

面向社交媒体多特征增强的药物不良反应检测

李浩, 邱云志, 林鸿飞*

大连理工大学, 计算机科学与技术学院, 大连, 116024
(lihao, yzqiu@mail.dlut.edu.cn, hflin@dlut.edu.cn)

摘要

社交媒体是药物不良反应(ADR)检测的重要途径之一。本文提出一个基于社交媒体的药物不良反应检测模型DMFE,以全面捕捉患者对药物使用的反馈信息。与传统的文本检测相比,社交媒体数据中通常会有语法不规范与单词拼写错误的问题。本文提取出社交媒体数据的抽象语义表示(AMR)使用图注意力网络(GAT)学习抽象语义特征提高模型对语义信息的理解,使用字符级卷积神经网络(charCNN)捕获字符特征以减少单词拼写错误带来的影响。此外,本文使用提示学习的方法融入MedDRA药物不良反应领域关键词进一步增强模型对领域知识的理解能力。经实验评估,本文模型DMFE在CADEC、TwiMed两个数据集上F1值与基线模型相比取得最优效果。

关键词: 社交媒体; 药物不良反应; 文本检测

Adverse drug reaction detection with multi-feature enhancement for social media

Hao Li, Yunzhi Qiu, Hongfei Lin*

School of Computer Science and Technology, Dalian University of Technology, Dalian, 116024
(lihao, yzqiu@mail.dlut.edu.cn, hflin@dlut.edu.cn)

Abstract

Social media is one of the important ways to detect adverse drug reactions (ADR). This paper proposes a social media-based adverse drug reaction detection model DMFE to capture patients' feedback on drug use. Compared with traditional text detection, social media data often contains grammatical irregularities and word spelling errors. In this paper, abstract semantic representation (AMR) of social media data is extracted, abstract semantic features are learned by graph attention network (GAT) to improve the model's understanding of semantic information, and character level convolutional neural network (charCNN) is used to capture character features to reduce the impact of word spelling errors. In addition, this paper uses the method of prompt learning to integrate key words in the field of MedDRA adverse drug reactions to further enhance the model's understanding of domain knowledge. According to the experimental evaluation, the F1 value of the DMFE model in this paper achieves the optimal effect compared with the baseline model on the CADEC and TwiMed datasets.

Keywords: Social media, Adverse drug reaction, Text detection

* 通讯作者

1 引言

药物不良反应 (Adverse Drug Reaction, 简称ADR) 是指正常剂量的药物用于预防、诊断、治疗疾病或调节生理机能时出现有害的、与用药目的无关的反应(National Center for ADR Monitoring, China, 2023)。这些反应是预期之外的, 会严重影响患者的生理状态与健康状况, 甚至会有致命的后果。因此及时准确地发现潜在的不良反应对于确保药物的安全性至关重要(Marano et al., 2005)。

随着社交媒体的快速发展, 大量用户选择在社交媒体平台上分享自己的用药经验, 其中包括处方药和非处方药的使用、副作用以及治疗效果, 社交媒体中的大量用药信息为在社交媒体中检测药物不良反应提供可能。在社交媒体中检测药物不良反应相较于普通文本检测任务主要有以下的特点:

(1) 在社交媒体中, 用户经常以非正式的方式分享交流, 导致数据语法层面规范性普遍缺失。例如, 在句子中省略主语、谓语、连词, 或者混用不同时态的动词, 这样的语言风格会对检测模型造成理解上的困扰。

(2) 使用社交媒体过程中, 用户追求表达速度导致单词拼写错误、使用缩写的现象十分常见, 例如将“their”误写成“there”, 用“meddizzy”来代替“medication”和“dizzy”等。另外对词性的不正确使用也是普遍问题, 用户可能在不经意间将名词用作动词, 将形容词用作名词, 例如将“affect”误用为名词形式“effect”, 以上都会给模型检测带来不确定性。

(3) 社交媒体数据中关于药物不良反应的讨论往往围绕着病痛、药物名称以及不良反应症状等关键词展开。用户在描述某种药物的副作用时可能会提到诸如“headache”、“nausea”、“insomnia”等病痛相关的单词; 如“aspirin”、“antibiotic”等药物名称相关单词; 如“skin allergy”、“rapid heartbeat”等不良反应的的症状相关单词, 增强模型对领域知识的理解有助于提高模型在社交媒体中药物不良反应检测的能力。

针对上述问题, 本文提出一个融合药物不良反应领域知识, 字符级别信息以及抽象语义表示 (AMR) 的模型: 面向社交媒体多特征增强的药物不良反应检测 (Adverse drug reaction detection with multi-feature enhancement for social media, 简称DMFE), 本文工作如下:

(1) 对MedDRA数据库进行预处理并完成词级编码以获取药物不良反应领域词典。将社交媒体数据在词典中检索, 并将检索到的关键词使用提示学习的方式与社交媒体数据共同融入Prompt模板。

(2) 使用字符级别的卷积神经网络 (charCNN) 捕捉社交媒体数据的字符级别特征。卷积神经网络可以捕获到单词内外部特征学习到不同单词之间的相似性从而解决单词字母拼写错误问题和词性使用错误的问题, 意思相近但词性不同的单词之间通常只在前后缀有区分。

(3) 社交媒体数据的口语化往往导致语法缺失, 传统使用句法依赖树增强模型理解语义的方式不再适用。本文使用图注意力网络 (GAT) 学习抽象语义表示 (AMR) 的有效信息, 提高模型对语义信息的理解。

2 相关工作

2.1 文本的分类检测

随着深度学习的发展, 基于深度学习的模型应用到文本检测任务上取得不错的效果。Chen (2015)提出的基于卷积神经网络的TextCNN模型, 首次将CNN应用到文本分类上。Zhang et al. (2015)提出charCNN, 将句子离散化为字符级别表示并使用CNN提取字符级信息, 字符级信息可以使模型自然而然学会拼写错误和表情符号等异常的字符组合。随着研究的进一步推进, 预训练语言模型被应用到自然语言处理的任务中, Sun et al. (2019)研究了Bert在文本分类任务上的不同微调方法, 包括文本的预处理、逐层学习率和灾难性遗忘等问题, 在文本分类任务上取得了较大的性能改进。

2.2 药物不良反应检测

药物不良反应检测任务早期研究工作集中在使用规则匹配以及机器学习的方法从生物医学文献和临床报告中进行检测(Benton et al., 2011; Kuhn et al., 2010; Yates and Goharian, 2013;

Patki et al., 2014)。尽管它们在一定程度上取得了不错的效果，这些方法严重依赖于领域知识库与手工特征制作，需要大量专业生物医学领域人员的介入，给检测工作带来额外成本。近年来，深度学习也广泛应用于药物不良反应检测，Huynh et al. (2016)提出了两种新的卷积神经网络模型CRNN、CNNA用于药物不良反应的检测。Zhang et al. (2021)提出了ANNSA模型结合情感注意力机制和对抗学习的检测模型。Alimova and Solovyev (2018)提出交互注意力网络(IAN)从用户评论中检测药物不良反应。Wu et al. (2018)提出了一种多头自注意力和分层推文表示的检测方法。Qiu et al. (2023)提出了融入领域关键词的KESDT模型，融入外部领域知识对药物不良反应检测增强，但没有考虑到数据中有大量不规范的语法与单词会造成很大噪声。目前没有检测模型既考虑到社交媒体数据不规范的问题又采用领域关键词的方式提高模型对领域知识的理解能力，本文提出DMFE使用多特征融合的方式解决以上问题，提高药物不良反应的检测能力。

3 模型结构

针对在社交媒体中检测药物不良反应的任务特点，本文提出的DMFE模型结构如图1所示，具体如下：（1）输入数据：模型的输入共分为两部分，一部分是需要检测的社交媒体数据，另一部分是从MedDRA预处理编码后的词典中检索到的药物不良反应领域关键词。（2）多特征提取层：分别包括基于Prompt的领域信息与社交媒体数据融合提取，使用charCNN进行的字符级信息提取与使用GAT网络进行的AMR语义信息提取。（3）特征融合层：将特征提取层提取到的三种不同维度特征进行融合。（4）目标检测层：将融合后的特征经过全连接层，然后通过SoftMax函数输出检测结果。

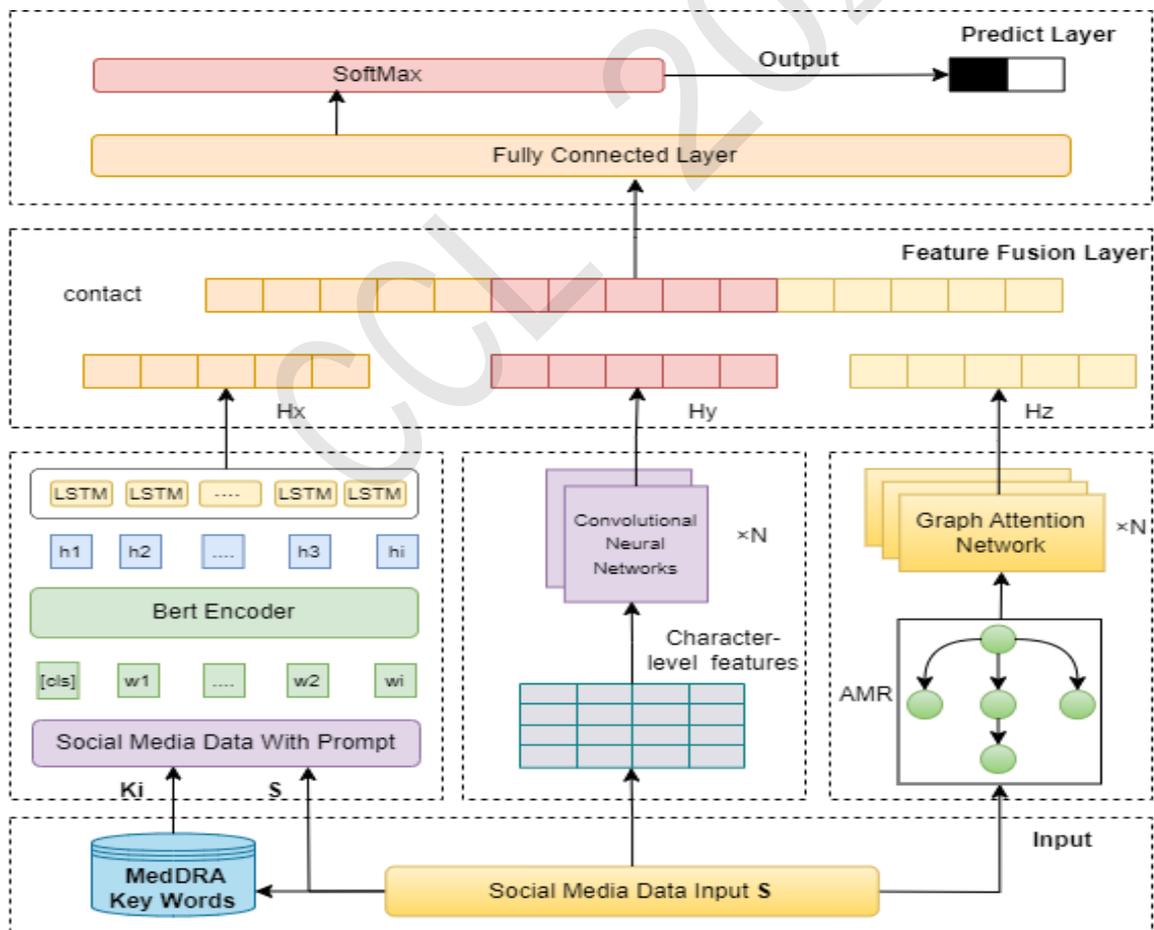


Figure 1: DMFE模型结构图

3.1 提示学习融合领域知识

3.1.1 MedDRA

MedDRA全称为监管活动医学词典 (Medical Dictionary for Regulatory Activities) (Mozzicato, 2009), 是由世界卫生组织主导发展的一种用于药物安全监测的临床术语分类系统。MedDRA广泛应用于药物开发和监管领域, 是药物不良事件报告的国际标准, 它包括数千个术语用于描述药物治疗过程中患者可能遇到的各种不良事件和副作用。MedDRA收录的医学术语范围包括: 体征、症状、疾病、诊断、适应证、各类检查的名称和定性结果等, 本文将MedDRA预处理并过滤编码为含有12033个药物不良反应领域关键词的词典。

3.1.2 提示学习

提示学习 (Prompt learning) (Liu et al., 2023) 可以通过关键信息的提供, 激发模型学习理解能力, 本文将社交媒体数据在药物不良反应词典中检索到领域关键词 \mathbf{K}_i , 和社交媒体数据 \mathbf{S} 按照模板构建提示学习语句, 从而融入药物不良反应关键词信息增强模型对领域知识理解。本文最终构建Prompt格式为: The key words: \mathbf{K}_i [SEP]Is there adverse drug reaction in the following sentence?[SEP] \mathbf{S} 。

3.1.3 编码层

本文使用Bert(Devlin et al., 2018)作为社交媒体数据与领域关键词构造Prompt之后的编码层。Bert是一种预训练语言模型, 其基于Transformer的Encoder架构, 通过在大规模文本语料库上进行预训练, 学习了丰富的语言表示, 其引入了预训练微调的范式, 允许在特定任务上微调模型, 从而更好地适应特定领域。

3.1.4 特征提取器

通过Prompt融合社交媒体数据与领域关键词之后输入信息长度大大增加, 为获取长序列信息的深层次维度特征, 本文使用LSTM模型作为编码层词向量的特征提取器。LSTM自循环的权重是不固定的, 根据上下文来确定权重值, 可以更好地处理长距离依赖问题, LSTM引入了输入门、遗忘门和输出门, 输入门控制当前计算的新状态以多大程度更新到记忆单元中; 遗忘门控制前一步记忆单元中的信息有多大程度被遗忘掉; 输出门控制当前的输出有多大程度上取决于当前的记忆单元, 进而实现了对较长序列的语义长短期记忆, 具体公式如(1)-(5)所示。

$$i_t = \sigma(\mathbf{W}_i x_t + \mathbf{U}_i h_{t-1} + \mathbf{V}_i c_{t-1}) \quad (1)$$

$$f_t = \sigma(\mathbf{W}_f x_t + \mathbf{U}_f h_{t-1} + \mathbf{V}_f c_{t-1}) \quad (2)$$

$$o_t = \sigma(\mathbf{W}_o x_t + \mathbf{U}_o h_{t-1} + \mathbf{V}_o c_{t-1}) \quad (3)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(\mathbf{W}_c x_t + \mathbf{U}_c h_{t-1}) \quad (4)$$

$$h_t = o_t \times \tanh(c_t) \quad (5)$$

i_t 代表输入门, f_t 代表遗忘门, o_t 代表输出门, h_t 代表隐藏状态节点, c_t 代表记忆单元, x_t 是当前时刻 t 的输入, σ 是logistic激活函数, $\mathbf{W}, \mathbf{U}, \mathbf{V}$ 是权重矩阵, 其中 $\mathbf{V}_i, \mathbf{V}_f, \mathbf{V}_o$ 是对角矩阵。

3.2 抽象语义信息提取

3.2.1 AMR抽象语义表示

抽象语义表示(abstract meaning representation, AMR)是句子语义的一种表示方法, 可以将社交媒体数据的语义抽象为一个单根有向无环图, 其中实词抽象为概念节点, 实词之间的关系抽象为带有语义关系标签的有向弧, 同时忽略虚词和形态变化体现的较虚的语义(如the、单复数等), 本文使用amrlib⁰对社交媒体数据进行解析得到相对应的抽象语义表示。以图2为例, A为AMR原始格式, B为直观抽象语义表示图, 使用penman库¹将抽象语义图格式化, 为

⁰<https://github.com/bjascob/amrlib/>

¹<https://pypi.org/project/Penman/1.3.0/>

保证AMR语义的词汇信息多样性表示并且加快模型训练推理速度，使用word2vec²对AMR语义图中每个节点进行词嵌入表示，无匹配词嵌入使用全0向量表示。

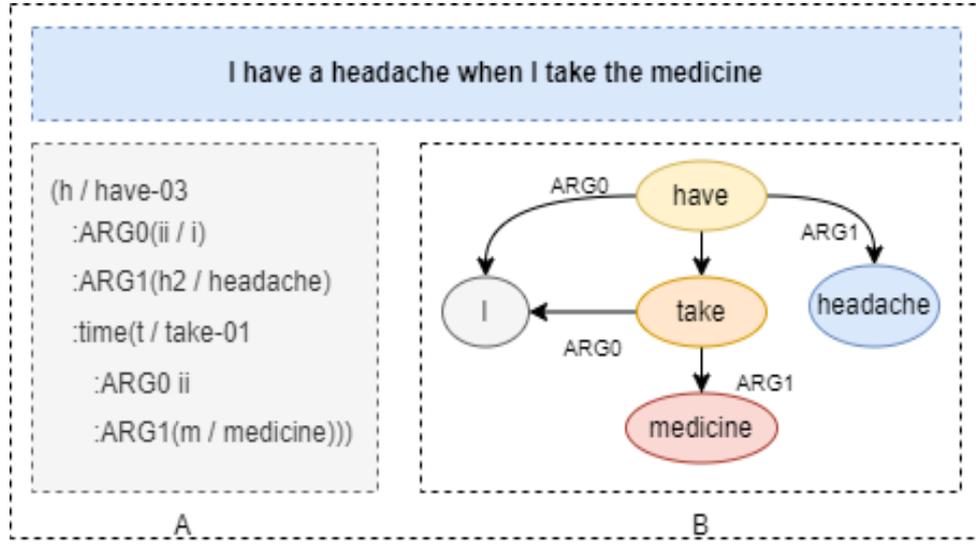


Figure 2: 社交媒体数据的抽象语义表示

3.2.2 语义特征学习

图注意力网络（GAT）可以动态捕获社交媒体数据AMR中各节点间的相关性，根据AMR语义图的不同，在聚合邻域特征时为各个节点分配不同权重，因此GAT比图神经网络（GCN）等基于静态图的方法具有更大的表示空间和更强的表示能力。单个节点更新过程如图3所示，中心节点 h_i 扫描所有邻域节点 h_j 对它们的特征向量进行加权求和得到聚合后的特征向量，其中加权求和的权重 α_{ij} 是由 h_i 对每个邻域节点 h_j 的注意力分数 e_{ij} 经过SoftMax后得到的， $h_i \in R^{d_h}$ 是第 i 个节点的输入特征向量， d_h 表示隐藏层的特征维度， $\mathbf{W} \in R^{d_h \times d_h}$ 和 $\mathbf{a} \in R^{2d_h}$ 表示可学习的参数， \parallel 表示拼接操作，如公式(6)(7)所示。

$$e_{ij} = \text{leakyReLU}(\mathbf{a}^T [\mathbf{W}h_i \parallel \mathbf{W}h_j]) \quad (6)$$

$$\alpha_{ij} = \text{SoftMax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i \cup \{i\}} \exp(e_{ik})} \quad (7)$$

聚合时的注意力系数是由中心节点的特征和邻域节点的特征决定的，其随特征的变化自适应地发生改变，因此具有更大的表示空间。如公式(8)所示， N_i 表示节点 h_i 的邻域节点集合， \cup 表示合并操作，Relu表示激活函数，聚合邻域节点特征得到中心节点 h_i 新的特征表示 h'_i 。

$$h'_i = \text{Relu} \left(\sum_{j \in N_i \cup \{i\}} \alpha_{ij} h_j \right) \quad (8)$$

为了进一步提高GAT对AMR语义的建模能力，本文设置多个并行的注意力头如公式(9)所示， K 表示注意力头的数量，每个注意力头都有独属于自己的参数， $h_i^{(k)}$ 表示第 k 个注意力头输出的特征向量， $\mathbf{W}_o \in R^{d_h \times Kd_h}$ 为参数矩阵，最后对更新后的所有AMR语义节点做平均池化得到抽象语义表示特征。

$$h'_i = \mathbf{W}_o \left[\parallel_{k=1}^K h_i^{(k)} \right] \quad (9)$$

²<https://github.com/mmhaltz/word2vec-GoogleNews-vectors>

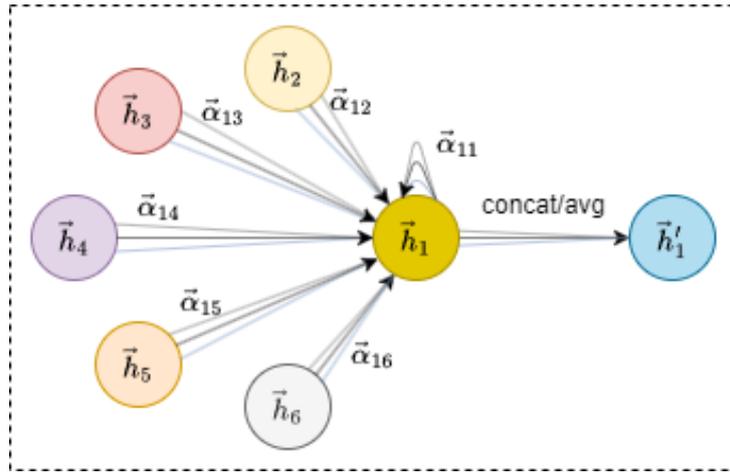


Figure 3: 社交媒体数据AMR语义信息的GAT节点状态更新

3.3 字符级信息提取

字符级别特征对解决社交媒体数据中单词表达不规范的问题有很大的优势。通过学习数据整体的字符级表示可纠正单词错误以及学习单词的共性特征，例如模型学习到字符特征后会理解前缀un和dis通常在英文中表示否定，而“their”因为和“there”单词形态相似，二者非常可能由一个误拼写为另一个。Zhang et al. (2021)提出对每个单词使用CNN进行编码提取词嵌入，这样很难利用不同单词之间形态相似关系，只能学习每个单词内部特征，本文使用charCNN(Zhang et al., 2015)直接对字符化的社交媒体数据提取全局字符级别的信息。

本文构建由符号和字母组成的字符表对应使用one-hot表示，共统计了69个字符包含26个英文字母、10个数字、33个其他的字符，当出现不存在字符时使用全0向量表示。将每个社交媒体数据中的字符处理成统一长度，通过多头charCNN和最大池化进行字符级特征提取，然后进行全连接操作生成字符级别的特征向量，过程如图4所示。

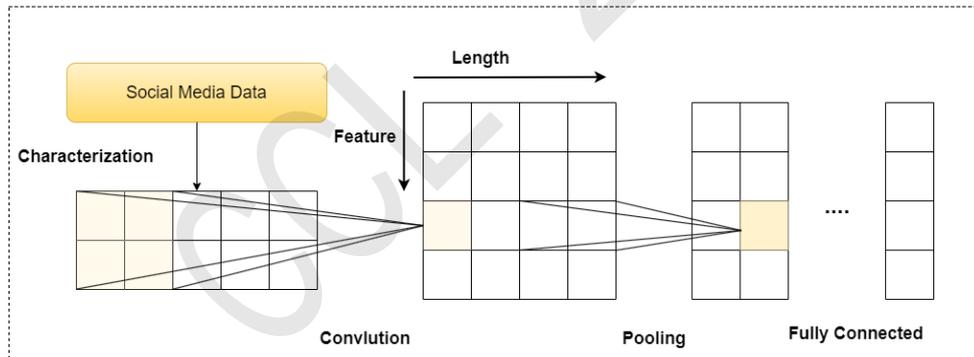


Figure 4: 社交媒体数据字符级别特征提取

3.4 多维度特征融合层

特征融合层将基于Prompt的领域信息与社交媒体数据融合特征 H_x 、抽象语义表示特征 H_y 、字符级信息特征 H_z 融合后得到特征融合表示 H_{concat} ，如公式(10)所示。

$$H_{concat} = \text{Concat}(H_x, H_y, H_z) \quad (10)$$

3.5 检测结果输出

3.5.1 损失函数

由于药物不良反应检测任务中存在严重的数据不平衡问题，本模型损失函数采用Focal Loss，相比交叉熵损失函数其对于分类不准确的样本损失没有改变，对于分类准确的样本损失

会变小。整体而言相当于增加了分类不准确样本在损失函数中的权重，具体分类函数如公式 (11) 所示。

$$L_{fl} = -(1 - p_t)^\gamma \log(p_t) \quad (11)$$

$\gamma > 0$ 为可调节因子， p_t 反映了预测值与真实标签值的接近程度， p_t 越大说明预测值越接近真实标签值分类越准确。

3.5.2 目标检测层

融合层向量 H_{concat} 经过全连接层通过权重矩阵 W_{out} 得到最终特征向量 Y 如公式 (12) 所示，然后通过 SoftMax 函数进行分类选择较大概率值作为分类结果，如公式 (13) 所示。

$$Y = W_{out} H_{concat} \quad (12)$$

$$\text{SoftMax}(y_i) = \frac{e^{y_i}}{\sum_{i=1}^n e^{y_i}} \in (0, 1) \quad (13)$$

4 实验分析

4.1 数据集介绍

为公平评估本文模型 DMFE 的性能，在两个数据集上对模型进行性能评估。数据集情况如表1所示，下面对使用的两个数据集进行介绍：

(1) CADEC (Karimi et al., 2015): 数据来自于医学论坛 AskaPatient 用户评论的帖子，每个帖子标有药物、副作用、症状和疾病，将包含副作用的帖子标记为 Yes，其余标记为 No。

(2) TwiMed (Alvaro et al., 2017): 数据由两部分组成，来自社交媒体平台 Twitter 的 TwiMed-Twitter 和来自生物医学文献的 TwiMed-PubMed。每个文档都附有疾病、症状、药物及其关系的注释，三种类型的关系为 Outcome-negative、Outcome-positive 和 Reason-to-use。当关系类型为 Outcome-positive 时，标记为 Yes，含有药物不良反应信息；其余标记为 No，不含药物不良反应信息。

数据集	ADR	No-ADR	Total	Max.length
CADEC	2478	4996	7474	46
TwiMed	426	1182	1608	241

Table 1: 社交媒体药物不良反应数据集介绍

4.2 实验参数

实验使用 Pytorch 框架，在 Ubuntu 系统下进行，使用 RTX3080 进行训练，优化器使用 Adam。在两个数据集上进行 30 个 epoch 的训练，batchsize 设置为 32，学习率设置为 1×10^{-5} ，Dropout 设置为默认参数 0.1，Bert 的隐藏层大小设置为 768 维，最大句子长度设置为 128，损失函数超参数 γ 设置为 3。

4.3 评价指标

本文使用常用的三个评价指标准确率 (P)、召回率 (R)、F1 值来对模型进行评价。其中 TP 表示药物不良反应正确分类的数据，FN 表示含有药物不良反应的数据预测为不含不良反应，FP 表示不含不良反应的数据预测为含有药物不良反应，如公式 (14) - (16) 所示。

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2PR}{P + R} \quad (16)$$

4.4 DMFE与基线模型对比实验

本文模型DMFE与7个基线模型进行对比实验分析，CADEC与TwiMed数据集上的对比实验结果如表2所示，基线模型介绍如下：

(1) CRNN(Huynh et al., 2016): 一种循环卷积神经网络模型，用于文本分类任务，该模型将卷积层叠加在循环神经网络之上。

(2) HTR+MSA(Wu et al., 2018): 该模型具有分层的多头自注意力机制结构，利用层级表示学习文本的多维度特征，并用于药物不良反应检测。

(3) CNN+corpus(Li et al., 2020): 该方法直接在每个数据集的训练集部分增加了ADE数据，实现对训练语料的扩充，然后在CNN方法上使用扩充后的训练集进行训练。

(4) CNN+transfer(Li et al., 2020): 该模型使用迁移学习方法，将源数据集与目标数据集共同训练，采用CNN学习两部分数据的共享特征，最终使用各自的文本分类器进行预测，得到两部分数据各自的预测结果。

(5) ATL(Li et al., 2020): 结合对抗训练和迁移学习，学习两部分的特征，实现不同数据之间的融合。

(6) ANNSA(Zhang et al., 2021): 融合情感感知的注意力机制的对抗网络模型。

(7) KESDT(Qiu et al., 2023): 融合了药物不良反应领域知识的多层Transformer模型，在Transformer层级之间融入领域关键词。

模型	CADEC			TwiMed		
	P(%)	R%	F1(%)	P(%)	R(%)	F1(%)
CRNN	61.26	65.96	63.52	68.52	66.43	67.46
HTR + MSA	60.67	61.70	61.18	66.58	63.62	65.07
CNN + corpus	52.75	61.28	56.69	60.51	61.50	61.00
CNN+ transfer	61.84	60.00	60.91	69.58	61.74	65.42
ATL	63.68	63.40	63.54	70.84	65.02	67.81
ANNSA	58.82	73.34	64.18	-	-	-
KESDT	88.16	87.63	87.82	71.22	69.15	68.63
DMFE	88.40	89.63	89.01	76.27	71.43	73.77

Table 2: 模型在CADEC、TwiMed数据集上的实验结果

通过实验数据可以看出DMFE考虑到单词拼写、语法不规范问题并且使用Prompt与领域关键词提高模型对领域知识理解之后，在两个数据集上的F1值都取得了比基线模型更好的效果。相较于没有引入预训练模型的工作DMFE有很大的提升，相较于KESDT模型本文除引入了药物不良反应领域知识外还引入字符级信息与AMR特征，解决了社交媒体数据的单词和语法不规范性的问题从而减少数据中的噪声，另外本文还尝试使用Llama2-base(Touvron et al., 2023)通过指令微调训练生成式的检测模型，但由于数据集分布不均衡，导致微调之后很难将社交媒体数据检测为含药物不良反应的数据，并且相较于DMFE占用大量的计算资源，故没有列为合格基线模型作为比较。

4.5 多特征提取层的消融实验

为进一步评估模型各维度特征提取的有效性，本文从模型中删除基于Prompt的领域信息、charCNN字符特征提取、AMR-GAT语义特征提取进行了消融实验，实验结果如表3、表4所示。消融之后模型与DMFE相比，在CADEC数据集上的F1值分别下降了1.72%、1.36%和1.04%，在TwiMed数据集上F1值分别下降了4.69%、1.69%和2.02%，充分证明了不同维度特征提取在社交媒体中进行药物不良反应检测起到重要作用。其中消融基于Prompt的领域知识后影响最大，证明Prompt和领域知识可以帮助模型更好地理解药物不良反应，从而提高检测效果。

模型方法	P(%)	R(%)	F1(%)
Bert(bert-base-uncased)	86.76	85.43	86.09
w/o prompt learning	87.02	87.56	87.29
w/o charCNN	86.32	89.02	87.65
w/o AMR-GAT	87.43	88.52	87.97
DMFE	88.40	89.63	89.01

Table 3: CADEC数据集上的消融实验

模型方法	P(%)	R(%)	F1(%)
Bert(bert-base-uncased)	66.71	62.38	64.47
w/o prompt learning	69.82	68.36	69.08
w/o charCNN	70.98	73.22	72.08
w/o AMR-GAT	71.49	72.02	71.75
DMFE	76.27	71.43	73.77

Table 4: TwiMed数据集上的消融实验

4.6 低资源情况下模型表现结果

社交媒体数据在药物不良反应方向数据集具有稀缺性，所以模型在低资源情况下的表现尤为重要。为验证DMFE在低资源场景下的有效性，分别取训练集的20%、40%、60%、80%数据量在DMFE上进行实验与Bert（Bert-base-uncased）模型进行对比，可以看出DMFE相较于Bert模型在低资源下有更好的性能，实验结果如图5所示。

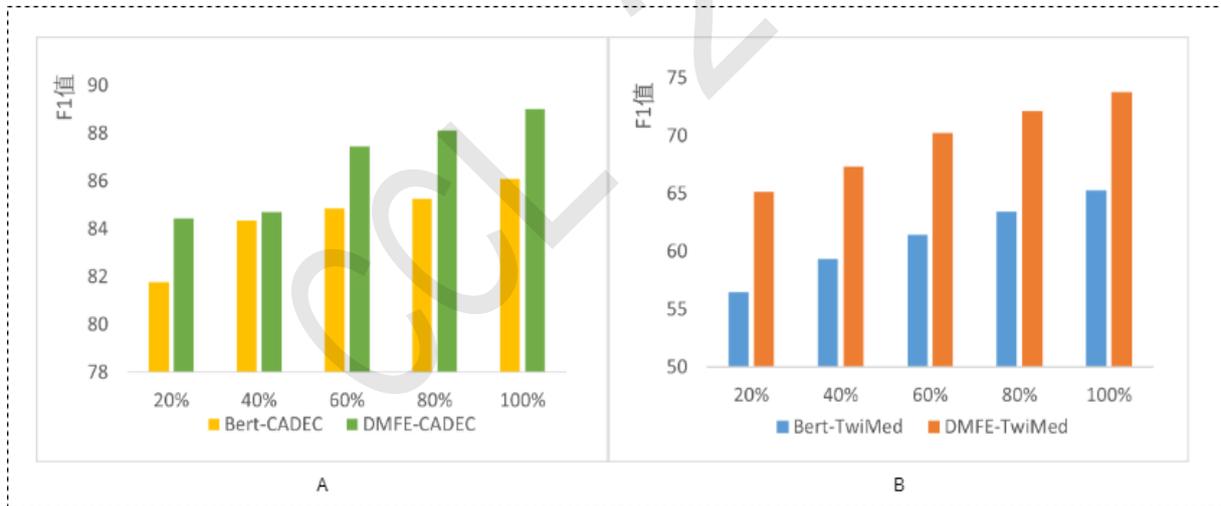


Figure 5: 低资源情况下DMFE与Bert的对比实验

4.7 不同Prompt对检测效果的影响

基于提示学习融入领域知识时不同Prompt的选取对模型检测能力影响不同，本文共选取三条不同的Prompt指令进行实验，在CADEC与TwiMed数据集上的实验结果如图6所示，不同Prompt对模型性能有不可忽略的影响。在CADEC数据集上有1.16%的差异，在TwiMed数据集上有1.93%的差异，如何选取更能促进模型理解的Prompt模板对模型检测能力的提升至关重要且值得进一步探讨。

Prompt指令	CADEC数据集F1值	Twimed数据集F1值
ADR key words: K_i [SEP] S	87.85	72.17
Is there ADR in the following sentence with the key words: K_i [SEP] S	88.75	74.10
The key words: K_i [SEP] Is there adverse drug reaction in the following sentence? [SEP] S	89.01	73.77

Figure 6: 不同Prompt对检测效果的影响

5 结束语

为提高在社交媒体中药物不良反应的检测能力，本文充分考虑数据与任务特点使用多特征融合的方式提高检测效果，提出了DMFE模型：利用提示学习的方式融入药物不良反应领域关键词信息，另外采用卷积神经网络charCNN捕获字符级别信息，解决用户单词拼写不准确的问题，结合图注意力网络（GAT）学习句子的抽象语义表示（AMR）提高模型对语义的理解能力，以解决数据语法不规范问题。对DMFE的消融实验证明了各维度特征提取的有效性，此外本文还探讨了低资源环境下模型的有效性，以及不同Prompt模板对实验结果的影响。

在未来，我们对社交媒体中药物不良反应的检测主要集中在两方面：一方面结合生成式模型对药物不良反应的检测模型进行改进，充分发挥生成式模型大量知识信息的优势；另一方面将尝试引入半监督的方式通过其他领域数据进行迁移学习，以解决药物不良反应数据集稀缺的问题。

参考文献

- Ilseayr Alimova and Valery Solovyev. 2018. Interactive attention network for adverse drug reaction classification. In *Artificial Intelligence and Natural Language: 7th International Conference, AINL 2018, St. Petersburg, Russia, October 17–19, 2018, Proceedings 7*, pages 185–196. Springer.
- Nestor Alvaro, Yusuke Miyao, Nigel Collier, et al. 2017. Twimed: Twitter and pubmed comparable corpus of drugs, diseases, symptoms, and their relations. *JMIR public health and surveillance*, 3(2):e6396.
- Adrian Benton, Lyle Ungar, Shawndra Hill, Sean Hennessy, Jun Mao, Annie Chung, Charles E Leonard, and John H Holmes. 2011. Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *Journal of biomedical informatics*, 44(6):989–996.
- Yahui Chen. 2015. Convolutional neural network for sentence classification. Master’s thesis, University of Waterloo.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- National Center for ADR Monitoring, China. 2023. 国家药品不良反应监测年度报告（2022年）. 中国病毒病杂志, 13:245–251.
- Trung Huynh, Yulan He, Alistair Willis, and Stefan Rüger. 2016. Adverse drug reaction classification with deep neural networks. *Coling*.
- Sarvnaz Karimi, Alejandro Metke-Jimenez, Madonna Kemp, and Chen Wang. 2015. Cadec: A corpus of adverse drug event annotations. *Journal of biomedical informatics*, 55:73–81.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *Eprint Arxiv*.
- Michael Kuhn, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. 2010. A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, 6(1):343.
- Zhiheng Li, Zhihao Yang, Ling Luo, Yang Xiang, and Hongfei Lin. 2020. Exploiting adversarial transfer learning for adverse drug reaction detection from texts. *Journal of biomedical informatics*, 106:103431.

- Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.
- Lara Magro, Elena Arzenton, Roberto Leone, Annette Rudolph, and Ugo Moretti. 2021. Identifying and characterizing serious adverse drug reactions associated with drug-drug interactions in a spontaneous reporting database. *Frontiers in Pharmacology*, 11:622862.
- Cinzia Marano, Laura Murianni, and Laura Sticchi. 2005. To err is human. building a safer health system. *Epidemiology Biostatistics Public Health*, 2(3-4).
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Patricia Mozzicato. 2009. Meddra: an overview of the medical dictionary for regulatory activities. *Pharmaceutical Medicine*, 23:65–75.
- Apurv Patki, Abeer Sarker, Pranoti Pimpalkhute, Azadeh Nikfarjam, Rachel Ginn, Karen O'Connor, Karen Smith, and Graciela Gonzalez. 2014. Mining adverse drug reaction signals from social media: going beyond extraction. *Proceedings of BioLinkSig*, 2014:1–8.
- Yunzhi Qiu, Xiaokun Zhang, Weiwei Wang, Tongxuan Zhang, Bo Xu, and Hongfei Lin. 2023. Kesdt: knowledge enhanced shallow and deep transformer for detecting adverse drug reactions. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 601–613. Springer.
- Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune bert for text classification? In *Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, proceedings 18*, pages 194–206. Springer.
- Kanchan Sushim and Gaidhane Abhay. 2023. Social media role and its impact on public health: A narrative review. *Cureus*, 15(1).
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- Chuhan Wu, Fangzhao Wu, Junxin Liu, Sixing Wu, Yongfeng Huang, and Xing Xie. 2018. Detecting tweets mentioning drug name and adverse drug reaction with hierarchical tweet representation and multi-head self-attention. In *Proceedings of the 2018 EMNLP workshop SMM4H: the 3rd social media mining for health applications workshop & shared task*, pages 34–37.
- Andrew Yates and Nazli Goharian. 2013. Adrtrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In *European Conference on Information Retrieval*, pages 816–819. Springer.
- Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28.
- Tongxuan Zhang, Hongfei Lin, Bo Xu, Liang Yang, Jian Wang, and Xiaodong Duan. 2021. Adversarial neural network with sentiment-aware attention for detecting adverse drug reactions. *Journal of Biomedical Informatics*, 123:103896.