

基于循环卷积多任务学习的多领域文本分类方法

谢金宝^① 李嘉辉*^② 康守强^② 王庆岩^② 王玉静^②

^①(广东科学技术职业学院机器人学院 珠海 519090)

^②(哈尔滨理工大学电气与电子工程学院 哈尔滨 150000)

摘要: 文本分类任务中,不同领域的文本很多表达相似,具有相关性的特点,可以解决有标签训练数据不足的问题。采用多任务学习的方法联合学习能够将不同领域的文本利用起来,提升模型的训练准确率和速度。该文提出循环卷积多任务学习(MTL-RC)模型用于文本多分类,将多个任务的文本共同建模,分别利用多任务学习、循环神经网络(RNN)和卷积神经网络(CNN)模型的优势获取多领域文本间的相关性、文本长期依赖关系、提取文本的局部特征。基于多领域文本分类数据集进行丰富的实验,该文提出的循环卷积多任务学习模型(MTL-LC)不同领域的文本分类平均准确率达到90.1%,比单任务学习模型循环卷积单任务学习模型(STL-LC)提升了6.5%,与当前热门的多任务学习模型完全共享多任务学习模型(FS-MTL)、对抗多任务学习模型(ASP-MTL)、间接交流多任务学习框架(IC-MTL)相比分别提升了5.4%, 4%和2.8%。

关键词: 多领域文本分类; 多任务学习; 循环神经网络; 卷积神经网络

中图分类号: TP391.1

文献标识码: A

文章编号: 1009-5896(2021)08-2395-09

DOI: [10.11999/JEIT200869](https://doi.org/10.11999/JEIT200869)

A Multi-domain Text Classification Method Based on Recurrent Convolution Multi-task Learning

XIE Jinbao^① LI Jiahui^② KANG Shouqiang^② WANG Qingyan^②
WANG Yujing^②

^①(School of Robotic, Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China)

^②(School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin 150000, China)

Abstract: In the text classification task, many texts in different domains are similarly expressed and have the characteristics of correlation, which can solve the problem of insufficient training data with labels. The text of different fields can be combined with the multi-task learning method, and the training accuracy and speed of the model can be improved. A Recurrent Convolution Multi-Task Learning (MTL-RC) model for text multi-classification is proposed, jointly modeling the text of multiple tasks, and taking advantage of multi-task learning, Recurrent Neural Network(RNN) and Convolutional Neural Network(CNN) models to obtain the correlation between multi-domain texts, long-term dependence of text. Local features of text are extracted. Rich experiments are carried out based on multi-domain text classification datasets, the Recurrent Convolution Multi-Task Learning(MTL-LC) proposed in this paper has an average accuracy of 90.1% for text classification in different fields, which is 6.5% higher than the single-task learning model STL-LC. Compared with mainstream multi-tasking learning models Full Shared Multi-Task Learning(FS-MTL), Adversarial Multi-Task Learning(ASP-MTL), and Indirect Communication for Multi-Task Learning(IC-MTL) have increased by 5.4%, 4%, and 2.8%, respectively.

收稿日期: 2020-10-09; 改回日期: 2021-02-03; 网络出版: 2021-03-01

*通信作者: 李嘉辉 maillijiahui@163.com

基金项目: 基于工业互联网的协作式智能机器人产教融合创新应用平台(2020CJPT004), 黑龙江省自然科学基金(LH2019E058), 智能机器人湖北省重点实验室开放基金(HBIR202004), 黑龙江省普通高校基本科研业务费专项资金(LGYC2018JC027)

Foundation Items: The collaborative intelligent robot production and education integrates innovative application platform based on the industrial Internet (2020CJPT004), The Natural Science Foundation of Heilongjiang Province (LH2019E058), The open fund projects of Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology) (HBIR 202004), The Fundamental Research Foundation for Universities of Heilongjiang Province (LGYC2018JC027)

Key words: Multi-domain text classification; Multi-task learning; Recurrent Neural Network(RNN); Convolutional Neural Network(CNN)

1 引言

文本分类是自然语言处理中一项重要的基础工作^[1], 具有广泛的应用场景, 例如垃圾邮件过滤、商品情感分类等。过去使用传统的机器学习方法进行文本分类。近年来, 循环神经网络(Recurrent Neural Network, RNN)和卷积神经网络(Convolutional Neural Network, CNN)等深度学习网络在文本分类任务中取得了很好的进展^[2]。与传统方法不同的是, 深度学习方法能自主学习文本的深层语义特征。长短时记忆网络(Long Short-Term Memory, LSTM)是一种改进的RNN, 可以有效地捕获文本的上下文信息, CNN能够捕获文本的局部特征。为每个领域的文本训练效果好的分类模型需要大量标记样本, 由于手工注释是很耗时的, 因此很难获得。为解决此问题, 很多研究者采用域适应方法^[3]有效利用标签数据进行文本分类。迁移学习过程中, 当源域和目标域数据分布不同但任务相同时, 这种特殊的迁移学习为域适应学习。在文本分类任务中, 一个领域的的数据不足时, 利用资源丰富的源域学习到的知识去学习, 减少领域之间的差异, 提高领域文本分类的准确率。不同的是, 假设每个领域的标记数据不足以训练出一个准确的单一领域情感分类器。我们研究多任务学习方法^[4], 从多个领域中提取共享的特征, 把各个任务的信息利用起来, 使用有限的标记样本训练一个比单一领域分类器性能更好的分类器。Liu等人^[5]开发出一种结合语义分类和信息检索任务的多任务学习模型, 采用词袋模型会丢失很多语义信息, 共享一个深度神经网络(Deep Neural Network, DNN)隐藏层很难得到文本的高层次语义信息。主流的多任务学习模型采用LSTM获取文本表示^[6,7], 但他们仅仅使用LSTM的最后一个隐藏层的输出作为文本的特征表示, 会导致文本特征被文本末尾单词所主导, 丢失很多关键的信息。常见的多任务学习模型有完全共享模型和共享私有模型。完全共享模型是多个任务共享一个模型提取共享特征; 共享私有模型则是多个任务通过共享一个模型提取共享特征, 同时每个任务输入到一个私有层提取特定任务的特征, 再将两种特征融合起来, 这样会对内存和时间造成更大的消耗。

基于以上结论, 本文提出一个基于循环卷积神经网络的多任务学习模型, 本文方法中不同领域的文本在一个多任务学习模型中联合训练, 能够利用

不同领域中的文本信息。考虑时间和内存问题, 本文采用完全共享模式, 利用一个共享层提取特征, 把相关任务的数据信息利用起来学习, 而不相关的部分可以作为噪声提升泛化效果。为了更全面地挖掘文本特征, 共享模型通过LSTM层获取文本的长短期依赖, 再通过CNN层利用不同窗口的卷积核提取句子不同位置的n-gram特征, 还可以将LSTM的所有隐藏层输出利用起来, 不至于被最后一个单词所主导。模型可以将多任务学习与LSTM, CNN两种深度学习网络的优势结合起来。为了评估本文所提模型性能, 本文的实验数据集为: 亚马逊的多领域情感分类数据集和电影情感分类数据集。实验结果表明, 本文所提模型优于单任务深度学习模型和主流的多任务学习模型, 能够有效地对多领域文本情感进行分类。

2 相关工作

2.1 深度学习方法

RNN在一个固定大小的隐藏层中逐字逐句分析文本, 保留前面的文本。LSTM作为一种改进的RNN模型可以提取文本的长期依赖关系, 并解决传统RNN存在的梯度消失问题^[8]。LSTM在文本预测和情感分类等序列任务中取得了不错的效果。CNN最初是为计算机视觉设计的。CNN利用具有卷积滤波器的层去提取局部特征, 在图像分类^[9]等计算机视觉任务中已经相当成熟, 在文本分类等自然语言处理任务中也取得了不错的进展。Kim^[2]将卷积神经网络应用于文本分类, 提出了一个具有不同窗口大小滤波器的CNN架构。Kalchbrenner等人^[10]提出了动态卷积神经网络得到句子的特征表示。他们在不进行任何特征工程的情况下, 实现了文本分类任务的高性能, 证明了模型的有效性。

2.2 多任务学习

近年来, 基于深度学习的多任务学习模型在自然语言处理领域变得非常流行^[11], 它们提供了一种方便的方式来组合来自多个任务的信息。Collobert等人^[12]用多任务学习处理词性标注、命名体识别、语义角色标注等几个传统的自然语言处理任务, 只有查找表部分是共享的。在多个领域的文本分类任务中, 早期有一些基于分类器组合的研究^[13]。在这些方法中, 每个领域训练的分类器使用不同的组合方案进行组合。他们的方法没有学习到更高层次的语义信息。Liu等人^[6]基于LSTM设计了多任务学习模型, 采用了两个多任务共享方式, 分别是完全共

享方式和共享私有方式。有研究者也在不断地研究新的多任务共享方式,去决定多个任务是如何共享的,比如外部记忆共享模型由多个任务共享的外部内存来增强神经模型^[14];图多任务学习框架可以将不同的任务有效地进行通信^[15];领域注意机制模型将所有领域的文本共享一个LSTM模型,再以每个领域的表示作为注意力从共享表示中提取与某领域中最相关的特征^[16]。

3 提出的模型

本文提出一种用于文本分类的循环卷积多任务学习模型(Recurrent Convolution Multi-Task Learning, MTL-RC),此模型将不同任务在一个共享深度学习模型中训练,共享模型将循环神经网络与卷积神经网络相结合可以从这两种结构中受益。本模型将不同领域的文本输入共享模型中,最后输入不同的Softmax层输出分类结果,模型包括输入层、词嵌入层、共享深度学习网络层和输出层。模型结构如图1所示。

3.1 输入层

多任务学习的目的是利于相关的任务之间的相

关性,并行学习任务进行分类,因此将不同任务的文本并行地输入到模型中。假设有 k 个任务,不同任务的数据集表示为 $D=(D_1, D_2, \dots, D_k)$, D_m 表示为包含 n 个样本的第 m 个任务的数据集, x_i^m 是任务 m 的一个样本, y_i^m 则是此样本的标签。如式(1)所示

$$D_m = \{(x_i^m, y_i^m)\}_{i=1}^n \quad (1)$$

3.2 词嵌入

考虑到每条训练样本的长度不同,为了保持维度大小一致,通常在词嵌入前通过取长补短把每条样本的长度变成相同的。把其中一个任务的一条数据的单词列表表示为 $\mathbf{x}^m=(x_1, x_2, \dots, x_n)$,其中 n 为一个样本单词的数量, x_t 为这段文本中的第 t 个单词,所有的单词都来自词汇表 V 中。在输入到模型之前,通过词向量矩阵 $\mathbf{W} \in \mathbf{R}^{d \times |V|}$ 将单词转变为词向量,其中 $|V|$ 为词汇表中单词的数量, d 为词向量的维度。词嵌入后每一条数据变为2维向量 $\mathbf{X} \in \mathbf{R}^{d \times n}$ 。

3.3 共享LSTM和CNN层

以两个任务为例,把两个任务文本词向量输入到共享LSTM层中,仅仅以LSTM的最后一个隐藏

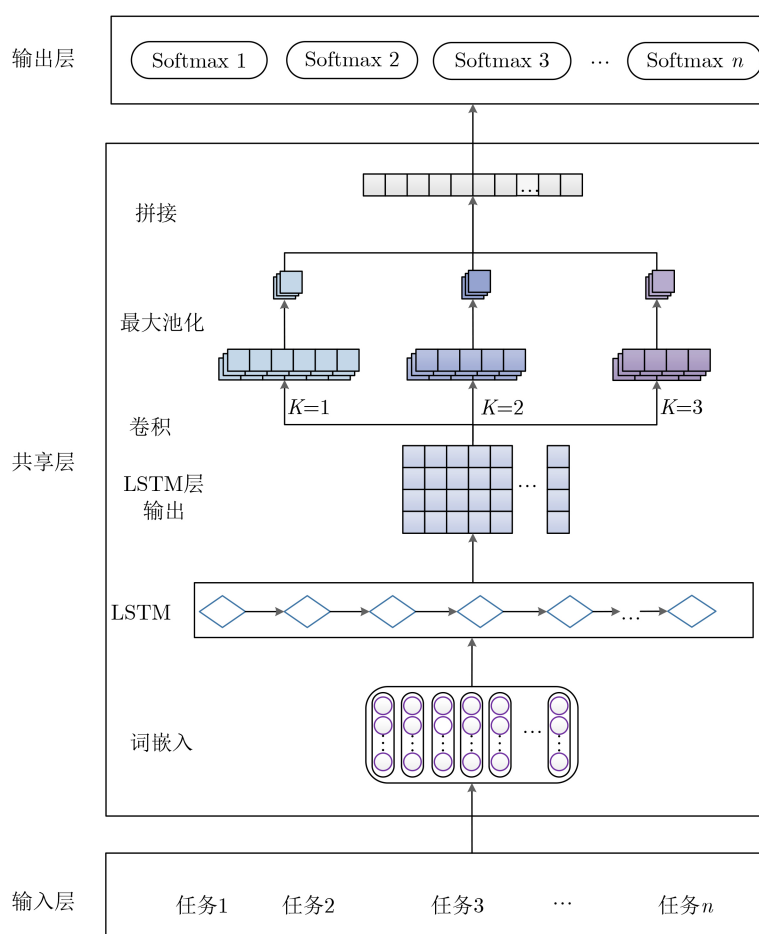


图1 MTL-RC多任务学习模型

层作为文本表示,不能获取丰富的语义信息,如图2所示。

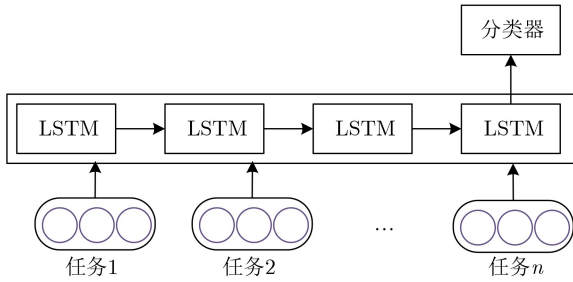


图2 共享LSTM层

本文将每一个任务的文本生成词向量后,都要经过一个由LSTM和CNN组成的共享层,LSTM是一种链式神经网络结构用来传播历史信息,CNN则是采用多个卷积核并行地提取文本的局部特征,LSTM和CNN结合起来可以更加全面地提取文本的深层语义特征。共享LSTM和CNN层如图3所示。下面主要介绍这两个深度学习网络层。

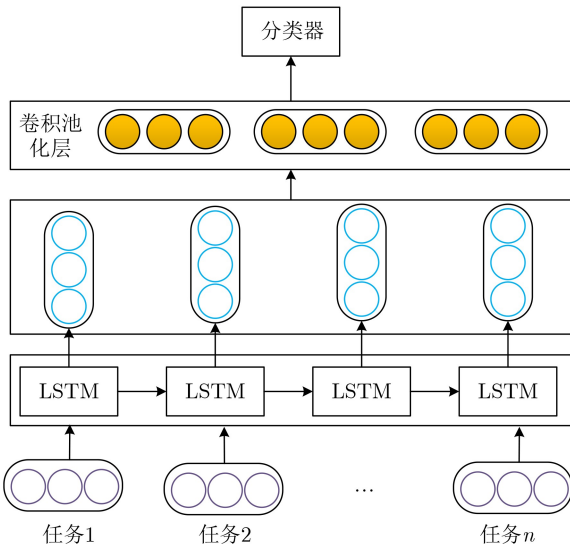


图3 共享LSTM和CNN层

3.3.1 LSTM层

LSTM对历史信息带有记忆功能,能够在长的序列中有好的表现^[8]。相比RNN只有一个传递状态,LSTM有两个传输状态,分别是细胞状态 c_t 、隐藏层状态 h_t 。虽然提出了很多LSTM变体,考虑模型的训练时间,在这项工作中采用标准的LSTM结构。

LSTM的核心思想是利用记忆细胞和门机制管理,其中令 x_t 为当前时间步的输入, h_{t-1} 为上一时刻隐藏层的输出,每一个LSTM单元包括3个门,分别是遗忘门 f_t 、输入门 i_t 、输出门 o_t 。而这3个门

的作用就是决定如何更新细胞状态 c_t 、隐藏层状态 h_t 。LSTM每个时间步的计算公式为

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(c_t) \quad (7)$$

上述公式中, W 表示权重矩阵, b 表示偏置值。 σ 是sigmoid函数,输出在 $[0,1]$ 之间, $\tanh$ 是双曲正切函数,输出在 $[-1,1]$ 之间,所有门的值和隐藏层的输出值在 $[0,1]$ 之间。

3.3.2 CNN层

LSTM的输出矩阵作为卷积层的输入,使用一个包括滤波向量的1维卷积在序列上滑动并在不同的位置检测特征。卷积核的宽度与LSTM输出特征向量的维度一样,本文使用不同窗口的多个卷积核提取丰富的文本局部特征。让 $h_i \in \mathbf{R}^d$ 为LSTM输出矩阵的第 i 个 d 维向量, $\mathbf{F} \in \mathbf{R}^{k \times d}$ 代表一个滤波器的矩阵。这样,一个特征被表示为

$$c_i = f(w \cdot x_{i:i+k-1} + b) \quad (8)$$

其中, f 为非线性激活函数,本文使用ReLU, b 为偏置项。

经过卷积操作后,可以得到一个 $n-k+1$ 维的向量 $\mathbf{C}$,形如

$$\mathbf{C} = [c_1, c_2, \dots, c_{n-k+1}] \quad (9)$$

在卷积层后进行池化操作,池化的特点之一就是固定卷积后矩阵维度的大小,还能降低输出结果的维度。本文使用了最大池化得到特征向量中最大的值,以这种方式,对于每个卷积核,得到了最显著的特征

$$\hat{c} = \max\{\mathbf{C}\} \quad (10)$$

经过池化之后,还需将不同卷积核得到的特征值拼接起来,得到最终的特征向量,表示为

$$\hat{\mathbf{C}} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_d] \quad (11)$$

3.4 输出层

最后一层是输出层,每一个任务的文本表示分别输入到特定任务的Softmax层中,进行情感分类。一共有 k 个任务,因此要输入到 k 个Softmax层中,令 m 为其中的第 m 个任务,利用Softmax函数计算第 m 个任务的样本每一个情感类别的概率,表示为

$$\hat{y}_c^{(m)} = \frac{\exp(W_c^{(m)}x^{(m)} + b_c^{(m)})}{\sum_{c=1}^C \exp(W_c^{(m)}x^{(m)} + b_c^{(m)})} \quad (12)$$

其中, $\mathbf{W}$ 为要学习的权重, $\mathbf{b}$ 为偏置, C 为样本情感分类的类别数。

更新网络参数不断地最小化交叉熵损失值, 某任务中某个样本的损失值表示为

$$L(\hat{y}, y) = - \sum_{c=1}^C y^c \lg(\hat{y}^c) \quad (13)$$

$\hat{y}^c$ 为某个样本的真实值, $\hat{y}^c$ 为预测概率值, C 为样本的类别数。

4 实验

4.1 实验设置

为了验证本文所提模型, 采用由王鑫等人^[9]收集的16个不同的情感分类数据集。其中14个数据集是来自不同领域的亚马逊商品评论, 包括books, electronics, DVD, kitchen, appareal, camera, health, music, toys, video, baby, magazines, software, sports, 目的是把商品评论情感分为积极的和消极的, 这些数据集是根据Blitzer等人^[3]的数据集收集的。另外2个数据集是关于电影评论的, 包括互联网电影资料库(Internet Movie DataBase, IMDB)和电影评论数据集(Movie Review, MR), IMDB首先被Maas等人^[17]提出, 它是用于情感分类的一个基准数据集, 是一个大型的电影评论数据集, 包含完整的评论。MR被Pang等人^[18]提出, 它包含从“烂番茄网”提取的积极和消极评论。这16个不同任务的数据集, 其中每个任务有2000条数据, 积极的和消极的分别为1000条。标记样本被随机分为训练集、验证集和测试集, 比例分别为70%, 20%和10%。

模型的参数最终根据模型的性能以及训练时间选择。本文使用Collobert等人^[12]训练的SENNa词向量, 维度是50维的, 在实验表现出不错的效果。考虑到时间问题, 仅仅使用了最原始的LSTM, 隐藏层的维度设为100。本文分别实验了不同的卷积核尺寸和数量, 最终卷积核尺寸设为(1, 2, 3), 每一个卷积核的个数为100。把dropout值设为0.7能有效地减小过拟合。使用了Adam优化器和反向传播策略对模型进行训练, 样本的批次设置为16, 学习率为0.0005。表1展示了本文所实验过的不同参数设置范围以及模型使用的最终的设置。

4.2 对比模型

LSTM: 长短时记忆网络, 单任务的文本分类模型, 分别用每个任务的样本进行训练和测试。

表1 参数设置

超参数	取值	选择
隐藏层状态维数	50/100/128	100
卷积核大小	1/2/3/4/5	1/2/3
过滤器个数	50/64/100/128/256	100
dropout	0.3/0.4/0.5/0.6/0.7/0.8	0.7
训练次数	10/20/30/40/50	40
批次	8/16/32	16
学习率	0.1/0.01/0.001/0.0005	0.0005

CNN: 卷积神经网络, 它也是单任务文本分类模型, 实验设置与本文相同

MTL-DNN: 共享DNN模型^[5], 使用词袋输入, 共享一个隐藏层。

MTL-CNN: 多任务卷积神经网络模型^[12], 其中查找表部分是共享的, 其他层比如CNN是基于特定任务的。

FS-MTL: 完全共享循环神经网络模型(Full Shared Multi-Task Learning, FS-MTL)^[6], 是将所有的任务共享一个LSTM模型, 每个任务从特定的输出层输出。

ASP-MTL: 对抗多任务学习模型(Adversarial Multi-Task Learning, ASP-MTL)^[7], 使用对抗训练和正交约束, 将共享特征和私有特征划分更细。

IC-MTL: 间接交流多任务学习框架(Indirect Communciation for Multi-Task Learning, IC-MTL)^[15], 是一种图多任务学习框架, 在这个框架中, 不同的任务可以相互通信, 共享层可以发送信息到某一任务中。

4.3 实验结果

4.3.1 模型对比

本文模型MTL-RC与当前热门的模型对比结果在表2中。为了更好地对比, 本文与对比实验采用同样的数据集, 分别是14个亚马逊商品评论数据集和2个电影评论数据集。其中比较流行的多任务学习模型ASP-MTL采用200维glove词向量, 学习率为0.01, 训练最小批次为16。

本文利用分类准确率对实验结果进行评价。其中前两种方法是单任务深度学习模型, 分别对每个任务建模进行情感分类。而其他的模型是多任务学习模型, 对多个任务的数据集进行共同建模, 多任务学习模型的实验结果文献^[7,15]。表2展示了16个任务使用不同的模型的准确率, 之后的讨论都是以每个模型的平均准确率进行对比。从表2可以看出, 除了MR数据集, 在其他数据集中, 本文所提模型MTL-RC是优于其他模型的。MR的提升之所

表 2 与其它模型准确率对比(%)

任务	LSTM	CNN	MTL-DNN	MTL-CNN	FS-MTL	ASP-MTL	IC-MTL	MTL-RC
books	77.8	79.8	82.2	84.5	82.5	84.0	86.2	89.0
electronics	79.8	80.3	81.7	83.2	85.7	86.8	88.5	92.8
DVD	78.0	77.8	84.2	84.0	83.5	85.5	88.0	89.5
kitchen	81.8	78.5	80.7	83.2	86.0	86.2	88.2	91.5
apparel	82.8	82.0	85.0	83.7	84.5	87.0	87.5	89.8
camera	82.5	83.0	86.2	86.0	86.5	89.2	89.0	93.0
health	83.3	84.3	85.7	87.2	88.0	88.2	89.5	92.0
music	77.0	77.8	84.7	83.7	81.2	82.5	85.7	88.0
toys	82.8	80.5	87.7	89.2	84.5	88.0	89.2	91.3
video	85.5	81.8	85.0	81.5	83.7	84.5	86.0	91.0
baby	84.8	81.0	88.0	87.7	88.0	88.2	88.7	91.8
magazines	90.5	86.5	89.5	87.7	92.5	92.2	92.2	93.3
software	84.0	81.5	85.7	86.5	86.2	87.2	87.2	93.3
sports	80.8	81.8	83.2	84.0	85.5	85.7	76.7	91.5
IMDB	76.3	77.5	83.2	86.2	82.5	85.5	86.5	89.0
MR	70.8	67.0	75.5	74.5	74.7	76.7	78.0	75.3
平均数	81.2	80.1	84.3	84.5	84.7	86.1	87.3	90.1

以不明显,甚至相比有的多任务模型略有下降,是因为MR的每条样本长度仅仅只有21,相对其他数据集太小,不容易获得很好的文本表示。相比单任务模型,多任务模型准确率都有一定的提升,单任务LSTM, CNN方法准确率分别为81.2%和80.1%。多任务FS-MTL和MTL-CNN方法准确率分别为84.7%和84.5%,比单任务模型分别提升3.5%和4.4%,而MTL-RC模型比单任务模型分别提升了8.9%和10.0%,由此可见,将多个任务的文本利用起来,能显著提升模型准确率。多任务模型MTL-RC比MTL-DNN, MTL-CNN和FS-MTL都有很大的提升,准确率分别提升了5.8%, 5.6%和5.4%,说明采用多任务学习的方法,将LSTM和CNN网络结合起来的的有效性,能获得更充分的文本表示,提升模型的性能。

在FS-MTL模型中,完全共享一个LSTM层,获取文本表示,每个任务的文本表示分别输入到不同的输出层进行情感分类。可以看出FS-MTL比MTL-DNN和MTL-CNN模型的准确率略有提升,一部分原因是MTL-DNN采用词袋输入丢失了语义信息,MT-CNN仅仅共享了词嵌入部分。ASP-MTL是对FS-MTL的改进,加入了一个私有层提取每个任务的私有特征,并且采用了对抗训练和正交约束将共享层和私有层精确划分,准确率达到了86.1%。与之前模型不同的是IC-MTL可以实现不同任务之间的通信。而本文的模型MTL-RC也是针对FS-

MTL的改进,在LSTM获取文本表示之后加入了不同窗口的多个卷积核并行提取文本的局部特征,因为卷积层是并行提取特征,所以模型的训练速度是很快,MTL-RC的准确率也比FS-MTL, ASP-MTL, IC-MTL分别提升了5.4%, 4%, 2.8%。以上分析表明本文方法优于这些对比方法,证明了本文方法的有效性。

4.3.2 多任务学习模型准确率和速度

本节将继续证明多任务学习模型的有效性,为每一个任务构建一个单任务深度学习模型STL-RC,同样在LSTM层之后采用卷积核提取特征,分别对每个任务进行训练,其他的设置保持不变。对STL-LC平均一次训练的所有任务的时间总和与MTL-RC的时间进行对比,同时也将两个模型平均每个任务的准确率进行对比。MTL-RC与STL-LC模型时间及准确率对比如表3所示。图4展示了其中4个商品评论数据集和2个电影评论数据集在本文提出模型MTL-RC和单任务学习模型STL-LC准确率的对比。

由表3可知,MTL-RC模型平均每个任务的分类准确率比STL-LC提升6.5%,证明了多任务学习能够提高模型的准确率。本文所提模型平均每次训练的时间比每个任务训练STL-RC模型时间总和快213.1 s,可以证明多任务学习方法能够提高模型的训练速度。由图4可以看出,相比STL-RC模型,本文所提MTL-RC模型每个任务的准确率都有

表 3 MTL-LC与STL-LC模型准确率与时间比较

方法	STL-LC	MTL-RC
平均准确率(%)	83.6	90.1
平均1次训练的时间(s)	483.4	270.3

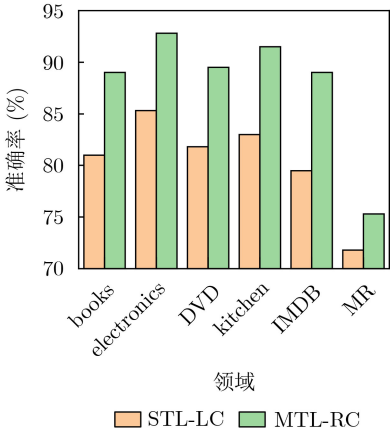


图 4 MTL-RC与STL-LC模型每个领域分类准确率的对比

显著的提升。总之，本文方法在时间和性能上都优于单任务方法。

4.3.3 卷积神经网络的影响

为了避免仅仅输出LSTM最后一个时间步的特征，被最后一个单词主导，本文提出把LSTM的全部隐藏层的输出再通过CNN提取特征的多任务学习模型MTL-RC与仅仅把LSTM最后一个隐藏层输出的模型MTL-LSTM模型进行对比分析，实验结果如图5所示。

由图5可以看出，本文所提模型MTL-RC与改进前的MTL-LSTM相比，每个领域的分类准确率都有一定幅度的提升，只有MR数据集的准确率较低，是因为其文本最短，得到的特征信息较少，改进不是很明显。由此可以说明本文所提模型MTL-RC先把LSTM的全部信息输出，再采用CNN提取文本的局部特征，可以有效地防止特征被最后一个单词所主导，同时能分别利用LSTM和CNN的优

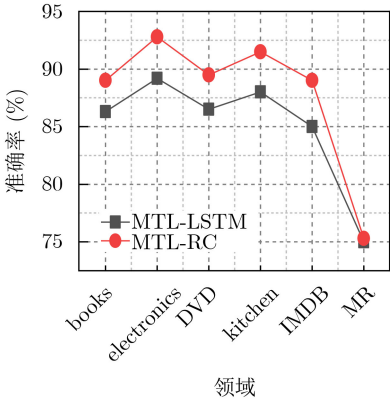


图 5 MTL-RC与MTL-LSTM模型每个领域分类准确率的对比

势，得到深层次的文本语义特征，取得更好的分类效果。

4.3.4 模型参数的影响

为了验证不同窗口卷积核对分类效果的影响，本节采用不同的卷积核窗口进行实验，实验结果如表4所示。

由表4可知，采用不同的卷积核窗口比只使用一种卷积核窗口的分类准确率有一定的提升，当窗口大小选(1, 2, 3)时，比只选用1, 2和3准确率分别提升1.5%, 0.6%和0.9%，也比其他的不同窗口组合准确率高，因此选择合适的卷积核窗口对模型的分性能会有一定的提升。卷积核用来提取文本的局部特征，采用具有相同窗口的不同数量的卷积核可以学习互补的特性。实验结果可以说明采用不同窗口的卷积核可以获得不同高度视野下的局部特征，因此可以获取更加丰富的文本特征。

4.3.5 领域数量对模型性能的影响

本节通过实验探索领域的数量对本文模型性能的影响。在本文实验中，将不同数量领域的文本输入到模型进行训练，以books, electronics, DVD和kitchen这4个领域为基准，领域的数量从1增加到16，步长为3。这4个领域在本文方法的准确率，如图6所示。

由图6所示，随着领域数量的增加，这4个领域的评论分类准确率也在逐渐增加，其中从1个领域到10个领域，准确率上升较为显著。可见领域的数量对本文模型准确率影响很大。因为多任务学习能够起到一个扩充数据量的作用，不同领域的样本很多表达是相似的，具有相关性。领域数量增多，样本的数量也会呈线性增加，可以有效解决有标签样

表 4 MTL-RC模型使用不同卷积核的准确率对比

卷积核	1	2	3	4	5	(3, 4, 5)	(2, 3, 4)	(1, 2, 3)
准确率(%)	88.6	89.5	89.2	89.2	89.4	89.3	89.5	90.1

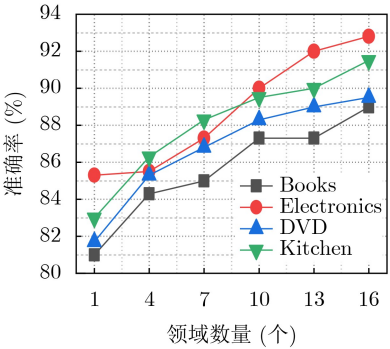


图 6 不同领域数量下模型的准确率

本的问题。多个任务在浅层共享,可以削弱网络的能力,一定程度上防止过拟合。同时多个领域的文本也会有不相关的部分,学习一个任务时与该任务不相关的部分可以作为噪声,提高模型的泛化能力。

5 结 论

本文提出一种基于循环卷积神经网络的多任务学习文本分类模型,将多个文本任务在一个深度学习模型中训练。现有常用的多任务学习模型包括完全共享模型和共享私有模型,共享私有模型例如ASP-MTL模型是加入了对抗网络防止共享私有特征混淆,IC-MTL是图多任务学习模型,使得不同任务共享层和私有层之间可以相互通信。完全共享模型例如MTL-DNN, MTL-CNN, FS-MTL则是分别共享不同的深度学习网络构建多任务学习实现不同的自然语言处理任务。本文则考虑内存和时间问题仅仅构建完全共享模型也获得了很好的效果。与过去的完全共享模型不同的是,本文模型通过LSTM和不同尺寸的卷积核提取丰富的语义信息,能够同时利用多任务学习及深度学习模型的优势。在多领域文本分类数据集上的实验结果表明,本文所提MTL-RC模型准确率比单任务STL-LC模型提高了6.5%,训练1次的时间快了213.1 s。单任务学习模型仅仅采用单个小样本数据集,因此标签数据量是比较少的,在同等实验参数和条件下,多任务学习模型比单任务学习准确率提高了很多而且有更快的速度。MTL-RC模型比最新的多任务学习模型FS-MTL, ASP-MTL, IC-MTL分别提升了5.4%, 4%和2.8%。因此本文所提模型在解决有标签数据不足的问题上准确率更高,有更好的成效。

参 考 文 献

- [1] 谢金宝, 侯永进, 康守强, 等. 基于语义理解注意力神经网络的多元特征融合中文文本分类[J]. 电子与信息学报, 2018, 40(5): 1258–1265. doi: [10.11999/JEIT170815](https://doi.org/10.11999/JEIT170815).
XIE Jinbao, HOU Yongjin, KANG Shouqiang, *et al.* Multi-feature fusion based on semantic understanding attention neural network for Chinese text categorization[J]. *Journal of Electronics & Information Technology*, 2018, 40(5): 1258–1265. doi: [10.11999/JEIT170815](https://doi.org/10.11999/JEIT170815).
- [2] KIM Y. Convolutional neural networks for sentence classification[C]. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014: 1746–1751. doi: [10.3115/v1/D14-1181](https://doi.org/10.3115/v1/D14-1181).
- [3] BLITZER J, DREDZE M, and PEREIRA F. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification[C]. The 45th Annual Meeting of the Association of Computational Linguistics (ACL), Prague, Czech Republic, 2007: 440–447.
- [4] CARUANA R. Multitask learning[J]. *Machine Learning*, 1997, 28(1): 41–75. doi: [10.1023/A:1007379606734](https://doi.org/10.1023/A:1007379606734).
- [5] LIU Xiaodong, GAO Jianfeng, HE Xiaodong, *et al.* Representation learning using multi-task deep neural networks for semantic classification and information retrieval[C]. The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), Denver, USA, 2015: 912–921.
- [6] LIU Pengfei, QIU Xipeng, and HUANG Xuanjing. Recurrent neural network for text classification with multi-task learning[C]. The Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI), New York, USA, 2016: 2873–2879.
- [7] LIU Pengfei, QIU Xipeng, and HUANG Xuanjing. Adversarial multi-task learning for text classification[C]. The 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada, 2017: 1–10.
- [8] HOCHREITER S and SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735–1780. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [9] 王鑫, 李可, 宁晨, 等. 基于深度卷积神经网络和多核学习的遥感图像分类方法[J]. 电子与信息学报, 2019, 41(5): 1098–1105. doi: [10.11999/JEIT180628](https://doi.org/10.11999/JEIT180628).
WANG Xin, LI Ke, NING Chen, *et al.* Remote sensing image classification method based on deep convolution neural network and multi-kernel learning[J]. *Journal of Electronics & Information Technology*, 2019, 41(5): 1098–1105. doi: [10.11999/JEIT180628](https://doi.org/10.11999/JEIT180628).
- [10] KALCHBRENNER N, GREFFENSTETTE E, and BLUNSOM P. A convolutional neural network for modelling sentences[C]. The 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Baltimore, USA, 2014: 655–665.
- [11] 刘宗林, 张梅山, 甄冉冉, 等. 融入罪名关键词的法律判决预测多任务学习模型[J]. 清华大学学报: 自然科学版, 2019, 59(7): 497–504. doi: [10.16511/j.cnki.qhdxxb.2019.21.020](https://doi.org/10.16511/j.cnki.qhdxxb.2019.21.020).
LIU Zonglin, ZHANG Meishan, ZHEN Ranran, *et al.* Multi-task learning model for legal judgment predictions with charge keywords[J]. *Journal of Tsinghua University: Science and Technology*, 2019, 59(7): 497–504. doi: [10.16511/j.cnki.qhdxxb.2019.21.020](https://doi.org/10.16511/j.cnki.qhdxxb.2019.21.020).
- [12] COLLOBERT R and WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]. The 25th International Conference on Machine Learning (ICML), Helsinki, Finland, 2008: 160–167.

- [13] LI Shoushan, HUANG Churen, and ZONG Chengqing. Multi-domain sentiment classification with classifier combination[J]. *Journal of Computer Science and Technology*, 2011, 26(1): 25–33. doi: [10.1007/s11390-011-9412-y](https://doi.org/10.1007/s11390-011-9412-y).
- [14] LIU Pengfei, QIU Xipeng, and HUANG Xuanjing. Deep multi-task learning with shared memory for text classification[C]. 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), Austin, USA, 2016: 118–127.
- [15] LIU Pengfei, FU Jie, DONG Yue, *et al.* Learning multi-task communication with message passing for sequence learning[C]. AAAI Conference on Artificial Intelligence (AAAI), Palo Alto, USA, 2019: 4360–4367.
- [16] YUAN Zhigang, WU Sixing, WU Fangzhao, *et al.* Domain attention model for multi-domain sentiment classification[J]. *Knowledge-Based Systems*, 2018, 155: 1–10. doi: [10.1016/j.knosys.2018.05.004](https://doi.org/10.1016/j.knosys.2018.05.004).
- [17] MAAS A L, DALY R E, PHAM P T, *et al.* Learning word vectors for sentiment analysis[C]. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL), Portland, USA, 2011: 142–150.
- [18] PANG Bo and LEE L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales[C]. The 43rd Annual Meeting on Association for Computational Linguistics (ACL), Ann Arbor, USA, 2005: 115–124.
- 谢金宝：男，1980年生，副教授，研究方向为自然语言处理、人工智能。
- 李嘉辉：男，1995年生，硕士生，研究方向为自然语言处理、人工智能。
- 康守强：男，1980年生，教授，研究方向为智能诊断、人工智能。
- 王庆岩：男，1984年生，副教授，研究方向为智能诊断、人工智能、智能图像处理。
- 王玉静：女，1983年生，副教授，研究方向为智能诊断、人工智能。

责任编辑：余 蓉