计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201200007

基于 DeepFM 和卷积神经网络的集成式多模态谣言检测方法

陈志毅　隋　杰
中国科学院大学工程科学学院　北京 100049
(18811722686@163.com)

摘　要　随着以微博为代表的社交媒体越来越流行,谣言信息借助社交媒体迅速传播,容易造成严重的后果,因此自动谣言检测问题受到了国内外学术界、产业界的广泛关注。目前,越来越多的用户使用图片来发布微博,而不仅仅是文本,微博通常由文本、图像和社会语境组成。因此,文中提出了一种基于深度神经网络,针对配文文本内容、图像以及用户属性信息的多模态网络谣言检测方法 DCNN。该方法由多模态特征提取器和谣言检测器组成,多模态特征提取器分为 3 部分,即基于 TextCNN 的文本特征提取器、基于 VGG-19 的图片特征提取器和基于 DeepFM 算法的用户社会特征提取器,分别用于学习微博不同模态上的特征表示,以形成重新参数化的多模态特征,特征融合后将该融合后的多模态特征作为谣言检测器的输入进行分类检测。在微博数据集上对该算法进行了大量实验,实验结果表明 DCNN 算法将识别准确率从 78.1%提高到了 80.3%,验证了 DCNN 算法和其中对社会特征建立特征交互方法的可行性与有效性。

关键词:多模态;谣言检测;DeepFM;卷积神经网络;社会特征;自然语言处理
中图法分类号　TP391

DeepFM and Convolutional Neural Networks Ensembles for Multimodal Rumor Detection

CHEN Zhi-yi and SUI Jie
School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract　With the increasing popularity of social media represented by Weibo, rumors spread rapidly through social media, which is more likely to cause serious consequences. The problem of automatic rumor detection has attracted widespread attention from academic and industrial circles at home and abroad. We have noticed that more and more users use pictures to post Weibo, not just text. Weibo usually consists of text, images and social context. Therefore, a multi-modal network rumor detection method DCNN based on deep neural network for the text content, image and user attribute information of the accompanying text is proposed. This method consists of a multi-modal feature extractor and a rumor detector. The multi-modal feature extractor is divided into three parts: a text feature extractor based on TextCNN, a picture feature extractor based on VGG-19, and a user social feature extractor based on DeepFM algorithm. These three parts learn feature representations on different modalities of Weibo to form re-parameterized multi-modal features, which are fused as input to the rumor detector classification detection. This algorithm has carried out a large number of experiments on the Weibo data set, and the experimental results show that the recognition accuracy of DCNN algorithm is improved from 78.1% to 80.3%, which verifies the feasibility and effectiveness of DCNN algorithm and feature interaction method for social characteristics.

Keywords　Multimodal, Rumor detection, DeepFM, Convolutional neural networks, Social feature, Natural language processing

1　引言

社交媒体的迅速发展极大地改变了人们获取信息的方式,越来越多的人利用社交媒体来追踪新闻事件的报导,了解实时的最新进展,同时,任何一个接入互联网的人都可以使用社交媒体发布信息。因此,在社交媒体中,谣言的传播速度更快、影响范围更广、监测难度更大、危害程度更深。根据新浪微博 2019 年发布的《微博辟谣 2018 年度报告》可知,三分之一的谣言始发于社交网络。2020 年伊始,突如其来的新冠肺炎在全国蔓延,由于疫情来势凶猛、猝不及防,部分民众一时处于极度恐慌之中,面对信息的不确定性,一时谣言蜂起,影响了疫情的防控和疫情期间的社会治理。谣言无节制地在网络上传播不仅会影响社会和谐与稳定,甚至会威胁到国家和地区安全。

到稿日期:2020-12-01　返修日期:2021-04-16　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:国家重点研发计划(2017YFB0803001);国家自然科学基金(61572459)
This work was supported by the National Key R&D Program of China(2017YFB0803001) and National Natural Science Foundation of China (61572459).
通信作者:隋杰(suijie@ucas.ac.cn)

研究表明，带图微博占全部微博的比例是51.6%，图片在限制为140个文字的微博中广泛存在。多媒体内容承载着更加丰富与直观的信息，能够更好地描述新闻事件，且更易广泛传播，带图片新闻的平均转发次数是纯文本新闻的11倍[1]。谣言的传播可能造成大规模的负面影响，有时会影响甚至操纵重要的公共事件。例如，在2016年美国总统大选的最后三个月内，许多人都相信了“图文并茂”的假消息，并在Facebook上分享了3 700多万次。因此，为了减小社交媒体中谣言传播带来的严重负面影响，迫切需要一种自动检测方法。

到目前为止，谣言检测方法包括传统的学习方法和基于深度学习的模型。现有的谣言自动检测方法大多是基于文本和社会语境的。基于分类的方法[2-4]和基于图的优化方法[5-7]被用于根据手工构建的文本和社会上下文特征来验证在线文本帖子的真假，另一部分工作以传播时间、传播结构、语言特征等因素为考量，提出了基于传播结构检测法以及时间序列检测的谣言检测方法。目前只有少数研究尝试基于包含图像信息的多媒体内容来发现谣言[8-9]，利用深度神经网络提取图像特征，并联合文本特征进行谣言检测。然而，这些工作忽略了社会特征类别间的联系。以微博用户为例，社会特征包括用户发帖时的情绪、粉丝数量和微博定位的次数等。这些社会特征内容丰富且异构，从不同的维度展现了不同的信息，而现有的对社会特征的利用是相当初步的，并没有建立社会特征中不同类别间的联系。另外，部分社会特征是连续型变量，建模时要对连续型变量进行转换，如何将连续型特征转换为离散值成为了另一项挑战。

为了解决上述问题，本文提出了一种基于DeepFM和卷积神经网络的多模态特征融合的谣言检测方法DCNN(DeepFM and Convolutional Neural Networks Ensembles for Multimodal Rumor Detection)。卷积神经网络被证明在学习精确的文本或视觉表征方面是有效的[10-11]。本文提出的模型主要由两部分组成：多模态特征提取和谣言检测。对于多模态特征提取器，采用卷积神经网络(CNN)从文章的文本和视觉内容中自动提取特征，然后用一个基于DeepFM算法的预训练模型来提取社会特征，将3个模态的特征融合，最后训练分类器完成谣言检测的任务。在微博数据集上的实验结果表明，所提出的DCNN模型优于最新的方法。

本文的主要贡献可以概括为：

(1)提出了一个联合消息文本特征、用户的图像特征与社会特征的多模态谣言检测模型DCNN。

(2)提出了一种先对连续型变量进行离散化处理，然后对离散化数据进行one-hot编码的特征编码方法，并基于DeepFM算法建立了高阶与低阶的社会特征交互。

(3)为了验证模型的有效性，在微博收集的多媒体数据集上进行了评估。结果表明，与现有的基于特征的方法和最新的神经网络模型相比，DCNN取得了更好的性能。

2 相关工作

2.1 基于文本模态的谣言检测方法

目前，大多数已有方法都将单模态谣言检测看成是一个文本分类的任务，这些方法大致可以分为3类：基于传播的方法、基于特征的手工特征分类方法和基于深度学习的方法。

2.1.1 基于传播的方法

谣言在社交媒体上传播，而社交媒体上的消息和事件之间存在一些潜在的相关性，为了利用社交网络的异构结构，该类算法将消息和用户链接到一个完整的网络，然后利用基于图形的优化方法，将整个网络图作为一个整体用于评估它们的可信度，这样一来，当网络收敛时可以获得未分类文本的类别。

2.1.2 基于特征的手工特征分类方法

通常谣言检测被定义为一个二分类问题[12]，因此该类方法遵循机器学习领域中常用的监督学习分类模式：从两个类中提取代表样本的有限元模型；用提供的样本训练合适的分类器；对未分类数据进行测试和评估[13-14]。这个程序的关键步骤就是如何提取突出的特征。

2.1.3 基于深度学习的方法

与传统分类器相比，深度神经网络(Deep Neural Networks，DNN)在许多机器学习问题(如物体检测、情感分类和语音识别)中具有明显的优势。基于DNN的方法旨在自动学习谣言数据的深度表示。根据神经网络的不同结构，可以将神经网络方法进一步分为两类。

(1)递归神经网络(Recurrent Neural Network，RNN)：这种方法将谣言数据建模为顺序数据。其关键是RNN中各个单元之间的连接形成了一个直接循环并创建了网络的内部状态，这可能使它能够捕获具有谣言扩散特性的动态时间信号[15-17]。

(2)卷积神经网络(Convolutional Neural Network，CNN)：由堆叠的卷积和池化层组成，其结构有助于对重要的语义特征进行建模。基于CNN的方法[18]不仅可以从输入实例中自动提取局部重要特征，而且可以揭示那些高层交互。

2.2 多模态谣言检测

在谣言检测任务中，如何利用不同模态的信息来区分谣言和非谣言，是当前面临的主要挑战。现有的方法大多集中在文本内容和社会语境上，社会语境指新闻在社会网络传播的过程中产生的信息。近年来，视觉信息已成为谣言检测的重要指标。随着多媒体内容的普及，研究者开始结合视觉信息来检测谣言。

为了从多个方面学习特征表示，深度神经网络已经成功地应用于各种任务，包括但不限于视觉问答、图像字幕和谣言检测。文献[19]提出了一种基于深度学习的谣言检测模型，该模型提取多模态和社会语境特征，并通过注意力机制进行融合。

受CNN强大的性能启发，大多数基于多媒体内容的现有作品使用预先训练好的深度卷积神经网络(如VGG19)，来获得一般的视觉表现，并将它们与文本信息融合。具体来说，Jin等[19]首先通过深度神经网络将多模态内容融入到社交网络中，以解决谣言检测问题；Wang等[20]提出了一种基于多模态特征的端到端事件对抗神经网络，用于检测新出现的谣言事件；Khattar等[21]提出了一种学习多模态共享表示的新方

法。这些工作主要集中在如何融合不同形态的信息,但一方面它们需要额外的信息,如说话人的身份、话语情感的序列,同时为它们的融合方案建模,在一般情况下,此附加信息可能不可用。另一方面,所有情况下的联合表示都比文本模态稀疏(包含更多的缺失值),以致学习机制不能有效地提取信息。

为了克服现有研究的局限性,本文提出了一种新的多模态谣言检测模型。基于 DeepFM 的网络通过稀疏的特征向量来获得特定模式的唯一信息,在联合表示无效的情况下,这种方法具有很强的鲁棒性,从而减小了对 DCNN 性能和信息发现的影响。对于公共的和唯一的网络,学习非共享的潜在表示可以确保高级表示的潜在嵌入不受低级表示(即无用表示的梯度)的影响。此限制强制延迟嵌入以获得补充信息,并在更高层执行融合时提供更高的表达能力。

3 多模态谣言检测模型 DCNN

本节首先提出了模型中的多模态特征提取器,它由 3 个部分组成,即文本特征提取器、视觉特征提取器和社会特征提取器,然后描述了如何集成这 3 个预训练模型来学习可转移的共享特征表示,最后输入谣言检测器进行分类。

3.1 模型架构

多模态网络谣言检测模型 DCNN 的目标是学习用于谣言检测的可转换和可鉴别特征表示,其总体框架示意图如图 1 所示。

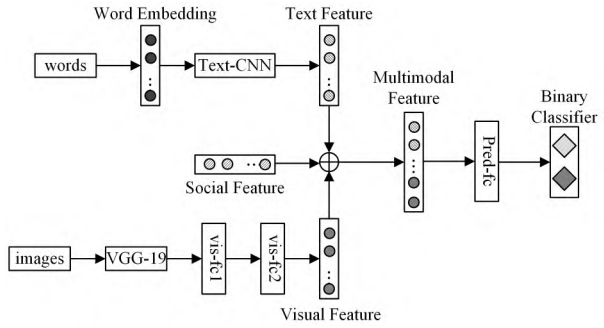

图 1 DCNN 模型的总体结构

Fig. 1 Overall framework of DCNN model

本文提出的 DCNN 模型包括多模态特征提取器和谣言检测器,多模态特征提取器集成了 3 个预训练模型。首先,由于社交媒体上的帖子通常包含不同形式的信息(如文本帖子和附加图像及用户的社会特征),因此多模式特征提取程序包括文本、视觉和社会特征提取程序,以处理不同类型的输入。在学习了文本、视觉和社会的潜在特征表示之后,它们被连接在一起形成最终的多模态特征表示。谣言检测器都是建立在多模态特征提取器的基础上的。谣言检测器将学习到的特征表示作为输入来预测帖子的真假。

3.2 文本特征提取

从文本内容中提取信息特征,采用卷积神经网络(CNN)作为文本特征提取器的核心模块。CNN 已经被证明在许多领域都是有效的,如计算机视觉和文本分类。图 1 中,在文本特征提取器中加入了一个改进的 CNN 模型,即 TextCNN,其网络结构如图 2 所示。

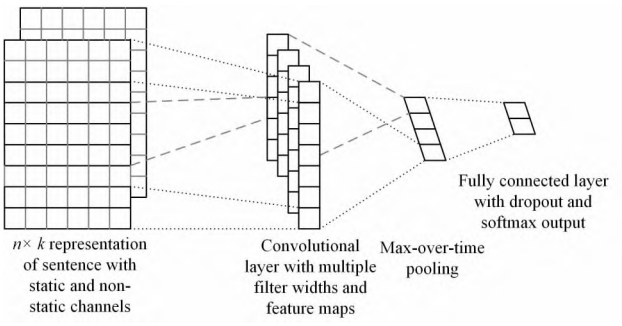

图 2 TextCNN 网络结构示意图[10]

Fig. 2 TextCNN network architecture diagram[10]

TextCNN 由多个卷积核组成的卷积层、池化层和全连接层拼接而成,它利用具有不同窗口大小的多个过滤器捕获不同粒度的特征来识别谣言。具体来说,文本中的每个单词都表示为一个单词嵌入向量。每个单词的嵌入向量用预先训练好的嵌入对给定的数据集上的单词进行初始化。对于句子中的第 i 个单词,对应的 k 维单词嵌入向量记为 $T_i \in \boldsymbol{R}_k$。因此,一个有 n 个单词的句子可以表示为:

$$T_{1:n} = T_1 \oplus T_2 \oplus \cdots \oplus T_n \tag{1}$$

其中,$\oplus$ 是连接操作符号。窗口大小为 h 的卷积核将句子中 h 个单词的连续序列作为输入,输出一个特征向量:

$$t = [t_1, t_2, \cdots, t_{n-h+1}] \tag{2}$$

对于每个特征向量 $\boldsymbol{t}$,用最大池化操作取最大值,从而提取出最重要的信息。重复这个过程,直到得到所有卷积核中的特征。在最大池化操作之后,使用一个全连接层来确保最终的文本特征表示与视觉特征表示具有相同的尺寸,这项操作可以被表示为:

$$R_T = (\boldsymbol{W}_{tf} \cdot R_{T_c}) \tag{3}$$

其中,$\boldsymbol{W}_{tf}$ 是该全连接层的权重矩阵。

3.3 视觉特征提取

将图片输入到视觉特征提取器中,记为 v。为了有效地提取视觉特征,使用预先训练好的 VGG-19。在 VGG-19 网络的最后一层上,添加一个全连接层来调整最终的视觉特征表示的尺寸 p,这项操作可以被表示为:

$$R_V = (\boldsymbol{W}_{vf} \cdot R_{V_{vgg}}) \tag{4}$$

其中,$R_{V_{vgg}}$ 是从预训练模型 VGG-19 中获得的视觉特征表示,$\boldsymbol{W}_{vf}$ 是视觉特征提取器中全连接层的权重矩阵。

另外,在与文本特征提取器联合训练的过程中,为了避免过拟合,预训练的 VGG-19 神经网络的参数保持静态。

3.4 社会特征提取

传统的多模态方法通常基于文本和视觉/音频功能。然而,对于微博上的谣言检测任务而言,社会特征已在文献[2-4]中用于有效的谣言检测。在社交媒体平台上可以观察到几个因素,这些因素对评估信息的可信度是有用的,主要包括:

(1)某些话题所产生的反应和用户讨论该话题时所表达的情绪,例如用户对该话题使用的意见表达,表示了正面或负面的情绪;

(2)用户传播信息的确定性水平,例如他们是否对提供给

他们的信息表示质疑；

(3)引用的外部来源，例如他们是否引用了一个特定的URL和他们传播的信息，以及该来源是否是一个流行的域名；

(4)微博用户本身的特征，例如每个用户在平台上的关注者数量。

这些社会特征既有连续型的，也有离散型的，内容丰富且异构，从不同的维度展现了不同的信息。目前的工作没有对连续型特征加以利用，并且忽略了社会特征类别间的联系。为了更好地利用微博用户的社会特征，我们基于DeepFM算法的思想，利用多层全连接神经网络自动学习特征间的高阶交互关系，建立了社会特征类别之间的联系，具体如图3所示。

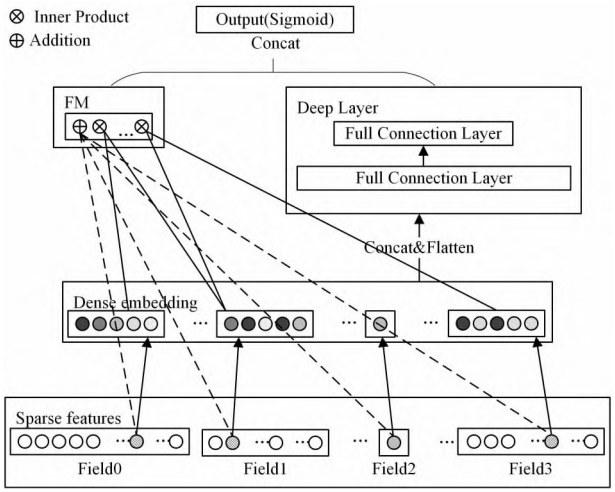

图3 DeepFM网络结构示意图[22]

Fig. 3 DeepFM network architecture diagram[22]

3.4.1 FM部分建立特征交互

在衡量特征 i 与特征 j 之间的交互信息时，需要两个特征同时存在于一条记录中，但是社会特征中的数据具有稀疏性，同时记录两个特征信息的数据非常少，因此不能进行有效的学习。

为了解决这一问题，我们采用了one-hot编码的方式将社会特征表示为向量。每个向量只包含一个位置，对应社会特征类别中的字符索引的值为1，其他所有位置都为0。向量的维度等于社会特征类别的大小，这样可以在更基本的层次上表示更多的信息，适用于社会特征这样的稀疏数据或有噪声的数据。我们又为每一个特征引入了一个具有低维与稠密特性的向量特征，并使用特征间向量特征的内积来衡量特征间的相关性，此时即使两个特征共同存在的数据很少甚至没有，也可以衡量两者之间的相关性，从而有效解决了社会特征中数据稀疏导致的难以计算特征交互的问题。

本文的思路是在LR模型后追加考虑任意两个特征之间的关系。模型等式初步定义为：

$$y=w_0+\sum_{i=1}^{n}w_ix_i+\sum_{i=1}^{n}\sum_{j=i+1}^{n}w_{ij}x_ix_j \tag{5}$$

其中，w_{ij} 是特征组合 x_ix_j 的交叉权重。虽然相比LR模型，上述模型引入了二阶特征组合，但由于 $x_ix_j=0$ 都会有 $w_{ij}=0$，因此在特征大规模稀疏的情况下，w_{ij} 最终都会变为0。为了解决这个问题，FM部分进行了以下改进：

$$y=w_0+\sum_{i=1}^{n}w_ix_i+\sum_{i=1}^{n}\sum_{j=i+1}^{n}\langle V_i,V_j\rangle x_ix_j \tag{6}$$

而 $\langle V_i,V_j\rangle$ 是两个大小为 k 的向量的点积：

$$\langle V_i,V_j\rangle\geqslant\sum_{f=1}^{k}V_{if}\times V_{jf} \tag{7}$$

通过上述改进，FM部分可以充分地捕获所有一阶和二阶特征组合，避免数据稀疏情况下模型无法训练的情况。其中，1) w_0 是全局的偏差值，且 $w_0\in R$；2) w_i 是第 i 个变量 x_i 的权重，且 $x_i\in R^n$；3) $\boldsymbol{V}_i$ 是第 i 个变量 x_i 对应的需要学习的embedding向量，且 $\boldsymbol{V}_i\in R^{1\times k}$；$k$ 为向量的长度，是一个重要的超参数，表示分解的维数，也反映了FM模型的复杂度；4) $\langle V_i,V_j\rangle$ 用来代替原来的 w_{ij}，表示二阶特征组合 x_ix_j 的权重，只要特征 x_i 和其他任意特征的组合出现过(即存在 $x_ix_j\neq0$)，就能通过训练学习到自己对应的embedding向量 $\boldsymbol{V}_i$。因此，在预测时，尽管训练数据中 x_i 与 x_k 的组合从未出现过，但可以通过计算它们对应embedding向量的内积作为权重。

3.4.2 Deep Layer学习特征交互

Deep Layer部分是一个多层的 $\boldsymbol{V}_i\in R^{(1\times k)}$，用来学习高阶特征组合信息的前馈神经网络，结构如图4所示。

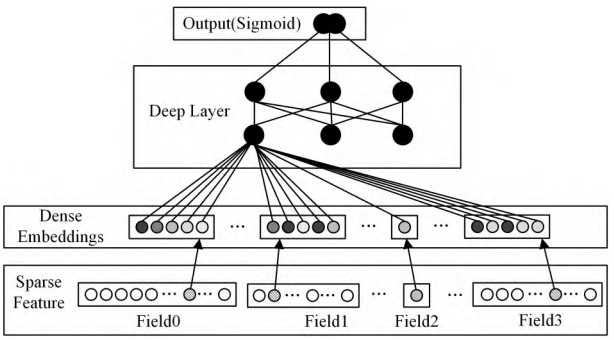

图4 Deep Layer网络结构示意图

Fig. 4 Deep Layer network architecture diagram

由于微博用户的社会特征具有高度稀疏、超高维、按字段分组等特点，因此需要在普通的前向神经网络之前加上一层embedding层来对输入长度不同的各个特征分组进行压缩，使之成为低维、密集、长度固定为 k 的向量。使用embedding层的另一个优点是Deep Layer部分可以和FM部分共享embedding向量，省去了单独训练embedding层的步骤。假设embedding层的输出为：

$$z^{(0)}=[e_1,e_2,\cdots,e_m] \tag{8}$$

然后，将 $z\wedge((0))$ 作为前馈神经网络的输入，前向传播的过程如下：

$$z^{(l+1)}=\mathrm{ReLU}(W^{(l)}z^{(l)}+b^{(l)}) \tag{9}$$

假设前馈神经网络隐藏层层数为 H，则最终DNN部分输出的结果为：

$$y=\mathrm{sigmoid}(W^{(H+1)}z^{(H)}+b^{(H+1)}) \tag{10}$$

3.5 共享表示

将文本特征表示 R_T、视觉特征表示 R_V 和社会特征表示 R_S 串联起来，形成多模态特征表示，记为 $R_F=R_T\oplus R_V\oplus R_S$，这是多模态特征提取器的输出。本文将整个多模态特征提取器

表示为$G_f(M;\theta_f)$,其中 M 表示消息集合中一条待判别的消息,是多模态特征提取器的输入,θ_f 表示多模态特征提取器中所有学习的参数。

3.6 谣言检测器

谣言检测器将文本与图像、社会特征的共享表示特征R_F作为输入,用于判别消息是否为谣言。将谣言检测器表示为$G_d(\cdot;\theta_d)$,其中θ_d表示谣言检测器中所有的参数,谣言检测器的输出为该消息是谣言的概率 P,则对于一条带有多模态特征的消息m_i来说,其是谣言的概率为:

$$P_\theta(m_i)=G_d(G_f(m_i;\theta_f);\theta_d) \tag{11}$$

将P_θ的值视为标签,标签为 1 表示消息m_i是假的,否则标签为 0 表示消息m_i是真的。为了将输出值限制在 0 到 1 之间,使用 Sigmoid 逻辑函数。因此,为了计算分类损失,采用交叉熵:

$$L_d(\theta_f,\theta_d)=-E_{(m,y)\backsim(M,Y_d)}[y\log(P_\theta(m))+(1-y)(\log(1-P_\theta(m)))] \tag{12}$$

通过寻找最优参数θ_f^*和θ_d^*来最小化分类损失函数$L_d(\theta_f,\theta_d)$,具体表示如下:

$$(\theta_f^*,\theta_d^*)=\arg\min_{\theta_f,\theta_d}L_d(\theta_f,\theta_d) \tag{13}$$

4 实验

本节首先介绍实验环境和实验所使用的大型微博数据集,然后介绍目前最先进的多模态谣言检测方法,最后分析所提模型的性能。

4.1 实验环境

代码在开源的深度学习框架 Pytorch1.6.0 上开发,底层为 Python3,所用操作系统为 Windows 10。实验所用服务器为拥有 12 GB 显存的 GeForce RTX2080Ti GPU,Intel(R) Core(TM) i7-9700K CPU。

4.2 数据集

在这个数据集中,真实的新闻来源于中国的权威新闻来源,如新华社。谣言抓取时间为 2012 年 5 月至 2016 年 1 月,经微博官方辟谣系统核实。该系统鼓励普通用户报告可疑帖子,并由可信用户委员会检查可疑帖子。根据之前的工作,该系统也是收集谣言消息的权威来源。在预处理这个数据集时,首先去除重复的和低质量的图像,以确保整个数据集的质量,然后应用单次聚类方法来发现新出现的事件,最后将整个数据集按 7:1:2 的比例分割成训练集、验证集和测试集,并确保它们不包含任何公共事件。表 1 列出了这个数据集的详细统计。

表 1 数据集的统计信息

Table 1 Statistics of experiment datasets

Statistic		Data Size
Trainning Set	Rumor	3748
	Non-rumor	3783
Testing Set	Rumor	1000
	Non-rumor	996

4.3 基线模型

为了验证所提模型的有效性,本文从以下 3 个类别中选择基线:单模态模型、多模态模型和所提模型的变体。

4.3.1 单模态模型

该模型利用文本和图像信息来检测谣言。对于每种模式,它也可以单独用于发现谣言。因此,提出以下两个简单的基线。

(1)文本模型(Textual)。使用来自所有帖子的文本内容的 32 维预训练的文字嵌入权重来初始化嵌入层的参数,然后使用 CNN 为每个帖子提取文本特征,最后使用 softmax 功能的全连接层来预测这篇文章的真假。使用了 20 个过滤器,窗口大小为 1~4,全连接层的隐藏尺寸是 32。

(2)图片模型(Visual)。输入是一个图像,使用预训练的 VGG-19 和全连接层提取视觉特征。然后将视觉特征传入全连接层进行预测。将全连接层的隐藏大小设为 32。

4.3.2 多模态模型

所有的多模态方法都考虑了来自多个模态的信息,包括 VQA[23],att-RNN 和 MSRD[24]。

(1)VQA。VQA (Visual Question Answer)模型的目的是在给定图像的基础上回答问题。原始的 VQA 模型是针对多类分类任务设计的,而本文主要集中在二分类任务上。因此,在实现 VQA 模型时,将最后的多类层替换为二进制类层。另外,为了公平比较,使用单层 LSTM,LSTM 的隐藏单元数大小设置为 32。

(2)att-RNN。att-RNN 是目前最先进的多模态谣言检测模型之一。它利用带有注意力机制的递归神经网络来融合文本、视觉和社会语境特征,输出联合表示。

(3)MSRD。MSRD 是目前最新的模态谣言检测模型之一。它利用了消息文本信息以及图像内文本信息与图像信息,并基于图像语义分割思想做了文本定位与识别,最后用高斯分布来获得多模态特征的共享表示。为了公平比较,在实验中删除了处理图像内文本信息的部分。

4.4 实验结果与分析

表 2 列出了基线模型以及 DCNN 方法的结果。从表中可以清楚地看到,DCNN 的性能优于基线模型。表 3 列出了 DCNN 方法分别在正负样例上的实验结果。

表 2 DCNN 模型与其他方法的对比实验结果

Table 2 Comparison of experimental results between DCNN model and other methods

Method	*Accuracy*	*Precision*	*Recall*	*F*1
Textual	0.763	0.797	0.683	0.748
Visual	0.622	0.578	0.591	0.585
VQA	0.736	0.797	0.634	0.706
att-RNN	0.779	0.778	0.799	0.789
MSRD	0.781	0.800	0.699	0.744
DCNN	**0.803**	**0.801**	**0.809**	**0.799**

表 3 DCNN 模型在正负样例上的实验结果

Table 3 Experimental results of DCNN model on non-rumor and romor samples

Type	*Accuracy*	*Precision*	*Recall*	*F*1
Non-rumor	0.803	0.803	0.787	0.795
Rumor	0.803	0.804	0.819	0.811

从表 2 中可以观察到,尽管视觉特征对谣言检测是有效

的，但是 Visual 的性能仍然比多模态方法差，这证实了集成式多模态谣言检测方法具有优越性。多模态模型 MSRD 和 att-RNN 的性能优于单模态模型和 VQA，这也佐证了这一点。在多模态模型中，att-RNN 的性能优于 VQA，这表明应用注意力机制有助于改善模型性能；MSRD 的性能也优于 VQA，表明挖掘图像内的文本信息对谣言检测也具有一定的作用。

本文提出的集成式多模态谣言检测方法 DCNN 在准确率、精度、召回值和 F1 得分等方面都优于所有基线模型，并将准确率从 78.1%提高到 80.3%，这验证了 DCNN 方法在检测社交媒体网络谣言方面的准确性。与以前的最佳基准相比，F1 分数从 78.9%提高到 79.9%，这表明 DCNN 方法的精确率和召回率的总体表现也更好。其原因在于，首先本文所使用的卷积神经网络可以更好地捕捉局部特征，对于文本来说，局部特征就是由若干单词组成的滑动窗口。卷积神经网络的优势在于能够自动地对特征进行组合和筛选，从而获得不同抽象层次的语义信息。通过引入已经训练好的词向量和构造更好的 embedding 层，可以更好地提取文本特征。其次，VGG-19 的结构非常简洁有效，整个网络都使用了同样大小的卷积核尺寸(3×3)和最大池化尺寸(2×2)，以多个小滤波器的组合来不断加深网络结构，从而提升性能。最后，用户的社会特征中包含了丰富的信息，有待挖掘。一方面，本文使用了 one-hot 对其中的稀疏数据进行编码，将社会特征向量化。另一方面，社会特征种类繁多，属性之间的关系难以明确，需要更深层次的关系挖掘才可以显现两者的关系。为了取得较好的检测效果，在进行社会特征提取时需要同时考虑低阶与高阶特征关系。而本文的基于 DeepFM 算法的思想，利用多层全连接神经网络自动学习特征间的深层交互关系，建立了社会特征类别之间的联系。

att-RNN 和 DCNN 模型同样利用了用户的文本特征、视觉特征和社会特征，而 DCNN 的谣言检测效果更好。这说明采用 one-hot 的特征编码方法，并基于 DeepFM 算法建立高阶与低阶的社会特征交互具有一定意义，会给谣言检测器的整体准确率带来提升。

从表 3 中可以看到，正负样例的准确率与预测精度都达到了较高水平，说明 DCNN 方法对谣言与非谣言消息均有良好的适应力。

结束语 本文提出了一种基于深度神经网络的针对配文文本内容、图像内容以及用户属性信息的多模态网络谣言检测模型 DCNN，该模型利用 one-hot 编码，基于 DeepFM 算法建立了用户社会特征交互，采用共享特征方法将文本特征、图像特征与社会特征进行了较好的融合表示，并将其用于谣言检测。本文在微博数据集上进行了实验验证，实验结果表明，DCNN 模型优于基线模型。

目前针对的问题是事件级的谣言检测任务，但是当形成事件级微博文本时，谣言已经具有了一定的传播程度，而谣言的早期检测对于谣言治理检测起着重要作用。因此，如何基于现有数据集，研究谣言的早期检测是一个值得继续深入研究的问题。另外，在未来的研究中，我们也将在其他的数据集上进行更广泛的实验。

参 考 文 献

[1] JIN Z W, CAO J, ZHANG Y D. Novel Visual and Statistical Image Features for Microblogs News Verification[J]. IEEE Transactions on Multimedia, 2016, 19(3): 598-608.

[2] CARLOS C, MARCELO M, BARBARA P. Information credibility on twitter[C]// 20th International Conference on World Wide Web (WWW). ACM, 2011: 675-684.

[3] SALLOUM R, REN Y, KUO C C J. Image splicing localization using a multi-task fully convolutional network (MFCN)[J]. Journal of Visual Communication and Image Representation, 2018, 51: 201-208.

[4] WU K, YANG S, ZHU K Q. False rumors detection on sina weibo by propagation structures[C]// IEEE International Conference on Data Engineering. ICDE, 2015: 651-662.

[5] MANISH G, ZHAO P, HAN J W. Evaluating Event Credibility on Twitter[C]// SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2012: 153-164.

[6] JIN Z W, CAO J, JIANG Y G, et al. News credibility evaluation on microblog with a hierarchical propagation model[C]// IEEE International Conference on Data Mining (ICDM). IEEE, 2014: 230-239.

[7] JIN Z W, CAO J, ZHANG Y D, et al. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016: 2-17.

[8] ADITI G, HEMANK L, PONNURANGAM K, et al. Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy[C]// Proceedings of the 22nd International Conference on World Wide Web Companion. 2013: 729-736.

[9] JIN Z W, CAO J, ZHANG Y Z, et al. Verifying Multimedia Use with a Two-Level Classification Model[C]// MediaEval Multimedia Benchmark Workshop. 2015.

[10] YOON K. Convolutional Neural Networks for Sentence Classification[C]// 2014 Conference on Empirical Methods in Natural Language Processing. 2014.

[11] KAREN S, ANDREW Z. Very Deep ConvNets for Large-Scale Image Recognition[C]// 3rd International Conference on Learning Representations. ACM, 2015: 358-406.

[12] SUN S, LIU H, HE J, et al. Detecting event rumors on sina weibo automatically[C]// Web Technologies and Applications. Springer, 2013: 120-131.

[13] YANG F, LIU Y, YU X H, et al. Automatic detection of rumor on sina weibo[C]// Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012: 1-7.

[14] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media[C]// IEEE 13th International Conference on Data Mining (ICDM). IEEE, 2013: 1103-1108.

[15] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// International Joint Conference on Artificial Intelligence. 2016: 3818-3824.

[16] CHEN T,LI X,ZHANG J,et al. Call attention to rumors:Deep attention based recurrent neural networks for early rumor detection[C]// Trends and Applications in Knowledge Discovery and Data. 2017:40-52.

[17] RUCHANSKY N,SEO S,LIU Y. CSI:A hybrid deep model for fake news detection [C] // Conference on Information and Knowledge Management. 2017:797-806.

[18] NGUYENT N,LI C,NIEDERÈE C. On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners [C]// International Conference on Social Informatics. Springer, 2017:141-158.

[19] JIN Z W,CAO J,GUO H,et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]// ACM on Multimedia Conference. 2017:795-816.

[20] WANG Y Q,MA F L,JIN Z W,et al. Eann:Event adversarial neural networks for multi-modal fake news detection[C]// 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018:849-857.

[21] KHATTAR D,GOUD J S,GUPTA M,et al. Mvae:Multimodal variational autoencoder for fake news detection [C] // World Wide Web Conference. ACM,2019:2915-2921.

[22] WANG R P,JIA Z,LIU C,et al. Deep Interest Factorization Machine Network Based on DeepFM[J]. Computer Science, 2021,48(1):226-232.

[23] STANISLAW A,AISHWARYA A,JIASEN L,et al. Vqa:Visual question answering [C]// IEEE International Conference on Computer Vision. 2015:2425-2433.

[24] LIU J S,FENG K,JEFF Z,et al. MSRD:Multi-Modal Web Rumor Detection Method[J]. Journal of Computer Research and Development,2020,57(11):2328-2336.

CHEN Zhi-yi, born in 1996, postgraduate. His main research interests include natural language processing and data mining.

SUI Jie, born in 1976, associate professor. Her main research interests include data mining and social network analysis.

(责任编辑:李亚辉)