

基于迁移学习的微博谣言检测方法

沈瑞琳, 潘伟民, 张海军

(新疆师范大学 计算机科学技术学院, 新疆 乌鲁木齐 830054)

摘要:为解决基于深度神经网络的微博谣言检测工作中带标签数据稀缺的问题,提出一种基于迁移学习的微博谣言检测方法。利用双层双向的门控循环单元和卷积神经网络组成的联合模型作为特征提取器,利用丰富的评论数据对联合神经网络进行预训练,将训练好的特征提取层迁移到微博谣言检测任务中,通过区分微调 and 斜三角学习率两种微调策略对特征提取层进行调整,使其适应于目标任务。实验结果表明,采用迁移学习方法的联合神经网络能有效提高微博谣言检测的准确率。

关键词: 谣言检测; 迁移学习; 微博; 深度神经网络; 预训练

中图分类号: TP389.1 **文献标识码:** A **文章编号:** 1000-7024 (2021) 12-3534-06

doi: 10.16208/j.issn1000-7024.2021.12.031

Detection method of micro blog rumor based on transfer learning

SHEN Rui-lin, PAN Wei-min, ZHANG Hai-jun

(School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China)

Abstract: To solve the problem of the scarcity of tagged data in the detection of micro blog rumors based on deep neural network, a method of micro blog rumors detection based on transfer learning was proposed. The joint model composed of two-layer bi-directional gating loop unit and convolutional neural network was used as feature extractor, rich review data were used to pre train the joint neural network, the trained feature extraction layer was migrated to the micro blog rumor detection task, and the feature extraction layer was adjusted by distinguishing the fine-tuning and the oblique triangle learning rate to make it suitable for the target task. Experimental results show that the combined neural network based on transfer learning method can improve the accuracy of rumor detection effectively.

Key words: rumor detection; transfer learning; Weibo; deep neural network; pre training

0 引言

由于网络的开放性、微博平台的言论自由性等特点,微博在方便人们日常生活的同时,也为谣言的产生提供了便利场所。本文中谣言指未经证实的信息,即在人与人之间传播,与公众关注的对象、事件或问题有关,并且在没有被权威机构证实的情况下流传的信息,因此,信息可能是真的,也可能是假的。

现有的谣言检测方法大致分为3种:一是基于人工的方法,主要依靠人的经验对事件的真实性做出判断,例如@微博辟谣、@谣言粉碎机,不仅耗费大量的人力和物力,

还导致了更长的延迟。二是基于传统机器学习的方法,根据消息内容、用户信息、传播模式等进行分析来人工构造特征,通过人工特征提取数据中的关键信息^[1-5]。三是基于深度学习的方法,该方法不需要特征工程,同时可以挖掘到不易被人们发现的深层特征^[6-11]。深度学习方法需要依赖大量的带标签数据才能学到更深层的特征表示,目前在微博谣言检测工作中仅有少量的带标签数据。因此,解决深度学习模型中带标签数据少的问题是如今热点研究问题之一。对于标注数据少的问题,研究者展开了大量研究。起初采用无监督的方法,但是由于数据没有标注,导致分类歧义性较高。近几年,随着迁移学习的应用,许多领域

收稿日期: 2020-08-05; **修订日期:** 2020-12-10

基金项目: 2020年新疆维吾尔自治区研究生教育改革创新计划基金项目(XJ2020G235);新疆师范大学重点实验室基金项目(XJNUSYS2019B13);国家自然科学基金委NSFC-新疆联合基金重点支持基金项目(U1703261)

作者简介: 沈瑞琳(1995-),女,河北保定人,硕士研究生,研究方向为计算机应用技术和网络信息安全;潘伟民(1963-),男,新疆乌鲁木齐人,硕士,教授,硕士生导师,研究方向为计算机应用技术和网络信息安全;张海军(1973-),男,吉林公主岭人,博士,教授,硕士生导师,研究方向为自然语言处理、人工智能、信息抽取和谣言检测。E-mail: panweiminss@163.com

标注数据少的问题开始得到有效解决。迁移学习是运用已有知识对不同但相关领域问题求解的一种机器学习方法^[12], 打破了传统机器学习中训练数据和测试数据必须满足独立同分布的假设, 并且解决了标注数据不足的问题。如可以用来辨识自行车的知识也可以用来提升识别摩托车的能力。采用迁移学习方法借助相关领域丰富的数据资源, 对于解决微博谣言检测中带标签数据少的问题提供了很好的研究思路。本文将迁移学习技术应用于微博谣言检测中, 利用相关领域中充足的带标签数据辅助微博中少量的标签数据, 进行谣言检测。

1 相关工作

社交媒体上的谣言自动检测一直是近年来一个研究热点。传统的谣言检测方法主要利用人工构造特征, 再采用机器学习模型学习文本的浅层特征。最早的自动谣言检测方法源于2011年Castillo等^[1]对Twitter中信息可信度的检测, 该方法首先利用特征工程构造特征, 然后采用支持向量机(SVM)对文本进行检测。Yang等^[2]在2012年提出基于微博的谣言检测方法, 该方法利用微博中涉及的地理位置、发文客户端信息、文本符号的情感极性特征, 采用SVM构造微博谣言分类器模型。后人在此基础上展开了对Twitter和微博中谣言检测的研究^[3-5]。以上方法都需要特征工程的参与, 不仅耗时费力, 且仅能学到文本的浅层特征。

随着深度神经网络模型在很多领域取得了不错的成果^[13,14], 研究者开始将深度神经网络应用到微博谣言检测领域。Ma等^[6]提出基于深度神经网络模型的微博谣言检测方法, 作者实现了tanh-RNN、长短期记忆网络(LSTM)、门控循环单元(GRU)、双层GRU这4种模型, 由于循环神经网络(RNN)存在梯度消失和梯度爆炸的问题, 其它3种模型的性能普遍高于tanh-RNN, 在单层网络模型中GRU比LSTM性能略好, 与双层网络模型比较, 由于双层GRU可以提取更深层的语义特征, 因此性能最优, 这也说明, 使用相同的神经网络, 网络层数越多模型性能越好。Sampson等^[7]通过利用少量会话之间的隐式链接提高了早期谣言的检测精度。Ruchansky等^[8]提出了一种将文章文本、用户的响应以及来源用户3种特征结合起来的混合模型, 结果优于仅使用单一特征和模型的方法。Yu等^[9]将各时间段文本向量拼接成事件的特征矩阵, 并采用卷积神经网络(CNN)学习事件的隐层表示。Zhou等^[10]通过强化学习实现谣言早期检测。Li等^[11]利用内容、用户可信度和传播信息在社交媒体上发现谣言。这些方法主要依靠公开数据集进行实验, 也有研究者对数据进行了扩充, 但都是有限的数据扩充。微博谣言检测仍然面临数据少的困境。虽然基于深度学习的方法在微博谣言检测任务取得了一定的进展, 但是深度学习模型对大量标注数据的需求也制约了

深度学习在该领域的进一步发展。

近几年, 随着迁移学习的不断发展^[12-16], 研究者开始将其应用于谣言检测领域。Ma等^[17]将多任务学习应用于Twitter谣言检测任务, 通过共享多个任务的通用知识, 使多个任务同时获得较好的效果。Wen等^[18]提出了一种基于跨语言跨平台的社交媒体谣言检测方法, 在谣言检测中加入其它平台与该事件相关的信息, 来提高检测结果的真实性。刘等^[19]将多任务学习应用于Twitter中的分领域谣言检测, 通过领域适配技术使源领域数据与目标领域的数据分布趋于相似。郭^[20]将模型迁移应用于Twitter谣言检测任务, 首先利用包含大量标签数据的评论数据集对模型进行训练, 然后利用模型迁移, 使模型适用于Twitter谣言检测任务, 提高了Twitter谣言检测任务的准确率, 同时验证了评论数据对谣言检测任务的有用性。

为解决带标签数据少和检测准确率不高的问题, 本文将模型迁移应用到微博谣言检测领域, 利用大量带标签的评论数据辅助微博谣言检测任务。在模型迁移中, 微调学习率的设置决定了迁移效果的好坏, 本文将区分微调和斜三角学习率两种微调策略相结合, 为每一层设置不同的学习率, 以保留先前的知识, 避免灾难性遗忘, 并根据目标任务的需求对学习率进行调整。

2 基于迁移学习的微博谣言检测模型

本文提出的基于迁移学习的微博谣言检测模型(transferring learn-BiGRU-2-CNN, TB₂GC)模型如图1所示。按照自下而上、自左到右的顺序对模型进行介绍, 大致有4个模块, 分为3个步骤。首先利用丰富的评论数据对模型进行预训练, 然后将训练好的特征提取层迁移到目标任务中, 再通过微调策略对特征提取模块进行调整, 使其适应于目标任务。

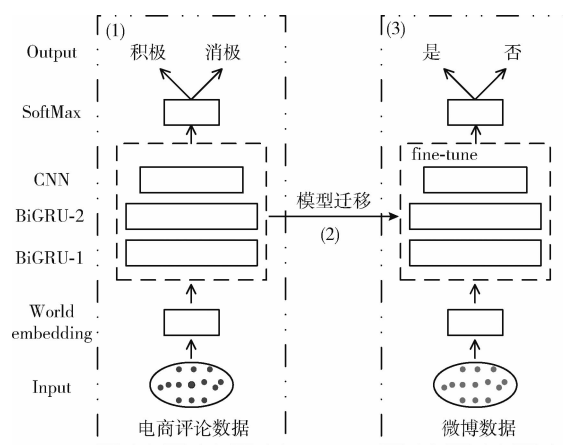


图1 TB₂GC模型结构

2.1 词嵌入

使用低维向量代替文本中词的表示是目前自然语言处

理中的常见方式。本文将微博文本数据输入到开源的 word2vec 模型对文本进行向量化,向量的维数为 300,该模型由 Google News 利用 1000 亿个单词训练而成,并使用字结构的连续文本进行训练^[21]。未出现在预先训练的词集合中的词是随机初始化的。

2.2 特征提取

鉴于 BiGRU₂ 和 CNN 各自的特点,本文采用双层 BiGRU 和 CNN 的联合模型作为特征提取器,特征提取网络模型如图 2 所示。将 Word2vec 输出的词向量输入到 BiGRU₂-CNN 联合神经网络中,提取文本数据的全局特征,提取的特征将用于后续分类器的分类工作。

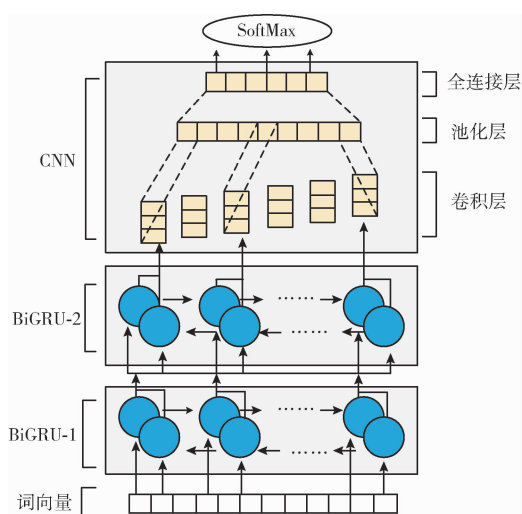


图 2 联合神经网络模型

双层双向门控循环单元(BiGRU₂):通过 BiGRU₂模型获取微博文本在时间序列上的深层特征。循环神经网络(RNN)中两个节点之间的连接形成了一个内部循环结构,这种结构使它能够捕捉文本的动态时间信号特征。由于 RNN 模型存在梯度消失和梯度爆炸的问题,通过改进得到了 LSTM 模型结构,LSTM 模型结构复杂、模型参数多、训练时间长。随着样本数量的增加,导致训练时间延长,参数变多,内部计算复杂度提高。对此研究者提出了 GRU 网络模型,GRU 模型不仅可以达到 LSTM 的效果,并且结构简单、参数少、收敛性好。GRU 模型由两个门组成,一个更新门和一个重置门,更新门决定了前一个输出隐藏层对当前层的影响程度,值越大,影响越大。重置门决定忽略以前隐藏层信息的范围,值越小,信息就越容易被忽略。

GRU 只能从前到后获取信息,不能从后到前对信息间的依赖关系进行提取。BiGRU 是由两个方向相反的 GRU 模型组成的双向网络结构,可以双向的获取前后文的依赖关系,这对获得更多与任务相关的特征非常有利。研究表明,深层网络结构有助于获取深层特征,可以提高分类的

效果,因此本文采用双层的 BiGRU 网络结构获取数据的全局特征。

CNN:通过 CNN 模型获取微博文本的局部特征。CNN 模型最初是为计算机视觉而发明的,后来被证明对自然语言处理(NLP)领域有效,已经在语义分析、搜索查询检索、句子建模和其它传统 NLP 任务中取得了优异的成果。CNN 利用由多个相互转换的层组成的计算模型来学习具有多个抽象级别的数据表示,通过发现大数据集中复杂的结构,极大地提高了图像识别、视觉对象识别和句子分类的技术水平。CNN 常用的体系结构包括卷积层、池化层和全连接层,本文将带滤波器的卷积层应用于局部特征的提取,将池化层用来提高模型的容错性,然后通过全连接层输出隐层特征,再利用 Softmax 函数进行分类结果的输出。

2.3 模型迁移

首先利用丰富的评论数据对 TB₂GC 神经网络模型进行预训练,预训练可以获取文本的通用语言信息,如情感倾向、上下与依赖关系、深层语义表示等。然后针对目标谣言检测任务对特征提取层进行微调,由于不同的层捕获不同类型的信息,因此应该根据情况为每一层设置不同的学习率,对此本文采用区分性微调策略。为了使模型在训练开始时就能够快速收敛到合适的参数空间,本文采用斜三角形学习率(slanted triangle learning rates, STLRL)^[22]策略。

区分性微调:与对模型的所有层使用相同的学习率不同,区分性微调能够用不同的学习率来调整每个层,从而根据不同层对目标任务的贡献设置不同的学习率,贡献大的层设置较小的学习率,贡献小的层设置较小的学习率。第 L 层模型的参数 θ 在时间 t 的更新如式(1)所示

$$\theta_t^L = \theta_{t-1}^L - \eta^L \cdot \nabla_{\theta} J(\theta) \quad (1)$$

式中: η 是学习率, $\nabla_{\theta} J(\theta)$ 是关于模型目标函数的梯度。在区分性微调中,将参数 θ 分解为 $\{\theta^1, \dots, \theta^L\}$,其中 θ^L 是第 L 层模型的参数, L 是模型的层数, η 分解为 $\{\eta^1, \dots, \eta^L\}$,其中 η^L 是第 L 层的学习率。

由于神经网络具有浅层网络提取文本的浅层特征,深层网络提取文本的深层隐含特征的特点,而且深层特征在 NLP 中更具有通用性,因此选择最后一层的学习率设为 η^L ,较低层的学习率为 $\eta^{L-1} = \eta^L / 2.3$ 。

斜三角形学习率:与微调过程中使用相同的学习率或仅递增或仅递减的方式不同,斜三角形学习率先线性增加学习率,然后再线性衰减,有助于模型快速收敛到合适的范围,并在学习率下降的过程中达到适应目标任务的最佳准确率,具体方案如式(2)所示

$$\begin{aligned}
 cut &= \lfloor T \cdot cut_frac \rfloor \\
 p &= \begin{cases} t/cut, & \text{if } t < cut \\ 1 - \frac{t-cut}{cut \cdot (ratio-1)}, & \text{otherwise} \end{cases} \quad (2) \\
 \eta_t &= \eta_{\max} \cdot \frac{1+p \cdot (ratio-1)}{ratio}
 \end{aligned}$$

式中: T 是训练迭代次数, cut_frac 是使学习率增加的迭代次数占总迭代次数的比例, cut 是学习率开始下降时的迭代次数, $ratio$ 指最小学习率与最大学习率的比值, η_t 是迭代 t 时的学习率。通常使用 $cut_frac = 0.1$, $ratio = 32$, $\eta_{\max} = 0.01$ 。

在学习率不断增加的过程中观察准确率的变化, 当准确率第一次出现下降时, 学习率也开始线性减小。即学习率出现拐点。

通过斜三角学习率和区分微调, 已经将初始模型的特征提取层有效迁移到了微博谣言检测任务中。

2.4 分类器

将微调后的特征提取层提取的特征输入到 $softmax$ 层, 神经元的激活函数使用线性修正单元函数 (rectified linear units, ReLU)。ReLU 函数定义为 $f(x) = \text{Softmax}(0, x)$, 该激活函数在具有深层体系结构的网络中通常会使网络学习的更快。最后输出对一条事件是否为谣言的检测结果。

3 实验与分析

3.1 实验数据集

源数据集选用 Zhang 等^[23] 在 2014 年收集的评论数据, 该数据集来自 DianPing.com, 包括 510 071 个用户对 209 132 个商家的 3 605 300 条评论。

目标数据集选用 Ma 等^[7] 在 2016 年公开的新浪微博数据, 该数据集包含微博和 Twitter 两部分, 微博谣言数据来自新浪微博平台已经确认的微博谣言事件, 作者按照谣言数据的数量利用网络爬虫在微信平台爬取了相似数量的非谣言数据。共包含 2313 个谣言和 2351 个非谣言。本文保留 10% 的事件作为验证集, 其余数据按照 3:1 的比例分割用于训练集和测试集。

为了提高数据的质量, 对源数据和目标数据集进行了去噪处理。利用正则表达式去除了数据中的 @ 符号、@ 的内容、空格、空行、URL 信息等。本文中并没有去掉表情符号, 因为如今表情符号已经成为人们在网络平台表达自己感情倾向的一种重要形式, 深度神经网络也可以根据表情符号挖掘深层情感特征, 因此, 这里保留了文本中的表情符号。

3.2 实验对比

(1) 本模型和其它基线模型对比

本文将 TB_2GC 模型方法与以下几个基线方法进行比较:

DT-Rank^[1]: 该方法通过对有争议的微博信息进行聚

类, 然后根据统计特征对聚类结果进行排序, 以识别趋势性谣言。

DTC^[4]: 该方法对 15 个评判特征进行分析, 并将 J48 决策树应用于谣言检测任务。

SVM-TS^[3]: 该方法利用时间序列对人工构造的特征集进行建模, 利用线性支持向量机分类器进行分类预测。

GRU、GRU-2^[6]: Ma 等在 2016 年提出的基于深度学习的模型中, 分别实现了 LSTM、单层 GRU 和双层 GRU, 证明了深度学习模型在谣言检测中的优势。

CNN^[8]: 该方法设计 3CAMI 模型, 将各时间段文本向量拼接成事件的特征矩阵, 采用 CNN 学习事件的隐层表示。

TB_2GC 模型与各基线模型的实验结果对比见表 1。

表 1 TB_2GC 模型与基线模型的实验对比结果

方法	准确率	精确率	召回率	F1
DT-Rank	0.732	0.738	0.715	0.726
DTC	0.831	0.847	0.815	0.831
SVM-TS	0.857	0.839	0.885	0.861
GRU	0.908	0.871	0.958	0.913
GRU-2	0.910	0.876	0.956	0.914
CNN	0.933	0.921	0.945	0.933
TB_2GC	0.962	0.953	0.963	0.958

表 1 展示了本文模型与各基线模型的对比结果。为了更全面分析传统机器学习方法、深度学习方法、迁移学习在谣言检测中的效果, 本文在传统机器学习方法和深度学习方法中各选取了 3 个基线模型, 表 1 从上到下依次为传统机器学习模型、传统深度学习模型、本文的迁移学习模型。

在 3 种传统机器学习模型中, SVM-TS 的效果最佳, 准确率达到 85.7%, 在 3 种深度学习方法中 CNN 的效果最佳, 准确率达到 93.3%。相比于 3 种深度学习方法, SVM-TS 的效果却是最差的, 深度学习模型 GRU 的准确率比 SVM-TS 高出 5.1 个百分点, 由此可见, 通过深度神经网络模型提取的特征优于人工构造的特征。本文提出的基于迁移学习的方法在准确率上比最好的基线模型 CNN 高出 2.9 个百分点, 在精确率上高出 3.2 个百分点, 在召回率上高出 1.8 个百分点, 在 F1 值上高出 2.5 个百分点。实验结果表明, 本文提出的神经网络模型表现出了良好的性能。原因可能在于深度学习基线方法中, 研究者仅在现有的公开数据集上进行研究, 忽略了数据集对深层特征提取的重要性, 因此效果不佳。

(2) 模型组合对比

为了验证提出的联合模型的组合方式的有效性, 将模型拆分为不同的形式, 再结合迁移学习进行实验, 与本文联合模型进行对比, 结果见表 2。

表 2 TB₂GC 模型与分解模型的实验对比结果

方法	准确率	精确率	召回率	F1
GRU	0.941	0.947	0.926	0.937
CNN	0.946	0.945	0.938	0.941
BiGRU	0.953	0.952	0.951	0.948
BiGRU-CNN	0.957	0.951	0.954	0.951
TB ₂ GC	0.962	0.958	0.963	0.961

表 2 通过对模型的拆分部分进行实验,验证了本文所提出的模型组合的有效性。可以看出,3 种使用单一神经网络模型的方法中 BiGRU 的效果最佳,准确率达到 95.3%,当增加 CNN 模块时,联合模型的准确率增加了 0.2 个百分点,因为 CNN 有利于提取文本中的局部特征,使特征提取更全面。当再增加第二层 BiGRU 时,准确率提高了 0.5 个百分点,由此可见,在数据量足够的情况下,深层神经网络模型对检测结果更有利。

(3) 源数据集的数量对迁移效果的影响

为了探究源数据集的数量对迁移效果的影响,随机抽取源数据中的 60 万条、120 万条、180 万条和 240 万条数据进行实验,观察不同数据量对迁移学习效果的影响,对比结果如图 3 所示。

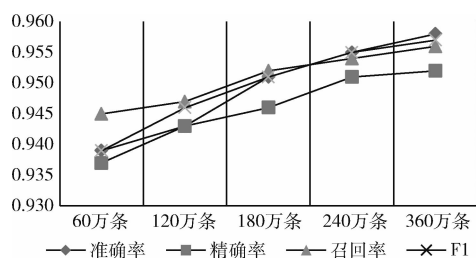


图 3 源数据的数量对迁移效果的影响

图 3 展示了不同量的源数据对迁移效果的影响,实验结果显示随着源数据集数据量的增加,准确率也在增加,表明使用大量带标签数据进行迁移学习的效果更好。

(4) 目标数据集的数量对迁移效果的影响

为进一步探究目标数据集的数量对迁移效果的影响,从目标数据集随机抽取 500 条和 2000 条数据进行实验,同样保留 10% 的事件作为验证集,其余数据按照 3:1 的比例分割用于训练和测试集。结果显示 500 条数据的准确率与本文最好的基线模型 CAMI 仅差 0.004 个百分点,于是又随机选取了 400 条数据进行实验,对比结果如图 4 所示。

图 4 展示了迁移学习在不同量的目标数据集集中的效果,实验结果显示随着目标数据量的减少,准确率也在减小,当数据量为 400 条时,准确率首次出现低于基准模型的现象,表明本文模型适用于目标数据大于 400 条的自然语言处理任务。

3.3 实验结果分析

通过分析 TB₂GC 模型与各基线模型的对比实验和联合

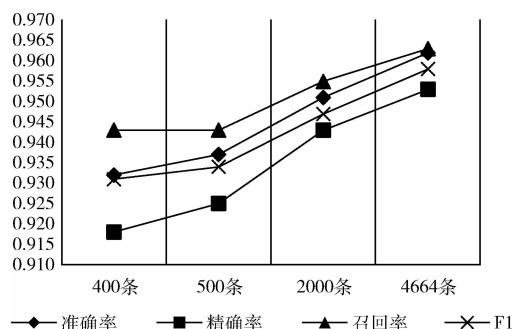


图 4 目标数据的数量对迁移效果的影响

神经网络的各种拆解模型的对比实验,表明采用联合神经网络模型比仅采用单一的神经网络能获取更全面的特征,并且表明更深层的神经网络模型可以提取更多的特征。迁移学习的应用则是有助于进一步加深神经网络的深度,这对学习更深层的特征表示提供了帮助,解决了基于深度学习中的微博谣言检测中带标签数据少的问题。实验结果表明,无论是迁移学习方法的应用,还是神经网络模型的组合,在微博谣言检测任务中都表现出了良好的效果。

此外,本文还对数据集的数量对迁移效果的影响进行了分析。通过将源数据集和目标数据集进行分割实验,结果表明,在本文提出的神经网络模型中,无论是源数据集还是目标数据集,更多的数据量,会使迁移效果更好。

4 结束语

本文将迁移学习方法应用到微博谣言检测领域,利用丰富的电商评论数据辅助微博谣言检测任务进行学习,解决了微博谣言检测领域带标签数据少的问题。实验结果显示基于迁移学习的方法在准确率、精确率和 F1 值 3 个方面都优于基线方法,表明使用相关数据集进行迁移是一种很好的策略。当然,所提出的方法还有很多不足,例如微调策略、源数据集的选择、特征提取网络的设计等方面都可以做进一步的调整。

参考文献:

- [1] Castillo C, Mendoza M, Poblete B. Information credibility on twitter [C] //Proc of International Conference on World Wide Web. Hyderabad: ACM Press, 2011: 675-684.
- [2] Yang Fan, Liu Yang, Yu Xiaohui, et al. Automatic detection of rumor on Sina Weibo [C] //Proc of the ACM SIGKDD Work Shop on Mining Data Semantics. Beijing: ACM Press, 2012: 13-20.
- [3] Ma Jing, Gao Wei, Wei Zhongyu, et al. Detect rumors using time series of social context information on microblogging web-sites [C] //Proc of the 24th ACM International Conference on Information and Knowledge Management. Melbourne: ACM Press, 2015: 1751-1754.

- [4] Zhao Zhe, Resnick P, Mei Qiaozhu. Enquiring minds: Early detection of rumors in social media from enquiry posts [C] // Proc of the 24th International Conference on World Wide Web. Florence: ACM Press, 2015: 1395-1405.
- [5] Ma Jing, Gao Wei, Wong KF. Detect rumors in microblog posts using propagation structure via kernel learning [C] // Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL Press, 2017: 708-717.
- [6] Ma Jing, Gao Wei, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C] // Proc of the 25th International Joint Conference on Artificial Intelligence, 2016: 3818-3824.
- [7] Sampson J, Morstatter F, Wu Liang, et al. Leveraging the implicit structure with in social media for emergent rumor detection [C] // Proc of the 25th ACM International on Conference on Information and Knowledge Management. ACM Press, 2016: 2377-2382.
- [8] Ruchansky N, Seo S, Liu Yan. Csi: A hybrid deep model for fake news detection [C] // Proc of the ACM on Conference on Information and Knowledge Management. ACM Press, 2017: 797-806.
- [9] Yu Feng, Liu Qiang, Wu Shu, et al. A convolutional approach for misinformation identification [C] // Proc of the 26th International Joint Conference on Artificial Intelligence, 2017: 3901-3907.
- [10] Zhou Kaimin, Shu Chang, Li Binyang, et al. Early rumour detection [C] // Proc of the Conference of the North American Chapter of the Association for Computational. Minneapolis: ACL Press, 2019: 1614-1623.
- [11] Li Quanzhi, Zhang Qiong, Si Luo. Rumor detection on social media by exploiting content, user credibility and propagation information [C] // Proc of the 13th International Workshop on Semantic Evaluation. Minneapolis: ACL Press, 2019: 855-859.
- [12] GUO Zengrong. Research and implementation of micro blog news rumor detection based on deep learning [D]. Beijing: Beijing University of Posts and Telecommunications, 2019: 1-65 (in Chinese). [郭增荣. 基于深度学习的微博新闻谣言检测研究与实现 [D]. 北京: 北京邮电大学, 2019: 1-65.]
- [13] Chen Huimin, Sun Maosong, Tu Cunchao, et al. Neural sentiment classification with user and product attention [C] // Proc of Conference on Empirical Methods in Natural Language Processing. Austin: ACL Press, 2016: 1650-1659.
- [14] Yang Zichao, Yang Diyi, Dyer C, et al. Hierarchical attention networks for document classification [C] // Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL Press, 2016: 1480-1489.
- [15] Misra I, Shrivastava A, Gupta A, et al. Cross-stitch networks for multi-task learning [C] // Proc of Conference of the Computer Vision and Pattern Recognition. NJ: IEEE, 2019: 3994-4003.
- [16] Xiao Liqiang, Zhang Honglun, Chen Wenqing, et al. Learning what to share: Leaky multi-task network for text classification [C] // Proc of the 27th International Conference on Computational Linguistics. NY: ACM, 2018: 2055-2065.
- [17] Ma Jing, Gao Wei, Wong KF. Detect rumor and stance jointly by neural multi-task learning [C] // Proc of the WWW '18 Companion. Lyon: ACM Press, 2018: 585-593.
- [18] Wen Weiming, Su Songwen, Yu Zhou. Cross-lingual cross-platform rumor verification pivoting on multimedia content [C] // Proc of the Conference on Empirical Methods in Natural Language Processing, 2018: 3487-3496.
- [19] LIU Kan, DUI Haochen. Research on Twitter rumor detection based on deep migration network [J]. Data Analysis and Knowledge Discovery, 2019, 34 (10): 47-55 (in Chinese). [刘勘, 杜好宸. 基于深度迁移网络的 Twitter 谣言检测研究 [J]. 数据分析与知识发现, 2019, 34 (10): 47-55.]
- [20] GUO Mengjie. Research on rumor detection mechanism based on transfer learning [D]. Inner Mongolia: Inner Mongolia University of Technology, 2019: 1-54 (in Chinese). [郭孟杰. 基于迁移学习的谣言检测机制研究 [D]. 内蒙古: 内蒙古工业大学, 2019: 1-54.]
- [21] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [C] // Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013: 3111-3119.
- [22] Howard J, Ruder S. Universal language model fine-tuning for text classification [C] // Proc of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1-12.
- [23] Zhang Yongfeng, Lai Guokun, Zhang Min, et al. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis [C] // Proc of the 37th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval. Gold Coast: ACM Press, 2014: 83-92.