文章编号：1000-8608(2020)05-0547-08

基于图嵌入的社交媒体药物不良反应事件检测方法

申　晨，林鸿飞*

（大连理工大学 计算机科学与技术学院，辽宁 大连 116024）

摘要：药物不良反应事件是造成患者发病、死亡的主要原因之一.传统的基于患者自发报告系统存在较为严重的漏报情况，近年来将推特等社交媒体作为数据来源进行药物不良反应事件检测的研究愈发受到重视.各种深度学习模型通常依赖于大量的训练样本，然而受限于数据来源的特质和耗时的数据标注工作，该领域的相关研究面临标注数据规模小、数据噪声大等问题，制约了这些模型发挥良好的效果.据此，在文本表示层面引入基于图嵌入数据增强和对抗训练两种正则化方法，提升模型在低资源高噪声下的药物不良反应事件检测效果.通过实验，具体分析和讨论两种方法的适用范围，结合卷积神经网络，提出一种同时发挥其优势的药物不良反应事件检测模型，实验结果显示其具有良好的适用性.

关键词：社交媒体；图嵌入；对抗训练；药物不良反应事件检测；深度学习

中图分类号：TP391　　**文献标识码：**A　　**doi：**10.7511/dllgxb202005011

0 引　言

药物不良反应事件(adverse drug events, ADEs)指患者在使用某种药物时，造成的与治疗目的无关的有害后果.其已经成为造成医源性发病、死亡和经济损失的主要原因之一.有研究统计，仅在英、美两国每年由于药物不良反应事件造成的损失已超过50亿美元[1].受限于临床试验阶段参与者的数量，通常在药物上市以前无法掌控其可能造成的所有不良反应，使得上市后药物预警(postmarketing surveillance, PMS)成为整个药物预警流程中的重要一环.传统上市后药物预警主要依赖于隶属各国卫生部门的自发报告系统，如2019年我国国家药品不良反应监测中心就收到“药品不良反应/事件报告表”超过150万份[2].然而更多的研究证据显示，这类自发报告系统的漏报率高达82%～98%[3].针对更多的药物预警数据来源进行监测仍是需要研究解决的重要问题.

推特(Twitter)作为社交媒体，在其逾300万活跃用户每天发布的500余万条信息中蕴含着可用于公共健康研究的宝贵资源[4].然而对其进行标注的工作耗费时间和物力，使得目前可用于研究的语料集规模较小.近年来有许多研究尝试进行基于推特文本的药物不良反应事件检测，如何在低标注资源的条件下获得高性能的检测模型是这些研究需要面对的主要问题之一.

这类检测工作通常需要从大量文本中学习并提取有效信息，为了使模型可以基于有限标注语料获得更高性能，相关研究中一种常用的方式是引入外部资源或专家设计丰富的特征.Stanovsky等[5]在研究中利用由维基百科构筑的知识图谱以及由专家参与标注的主动学习方法进行不良反应识别.Lee等[6]设计了7种基于规则和特征的分类方法，采用投票方式将其作为一个集成学习架构进行预测.这类方法在提高模型性能时，往往需要更多的领域知识参与，投入人力、物力.同时针对特定语料设计的特征也会限制模型的泛化能力.

Li等[7]将迁移学习方法引入不良反应事件检测任务，通过从不同标注资源中学习具有共性

收稿日期：2020-05-25；　**修回日期：**2020-07-05.
基金项目：国家自然科学基金资助项目(61772103,61572102).
作者简介：申　晨(1986-)，男，博士生，E-mail：shenchenioi@mail.dlut.edu.cn；林鸿飞*(1962-)，男，博士，教授，E-mail：hflin@dlut.edu.cn.

的特征，提升样本数较少的语料集上的模型性能. Chowdhury 等[8]、Gupta 等[9]的研究采用了多任务学习方法，在药物不良反应事件检测的同时进行实体识别任务，提升事件检测效果. 这些方法被证实可以提升小规模语料集上的模型性能，但是需要引入额外的数据集或标签进行训练.

针对上述研究过于依赖标注工作的不足，本文提出一种基于图嵌入的药物不良反应事件检测方法. 该方法不依赖额外的标注工作，在文本表示层面通过将推特文本中的药物、潜在的不良反应描述与其同类文本作为相同的节点处理，构建文本的网络结构，从而在其后的图嵌入学习过程中得到更为丰富的节点结构特征. 同时针对推特文本噪声较高，仅凭借少数样本训练模型的过拟合问题，本文引入另一种正则化方法——对抗训练，用于避免模型依赖与正确预测无关的特征.

1 研究方法

本文针对该研究领域中标注语料规模较小、社交媒体文本噪声较高等问题，在药物不良反应事件检测任务中将基于图嵌入的文本表示和对抗训练两种方式用于深度学习模型的改进.

1.1 模型构建

本文提出的药物不良反应事件检测模型如图 1 所示，主要分为 4 个部分：(1)向量化表示部分通过结合顺序和图结构两种方式训练的词的分布式向量表示，为模型提供更为丰富和多样化的原始特征输入. (2)对抗训练部分通过模型损失对于输入样本的梯度计算最优对抗噪声，将这种微小扰动叠加在原始文本的向量表示空间生成对抗样本. 通过将生成的对抗样本用于模型训练提升模型对带有噪声数据的特征抽取能力. (3)特征抽取部分采用多通道卷积神经网络结构，不同通道选择不同大小的卷积窗口，用于获取不同窗口内文本向量化表示中具有判别能力的高级别特征，再通过池化操作降维. (4)事件检测器部分由带有 softmax 激活函数的全连接层构成，用于预测输入样本是否包含药物不良反应事件.

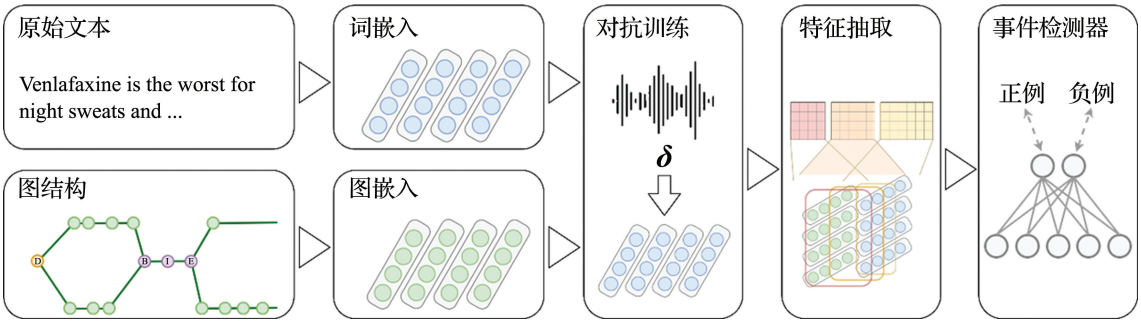

图 1 模型总体架构

Fig. 1 The overall model architecture

1.2 构造图结构

标注资源规模较小是制约深度学习方法在社交媒体药物不良反应事件检测任务中取得更好效果的主要原因之一. 为了更加有效地利用有限的标注语料，设计了一种将图结构用于数据增强的方式. 通过来自 SIDER 数据库[10]的药名和 ADR 词典[11]的词条与采集自推特的文本进行匹配，将得到的文本中的药物和药物不良反应候选词对应到相应节点，如图 2 所示.

其中药名均由单个单词构成，匹配后对应 D 类节点. 药物不良反应候选词根据自身在词组中

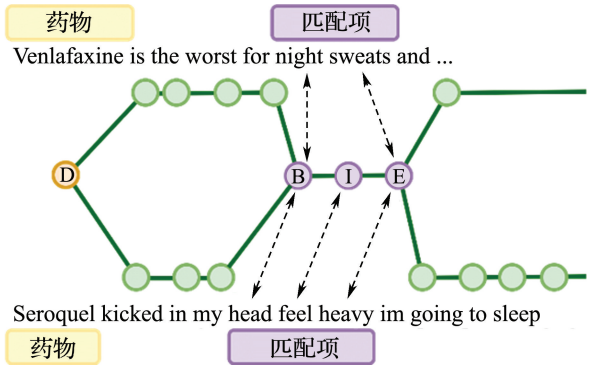

图 2 图结构构造方法

Fig. 2 The construction method of graph structure

的位置，分别对应 B、I、E、S 等类型节点；词组中的首个单词对应 B 类节点，末尾单词对应 E 类节点，词组中间位置的单词对应 I 类节点；单个单词的独立候选词对应 S 类节点. 其他未匹配的词以自身为节点，仍以原词序顺序连接.

通过上述构图方法，可以将采集自推特的包含药物名称的文本构建为一张连通的有向图结构. 通过该图进行遍历时会产生原本并不包含在样本中的实例，例如从图中药物 Venlafaxine 出发，可以遍历至药物不良反应候选词 head feel heavy，从而实现数据增强效果. 但这样生成的文本并不总是满足语法，同时药物与相关药物不良反应间的对应关系，也即图中的边可能并不客观存在，直接将其引入模型会带来过高噪声. 本文并不据此构造特征，或使用这些边作为连接方式进行诸如图神经网络等模型的训练，而是将其用于基于图嵌入方法的表示学习. 通过图嵌入方法从构造图中随机进行序列采样，预训练得到的向量化表示将包含药物、药物不良反应候选词的上下文信息以及结构信息.

1.3　向量化表示

分布式的文本向量化表示是当前自然语言处理任务中各种深度学习模型的主要输入形式，能够将高维离散的原始文本输入转化为相对低维稠密的向量，同时尽可能保留文本信息. 本文应用的向量化表示方法包括词嵌入和图嵌入两个部分，预训练语料来自采用 SIDER 数据库中的药名作为关键字爬取的推特中的用户生成文本.

本文方法中的词嵌入部分用于捕获文本中的语义信息，图嵌入部分用于捕获与药物和潜在药物不良反应相关上下文的结构信息. 其中在词嵌入部分，本文采用 GloVe 方法[12]进行预训练. 另一种被广泛应用的方法是 word2vec 方法，其特征提取是基于滑动窗口的，在局部文本中进行训练，而 GloVe 方法首先计算全局词共现频率，用于优化分布式向量表示，更善于利用文本的全局统计信息.

在图嵌入部分，本文采用的预训练方法为 SDNE[13]，它使用相对此前 DeepWalk、node2vec 等常用图嵌入方法更加深层的自编码器结构学习图结构中节点的一阶和二阶近邻关系. 其中节点一阶近邻关系基于自编码器中间层神经元，也即编码器的输出，进行优化. 这使得节点的特征表示与其相邻节点相似，一阶近邻的损失函数记为

$$L_1=\sum_{i,j=1}^{n}s_{i,j}\|\boldsymbol{y}_i'-\boldsymbol{y}_j'\|_2^2$$

其中 n 为图结构节点总数，$\boldsymbol{y}_i'$、$\boldsymbol{y}_j'$ 分别为节点 i、j 对应的编码器输出，$s_{i,j}$ 为两个节点间边的权重. 节点二阶近邻关系基于自编码器的输入 $\boldsymbol{x}'$ 和对应的重构输出 $\hat{\boldsymbol{x}}'$ 构建，使得具有类似邻居节点的特征表示相似，相关的损失函数记为

$$L_2=\sum_{i=1}^{n}\|(\boldsymbol{x}_i'-\hat{\boldsymbol{x}}_i')\odot\boldsymbol{b}_i\|_2^2$$

其中 $\boldsymbol{x}_i'$、$\hat{\boldsymbol{x}}_i'$ 分别是节点 i 对应的自编码器模型的输入和重构的输出，$\odot$ 为按位相乘运算，$\boldsymbol{b}_i$ 为惩罚项.

1.4　对抗训练

对抗训练方法是本文模型采用的另一种正则化方法. 在模型训练过程中，通过在样本的向量化表示中增加微小扰动，形成对抗样本. 利用对抗样本和原始样本共同进行模型训练，通过这样的方式防止模型过拟合，提升模型的鲁棒性和泛化能力[14]. 本文模型采用对抗训练方法提升模型对带噪声原始数据的适应能力，优化模型性能. 对抗训练最初在图像识别任务中被提出，其基本原理可表示为

$$\min_{\boldsymbol{\theta}}E_{(\boldsymbol{x},y)\sim D}\left[\max_{\|\boldsymbol{\delta}\|\leqslant\varepsilon}L(f_{\boldsymbol{\theta}}(\boldsymbol{x}+\boldsymbol{\delta}),y)\right]$$

式中：$f_{\boldsymbol{\theta}}$ 为神经网络模型；$E_{(\boldsymbol{x},y)\sim D}$ 为样本经验误差；$\boldsymbol{x}$、y 分别为样本及其对应的标签；$\boldsymbol{\delta}$ 为在样本上叠加的微小扰动；L 为模型损失函数；$\boldsymbol{\theta}$ 为模型参数；ε 为超参数，限制扰动的上界.

本文方法将扰动叠加于表示原始文本的向量空间，扰动的尺度可以采用快速梯度方法[15]获得. 对于输入样本 $\boldsymbol{x}=(\boldsymbol{v}_1\quad\boldsymbol{v}_2\quad\cdots\quad\boldsymbol{v}_n)$，其中 n 为输入文本序列长度，$\boldsymbol{v}$ 为词的向量化表示，需要叠加的扰动可由下式求得：

$$\boldsymbol{\delta}=\varepsilon\cdot\frac{\boldsymbol{g}}{\|\boldsymbol{g}\|_2},\ \boldsymbol{g}=\nabla_{\boldsymbol{x}}L(\boldsymbol{\theta},\boldsymbol{x},y)$$

其中 $\boldsymbol{g}$ 为模型损失函数在输入样本 $\boldsymbol{x}$ 上的梯度. 由此得到的对抗样本可记为 $(\boldsymbol{x}+\boldsymbol{\delta},y)$，通过使用对抗样本与原始样本共同进行模型训练，提升模型对带噪声推特文本的适应能力.

1.5　特征抽取和模型训练

本文方法采用类似 Kim 提出的多通道卷积

神经网络模型[16](TextCNN)进行特征抽取.卷积方法为一维卷积,卷积核的宽度与词向量维度相同,卷积通道数量与卷积窗口大小为超参数.不同于原始多通道卷积神经网络所采用的池化操作max-overtime-pooling,本文方法采用 k-max-pooling,提取特征图中最具判别能力的 k 个特征.对于维度为 $n\times k$ 的池化层输出矩阵 $\boldsymbol{p}$ 中的某一行 $\boldsymbol{r}$,池化运算表示为

$$\boldsymbol{p}_{r,*}=f(\boldsymbol{m}_{r,j});\ j=1,2,\cdots,l$$

其中 f 函数用于获取序列中最大的 k 个数值,$*$ 代表对于矩阵中某一行的所有值均进行该计算,$\boldsymbol{m}$ 为卷积层生成的特征图,l 为特征图宽度.

模型的事件检测器部分由全连接层、Dropout 方法和 softmax 激活函数构成.全连接层用于将池化层输出向量的维度降至 2.Dropout 方法用于在训练过程中屏蔽一部分神经元的输出,防止模型过拟合.softmax 激活函数产生模型对于输入样本对应标签的预测,事件检测器形式化描述如下:

$$y=\mathrm{softmax}(\boldsymbol{w}\cdot(\boldsymbol{s}\odot\boldsymbol{d})+\boldsymbol{b})$$

其中 $\boldsymbol{s}$ 为池化层输出的特征,$\boldsymbol{w}$ 和 $\boldsymbol{b}$ 分别为可训练的全连接层权重和偏差,$\odot$ 代表向量按位相乘操作,$\boldsymbol{d}$ 为与 $\boldsymbol{s}$ 等长的伯努利随机变量.

在训练过程中,L_2 正则化项和对抗正则化项同时用于计算损失,模型的损失函数为

$$\tilde{L}(\boldsymbol{\theta},\boldsymbol{x},y)=L(\boldsymbol{\theta},\boldsymbol{x},y)+\lambda_{\mathrm{adv}}L(\boldsymbol{\theta},\boldsymbol{x}+\boldsymbol{\delta},y),$$

$$L(\boldsymbol{\theta},\boldsymbol{x},y)=\sum_{i=1}^{c}t_i\log y_i+\lambda_{L_2}\sum\boldsymbol{\theta}^2$$

其中 $L(\boldsymbol{\theta},\boldsymbol{x},y)$ 为不使用对抗样本训练的损失,c 为测试样本数量,t 为样本真实标签,$\boldsymbol{\theta}$ 为模型参数,λ_{adv} 和 λ_{L_2} 分别为对抗正则项系数和 L_2 正则项系数.

2 实验结果与分析

本文在 TwiMed[17] 和 TwitterADR[18] 两个语料集上进行实验.两个语料集均标注自推特,提供推文对应的 ID 和分类标签供使用者自行从推特爬取.由于其后推特用户注销或将发布内容删除等,目前仅部分推文可以爬取,这无疑增加了模型取得更好实验结果的难度.语料集统计信息详见表 1,两个语料集上可获取的样本比例均约为 60%;TwiMed 语料集在标注时考虑了正负样本比例,相对更为均衡,约为 1∶1.6;TwitterADR 语料集规模相比 TwiMed 更大,正负样本比例约为 1∶8.1.

表 1 语料集统计信息

Tab.1 The statistics of corpora

语料集名称	发布时样本数	可获得样本数	正例	负例
TwiMed	1 000	608	234	374
TwitterADR	10 822	6 966	765	6 201

实验的评价指标使用准确率(P)、召回率(R)和 F_1 值,其中 F_1 值用于综合衡量模型的 P 和 R,为实验中使用的主要评价指标.实验中使用的词嵌入维度为 100,图嵌入维度为 50,CNN 通道数为 3,卷积窗口大小分别为 3、4、5,学习率为 0.000 3,对抗噪声参数 ε 为 0.07.

在规模较小的语料集 TwiMed 上的实验结果如表 2 所示.其中 BLSTM 方法[19] 是一种将正向 LSTM 的输出和反向 LSTM 的输出拼接后用于药物不良反应事件检测的方法;KB-Embedding 方法[5] 利用来自维基百科的外部资源帮助模型预测;ATL 方法[7] 通过迁移学习,每次利用两个语料集进行模型训练;MTL 方法[9] 在进行检测任务的同时,通过实体识别任务进行多任务学习;Seq2seqAttn 方法[20] 是一种基于编码-解码架构和注意力机制的预测模型;TextRNN 方法[21] 以双向 LSTM 进行特征抽取;FastText 方法[22] 利用 n-gram 特征和浅层的神经网络进行预测.对于提供源代码的 Seq2seqAttn、TextRNN、FastText 等方法,使用其开源代码,在采用与本文方法相同词嵌入的条件下进行实验;其他基线方法采用相关论文中汇报的实验结果进行比较.

表 2 在 TwiMed 语料集上的实验结果对比

Tab.2 Experimental results comparison on TwiMed corpus

方法	准确率/%	召回率/%	F_1 值/%
BLSTM[19]	61.20	51.49	56.01
KB-Embedding[5]	59.60	61.44	60.42
ATL[7]	63.68	63.40	63.54
MTL[9]	66.56	63.80	64.82
Seq2seqAttn[20]	67.13	65.64	65.67
TextRNN[21]	69.58	68.52	68.50
FastText[22]	71.06	70.27	70.47
本文方法	76.14	75.26	75.25

实验结果显示，在 TwiMed 语料集上本文提出的模型在准确率、召回率和 F_1 值 3 项指标上均超过了各种基线方法，其中主要评价指标 F_1 值相较 FastText 模型提升了 4.78%. 可见对于仅有数百样本的小规模语料集，相对复杂的神经网络模型往往难以达到理想效果. 通过本文采用的两种正则化方法，配合相对浅层的神经网络模型能够适应较少的训练样本. 本文模型特征抽取部分使用的卷积神经网络，可训练参数少，更加适合小规模标注语料以及限定句子最大长度的推特文本.

为了进一步验证本文方法的有效性，在另一个规模稍大的语料集——TwitterADR 上与多种方法进行对比，见表 3. 其中 ME-TFIDF 方法[23]采用 TFIDF 作为原始特征表示，采用最大熵模型进行预测；CRNN 方法[23]在循环神经网络层后设置卷积层提取文本特征；SemiEnsemble 方法[6]通过 7 种模型的多数投票进行预测；MT-Atten 方法[8]在进行药物不良反应事件检测任务的同时，对句子中的适应症等实体进行标注，通过多任务学习提升模型的整体性能. 基线方法的结果来自相关论文中的汇报.

表 3　在 TwitterADR 语料集上的实验结果对比

Tab. 3　Experimental results comparison on TwitterADR corpus

方法	准确率/%	召回率/%	F_1 值/%
ME-TFIDF[23]1)	33.00	70.00	45.00
CRNN[23]1)	49.00	55.00	51.00
SemiEnsemble[6]	70.21	59.64	64.50
MT-Atten[8]	72.88	70.54	70.69
FastText[22]	74.85	70.28	72.15
Seq2seqAttn[20]	75.10	71.12	72.65
TextRNN[21]	74.67	73.55	73.69
本文方法	80.19	71.23	74.49

注：1）该方法论文中汇报的实验结果只提供了两位有效数字.

实验结果显示，在 TwitterADR 语料集上本文提出的模型在准确率和 F_1 值两项指标上超过了各种基线方法，其中主要评价指标 F_1 值相较 TextRNN 模型提升了 0.80%. 通过对比发现，TextRNN、Seq2seqAttn 模型相对其在小规模语料集上的表现提升明显，分别为 5.19% 和 6.98%，而本文方法性能则与小规模语料集上相当. 这说明本文方法在仅有数百条标注样本时就可以得到较为充分的训练，而 Seq2seqAttn 等模型在提供更多标注数据时可能仍具有更好的潜力. 采用 TFIDF 作为特征输入的最大熵模型获得了较高的召回率，但其准确率远低于各种采用词嵌入作为输入的模型；CRNN 模型性能落后于其他神经网络模型，主要原因可能在于其使用的词嵌入预训练语料并非源自社交媒体；SemiEnsemble 方法中大量利用了人工设计的特征，其准确率相对较高，但召回率较低；MT-Atten 方法通过在同一个语料集上构建分类和序列标注等多个生物医学领域任务，通过参数共享的形式使得模型学习到了关于原始文本更为丰富的特征，但同时也需要为模型提供额外的标注信息. 总体而言，各种神经网络方法通过自动学习有效的特征表示，相对传统机器学习方法，在准确率和召回率之间取得了较好的均衡. 本文方法在不使用人工构造特征和额外标注信息的条件下，相较各基线方法在药物不良反应事件检测任务中取得了更好的总体性能.

为了进一步验证本文所提两种方法的有效性和适用范围，将其应用于其他深度学习模型，并通过消融实验在不同规模的两个语料集上进行了对比. 在 TwiMed 语料集上的实验结果如表 4 所示.

表 4　在 TwiMed 语料集上的模型性能分析

Tab. 4　The analysis of model performance on TwiMed corpus

模型	Seq2seqAttn[20]			TextRNN[21]			FastText[22]			本文方法		
	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%
原始模型	67.13	65.64	65.67	69.58	68.52	68.50	71.06	70.27	70.47	73.61	71.34	71.31
＋对抗训练	68.80	67.71	67.93	72.06	69.49	69.48	71.65	71.26	71.06	74.80	73.35	73.43
＋图嵌入	68.88	68.28	66.77	72.34	70.90	70.10	71.76	71.48	71.02	74.43	74.30	74.22
＋两种方法	69.38	68.56	68.07	73.10	72.78	71.69	72.41	72.32	72.01	76.14	75.26	75.25

由实验结果可知，在数据规模较小时，两种方法在 4 种深度学习模型上均能够有效提升模型性能. 其中单独加入对抗训练方法时，在所有评价方法下的模型性能均超越未使用对抗训练的原始模型，F_1 值平均提升了 1.49%. 同样在单独加入图嵌入方法的情况下，各项指标也均超越原始模型，而且在总体上提升幅度高于对抗训练方法，F_1 值平均提升了 1.54%. 同时采用两种方法的模型在所有指标上均获得了最佳结果，F_1 值平均提升了 2.77%，与两种方法分别加入时提升的总和相近，说明这两种方法能够通过不同方式提升模型性能且同时使用时其优势能够互补. 本文方法在不同深度学习模型下均发挥了作用，表明本文方法具有良好的适用性. 两种方法在对本文基于卷积神经网络的模型带来了 3.94%的性能提升，高于在 4 种模型上的平均提升幅度，验证了本文模型整体上的合理性.

两种方法在更大规模语料集上的性能表现见表 5. 在单独应用两种方法时，模型在大部分性能指标上超越了用作对比的原始模型. 在单独加入对抗训练方法时，F_1 值平均提升了 0.59%. 单独加入图嵌入方法时，F_1 值平均提升了 0.55%. 可见在有更多标注样本进行训练的情况下，两种方法仍能提升模型性能，但提升幅度小于在小规模语料集上的表现. 其中图嵌入方法带来提升较其在小数据集上的表现差距更大，这说明模型在通过更多样本的词嵌入训练时能够在一定程度上学习到与药物以及药物不良反应候选词相关的上下文结构信息. 在同时加入两种方法时，4 种模型的 F_1 值平均提升了 1.19%，大于分别单独加入两种方法带来提升的总和. 本文模型仅取得了与单独加入图嵌入方法时类似的性能提升，而 Seq2seqAttn 的 F_1 值相较原始模型提升了 1.60%. 同时对比该模型在 TwiMed 上加入两种方法的实验结果，提升幅度为 6.18%，说明该模型在有更多训练样本的情况下性能有更大提升.

表 5 在 TwitterADR 语料集上的模型性能分析

Tab. 5 The analysis of model performance on TwitterADR corpus

模型	Seq2seqAttn[20]			TextRNN[21]			FastText[22]			本文方法		
	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%	准确率/%	召回率/%	F_1 值/%
原始模型	75.10	71.12	72.65	74.67	73.55	73.69	74.85	70.28	72.15	79.07	70.70	73.75
+对抗训练	77.26	71.58	73.34	76.34	73.04	74.66	75.50	70.01	72.22	79.10	71.51	74.38
+图嵌入	75.50	72.28	73.53	76.51	72.31	73.92	75.27	70.58	72.50	79.19	71.76	74.50
+两种方法	76.72	72.73	74.25	75.85	74.91	75.11	76.00	71.02	73.07	79.62	71.53	74.57

3 结 语

本文针对在推特文本中进行药物不良反应事件检测任务时面对的标注资源稀缺问题，将一种能够为模型训练提供更加丰富信息的图嵌入方法引入该任务. 将来自推特的语料构建成图结构进行预训练，生成的图嵌入能够在不同规模的模型训练过程中带来性能提升. 同时本文引入的另一种正则化方法通过在训练数据中添加微小扰动的方式提升模型的鲁棒性，实验结果显示该方法能够较好地与图嵌入方法互补. 在不同深度学习模型上的实验结果表明本文方法具有良好的适用性，尤其是在小规模语料集中发挥了更好的性能. 在未来的研究中，可以探索对抗迁移方法在该任务中的应用，尝试构建能够从不同语料集中获取共性特征的模型，从而进一步减少模型训练对标注样本数量的依赖.

参考文献：

[1] PLUMPTON C O, ROBERTS D, PIRMOHAMED M, *et al*. A systematic review of economic evaluations of pharmacogenetic testing for prevention of adverse drug reactions [J]. **Pharmacoeconomics**, 2016, **34**(8): 771-793.

[2] 国家药品不良反应监测中心. 国家药品不良反应监测年度报告(2019 年)[R/OL]. (2020-04-10)[2020-05-05]. http://www.cdr-adr.org.cn/tzgg_home/202004/t20200410_47300.html.
National Center for ADR Monitoring, China. Annual Report of National for ADR Monitoring,

China (2019) [R/OL]. (2020-04-10) [2020-05-05]. http://www.cdr-adr.org.cn/tzgg_home/202004/t20200410_47300.html. (in Chinese)

[3] HAZELL L, SHAKIR S A W. Under-reporting of adverse drug reactions: A systematic review [J]. **Drug Safety**, 2006, **29**(5): 385-396.

[4] SINNENBERG L, BUTTENHEIM A M, PADREZ K, *et al*. Twitter as a tool for health research: A systematic review [J]. **American Journal of Public Health**, 2017, **107**(1): e1-e8.

[5] STANOVSKY G, GRUHL D, MENDES P N. Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models [C] // **15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference**. Valencia: Association for Computational Linguistics, 2017: 142-151.

[6] LEE K, QAQIR A, HASAN S A, *et al*. Adverse drug event detection in tweets with semi-supervised convolutional neural networks [C] // **26th International World Wide Web Conference, WWW 2017**. Perth: International World Wide Web Conferences Steering Committee, 2017: 705-714.

[7] LI Zhiheng, YANG Zhihao, LUO Ling, *et al*. Exploiting adversarial transfer learning for adverse drug reaction detection from texts [J]. **Journal of Biomedical Informatics**, 2020, **106**: 103431.

[8] CHOWDHURY S, ZHANG Chenwei, YU P S. Multi-task pharmacovigilance mining from social media posts [C] // **The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018**. Lyon: Association for Computing Machinery, Inc., 2018: 117-126.

[9] GUPTA S, GUPTA M, VARMA V, *et al*. Multi-task learning for extraction of adverse drug reaction mentions from tweets [C] // **Advances in Information Retrieval - 40th European Conference on IR Research, ECIR 2018, Proceedings**. Grenoble: Springer Verlag, 2018: 59-71.

[10] KUHN M, LETUNIC I, JENSEN L J, *et al*. The SIDER database of drugs and side effects [J]. **Nucleic Acids Research**, 2016, **44**(D1): D1075-D1079.

[11] NIKFARJAM A, SARKER A, O'CONNOR K, *et al*. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features [J]. **Journal of the American Medical Informatics Association: JAMIA**, 2015, **22**(3): 671-681.

[12] PENNINGTON J, SOCHER R, MANNING C D. GloVe: Global vectors for word representation [C] // **EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference**. Doha: Association for Computational Linguistics (ACL), 2014: 1532-1543.

[13] WANG Daixin, CUI Peng, ZHU Wenwu. Structural deep network embedding [C] // **KDD 2016 - Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. San Francisco: Association for Computing Machinery, 2016: 1225-1234.

[14] SZEGEDY C, ZAREMBA W, SUTSKEVER I, *et al*. Intriguing properties of neural networks [C] // **2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings**. Banff: International Conference on Learning Representations, ICLR, 2014.

[15] MIYATO T, DAI A M, GOODFELLOW I. Adversarial training methods for semi-supervised text classification [C] // **5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings**. Toulon: International Conference on Learning Representations, ICLR, 2017.

[16] KIM Y. Convolutional neural networks for sentence classification [C] // **EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference**. Doha: Association for Computational Linguistics (ACL), 2014: 1746-1751.

[17] ALVARO N, MIYAO Y, COLLIER N. TwiMed: Twitter and PubMed comparable corpus of drugs, diseases, symptoms, and their relations [J]. **JMIR Public Health and Surveillance**, 2017, **3**(2): e24.

[18] SARKER A, GONZALEZ G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training [J]. **Journal of Biomedical Informatics**, 2015, **53**: 196-207.

[19] COCOS A, FIKS A G, MASINO A J. Deep learning for pharmacovigilance: recurrent neural

network architectures for labeling adverse drug reactions in Twitter posts [J]. **Journal of the American Medical Informatics Association**, 2017, **24**(4): 813-821.

[20] DU C, HUANG L. Text classification research with attention-based recurrent neural networks [J]. **International Journal of Computers, Communications & Control (IJCCC)**, 2018, **13**(1): 50-61.

[21] LI Fei, ZHANG Meishan, FU Guohong, *et al*. A Bi-LSTM-RNN model for relation classification using low-cost sequence features [Z/OL]. [2020-02-02]. https://arxiv.org/abs/1608.07720v1.

[22] JOULIN A, GRAVE E, BOJANOWSKI P, *et al*. Bag of tricks for efficient text classification [C] // **15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference**. Valencia: Association for Computational Linguistics (ACL), 2017: 427-431.

[23] HUYNH T, HE Yulan, WILLIS A, *et al*. Adverse drug reaction classification with deep neural networks [C] // **COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers**. Osaka: Association for Computational Linguistics, ACL Anthology, 2016: 877-886.

Detection method of adverse drug events from social media based on graph embeddings

SHEN Chen, LIN Hongfei*

(School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China)

Abstract: Adverse drug event is a major cause of morbidity and mortality of hospitalized patients. Since the traditional spontaneous self-reporting system is experiencing serious underreporting issues, and the research on the use of social media, such as Twitter, as a data source for the detection of adverse drug events has received increasing attention in recent years. Deep learning models usually require many training samples. However, due to the characteristics of user-generated contents and time-consuming data annotation process, related research is faced with the problems caused by small scale annotation but highly noisy datasets, which restricts deep learning models to achieve good results. Accordingly, two regularization methods are introduced at the representation level, which are graph embedding-based data augment and adversarial training, to improve the performance of detecting adverse drug events under such condition. The applicable scopes of these two methods are analyzed and discussed through experiments. Combining with the convolutional neural network, an adverse drug event detection scheme that can make full use of the two methods is proposed, and the experimental results testify the feasibility.

Key words: social media; graph embeddings; adversarial training; adverse drug event detection; deep learning