文章编号: 1007-5321(2023)04-0097-06 DOI: 10.13190/j.jbupt.2022-176

基于层次门控交互融合网络的谣言检测方法

苏 兴, 禹 可, 吴晓非
(北京邮电大学 人工智能学院, 北京 100876)

摘要: 针对现有谣言检测方法对多特征做处理时因特征间差异导致特征冲突的问题,提出了一种基于层次门控交互融合网络的谣言检测方法。首先,利用一阶门控对原贴和评论的语义特征和情感特征做特征增强,然后,利用二阶门控对增强特征做跨语义特征融合,以解决特征融合时由于不同特征之间的差异引入噪声的问题。在公开的 Weibo 数据集和自建的 Weibo22 数据集上,所提方法的检测正确率分别为 96.71% 和 97.36%。与检测性能最好的基线方法相比,检测正确率分别提高了 0.84% 和 1.31%,训练时间分别减少了 53% 和 46%。

关 键 词: 谣言检测; 门控网络; 特征融合
中图分类号: TP183 **文献标志码:** A

Gated Interactive Fusion Network for Rumor Detection

SU Xing, YU Ke, WU Xiaofei
(School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: To address the issue of feature conflicts caused by differences between features when the existing rumor detection methods deal with multiple features, a hierarchical gated interactive fusion network-based rumor detection method is proposed. First, the first-order gate unit is conducted to obtain the enhanced semantic and sentiment features of original posts and comments. Then, the second-order gate unit is used to perform cross-semantic feature fusion on the enhanced features to solve the problem of introducing noise due to differences between different features during feature fusion. On the public Weibo dataset and the self-built Weibo22 dataset, the detection accuracy of the proposed method is 96.71% and 97.36%, respectively. Compared with the baseline methods with the best detection performance, the detection accuracy of the proposed method is improved by 0.84% and 1.31%, respectively, and the training time is reduced by 53% and 46%, respectively.

Key words: rumor detection; gated network; feature fusion

移动互联网时代,社交媒体凭借其信息共享、实时性、交互性、内容多样化等特征渐渐成为人们日常生活不可或缺的一部分。由于信息传播低成本及监督制度不够健全,大量谣言在社交媒体上快速传播,对社会经济、民众生活造成了严重危害。因此,如何有效地检测社交媒体上的谣言已成为学术界和工业界广泛关注的关键问题。

早期的谣言检测研究中,大多数学者聚焦于手工从消息中提取关于消息的文本内容、用户资料、传播结构等统计特征,然后利用机器学习算法进行分

收稿日期: 2022-07-13
基金项目: 国家自然科学基金项目(61601046)
作者简介: 苏 兴(1997—),男,硕士生。
通信作者: 禹 可(1977—),女,副教授,博士生导师,邮箱: yuke@bupt.edu.cn。

类检测。Castillo 等[1] 通过提取用户的平均发微博数、平均粉丝数、平均好友数、平均注册时间等信息作为用户特征，利用决策树进行分类检测。Ma 等[2] 将事件的传播模拟成时序序列，然后利用特征工程提取事件传播的时序特征来进行谣言检测。但这些利用手工提取特征的方法十分依赖特征工程，且这类方法大多聚焦于提取消息的统计特征而无法提取消息的语义特征，因此这也大大限制了这类方法的检测性能。

针对手工提取特征的缺陷，为了提取更高层的语义特征，学者提出了基于深度神经网络的方法。Ma 等[2] 利用循环神经网络从时序内容特征中提取隐藏表示对谣言进行检测。在此基础上，Ma 等[3-4] 又提出了一种树结构的递归神经网络，从谣言数据中提取隐藏表征，通过对不同类型的传播树结构的相似性分析进行谣言检测。Ruchansky 等[5] 提出了一种句子嵌入的方法，通过对消息内容和评论做嵌入处理，然后计算消息内容和所有评论的余弦相似度，通过设置相应的阈值过滤与内容不相关的评论，最后通过加权的方式得到最终的特征表示。Zhang 等[6] 利用双向长短时记忆神经网络和贝叶斯网络分别提取新闻和评论的语义特征，然后通过多层感知机将新闻特征和评论特征做融合输入分类器进行分类。

最近，利用图神经网络做谣言检测的研究吸引了越来越多学者的注意，利用图神经网络可以构建基于消息原贴与其评论、转发的交互关系。Bian 等[7] 将原贴及其回复输入到双向图卷积神经网络(Bi-GCN，bi-directional graph convolutional networks) 来获取谣言传播和散布的高级表征。Lu 等[8] 通过卷积图神经网络来提取用户特征，进而实现谣言检测目的。

总体来说，当前谣言检测领域主流的方法都是利用各种神经网络模型来提取不同的特征，然后利用特征融合策略来学习特征的增强表示[9-10]，最后将表示向量输入分类器实现谣言分类。然而，当前社交媒体平台上的很多消息都存在文本描述偏短、文本描述与实际内容不符、传播过程中存在水军空转发、发帖人身份造假等现象[9]，这对以传播和交互为特征的检测方法带来了挑战。同时，现有的特征拼接、特征相加或基于简单神经网络进行特征空间映射等特征融合方法很难准确地提取深层特征，而且这些方法提取的特征都包含很大的噪声。

针对以上问题，笔者提出了一种基于层次门控交互融合网络(GIFN，gated interactive fusion networks) 的谣言检测方法。首先，使用语义-情感融合门控(SSFG，semantic-sentiment fusion gate) 来增强原帖和评论的内容特征表示；然后，使用跨语义融合门控(CSFG，cross semantic fusion gate) 在原帖与评论之间构建特征关联；最后，在公开数据集 Weibo 数据集和自建的 Weibo22 数据集上的实验结果表明，与目前较先进的检测方法相比，GIFN 方法拥有更好的检测性能以及更短的训练时间。

1 GIFN

GIFN 的结构如图 1 所示，其主要包含特征编码、SSFG、CSFG 和谣言分类器 4 部分。

1.1 问题定义

定义给定谣言数据集为一系列的事件集合 $S=\{S_1,S_2,\cdots,S_n\}$，其中 S_i 为第 i 个事件，$i\in\{1,2,\cdots,n\}$，n 为数据集所有的事件数量。任意 $S_i\in S$ 都包含一条微博原贴和一系列相关评论，即 $S_i=\{s_i,c_{i1},c_{i2},\cdots,c_{im}\}$，其中 s_i 为原贴，c_{ij} 为关于 s_i 的第 j 条评论。每个事件 S_i 都附带一个标签 $y_i\in\{0,1\}$，其中 $y_i=1$ 表示该事件为谣言，$y_i=0$ 表示该事件为非谣言。研究目标为训练一个网络 f 判断给定事件 S_i 是否为谣言，可将其看成是一个二分类问题。

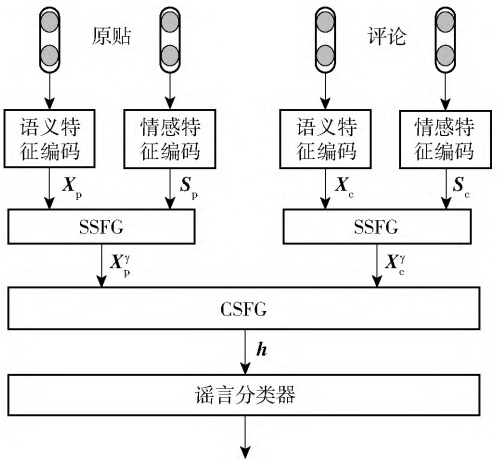

图 1 GIFN 结构

1.2 特征编码

在特征编码阶段，主要是提取原贴和评论的语义特征和情感特征。在提取语义特征时，使用基于转换器的双向编码器(BERT，bidirectional encoder representations from Transformers) 作为特征提取器，在提取情感特征时，基于情感词典设计了一种新颖

的特征提取方法。

BERT 在对事件的微博原帖提取语义特征时，将事件原贴描述直接输入 BERT，然后对最后一层所有词例(即 token)的隐藏状态做平均池化作为原贴的语义特征 $\boldsymbol{X}_{\mathrm{p}}$，$\boldsymbol{X}_{\mathrm{p}} \in \mathbb{R}^{d}$，$d$ 为特征维度。在对评论提取语义特征时，由于 BERT 限制最大输入长度为 512 个词例，因此将评论按时间顺序排序并拼接所有评论，当拼接的评论序列长度超过 512 时截取之前所有评论作为 BERT 输入，同样提取最后一层所有词例的隐藏状态做平均池化作为评论的语义特征 $\boldsymbol{X}_{\mathrm{c}}$。

在对微博事件的原贴和评论的情感特征编码时，先利用情感词典来计算原贴和每条评论的情感得分，在实际研究中选择使用波森情感词典来计算情感得分。在波森情感词典中，每个词有一个情感极性得分。得分大于 0，表示正向情感倾向，得分越高，倾向越强；得分小于 0，表示负向情感倾向，得分越低，倾向越强。利用波森情感词典计算原贴的情感得分 s_{p} 和所有评论的情感得分均值 s_{c} 后，借助神经网络分别提取原帖和评论的情感特征 $\boldsymbol{S}_{\mathrm{p}}$ 和 $\boldsymbol{S}_{\mathrm{c}}$。原帖的情感特征的提取过程为

$$\boldsymbol{S}_{\mathrm{p}} = \boldsymbol{W}^{\mathrm{s}} \boldsymbol{x}^{\mathrm{T}} \tag{1}$$

$$\boldsymbol{x} = \mathrm{sigmoid}\ s_{\mathrm{p}} \parallel (1 - \mathrm{sigmoid}\ s_{\mathrm{p}}) \tag{2}$$

其中：$\boldsymbol{W}^{\mathrm{s}} \in \mathbb{R}^{d \times 2}$ 为可学习参数矩阵，$\cdot \parallel \cdot$ 为将两个向量拼接。$\boldsymbol{S}_{\mathrm{c}}$ 也可通过相同的处理获得。

1.3　SSFG

为了提取情感特征中包含的与语义特征相关的信息，并将这些辅助信息融入进语义特征中，使用 SSFG 对语义信息进行特征增强，SSFG 结构如图 2 所示。

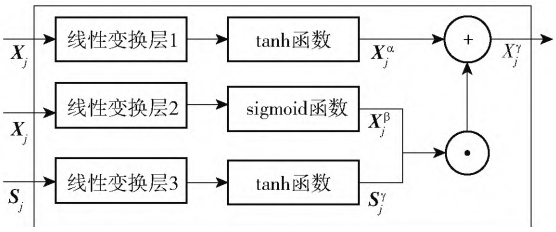

图 2　SSFG 结构

SSFG 的具体操作如下。首先，将语义特征 $\boldsymbol{X}_j (j = \mathrm{p}, \mathrm{c})$ 输入线性变换层 1 和 2 进行特征投影；将情感特征 $\boldsymbol{S}_j$ 输入线性变换层 3 进行特征投影；然后，使用非线性函数来增强投影后特征的非线性特征，以此得到隐藏特征 $\boldsymbol{X}_j^{\alpha}$，$\boldsymbol{X}_j^{\beta}$ 和 $\boldsymbol{S}_j^{\gamma}$ 分别为

$$\boldsymbol{X}_j^{\alpha} = \tanh\left(\boldsymbol{W}^{\alpha} \boldsymbol{X}_j + \boldsymbol{b}^{\alpha}\right), \quad j = \mathrm{p}, \mathrm{c} \tag{3}$$

$$\boldsymbol{X}_j^{\beta} = \mathrm{sigmoid}\left(\boldsymbol{W}^{\beta} \boldsymbol{X}_j + \boldsymbol{b}^{\beta}\right), \quad j = \mathrm{p}, \mathrm{c} \tag{4}$$

$$\boldsymbol{S}_j^{\gamma} = \tanh\left(\boldsymbol{W}^{\gamma} \boldsymbol{S}_j + \boldsymbol{b}^{\gamma}\right), \quad j = \mathrm{p}, \mathrm{c} \tag{5}$$

其中：$\boldsymbol{W}^{\alpha}, \boldsymbol{W}^{\beta}, \boldsymbol{W}^{\gamma} \in \mathbb{R}^{d \times d}$ 为可学习参数矩阵；$\boldsymbol{b}^{\alpha}, \boldsymbol{b}^{\beta}, \boldsymbol{b}^{\gamma} \in \mathbb{R}^{d}$ 为可学习偏置向量。

将 $\boldsymbol{X}_j^{\beta}$ 和 $\boldsymbol{S}_j^{\gamma}$ 进行逐元素相乘提取两个向量在每一维度上相似的特征，并将上述结果与 $\boldsymbol{X}_j^{\alpha}$ 相加得到融合语义和情感的增强语义特征为

$$\boldsymbol{X}_j^{\gamma} = \boldsymbol{X}_j^{\alpha} + \left(\boldsymbol{X}_j^{\beta} \odot \boldsymbol{S}_j^{\gamma}\right), \quad j = \mathrm{p}, \mathrm{c} \tag{6}$$

其中 $\odot$ 为向量点对点相乘。使用该门控融合原贴的语义特征和情感特征得到增强型原贴语义特征 $\boldsymbol{X}_{\mathrm{p}}^{\gamma}$，融合评论的语义特征和情感特征得到增强型评论语义特征 $\boldsymbol{X}_{\mathrm{c}}^{\gamma}$。

1.4　CSFG

在与事件原贴相关的评论中经常会有评论者对事件的真假给出评价甚至直接指出事件是谣言事件或非谣言事件，因此理解评论的语义信息可以更好地判定事件的真伪。为了将评论语义特征融入进原贴语义特征，让原贴语义具备更充分而详细的信息，使用 CSFG 在增强型原贴语义特征和增强型评论语义特征之间构建关联，将两个语义特征做深度特征融合。通过 CSFG 得到原贴评论融合特征为

$$\boldsymbol{h} = \left(\boldsymbol{Z} \odot \boldsymbol{h}^{\mathrm{p}}\right) + \left(1 - \boldsymbol{Z}\right) \odot \boldsymbol{h}^{\mathrm{c}} \tag{7}$$

其中：$\boldsymbol{Z}, \boldsymbol{h}^{\mathrm{p}}, \boldsymbol{h}^{\mathrm{c}}$ 均为隐藏特征，分别表示为

$$\boldsymbol{Z} = \mathrm{sigmoid}\left(\boldsymbol{W}^{z}\left[\boldsymbol{X}_{\mathrm{p}}^{\gamma} \parallel \boldsymbol{X}_{\mathrm{c}}^{\gamma}\right] + \boldsymbol{b}^{z}\right) \tag{8}$$

$$\boldsymbol{h}^{\mathrm{p}} = \tanh\left(\boldsymbol{W}^{\mathrm{p}} \boldsymbol{X}_{\mathrm{p}}^{\gamma} + \boldsymbol{b}^{\mathrm{p}}\right) \tag{9}$$

$$\boldsymbol{h}^{\mathrm{c}} = \tanh\left(\boldsymbol{W}^{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^{\gamma} + \boldsymbol{b}^{\mathrm{c}}\right) \tag{10}$$

其中：$\boldsymbol{W}^{z} \in \mathbb{R}^{d \times 2d}$，$\boldsymbol{W}^{\mathrm{p}}, \boldsymbol{W}^{\mathrm{c}} \in \mathbb{R}^{d \times d}$ 为可学习参数矩阵；$\boldsymbol{b}^{z}, \boldsymbol{b}^{\mathrm{p}}, \boldsymbol{b}^{\mathrm{c}} \in \mathbb{R}^{d}$ 为可学习偏置向量。

1.5　谣言分类器

将 $\boldsymbol{h}$ 输入前馈网络，该网络由一层具有 ReLU 激活函数的全连接层和另一层全连接层组成，前馈网络的输出结果经由 softmax 函数处理得到事件 i 的概率预测分布为

$$P_i = \mathrm{softmax}\ \boldsymbol{g} \tag{11}$$

$$\boldsymbol{g} = \boldsymbol{W}_2 \max\left(\boldsymbol{W}_1 \boldsymbol{h} + \boldsymbol{b}_1, \boldsymbol{0}\right) + \boldsymbol{b}_2 \tag{12}$$

其中：$\boldsymbol{g}$ 为前馈网络的输出；$\boldsymbol{W}_1 \in \mathbb{R}^{d \times d}$，$\boldsymbol{W}_2 \in \mathbb{R}^{2 \times d}$ 为可学习参数矩阵；$\boldsymbol{b}_1 \in \mathbb{R}^{d}$，$\boldsymbol{b}_2 \in \mathbb{R}^{2}$ 为可学习偏置向量。

P_i 与真实标签 y_i 的交叉熵损失为

$$L = -\sum y_i \mathrm{lb}\left(P_i\right) \tag{13}$$

通过对 L 最小化实现目标优化。

2 实验结果与分析

实验的硬件环境为: Xeon(R) Gold 5320 CPU(内存32G) + RTX 2080 GPU(显存16G);软件环境为: python3.8 + Pytorch1.7.0 + Tensorflow1.14 + Ubuntu 18.04.4 LTS。

2.1 数据集

Weibo数据集是由Ma等[4]收集的2016年之前在新浪微博上发布的4 664条中文帖子,是目前使用最多的中文谣言数据集。

考虑到社交媒体在近年来发展十分迅速,Weibo数据集中的数据特征与当前社交媒体上的信息存在较大差异,为了检测GIFN在当前社交媒体中的检测性能,基于新浪微博构建了Weibo22数据集。与Weibo数据集相比,Weibo22数据集在事件数量上有所增加,同时事件涉及社会、科技、党政、军事、娱乐等16个主题;在事件来源上,除了选取部分可信度很高的官媒微博之外,还选取了很多个人用户发表的微博,以增强数据集语义的多样性并更符合实际场景。2个数据集的统计信息如表1所示。

表1 2个数据集的统计信息

统计信息	Weibo数据集	Weibo22数据集
事件数量	4 664	5 846
所有帖子数量	3 805 656	4 507 656
谣言事件数量	2 351	2 923
非谣言事件数量	2 313	2 923

2.2 基准方法与评价指标

为验证GIFN的优越性,选取了7个基线方法作为对比,这些基线模型按方法类型可分为传统机器学习方法和深度学习方法,其中机器学习方法包括决策树分类器(DTC,decision tree classifier)[1]和支持向量机-径向基函数(SVM-RBF,support vector machine-radial basis function);深度学习方法包括树结构循环神经网络(TSRNN,tree-structured recursive neural networks)[3]、双层门控循环单元(GRU-2,two layers gated recurrent unit)[4]、Bi-GCN[7]、基于层级细粒度注意力掩蔽的堆叠网络(STANKER,stacking network based on level-grained attention-masked)[11]和后层次注意力模型(PLAN,post-level attention model)[12]。

实验将谣言检测任务视为二分类问题,采用准确率、精度、召回率、F1值和训练时间作为谣言检测方法的性能评价指标。

为了确保比较的公平性,实验时将数据集随机打乱5次,每次利用分层抽样按4∶1的比例划分训练集和测试集,同时使用与基线方法原文献中相同的参数设置进行实验,取5次实验的均值作为最终结果。实验使用scikit-learn框架实现DTC和SVM-RBF,使用Pytorch框架实现GRU-2、PLAN、TSRNN、Bi-GCN和GIFN,使用TensorFlow框架实现STANKER。

2.3 与其他方法的对比

所提方法与对比方法在2个数据集上的性能对比如表2和表3所示。由表2和表3可知,机器学习检测方法的性能表现远低于深度学习方法,这也证明神经网络可以提取到更高维的谣言特征,以及利用深度学习方法实现高准确率谣言检测的必要性。在Weibo数据集和Weibo22数据集中,GIFN在所有性能指标上均优于所有的基准方法,这不仅证明了GIFN的有效性,也表明了GIFN针对当前社交媒体上的谣言也具备十分优异的检测性能。

表2 各方法在Weibo数据集上的性能对比

模型	准确率	精度	召回率	F1值
DTC	0.821 3	0.832 1	0.795 4	0.811 3
SVM-RBF	0.854 2	0.846 2	0.860 5	0.857 1
TSRNN	0.898 1	0.901 2	0.891 5	0.896 2
GRU-2	0.882 3	0.886 1	0.875 3	0.880 2
Bi-GCN	0.953 4	0.953 7	0.951 4	0.952 5
STANKER	0.958 7	0.961 2	0.959 6	0.960 8
PLAN	0.913 3	0.911 3	0.908 4	0.909 1
GIFN	0.967 1	0.968 4	0.964 1	0.966 2

表3 各方法在Weibo22数据集上的性能对比

模型	准确率	精度	召回率	F1值
DTC	0.830 2	0.822 1	0.833 5	0.827 8
SVM-RBF	0.851 6	0.862 1	0.859 3	0.860 7
TSRNN	0.902 4	0.911 2	0.911 5	0.911 3
GRU-2	0.887 6	0.883 2	0.894 6	0.883 9
Bi-GCN	0.957 2	0.957 9	0.958 1	0.958 0
STANKER	0.960 5	0.960 2	0.960 8	0.960 5
PLAN	0.920 8	0.896 8	0.945 7	0.920 6
GIFN	0.973 6	0.969 4	0.978 9	0.974 1

所提方法和对比方法在 2 个数据集上的训练时间对比如表 4 所示。由表 4 可知,DTC 和 SVM-RBF 由于手工构造特征较为简单,因此无需太多训练时间; TSRNN 和 Bi-GCN 由于需要基于评论构建传播树,因此所需训练时间较长。

表 4　模型训练所需时间　　h

模型	Weibo 数据集	Weibo22 数据集
DTC	0.03	0.05
SVM-RBF	0.03	0.05
TSRNN	1.34	1.49
GRU-2	0.08	0.09
Bi-GCN	2.73	2.93
STANKER	1.22	1.28
PLAN	0.35	0.42
GIFN	0.57	0.69

由表 2 ~ 表 4 综合来看,STANKER 是检测性能最优的基线方法,与 STANKER 相比,GIFN 不仅在各性能上表现更优,在训练时间上也具备一定的优势。在 2 个数据集上,GIFN 的训练时间比 STANKER 分别减少了 53.28% 和 46.09%。

2.4　消融实验

为了验证 GIFN 中各模块对最终检测结果的影响,通过消融实验来进行实验验证。在完整 GIFN 的基础上,设置以下 4 组消融方法: 消融 1: 去除 SSFG,将语义特征和情感特征直接做特征拼接(GIFN-SSFG); 消融 2: 去除 CSFG,将经过 SSFG 处理的原贴特征和评论特征直接做特征拼接(GIFN-CSFG); 消融 3: 去除 SSFG,将语义特征和情感特征利用注意力机制做特征融合(GIFN-SSFG + Att),具体操作上,将语义特征当成注意力机制中的 Q 向量,计算其对自身特征和情感特征的权重,权重乘上对应的 V 向量,得到最终的增强型语义特征和增强型评论特征; 消融 4: 去除 CSFG,将经过 SSFG 处理的原贴特征和评论特征利用注意力机制做特征融合(GIFN-CSFG + Att),具体操作上将增强型原贴语义特征当成注意力机制中的 Q 向量,计算出其对自身和增强型评论特征的权重,将权重乘以对应的 V 向量,得到最终的原贴评论融合特征。在 2 个数据集上将 GIFN 与 4 组消融方法进行性能对比,结果如图 3 所示。

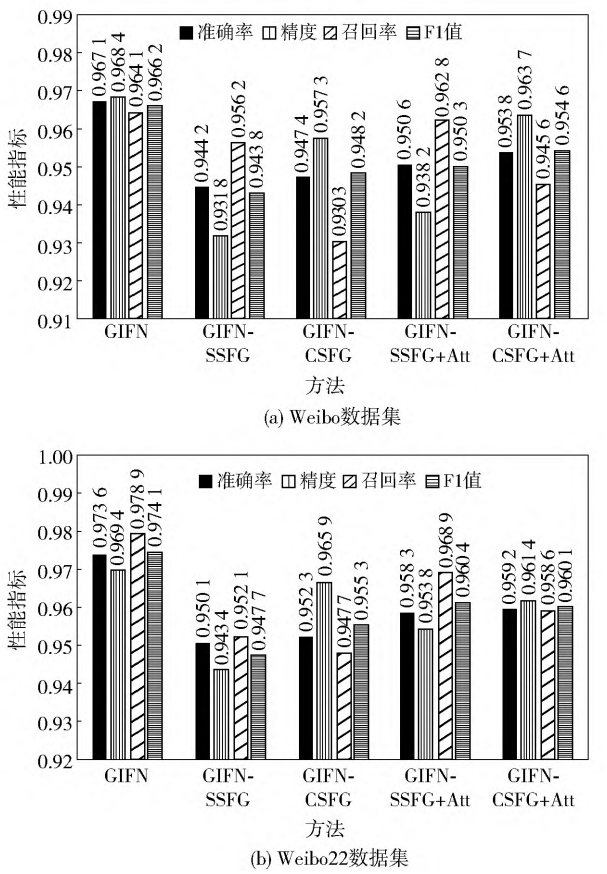

图 3　GIFN 与 4 组消融方法在 2 个数据集上的性能对比

由图 3 可知,与 GIFN 相比,GIFN-SSFG 在 2 个数据集上的准确率上分别降低了 2.37% 和 2.41%,这表明 SSFG 可以从情感特征中捕获额外特征对语义特征做补充,增强语义特征的特征性。与 GIFN 相比,GIFN-CSFG 在 2 个数据集上的准确率分别下降了 2.04% 和 2.19%,这表明 CSFG 在对来自 2 个不同内容的特征做特征融合的有效性。

与 GIFN 相比,GIFN-SSFG + Att 在 2 个数据集上的准确率分别降低了 1.71% 和 1.57%,与 GIFN 相比,GIFN-CSFG + Att 在 2 个数据集上的准确率分别降低了 1.38% 和 1.48%。由这 2 个实验结果可知,相比于直接对特征做特征拼接的处理方法,使用注意力机制做特征融合确实可以得到更有效的特征,但是由于注意力机制会将注意力更多聚焦于自身特征,缺乏将所有特征进行全局交互,因此当使用注意力机制做特征融合时容易引入额外的噪声或无关特征。GIFN 通过门控单元对不同特征进行全局交互,从而有选择性地从不同特征中提取相似特征并将这些特征融合成为一个增强型特征。这样的处理方式一方面解决了不同特征间由于特征差异而导致的特征冲突问题,另一方面也保证了能公平地将

注意力聚焦于每一个特征上，从而保证了融合后特征的有效性。

3 结束语

提出了一种基于GIFN的谣言检测方法，该方法聚焦于原贴和评论的语义特征和情感特征，先利用SSFG从情感特征中提取额外特征对语义特征做补充，增强语义特征的特征表示；然后，利用CSFG从增强型评论语义特征中提取辅助信息，将其与增强型原贴语义特征做融合，得到最终的特征表示。所提方法在公开数据集上和基于当前社交媒体构建的新数据集上均取得了优异的实验结果，与其他方法相比，在所有评测指标上均有明显的提升。

综上所述，寻求高效的特征融合方法对于提升方法的检测性能是切实有效的途径，高效的特征融合方法并非局限于谣言检测场景，也可应用至其他应用领域，促进其他领域的发展。

参考文献：

[1] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C] // Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 675-684.

[2] MA J, GAO W, JOY S, et al. An attention-based rumor detection model with tree-structured recursive neural networks [J]. ACM Transactions on Intelligent Systems and Technology, 2020, 11(4): 1-28.

[3] MA J, GAO W, WONG K F. Rumor detection on Twitter with tree-structured recursive neural networks [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2018: 1980-1989.

[4] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C] // Proceedings of the 25th International Joint Conference on Artificial Intelligence. Burlington: Morgan Kaufmann, 2016: 3818-3824.

[5] RUCHANSKY N, SEO S, LIU Y. CSI: a hybrid deep model for fake news detection [C] // Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: ACM Press, 2017: 797-806.

[6] ZHANG Q, LIPANI A, LIANG S, et al. Reply-aided detection of misinformation via Bayesian deep learning [C] // The World Wide Web Conference. New York: ACM Press, 2019: 2333-2343.

[7] BIAN T, XIAO X, XU T Y, et al. Rumor detection on social media with bi-directional graph convolutional networks [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 549-556.

[8] LU Y J, LI C T. GCAN: graph-aware co-attention networks for explainable fake news detection on social media [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 505-514.

[9] WANG W Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2017: 422-426.

[10] JIN Z W, CAO J, GUO H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs [C] // Proceedings of the 25th ACM International Conference on Multimedia. New York: ACM Press, 2017: 795-816.

[11] RAO D N, MIAO X, JIANG Z H, et al. STANKER: stacking network based on level-grained attention-masked BERT for rumor detection on social media [C] // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2021: 3347-3363.

[12] KHOO L M S, CHIEU H L, QIAN Z, et al. Interpretable rumor detection in microblogs by attending to user interactions [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 8783-8790.