5	49.7
ID-LSTM	81.6	50.0
HS-LSTM	82.1	49.8
LCA	82.1	50.5

对于多标记模型,其与单标记模型的区别在于注意力层,本文将多标记中的主题信息与注意力层融合,为验证算法的有效性,选取了细粒度方面级情感分析模型进行对比。方面级情感分析任务是给定“句子-方面”对,对给定句子的每个方面进行分类。分别在 Restaurant 2014 和 Restaurant-Large 数据集上设计了对比实验,对比模型包括 ATAE^[15], HEAT^[16] 和 GCAE^[17],评价标准为 4.2 节提出的 micro-F1 值,取 5 次实验的均值和方差,实验结果如表 3 所列。

表 3 Restaurant 2014 和 Restaurant-Large 数据集上的比较实验

Table 3 Comparative experiments on Restaurant 2014 and

Restaurant-Large datasets

Model	Restaurant 2014	Restaurant-Large
ATAE	80.41±0.46	76.06±0.40
HEAT	81.08±0.62	78.03±0.23
GCAE	79.97±0.48	78.02±0.26
LCA	80.72±0.19	78.38±0.17

由表 3 可知,在 Restaurant 2014 数据集上,由于数据集规模较小,模型训练不充分,多数模型表现都不太稳定。在 Restaurant-Large 数据集上, LCA 模型的起伏不超过 0.17。此外,多标记数据集样本类别极不均衡,随着数据集的增大, LCA 模型在两个多标记数据集上的稳定性均优于其他模型,在大数据集上的 micro-F1 性能同样优于其他模型。

结束语 本文通过融合各级特征,对单标记和多标记两个文本分类任务进行研究,提出了一种循环卷积注意力网络 LCA 模型,其通过利用循环神经网络的上下文信息、卷积神经网络的局部信息和注意力机制的特征融合能力,较好地提取了文本的特征信息,并成功应用于文本情感分析任务。单标记分类模型的注意力层采用随机向量,而多标记的注意力查询项采用主题向量,通过两种不同的注意力方式,对文本进行建模。最终,在单标记情感分析任务中, LCA 模型在精度上优于已有模型;在多标记情感分析任务中, LCA 性能接近

已有的模型,在大规模数据集上表现最优,同时稳定性最佳,且比其他模型更适合非均衡数据。下一步工作将考虑各层信息融合,而不仅仅是利用最终注意力和卷积表示。另外,新颖的模型框架以及文本信息表示方法也是下一步研究的重点。

参 考 文 献

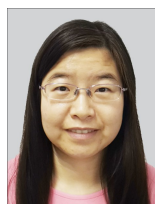
- [1] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv:1412.3555,2014.
- [2] YIN W, KANN K, YU M, et al. Comparative study of cnn and rnn for natural language processing [J]. arXiv:1702.01923, 2017.
- [3] DOMINIK S, ANDREAS M, SVEN B. Evaluation of pooling operations in convolutional architectures for object recognition [C]//International Conference on Artificial Neural Networks. 2010:92-101.
- [4] RONAN C, JASON W, LEON B, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12(1): 2493-2537.
- [5] ZHU J, HASTIE T. Kernel logistic regression and the import vector machine[C]// International Conference on Neural Information Processing Systems: Natural & Synthetic. 2001: 1081-1088.
- [6] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [7] LI Y, WEI B, LIU Y, et al. Incorporating knowledge into neural network for text representation[J]. Expert Systems with Applications, 2018, 96(4): 103-114.
- [8] KIM Y. Convolutional neural networks for sentence classification [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1746-1751.
- [9] ALEX G. Generating Sequences with Recurrent Neural Networks[J]. arXiv:1308.0850,2013.
- [10] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473,2016.
- [11] TANG D, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification [C] // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015:1422-1432.
- [12] XU J, CHEN D, QIU X, et al. Cached long short-term memory neural networks for document-level sentiment classification [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:1660-1669.
- [13] TIAN Z, RONG W, SHI L, et al. Attention Aware Bidirectional Gated Recurrent Unit Based Framework for Sentiment Analysis [C]//Proceedings of the 2018 Conference of the International Conference on Knowledge Science, Engineering and Management. Springer. Cham, 2018: 67-78.
- [14] DELVIN J, CHANG M W, LEE K, et al. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding [C]//Proceedings of the 2019 Conference of the Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2019:4171-4186.
- [15] WANG Y, HUANG M, ZHAO L, et al. Attention-based lstm

for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:606-615.

- [16] CHENG J, ZHAO S, ZHANG J, et al. Aspect-level sentiment classification with HEAT (hierarchical attention) network [C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017:97-106.
- [17] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 2514-2523.
- [18] JIANG M, ZHANG W, ZHANG M, et al. An LSTM-CNN attention approach for aspect-level sentiment classification[J]. Journal of Computational Methods in Sciences and Engineering, 2019, (19): 859-868.
- [19] QUOC L, TOMAS M. Distributed representations of sentences and documents [C] // International Conference on Machine Learning. 2014: 1188-1196.
- [20] JEFFREY P, RICHARD S, CHRISTOPHER M. Glove: Global vectors for word representation[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1532-1543.
- [21] KAI S, RICHARD S, CHRISTOPHER M. Improved semantic representations from tree-structured long short-term memory networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Association for Computational Linguistics (ACL). 2015:1556-1566.
- [22] RICHARD S, ALEX P, JEAN W, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C] // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013:1631-1642.
- [23] ZHANG R, HONGLAK L, DRAGOMIR R. Dependency sensitive convolutional neural networks for modeling sentences and documents[J]. arXiv:1611.02361,2016.
- [24] ZHANG T, HUANG M, ZHAO L. Learning structured representation for text classification via reinforcement learning[C]// proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA, 2018.
- [25] YIN W, SCHUTZE H. Multichannel variable-size convolution for sentence classification[J]. arXiv: 603.04513,2016.



CHEN Qian, born in 1983, Ph.D, associate professor, master supervisor, is a member of China Computer Federation. His main research interests include topic detection and evolution, machine reading comprehension and natural language processing.



GUO Xin, born in 1982, Ph.D, associate professor, master supervisor, is a member of China Computer Federation. Her main research interests include feature learning and natural language processing.