

引用格式:黄磊,杜昌顺. 融合群稀疏与排他性稀疏正则项的神经网络压缩情感分析方法[J]. 北京化工大学学报(自然科学版), 2019, 46(2): 103-112.

HUANG Lei, DU ChangShun. A neural network compressed sentiment analysis method based on combined group and exclusive sparsity regularization[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2019, 46(2): 103-112.

融合群稀疏与排他性稀疏正则项的神经网络 压缩情感分析方法

黄磊 杜昌顺*

(北京交通大学 经济管理学院, 北京 100044)

摘要: 文本情感分析是目前网络环境下舆情监控、服务评价及满意度分析等领域的重要任务,一些基于深度神经网络的方法已被用于此类任务。规模庞大的深度神经网络模型结构赋予了深度学习模型强大的非线性拟合的能力,大规模的数据资源为训练这样大规模的模型并保证其泛化能力提供了可能性。然而,在实际应用中,深度模型的时间和空间开销仍然制约着这些方法的落地。针对上述问题,提出一种融合群稀疏与排他性稀疏正则项的神经网络压缩情感分析方法,首先分别构建循环-卷积神经网络与卷积-循环神经网络,通过门控单元融合两种网络组成的分析模型,在模型中引入群稀疏与排他性稀疏正则项,剪除冗余神经元或链接,压缩模型规模。在不同数据集上的实验结果验证了本文方法的有效性。

关键词: 情感分析; 卷积神经网络; 循环神经网络; 门控单元; 模型压缩

中图分类号: TP391 **DOI:** 10.13543/j.bhxbzr.2019.02.016

引言

情感分析的基本任务是在文档、句子或特征/方面级别对给定文本的极性进行分类,并确定文档、句子或实体特征/方面中表达的意见是积极的、消极的还是中立的,而文本信息的复杂性导致检测普通文本中的情绪成为一项具有挑战性的任务。目前效果较好的情感分析算法都是基于统计学习的方法,该方式性能的好坏取决于特征提取的质量,而良好的特征工程需要较高的专家经验且费时费力,可迁移性较差;采用神经网络的方法能够减少对特征工程的依赖。

当前,有一些基于神经网络的方法已经被用于情感分类任务。Socher等^[1-5]提出利用递归神经网络(recursive neural network)进行建模,这些方法已被证明在构造句子的表示方面是有效的。然而,递

归神经网络首先需要将文本构造为一个树结构的数据结构表示方式,以便捕获句子的语义,此类模型应用于较长的句子或文档时,需要较大的时间开销。

在自然语言处理任务的应用上最为有效的循环神经网络(recurrent neural network, RNN)具有良好的捕获上下文信息的能力,这对长文本进行建模进而获取其语义十分有用。然而,循环神经网络是一个有偏的模型,越接近展开末尾的单词比之前的单词保留的信息更多。众所周知,与情感分析有关的关键特征可能出现在文档的任何位置,而不仅仅是在文本的末尾。因此,使用循环神经网络捕获整个输入文本的语义特征时,其有效性将会因关键特征的位置改变而出现不同程度的降低,甚至可能会将重要信息完全忽略,导致模型的性能极大退化。

同时,还有一些工作使用卷积神经网络(convolution neural network, CNN)进行情感分类^[6]。CNN被引入自然语言处理任务中,一个原因在于它能够解决词语等特征的顺序偏差问题,因为在使用最大值池化等池化操作时,其具有平移不变性,与特征所在的位置无关。然而,之前关于CNN的研究工作倾

收稿日期: 2018-12-20

第一作者: 男, 1965年生, 教授

* 通信联系人

E-mail: summer2015@bjtu.edu.cn

向于使用诸如固定窗口大小的简单卷积核^[7-8]。小窗口可能导致重要信息丢失,而大窗口则会导致巨大的参数空间,使得网络的训练变得极为困难。同时,基于 CNN 的文本分析方法能够通过池化获得文本的重要特征,但较难获得上下文信息。Cliche^[9]使用卷积神经网络和长短期记忆网络分别对句子进行建模,虽然该方法可以在实验结果上得到一定的提升,但仍然无法克服卷积神经网络和循环神经网络的上述缺陷。

深度网络模型在细分领域的情感分析任务上与传统方法相比取得了较大提升。概括来说,就是规模庞大的深度神经网络模型结构赋予了深度学习模型强大的非线性拟合的能力,而大规模的数据资源为训练这样大规模的模型并保证其泛化能力提供了可能性。然而,在实际应用中,较大的时间和空间开销仍然制约着这些方法的实施。层次较多且每一层神经元众多的大规模深度网络模型不论是训练还是预测,运算量都十分巨大,即使借助图形处理器(graphics processing unit, GPU)进行加速计算^[6],其较差的时效性也很难满足许多应用场景的需求。同时,对于大规模的深度学习模型,即使是卷积神经网络等使用了参数共享的模型,模型参数也要占用大量的内存空间,在实际运用中对设备的要求很高,对于手机等移动设备来说更加难以部署。因此,如何在不影响深度神经网络模型性能的前提下,压缩深度神经网络模型具有重要意义。

基于神经网络的情感分析方法为了保证模型能够取得良好的效果,使用了较大规模的卷积神经网络,导致模型的参数规模较大。而在实际运用中,带有情感标注信息的细分领域文本数据较少,存在数据问题,这就可能导致上述深度神经网络模型得不

到较为充分的训练。一方面,模型存在过参数化问题;另一方面,参数得不到充分训练,模型的泛化性能就难以保证。同时,为了能够快速分析出现的舆情,及时应对客户的反馈,实现高效客情关系管理,情感分析系统需要具有较高的时效性。

为了解决上述问题,提升网络预测时的效率,同时保证模型的性能不会有较大的下降,本文提出了基于门控单元的循环-卷积神经网络(RCNN)与卷积-循环神经网络(C-RNN)集成情感分析方法^[10-12],并利用剪枝的方法对本文提出的模型进行压缩。在卷积神经网络(包括全连接神经网络)的训练过程中,利用群稀疏与排他性稀疏正则项对模型进行剪裁,从而达到减少模型参数数目的效果。在模型预训练过程中,首先利用稀疏性正则项剪除权值较小的边,并去除连接稀疏的神经元节点,然后继续训练剪枝后的模型。在不同的数据集上进行了充分实验,验证了压缩方法的有效性,且保证了模型在情感分析任务上的性能。

1 基于门控单元的 RCNN 与 C-RNN 集成情感分析方法

1.1 基于 RCNN 的文本情感特征提取模型

首先,本文提出一种基于循环-卷积神经网络的深度神经模型来捕获文本的语义,将获得的语义特征作为情感分析器的特征进行情感倾向分析。图 1 显示了该模型的网络结构。网络的输入是文本 S , 由文本的单词序列 w_1, w_2, \dots, w_n 构成。

1.1.1 单词的语义特征提取

文本中每一个语言符号所表达的意义都与它所处的上下文环境有关,单词的上下文背景能够帮助模型获得词语在特定场景下更加精确的意义。为了

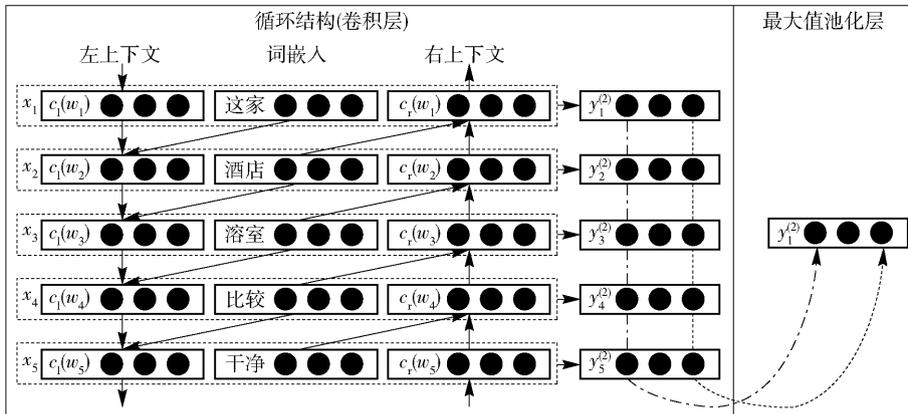


图 1 循环-卷积神经网络结构示意图

Fig. 1 Structure of the recurrent-convolutional neural network

使得神经网络提取到的语义特征能够融合单词的上下文环境,利用循环-卷积神经网络模型来对客户的评论文本进行特征提取。模型首先使用一个双向循环神经网络,来捕捉词语的上下文信息并结合单词本身的词嵌入表示来获得单词的强化表示。定义 $\mathbf{c}_l(w_i)$ 、 $\mathbf{c}_r(w_i)$ 分别表示单词 w_i 的左侧上下文信息(循环由前向后展开)特征以及右侧上下文信息(由后向前展开)特征, $\mathbf{c}_l(w_i)$ 与 $\mathbf{c}_r(w_i)$ 都是维度为 c 的实值稠密向量,计算过程分别如式(1)、(2)

$$\mathbf{c}_l(w_i) = f(\mathbf{W}^l \mathbf{c}_l(w_{i-1}) + \mathbf{W}^{sl} \mathbf{e}(w_{i-1})) \quad (1)$$

$$\mathbf{c}_r(w_i) = f(\mathbf{W}^r \mathbf{c}_r(w_{i+1}) + \mathbf{W}^{sr} \mathbf{e}(w_{i+1})) \quad (2)$$

式中, $\mathbf{e}(w_i)$ 为输入文本序列中第 i 个单词 w_i 的词嵌入表示; $\mathbf{c}_l(w_{i-1})$ 为前一个单词 w_{i-1} 的左侧上下文信息; \mathbf{W}^l 是将双向循环神经网络的前向展开的隐藏层状态(左上下文信息特征)转换为下一个隐藏层状态的参数矩阵; \mathbf{W}^{sl} 是将当前单词的词嵌入表示添加到下一个单词的左上下文信息特征的参数矩阵; f 为非线性激活函数。右上下文信息特征 $\mathbf{c}_r(w_i)$ 的计算方式与之类似。利用式(1)、(2)对文本序列进行计算后,当前单词的左侧上下文信息特征捕获了由前向后展开至单词的所有语义信息,右侧上下文信息特征则捕获了由后向前展开至单词的所有语义信息。在获得当前单词的上下文语义特征后,将上下文语义特征与单词本身的词嵌入表示融合,获得当前单词的最终表示,其形式化表示为

$$\mathbf{x}_i = [\mathbf{c}_l(w_i); \mathbf{e}(w_i); \mathbf{c}_r(w_i)] \quad (3)$$

如式(3)所示,单词 w_i 的最终表示 \mathbf{x}_i 由左侧上下文向量 $\mathbf{c}_l(w_i)$ 、单词的词嵌入 $\mathbf{e}(w_i)$ 和右侧上下文向量 $\mathbf{c}_r(w_i)$ 拼接组成。通过这种方式,模型学习到的单词表示中蕴含了丰富的单词上下文信息,与仅使用固定窗口大小的传统语言模型或神经网络模型相比(这些模型对每一个语言符号进行处理时,仅使用单词周围的部分文本信息),循环-卷积神经网络模型可以对整个文本序列进行建模并很好地保留词语的顺序等结构信息,利用上下文信息能够更好地消除词语 w_i 含义的歧义;同时,循环神经网络结构正向扫描获得所有左侧上下文语义特征 \mathbf{c}_l ,后向扫描获得所有右侧上下文特征 \mathbf{c}_r 。

1.1.2 从单词语义特征中获取全文的语义特征

直接使用文本中单个词语的表示对于分析整个评论的情感倾向仍然十分困难,因此还需要利用这些单词的表示计算获取整个评论文本的表示。从卷积神经网络的角度看,可以将循环神经网络结构看

作是卷积神经网络的卷积层。当计算出所有单词的表示时,利用最大值池化操作从评论文本序列的所有单词表示中获取全文的语义特征,计算过程如式(4)所示

$$y_i^{(3)} = \max_{i=1}^{|\mathcal{S}|} y_i^{(2)} \quad (4)$$

在循环-卷积神经网络模型中,最大值池化操作是对整个评论文本所有的词语表示逐语义因素提取最大值,即选取所有词语语义特征 $y_i^{(2)}$ 第 k 维的最大值作为全文语义特征 $y_i^{(3)}$ 第 k 维语义元素的值。经过最大值池化操作后,长度不同的评论文本转换为固定长度的全文语义向量,该向量能够较好地保留整个文本的重要语义信息。如前文所述,在整个评论中,某些与情感相关的单词及其组合语义是与情感分析任务关联性最强的特征,对于判断评论的情感倾向十分有效,提取这些重要特征可以提升情感分析模型的性能,因此在本节的模型中,依然采用最大值池化操作。

1.2 基于 C-RNN 的文本情感特征提取模型

1.2.1 卷积操作分段提取评论文本局部情感特征

标准的卷积神经网络能够很好地提取词语等的组合特征,且具有平移不变性,无论这些组合特征在文本的何种位置,卷积操作都能够提取到。在本文提出的模型中,输入是由文本的词嵌入构成的矩阵,矩阵的每一行是一个语言符号的向量化表示,卷积核沿着文本展开的方向滑动,令输入中第 i 个语言符号的向量化表示为 \mathbf{s}_i ,输入的文本可以由矩阵 $\mathbf{S} = (\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_{|\mathcal{S}|}^T)$ 来表示。则卷积操作可以表示为

$$\mathbf{c}^j = \mathbf{W} \otimes \mathbf{S}_{j:j+w-1} \quad (5)$$

式中, $1 \leq j \leq |\mathcal{S}| - w + 1$, \mathbf{c}^j 为卷积核从词 j 开始到卷积核高度为窗口的词之间进行卷积操作提取的特征值。

仅依靠一个卷积核对文本进行特征提取获得的有效特征是不全面的,为了能够从文本中提取更加丰富的信息,本文模型使用多个不同的卷积核。

输入的文本经过第 i 个卷积核卷积操作后得到特征向量 $\mathbf{c}_i = \{c_i^1, c_i^2, \dots, c_i^{|\mathcal{S}| - w + 1}\}$,则所有的卷积核共能得到 n 个特征向量。在本节提出的模型中,所有的卷积核都采用同样的大小,因此同一个评论文本经过卷积操作后将会得到 n 个维度相同的蕴含着评论文本局部语义特征的向量。这些特征向量相同的维度可以看作是评论中同一个局部的不同类型的特征,也就是每一个局部语义特征 $\hat{\mathbf{e}}_i$ 都由 n 个特征

组合而成,对于不同局部的语义特征这 n 个特征强度可能有所不同。与传统的卷积神经网络模型不同,本节的卷积神经网络模型不再使用池化层,而是将卷积操作后的特征向量拼接为序列特征输出到下一层,如图 2 所示。

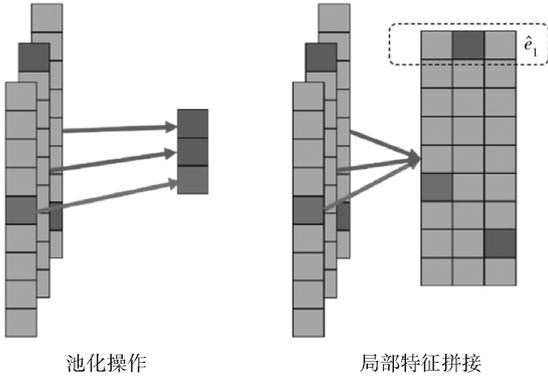


图 2 局部特征拼接示意图

Fig. 2 Local feature concatenating

1.2.2 循环神经网络融合文本句子结构特征

在利用卷积层获得评论文本的局部特征后,将局部特征看作文本序列输入到循环神经网络,可以对评论文本的长距离依赖建模,在文本的局部特征中融入句子的结构特征。

本文使用基于长短时记忆网络计算节点的双向循环神经网络对拼接得到的特征序列进行文本的表示学习,然后将两个方向学习到的特征向量拼接在一起,作为文本的向量表示。相对于单向的循环神经网络,其特征向量所表示的语义更加全面和丰富。令拼接后得到的第 t 个局部特征表示为 \hat{e}_t ,则经过循环神经网络融入文本句子结构特征的单向计算过程可以表示如下。

评论文本 S 的局部特征前向展开经过循环神经网络后得到融入前向句子结构特征的隐藏层状态 \vec{h}_s ,后向展开得到融入后向句子结构特征的隐藏层状态 \overleftarrow{h}_s ,拼接后得到代表输入的评论文本的最终特征向量 h_s ,即 $h_s = [\vec{h}_s, \overleftarrow{h}_s]$ 。

因此,本文提出的基于卷积-循环神经网络的文本情感特征提取模型总体结构可以表示为图 3。

1.3 基于门控单元的 RCNN 与 C-RNN 集成情感分析模型

经过循环-卷积神经网络和卷积-循环神经网络分别提取评论文本的情感特征后,需要将两个网络提取的特征进行融合集成,输入到情感倾向分析器,得到最终的情感分析结果。

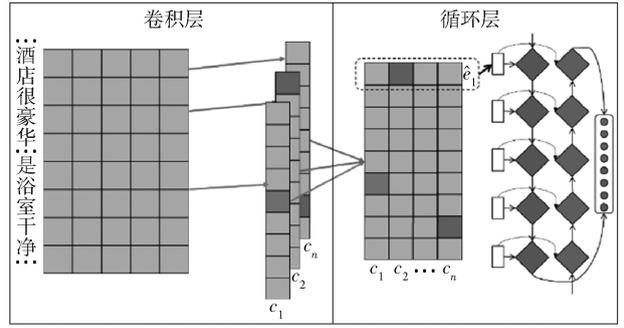


图 3 基于卷积-循环神经网络的文本情感特征提取模型示意图

Fig. 3 Text sentiment analysis model based on C-RNN

对于不同细分领域的评论,与情感倾向有关的特征会有区别,也就是情感倾向具有领域依赖性;同一个细分领域的评论文本在行文风格上具有差异性,因此情感倾向特征也会存在一定程度上的差异。本文提出的两种循环神经网络与卷积神经网络的组合模型提取到的特征较为丰富,但是在不同领域,相同的特征对情感倾向的判别具有不同的作用,过多的特征反而会给情感倾向判别引入噪声,使得分析器的性能发生退化。然而,依靠人工对这些特征进行筛选是十分困难的,因此需要一种自动的方法对这些特征加以选择,使其具有领域、文本适应性,可以去除冗余信息,保证情感倾向分析器的有效性。

本文提出利用门控单元自动对 RCNN 与 C-RNN 提取的两种特征进行筛选融合,融合过程可以表示为图 4。

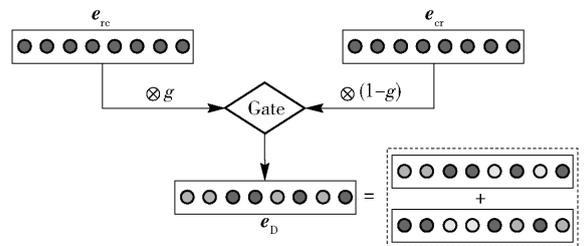


图 4 门控单元融合 RCNN 与 C-RNN 提取特征示意图

Fig. 4 Extract features with combined RCNN and C-RNN based on a gate unit

门控单元根据当前输入的两中特征计算得到有多少比例的特征能够通过融合门控单元,也就是输出的特征中两种特征分别需要保留多少原始特征。为了表示的方便,令循环-卷积神经网络提取的特征表示为 e_{rc} ,卷积-循环神经网络提取的特征表示为 e_{cr} ,则门控单元对特征通过的控制比例 g 的计算公式为

$$g = \sigma(U^g e_{rc} + W^g e_{cr} + b^g) \tag{6}$$

式中, \mathbf{U}^g 与 \mathbf{W}^g 分别为 \mathbf{e}_{rc} 与 \mathbf{e}_{cr} 连接门控单元的权重参数矩阵, \mathbf{b}^g 为偏置。为了简化计算, 门控单元对两种提取到的特征使用同一个控制比例进行特征筛选, 则特征的融合计算过程可以表示为

$$\mathbf{e}_D = g \otimes \mathbf{e}_{rc} + (1 - g) \otimes \mathbf{e}_{cr} \quad (7)$$

式中, \mathbf{e}_D 为经过门控单元融合两种特征得到的表征输入的评论文本情感特征的特征向量。

在得到评论文本的情感特征 \mathbf{e}_D 后, 利用 softmax 分类器基于该特征向量对评论的情感倾向进行分类, 获得评论的情感倾向。

2 融合群稀疏与排他性稀疏正则项的情感分析模型压缩

对深度模型压缩的主要目标是实现一个参数数量明显少于原始非稀疏网络的稀疏深度神经网络, 同时获得与原始模型相当甚至更好的性能。为了将重点放在对模型的压缩上, 本节将提出的情感分析神经网络模型的训练目标表示为

$$\min_{\{\mathbf{W}\}} \mathcal{L}(\{\mathbf{W}, D\}) + \lambda \sum_{l=1}^L \Omega(\mathbf{W}^{(l)}) \quad (8)$$

式中, D 为带有情感类别标记的数据集, $\{\mathbf{W}^{(l)}\}$ 为所有隐藏层参数的集合, $\mathcal{L}(\{\mathbf{W}, D\})$ 表示模型的情感分类损失, L 表示使用模型压缩的隐藏层数目, $\mathbf{W}^{(l)}$ 表示 l 隐藏层的参数, 则 $\Omega(\mathbf{W}^{(l)})$ 表示作用在这些隐藏层的正则项罚函数, λ 为调节情感分类损失与正则项之间的平衡参数。

在以往的机器学习方法中, 最常用的正则项是 L1 范数和 L2 范数正则项, 也即

$$\Omega(\mathbf{W}^{(l)}) = \|\mathbf{W}^{(l)}\|_1 \quad (9)$$

$$\Omega(\mathbf{W}^{(l)}) = \|\mathbf{W}^{(l)}\|_2^2 \quad (10)$$

L2 范数正则化能够通过增加一个偏差项, 限制模型参数的大小来减少模型的方差, 从而降低泛化误差, 然而 L2 范数很难使模型的参数达到稀疏化的效果; 当目标函数带有 L1 范数正则项时, 只有目标函数与 L1 范数在 L1 范数的顶点处相交, 整体的目标函数值才能达到最优。因此 L1 范数正则化能够得到一个稀疏权值矩阵, 也即产生稀疏模型。当大多数特性与目标无关时, L1 范数带来的稀疏性是十分有效的, 可以消除神经网络提取的与情感分析不相关的冗余特征。虽然 L1 范数正则化能够减少模型的参数从而节约存储空间, 但是通常不会在实际的网络架构(如 CNN)中带来有意义的速度提升, 因为卷积神经网络的速度瓶颈在于卷积操作, 当卷

积操作的滤波器数量与未采用 L1 范数正则项的方法相同时, 神经网络的运算数量不会有明显的减少。

因此, 本文结合 L1 与 L2 范数正则的特点, 采用群稀疏化对网络进行剪枝, 达到压缩神经元规模的效果。

群体稀疏性可以通过整体消除神经元或卷积滤波器来帮助降低模型的内在复杂性, 从而在深层神经网络中获得实际的加速^[13-14]。因此本文采用定义如下的组稀疏正则化进行神经元的减除

$$\Omega(\mathbf{W}^{(l)}) = \sum_g \|\mathbf{W}_g^{(l)}\|_2 = \sum_g \sqrt{\sum_i \mathbf{w}_{g,i}^{(l)2}} \quad (11)$$

式中, $g \in \mathcal{G}$ 为网络参数群, $\mathbf{W}_g^{(l)}$ 为某一群参数, i 为参数群中的第 i 个参数的索引。

上述稀疏正则项可以看作是先在群内利用 L2 范数正则约束, 再对不同的群利用 L1 范数正则约束。由于 L2 范数具有分组效应, 可以使得群内相关特征的权重参数相似, 而 L1 范数会驱动群之间获得稀疏的结果, 这将导致某些组的参数整体趋近于 0, 从而移除一些输入神经元。因此, 模型可以自动在每一层使用尽量少的神经元来达到最优的情感分析结果, 也就是对神经网络进行了剪枝, 群稀疏的效果如图 5 所示。

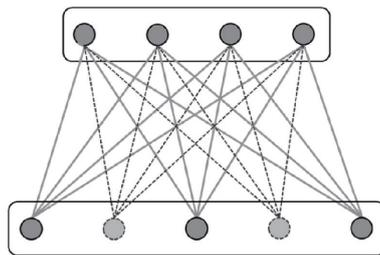


图 5 群稀疏效果示意图

Fig. 5 Effect of group sparsity

虽然利用群稀疏化能够剪除部分神经元, 使得网络的计算有所加快, 但是群稀疏性并不能最大限度地利用神经网络对特征的提取表示能力, 因为根据不同连接所选择的相同的上下文输入进行提取的特征之间仍然可能存在冗余。因此, 在本文模型中引入排他性稀疏(exclusive sparsity)正则化方法, 使得情感分析模型的卷积操作能够学到一个稀疏的网络权值矩阵, 同时最小化网络权值之间的冗余, 以更好地利用网络特征表示能力, 压缩模型的规模。

排他性稀疏最早是在多任务学习环境中引入的^[15]。排他性稀疏正则项可以定义为

$$\Omega(\mathbf{W}^{(l)}) = \frac{1}{2} \sum_g \|\mathbf{W}_g^{(l)}\|_1^2 = \frac{1}{2} \sum_g \left(\sum_i |w_{g,i}^{(l)}| \right)^2 \quad (12)$$

排他性稀疏范数通常称为(1,2)-范数,在一个参数群内,首先利用 L1 范数对参数进行约束,再利用 L2 范数对所有的群进行约束。在群内使用 L1 范数的基础上应用 L2 范数正则项会使组间参数权重的总量相等,因此所有群的非稀疏权值的个数相近,没有一个群内会出现大量的非稀疏权值而其他组都是稀疏权值。在这种情况下,(1,2)-范数强制情感分析器选择对细分领域情感分类最为有效的最少特征来对文本进行预测。对于本文提出的情感分析模型中的卷积层与全连通层,应用排他性稀疏正则。将不同卷积滤波器得到的特征图对应相同位置的特征定义为同一群权重,这将强制卷积滤波器彼此之间尽可能不同,消除提取的情感特征之间的冗余。排他性稀疏正则项的效果如图 6 所示。

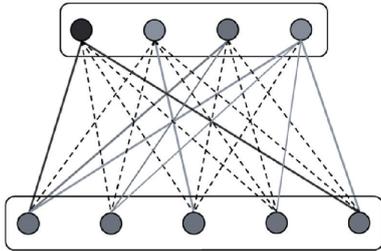


图 6 排他性稀疏效果示意图

Fig. 6 Effect of exclusive sparsity

融合群稀疏与排他性稀疏的正则项可以由式(13)表示

$$\Omega(\mathbf{W}^{(l)}) = \sum_g \left((1 - \mu_l) \|\mathbf{W}_g^{(l)}\| + \frac{\mu_l}{2} \|\mathbf{W}_g^{(l)}\|_1^2 \right) \quad (13)$$

模型除正则项以外的目标函数,可以直接求梯度,而对正则项的优化由于不能直接计算梯度,因此需要求解近似梯度(proximal gradient)。近似梯度正则化算法为首先计算不带正则项目标函数对参数的梯度,获得中间解参数 $\mathbf{W}_{t+1/2}$,然后优化正则化项。求解形如式(14)的优化问题

$$\min_{\mathbf{W}_{t+1}} \Omega(\mathbf{W}_{t+1}) + \frac{1}{2\lambda s} \|\mathbf{W}_{t+1} - \mathbf{W}_{t+1/2}\|_2^2 \quad (14)$$

可以求解得到群稀疏正则项的近似梯度为

$$\text{prox}_{\text{GL}}(\mathbf{W}) = \left(1 - \frac{\lambda}{\|\mathbf{W}_g\|_2} \right)_+ \mathbf{w}_{g,i} \quad (15)$$

排他性稀疏正则项的近似梯度为

$$\text{prox}_{\text{EL}}(\mathbf{W}) = \left(1 - \frac{\lambda \|\mathbf{W}_g\|_1}{|\mathbf{w}_{g,i}|} \right)_+ \mathbf{w}_{g,i} = \text{sign}(\mathbf{w}_{g,i}) (|\mathbf{w}_{g,i}| - \lambda \|\mathbf{W}_g\|_1)_+ \quad (16)$$

融合的正则项只需在每一次迭代时,利用不带正则项的损失函数更新得到中间参数,而后连续应用两个近似梯度算子求出近似梯度更新参数即可优化网络。

3 实验设置

3.1 数据集

实验中主要使用 3 个数据集。第一个是中文酒店数据集(简称携程酒店),语料规模为 10 000 篇,其中正面评价样本 7 000 篇,负面评价样本 3 000 篇。第二个数据集是英文数据集,来自斯坦福大学的情感树库(Stanford sentiment treebank, SST)^[16],它标注了句子中每个短语和整个句子的情感极性,本文中只抽取其句子的情感极性并进行分类。该数据集中共包含了 11 855 个句子,其中的训练集、验证集和测试集分别包含 8 544、1 101 和 2 210 个样本。每个句子的情感极性取值区间为[0,1],分值越小表示情感越倾向于负面,否则情感倾向于正面,数据集中所有句子的情感得分均由人工标注,然后取平均值,具有较高的可靠性。为了使实验更加贴近真实的生产环境,以及验证本文提出的基于门控单元的循环-卷积神经网络与卷积-循环神经网络情感分析方法的有效性,使用的第三个数据集是从生活服务类网站—大众点评上爬取的不同领域的评论文本(简称点评),包括美食、酒店、电影、休闲娱乐、结婚、家装 6 个细分领域的的数据,并依据评论中的打分信息将评论划分为不同的情感倾向,本文选取其中的 30 000 份评论作为训练集,10 000 份作为测试集。

3.2 数据前处理

对于中文数据集,首先使用中科院计算所开发的中文分词软件包 NLPIR 进行中文分词。英文数据本身就是独立的单词,因此不需要进行分词操作。由于在训练过程中使用 minibatch 训练模型,因此需要统一样本长度;同时,利用卷积-循环神经网络提取不同文本的特征需要保证维度的统一,卷积后拼接才能保证特征维度,因此需要对文本的长度进行补齐操作。由于评论的文本的长度是不一致的,先计算出最长的句子长度 l_{max} ,对于句子长度小于 l_{max} 的句子,统一使文本用(\s)符号补齐到长度

l_{\max} ($\langle \backslash s \rangle$ 的向量始终设置为 0)。同时,为了在卷积过程中保证提取到文本首尾的特征,还需要在最长文本的首尾添加与卷积核相适应的一定数量的 $\langle \backslash s \rangle$ 作为 Padding。

3.3 词嵌入的预训练

在正式进行模型训练前,需要获得词嵌入。词嵌入作为单词的一种分布式表示,能够很好地适应神经网络。同神经网络方法类似,当前许多研究都表明在大规模语料上进行词嵌入预训练,再将获得的词嵌入运用到后续的训练中,能够加大神经网络模型收敛速度并达到一个较好的局部最优解。本文使用 word2vec 算法预训练词嵌入,此算法学习的词嵌入在许多自然语言处理任务中都表现出较好的性能,同时具有较高的效率。选择 Skip-gram 模型和负采样 (negative sampling) 模型预训练中文和英文单词的词嵌入,中文词嵌入的预训练使用在百度百科上爬取的文本内容,英文词向量的预训练在 New York Times 语料上进行。

3.4 实验参数的设置

在模型的训练优化过程中,采用 Adam 优化器对模型的参数进行训练优化,Adam 优化器的参数设置采用作者推荐的参数值^[17]。在本文中,模型主要有以下超参数:词向量的维度 d ,卷积-循环神经网络中卷积核的个数 n ,循环-卷积神经网络中循环结构中隐藏状态的维度 N_{cc} 和卷积-循环神经网络中隐藏状态的维度 N_{cr} 。为了获得最优的超参数设置,使用网格搜索 (grid-search) 确定部分超参数的取值。循环-卷积神经网络中循环结构隐藏状态的维度 N_{cc} 和卷积-循环神经网络中隐藏状态的维度 N_{cr} 从 $\{50, 100, 200, 300\}$ 中选择取值;卷积核的个数 n 在 $\{100, 150, 200\}$ 中取值。在本文中,使用这些参数进行多次实验,最后得到结果的平均值。

4 结果与分析

为了验证本文提出的融合群稀疏正则项与排他性稀疏正则项的情感分析模型的有效性以及模型压缩的效果,将本文方法与一些主要的情感分析基线方法进行比较。

4.1 对比方法

为了验证本文提出的情感分析压缩模型的有效性,选取带有词义变换矩阵的递归神经网络 (MV-RNN)、基于张量的循环神经网络 (RNTN) 以及融合分段卷积网络与生成对抗网络的情感分析 (PCNN +

GAN) 作为基线方法进行比较。为了说明模型的压缩效果,与未压缩的模型从收敛速度和参数数量等方面进行对比。

4.2 实验结果分析

从表 1~3 的结果可以看出,在两个数据集上本文提出的未剪枝的 PCNN 与 RCNN&C-RNN 方法都取得了最好的效果。融合群稀疏与排他性稀疏正则项的情感分析压缩模型在性能上略微有所降低,情感预测准确率只下降 0.1%~0.3%,下降幅度都较小。对比基线方法,剪枝后的 Pruning-PCNN 模型只有在斯坦福英文情感树库数据集上准确率比 RNTN 模型低了 0.1%~0.3%,说明即使经过了压缩,本文提出的压缩模型仍然具有很好的性能,证明了压缩模型的有效性。

点评数据来自于不同的领域,从表 4~6 中可以看出本文提出的未剪枝的 PCNN 与 RCNN&C-RNN 方法在点评数据上都取得了最好的效果,融合群稀疏与排他性稀疏正则项的情感分析压缩模型在性能上略微有所降低,情感预测准确率下降 0.2%~0.7%,下降幅度都较小。从大众点评爬取的评论中包含不同的细分领域,每一个领域的评论文本相对较少,并且评论文本本身相对较短,因此不同领域的样本都存在一定程度的样本稀疏问题,需要模型能够有效地从文本中抽取与情感有关的特征才能正确对文本的情感倾向做出判断。本文还缩减每个领域中的样本数量,使数据稀疏更加严重,可以看出所有方法都不同程度地出现了性能退化的问题,但是本文提出的融合群稀疏与排他性稀疏正则项的情感分析压缩模型依旧取得了较好的效果,充分说明压缩后的模型在多领域数据稀疏情况下可以保持良好的稳定性以及特征提取的有效性,充分利用了深度神经网络的特征表示效果。

表 1 不同数据集上的情感分析精确率

Table 1 Precision of sentiment analysis of different datasets

| 模型 | 精确率/% | | |
|--------------------|---------------|---------------|------|
| | 二级情感 (SST) | 五级情感 (SST) | 携程酒店 |
| MV-RNN | 82.9 | 44.4 | 87.6 |
| RNTN | 85.4 | 45.7 | 89.3 |
| PCNN | 85.4 | 45.9 | 89.7 |
| RCNN&C-RNN | 85.8 | 46.1 | 89.8 |
| Pruning-PCNN | 85.1 | 45.6 | 89.3 |
| Pruning-RCNN&C-RNN | 85.5 | 45.8 | 89.6 |

表2 不同数据集上的情感分析召回率

Table 2 Recall of sentiment analysis on different datasets

| 模型 | 召回率/% | | |
|--------------------|---------------|---------------|------|
| | 二级情感 (SST) | 五级情感 (SST) | 携程酒店 |
| MV-RNN | 81.6 | 45.4 | 84.9 |
| RNTN | 86.3 | 48.6 | 87.3 |
| PCNN | 86.4 | 49.4 | 88.9 |
| RCNN&C-RNN | 87.8 | 50.2 | 89.7 |
| Pruning-PCNN | 86.2 | 48.9 | 88.2 |
| Pruning-RCNN&C-RNN | 87.4 | 49.7 | 88.9 |

表3 不同数据集上的情感分析 F1 值

Table 3 F1 score of sentiment analysis on different datasets

| 模型 | F1 值 | | |
|--------------------|---------------|---------------|------|
| | 二级情感 (SST) | 五级情感 (SST) | 携程酒店 |
| MV-RNN | 82.2 | 44.9 | 86.2 |
| RNTN | 85.8 | 47.1 | 88.3 |
| PCNN | 85.9 | 47.6 | 89.3 |
| RCNN&C-RNN | 86.8 | 48.1 | 89.7 |
| Pruning-PCNN | 85.6 | 47.2 | 88.7 |
| Pruning-RCNN&C-RNN | 86.4 | 47.7 | 89.2 |

表4 点评数据集上的情感分析精确率

Table 4 Precision of sentiment analysis on DianPing dataset

| 模型 | 精确率/% | | |
|--------------------|-------|------|----------------|
| | 二级情感 | 五级情感 | 二级情感 (缩减样本) |
| MV-RNN | 78.0 | 42.2 | 73.1 |
| RNTN | 80.3 | 42.6 | 75.7 |
| PCNN + GAN | 83.3 | 47.9 | 80.1 |
| RCNN&C-RNN | 83.5 | 47.6 | 80.6 |
| Pruning-PCNN + GAN | 82.8 | 47.1 | 79.5 |
| Pruning-RCNN&C-RNN | 83.1 | 47.4 | 79.9 |

表5 点评数据集上的情感分析召回率

Table 5 Recall of sentiment analysis on DianPing dataset

| 模型 | 召回率/% | | |
|--------------------|-------|------|----------------|
| | 二级情感 | 五级情感 | 二级情感 (缩减样本) |
| MV-RNN | 79.1 | 43.2 | 69.3 |
| RNTN | 81.4 | 45.6 | 74.7 |
| PCNN + GAN | 83.7 | 47.8 | 77.2 |
| RCNN&C-RNN | 83.9 | 48.2 | 80.1 |
| Pruning-PCNN + GAN | 83.2 | 47.1 | 76.7 |
| Pruning-RCNN&C-RNN | 83.5 | 47.9 | 79.6 |

表6 点评数据集上的情感分析 F1 值

Table 6 F1 score of sentiment analysis on DianPing dataset

| 模型 | F1 值 | | |
|--------------------|------|------|----------------|
| | 二级情感 | 五级情感 | 二级情感 (缩减样本) |
| MV-RNN | 78.5 | 42.7 | 71.1 |
| RNTN | 80.8 | 44.0 | 75.2 |
| PCNN + GAN | 83.5 | 47.8 | 78.6 |
| RCNN&C-RNN | 83.7 | 47.9 | 80.3 |
| Pruning-PCNN + GAN | 83.0 | 47.1 | 78.1 |
| Pruning-RCNN&C-RNN | 83.3 | 47.6 | 79.7 |

为了进一步说明融合群稀疏与排他性稀疏正则项情感分析压缩模型的有效性,从点评数据集中选取一部分样本,在样本中加入否定词,使样本的情感倾向转换为对立的情感,称为否定积极和否定消极。例如将“这家餐厅的菜很好吃”由积极转变为否定积极,既“这家餐厅的菜很不好吃”;将“这部电影难看”由消极转变为否定消极,既“这部电影不难看”。可以看出这样的转换对于评论文本的改变十分细微,情感倾向积极的样本加入否定词后情感倾向发生了强烈反转;而消极的文本变为否定消极后,评论的消极程度有所降低,但是没有像积极的样本一样强烈反转。这些样本能够测试情感分析方法对于情感提取的有效性和稳定性,测试模型是否能够捕获文本中的情感细节。

测试结果如表7所示。实验结果表明在加入否定词后,本文提出的融合群稀疏与排他性稀疏正则项情感分析压缩模型依旧能够很好地从细节上捕获评论文本的情感特征。

表7 点评数据集上加入否定词后的评论情感分析精确率

Table 7 Precision of sentiment analysis aggregating negative word on DianPing dataset

| 模型 | 精确率/% | |
|--------------------|-------|------|
| | 否定积极 | 否定消极 |
| MV-RNN | 62.4 | 56.2 |
| RNTN | 81.4 | 72.6 |
| PCNN + GAN | 81.4 | 76.3 |
| RCNN&C-RNN | 81.7 | 77.2 |
| Pruning-PCNN + GAN | 81.4 | 75.9 |
| Pruning-RCNN&C-RNN | 81.6 | 77.0 |

为了说明本文提出的融合群稀疏与排他性稀疏正则项情感分析压缩模型的计算加速效果以及对模

型规模的压缩效果,在点评数据集上进行剪枝与未剪枝的模型训练收敛速率对比实验以及模型的压缩程度实验,结果如图 7 所示。

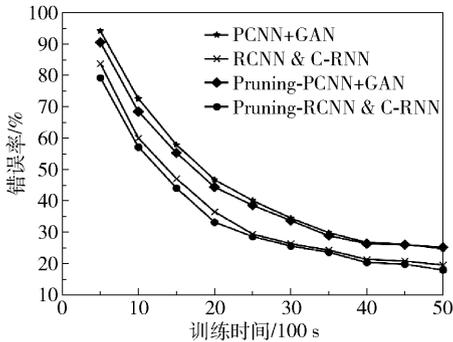


图 7 未剪枝模型与剪枝模型训练时间开销对比

Fig. 7 Comparison of training time cost between unpruned model and pruned model

从图 7 可以看出,本文提出的压缩模型比未剪枝的情感分析模型能够更快地收敛,说明了融合群稀疏与排他性稀疏正则项能够有效减少模型中卷积与全连接部分的计算量。

通过调节正则项与情感分类损失的平衡参数来改变模型中参数的数量,进而衡量性能与压缩程度的关系。从表 8 可以看出,压缩后的 PCNN + GAN 与 RCNN&C-RNN 在剪除超过 20% 的参数后,性能损失最多也只有 1.2%,说明了融合群稀疏与排他性稀疏正则项的情感分析压缩模型能够在大幅度压缩模型规模的同时保证模型的精确性。

表 8 不同参数量下点评数据集上评论的情感分析精确率

Table 8 Precision of models with different parameters on DianPing dataset

| 模型 | 精确率/% | | 参数量/ % |
|----------------------|-------|------|-----------|
| | 二级情感 | 五级情感 | |
| PCNN + GAN | 83.3 | 47.9 | 100 |
| Pruning-PCNN + GAN-1 | 82.8 | 47.1 | 76.56 |
| Pruning-PCNN + GAN-2 | 82.1 | 46.7 | 70.39 |
| RCNN&C-RNN | 83.5 | 47.6 | 100 |
| Pruning-RCNN&C-RNN-1 | 83.1 | 47.4 | 78.37 |
| Pruning-RCNN&C-RNN-2 | 82.9 | 47.2 | 73.24 |

5 结论

针对 RNN 能够获得上下文信息但是文本中特征出现的顺序不同会导致模型建模出现偏置,基于 CNN 的文本情感分析方法能够通过池化获得文本的重要特征但较难获得上下文信息,对于长距离依

赖建模能力较差等问题,提出了一种基于门控单元的融合 RCNN 与 C-RNN 集成情感分析方法模型。由于模型参数规模较大,带有情感标注信息的细分领域文本数据较少,存在数据稀疏问题,可能导致上述深度神经网络模型得不到较为充分的训练的问题;以及为了满足情感分析系统较高的时间要求,需要能够加速模型的计算。

为了解决上述问题,即提升网络在进行预测时的效率,同时保证模型的性能不会有较大的下降,提出利用剪枝的方法对本文提出的模型进行压缩。在模型中引入群稀疏正则项与排他性稀疏正则项,通过融合上述稀疏正则项约束神经网络的卷积与全连接部分的参数,剪除冗余的神经元或链接,压缩了情感分析模型规模,减少了计算量。在不同的数据集上进行充分实验的结果验证了压缩方法的有效性,且保证了模型在情感分析任务上的性能。

参考文献:

- [1] SOCHER R. Recursive deep learning for natural language processing and computer vision [D]. Palo Alto: Stanford University, 2014.
- [2] SOCHER R, CHEN D Q, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion [C] // Advances in Neural Information Processing Systems. Lake Tahoe, 2013: 926-934.
- [3] SOCHER R, PERELYGIN A, WU J Y, 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. Seattle, 2013: 1631-1642.
- [4] LUONG M T, SOCHER R, MANNING C D. Better word representations with recursive neural networks for morphology [C] // Proceedings of the Seventeenth Conference on Computational Natural Language Learning. Sofia, 2013: 104-113.
- [5] SOCHER R, HUVAL B, MANNING C D, et al. Semantic compositionality through recursive matrix-vector spaces [C] // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, 2012: 1201-1211.
- [6] LE Q V, NGIAM J, COATES A, et al. On optimization methods for deep learning [C] // Proceedings of the 28th International Conference on Machine Learning. Bellevue: Omni Press, 2011: 265-272.

- [7] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. *Journal of Machine Learning Research*, 2011, 12: 2493–2537.
- [8] KALCHBRENNER N, BLUNSOM P. Recurrent convolutional neural networks for discourse compositionality[J]. arXiv preprint arXiv: 1306. 3584, 2013.
- [9] CLICHE M. BB_twtr at SemEval-2017 Task 4: Twitter sentiment analysis with CNNs and LSTMs[J]. arXiv preprint arXiv: 1704. 06125, 2017.
- [10] SHI B G, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(11): 2298–2304.
- [11] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences [J]. arXiv preprint arXiv: 1404. 2188, 2014.
- [12] LAI S W, XU L H, LIU K, et al. Recurrent convolutional neural networks for text classification [C] // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, 2015: 2267–2273.
- [13] WEN W, WU C P, WANG Y D, et al. Learning structured sparsity in deep neural networks[C] // 30th Conference on Neural Information Processing Systems. Barcelona, 2016: 2074–2082.
- [14] ALVAREZ J M, SALZMANN M. Learning the number of neurons in deep networks[C] // 30th Conference on Neural Information Processing Systems. Barcelona, 2016: 2270–2278.
- [15] ZHOU Y, JIN R, HOI S C H. Exclusive lasso for multi-task feature selection[C] // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Sardinia, 2010: 988–995.
- [16] Stanford Press Release. Stanford sentiment tree-bank [EB/OL]. [2018-11-01]. <https://nlp.stanford.edu/sentiment/index.html>.
- [17] KINGMA D P, BA L J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv: 1412. 6980, 2014.

A neural network compressed sentiment analysis method based on combined group and exclusive sparsity regularization

HUANG Lei DU ChangShun*

(School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China)

Abstract: In the current internet environment, text sentiment analysis is one of the most important tasks in the field of public opinion monitoring and analysis of customer service satisfaction. Deep network models offer a great improvement over the traditional methods in sentiment analysis tasks, since large-scale networks enable a deep model to fit nonlinear data effectively and large-scale data resources provide the possibility to train such a large-scale model and guarantee its generalization ability. However, the time and space costs of the deep model still restrict its application. This paper proposes a neural network compressed sentiment analysis method that combines group sparsity and exclusive sparsity regularization. We first construct an analysis model composed of RCNN and C-RNN which is integrated by a gate unit. Group sparsity and exclusive sparsity regularization are then introduced in order to make the model compressed. Experiments on different data sets verify the effectiveness of our method.

Key words: sentiment analysis; convolutional neural network; recurrent neural network; gate unit; model compression

(责任编辑:吴万玲)