



计算机科学

Computer Science

ISSN 1002-137X, CN 50-1075/TP

《计算机科学》网络首发论文

题目：基于语义扩充和 HDGCN 的虚假新闻联合检测技术
作者：张明道，周欣，吴晓红，卿粼波，何小海
网络首发日期：2023-12-01
引用格式：张明道，周欣，吴晓红，卿粼波，何小海. 基于语义扩充和 HDGCN 的虚假新闻联合检测技术[J/OL]. 计算机科学.
<https://link.cnki.net/urlid/50.1075.TP.20231201.0855.004>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于语义扩充和 HDGCN 的虚假新闻联合检测技术

张明道 周欣 吴晓红 卿粼波 何小海

四川大学电子信息学院 成都 610065

(1950187531@qq.com)

摘要 虚假新闻检测的方法有很多种，单一的方法通常只关注新闻内容、社交上下文或外部事实等信息；而联合检测方法则通过整合多种模式信息以达到检测目的。Pref-FEND 即为一种整合新闻内容与外部事实的联合检测方法，它从新闻内容和外部事实中提取 3 种词语表示，利用动态图卷积网络获得词节点之间的关系。但在如何让两种模式更加专注于自己的偏好部分方面仍存在不足。因此，对 Pref-FEND 模型进行了改进，利用语义挖掘扩充新闻中的风格词，利用实体链接扩充新闻中的实体词，共得到 5 种词语作为图网络的节点表示，从而更有效的建模图神经网络的节点表征；同时，引入深度异构图卷积网络(HDGCN)进行偏好学习，它的深度策略和多层注意力机制可以让两种模型更加专注于自身需要的偏好感知并减少冗余信息。实验结果表明，在公开数据集 Weibo 和 Twitter 上，与当前主流的基于内容的单一模型 LDAVAE 相比，改进后的框架 F1 值分别高出 2.8% 和 1.9%；与基于事实的单一模型 GET 相比，F1 值分别高出 2.1% 和 1.8%；同时，在 LDAVAE+GET 联合检测情况下，比 Pref-FEND 的 F1 值分别高出 1.1% 和 1.3%。实验结果验证了所改进模型的有效性。

关键词： 虚假新闻；图卷积网络；实体抽取；注意力机制；自然语言处理

中图法分类号 TP183

Unified Fake News Detection Based on Semantic Expansion and HDGCN

ZHANG Ming-dao, ZHOU Xin, WU Xiao-hong, QING Lin-bo and HE Xiao-hai

College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

Abstract There are many methods for detecting fake news. The single method typically focuses only on information such as news content, social context, or external facts. On the other hand, joint detection methods integrate multiple modalities of information to achieve the detection goal. Pref-FEND is an example of a joint detection method that integrates news content and external facts. It extracts three types of word representations from news content and external facts, and uses dynamic graph convolutional networks to capture relationships between word nodes. However, there are still shortcomings in how to make each modality more focused on its preferred aspect. Therefore, the Pref-FEND model has been improved by using semantic mining to expand style words in news and entity linking to expand entity words in news. This results in five types of word as node representations in the graph neural network, enabling a more effective modeling of the node representation of the graph neural network. Additionally, a Deep Heterogeneous Graph Convolutional Network (HDGCN) is introduced for preference learning. Its deep strategy and multi-layer attention mechanism allow both models to focus more on their own preferred perception and reduce redundant information. The experimental results demonstrate the effectiveness of the improved framework. On the public datasets Weibo and Twitter, compared to the current state-of-the-art content-based single model LDAVAE, the improved framework achieves an F1 score improvement of 2.8% and 1.9% respectively. Compared to the fact-based single model GET, the F1 score improvement is 2.1% and 1.8% respectively. In the case of joint detection using LDAVAE+GET, the F1 scores are respectively higher than Pref-FEND by 1.1% and 1.3%. The experimental results validate the effectiveness of the improved model.

Keywords Fake News, Graph Convolutional Networks, Entity Extraction, Attention Mechanism, Natural Language Processing

1 引言

近年来,随着社交媒体的不断发展,虚假新闻的传播也越来越迅速和广泛。《科学》杂志上的研究表明,在 Twitter 上,假新闻被转发的次数要比真实消息高出 70%,而取得同样的转发次数,真实消息需要的时间是假新闻的 20 倍。

假新闻总是通过对正确的说法进行一些修改,具有很强的欺骗性和不可区分性。假新闻会带来很多危害,比如在政治^[1]和公共卫生^[2]等不同领域的广泛传播,对网络安全和人类社会构成了巨大威胁。因此,对虚假新闻检测的研究需求越来越大。

虚假新闻检测技术的研究方法有很多。基于内容的检测方法通常关注新闻的行文风格以及情感特征,包括新闻作者及评论的态度,如赞赏、愤怒等情绪。基于用户可信度的检测方法利用新闻作者的可信度来辅助检测,可信度低的用户发表的新闻更有可能是虚假新闻。Dou^[3]等人利用用户的发文历史来识别用户可信度,将其作为内因。同时该工作将新闻的传播情况作为外因,利用内因和外因共同进行虚假新闻检测。Zhu^[4]等人的研究表明,用户的可信度可以从他们的行为和与其他用户的联系中反映出来,通过分析参与者的属性及社交网络可以强化检测效果,但同时也可能会被社交机器人美化后的身份或恶意操纵者给出的有益的虚假评论所欺骗。基于传播信息的检测方法将新闻的传播过程建模为树形结构,源新闻作为树的根节点,转发以及评论信息作为分支节点和叶子节点。Dou 等人收集用户历史帖子以对用户内生偏好进行隐式建模,并利用社交媒体上的新闻传播作为用户的社交背景。该研究表明,用户的历史、社交活动(例如帖子)提供了用户对相关新闻偏好的丰富信息,并且具有辅助假新闻检测的巨大作用,当用户对某一新闻具有明确偏好时,其更有可能传播一条假新闻;Jiang^[5]等人提出了广度模体度与深度模体度的概念来量化传播网络的结构,通过构建的二维模体度量化指标,分析和揭示了微博和 Twitter 网络中虚假信息的结构特性与传播机制。基于外部事实的检测方法聚焦于客观事实中丰富的语义信息,如将新闻与事实进行对比,或者融合新闻与外部知识来辅助检测。本文重点关注基于内容和基于外部事实的虚假新闻检测方法。

基于新闻内容的方法。这种方法只考虑新闻本身所涉及文本模式以及行文风格。基于新闻内容的相关研究有很多,比如 Castillo^[6]等人重点关注新闻中基于文本的特征,如是否有感叹号或问号等。Volkova^[7]等人将主观性、心理语言学和道德基础特征注入深度神经网络(CNN 和 RNN)进行训练。Przybyla^[8]等人的工作专注于写作风格和行文模式。Nan 等人通过区分多个主题类别的模式^[6]来辅助检测。Popat^[10]等人根据文本特征和文章立场将新闻分类。此外, Lee 等人通过新闻反馈中的情绪立场来验证真实性,如转发、点赞和评论^[11]。Potthast 等人的工作聚焦于情绪模式挖掘,他们认为假新闻中可能存在明显的情绪偏见^[12]。

基于外部事实的方法。这种方法通过检验证据中的语义相似性(冲突),以检查新闻的真实性。证据通常是从知识图谱^[13]或事实核查网站^[14]中提供的未经核实的声明作为查询来检索得到的。基于外部事实的相关研究有很多,DeClarE^[15]是第一个在假新闻检测中利用证据的文章,它使用 BiLSTM 来嵌入证据的语义,并通过平均池来获得声明的句子级表示。另一些方法利用顺序模型获得语义嵌入,然后利用注意力机制对不同的证据进行融合;如 HAN^[16]等人通过计算新闻和证据之间的句子级连贯性和隐含得分来作为评判标准;EHIAN^[17]采用自注意力机制来获得单词和证据之间的互动得分。Nguyen^[18]等人将单词和句子的交互分层集成到证据中进行验证。

然而,基于内容的方法通常具有较差的泛化能力和可解释性;基于外部事实的方法需要有效的发现并整合给定证据中的有用信息,才能进行事实验证。因此,本文考虑融合两种方法来提升虚假新闻检测的准确度,其中的关键就是如何让模型更有效的学习对自己有利的偏好部分,并且尽可能消除其它冗余信息。具体来说,通过让基于内容的模型更关注新闻文本中的风格词(如愤怒、震惊等),让基于事实的模型更关注实体词(如确切的时间、人名、地名等)^[19],实现对两种偏好特征的有效提取,从而提升融合模型的检测效果。

为此,本文对现有工作实现了如下改进。一方面,风格词和实体词分别代表了基于内容的模型和基于事实的模型对应的特

征, 两者的语义信息可以帮助模型更好的学习对应偏好, 因此本文对新闻中的实体词和风格词进行了扩充。具体来说, 对于风格词, 利用其潜在的语义信息, 通过语义挖掘获得相关上下义词、同义词、义原等, 进而扩充对应的语义知识词; 对于实体词, 利用实体链接将文本中的实体词和外部知识源中的目标实体对应起来, 进而扩充新闻实体中的信息量, 为新闻判别做信息补充。通过语义扩充, 可以更为有效的建模图神经网络的节点表征, 从而提升模型的学习效果。另一方面, 本文引入深度异构图卷积网络 (Heterogenous Deep Graph Convolutional Network, HDGCN)^[20]。HDGCN 能使不同类型节点 (词类型) 的信息动态、充分地交互, 并且它的节点相关性可以动态调整, 从而帮助模型有效进行偏好学习。相比传统的图卷积网络, HDGCN 在不同特征的提取和融合方面效果提升明显。一方面, HDGCN 利用了层次注意力机制, 该机制可以捕捉不同关系、不同节点类型和不同层对节点嵌入的对应贡献; 另一方面, HDGCN 引入了相邻域采样策略来控制相邻聚合节点的数量, 可以有效缓解训练过程中的“邻域爆炸”和过度平滑问题。具体来说, 通过 HDGCN 的多层训练并结合词特征的关系矩阵, 分别提取文本中的新闻内容和外部事实对应的偏好, 同时动态调整不同类型节点的相关性, 实现信息间的动态、充分交互。综上, 利用语义挖掘扩充新闻中的风格词, 利用实体链接扩充新闻中的实体词, 共同作为图网络的节点表示, 并将其送入 HDGCN 中进行训练。

2 Pref-FEND 模型的改进

Pref-FEND (Preference-aware Fake News Detection Framework, Pref-FEND) 为 Qiang 等人提出的一种联合检测模型, 它构建了一个基于偏好感知的假新闻检测框架, 通过整合新闻内容和外部事实进行联合检测。Pref-FEND 从待测新闻中提取模式信息触发词、事实信息触发词和普通词, 共 3 种词语表示作为输入特征; 同时引入动态图卷积网络 (Heterogeneous graph attention net-works, HGCN) 来生成相应的偏好映射, 利用这些映射来指导内容和事实模型的联合学习, 以进行最终的预测。但 Pref-FEND 在如何让不同模型更加专注于自己的偏好部分方面

仍有提升空间, 通过改进可以更有有效的建模偏好学习。

本章将介绍对联合检测模型 Pref-FEND 的改进, 主要的改进点为对词语表征进行了扩充, 以及引入 HDGCN 来对偏好进行更有效的学习。本文将改进后的模型命名为 Pfnd (Preference Learning Fake News Detection, Pfnd)。Pfnd 模型的结构如图 1 所示。

Pfnd 模型的整体构建流程如下:

首先, 识别并提取新闻文本中的风格词和实体词。同时, 为了获得更多潜在的语义信息来丰富后续的图节点特征, 进而更好的学习偏好感知, 分别挖掘了风格词对应的相关语义词以及实体词对应的外部链接词, 共同作为图节点的表征。其次, 为了让模型有效的学习不同的偏好感知, 引入深度异构图卷积网络, 一方面它包含多种注意力机制, 可以更好的对图节点进行语义分析; 另一方面, 它的深度策略可以有效避免图节点的过渡平滑等问题, 有利于模型的训练。最后, 融合内容和外部事实对应的特征

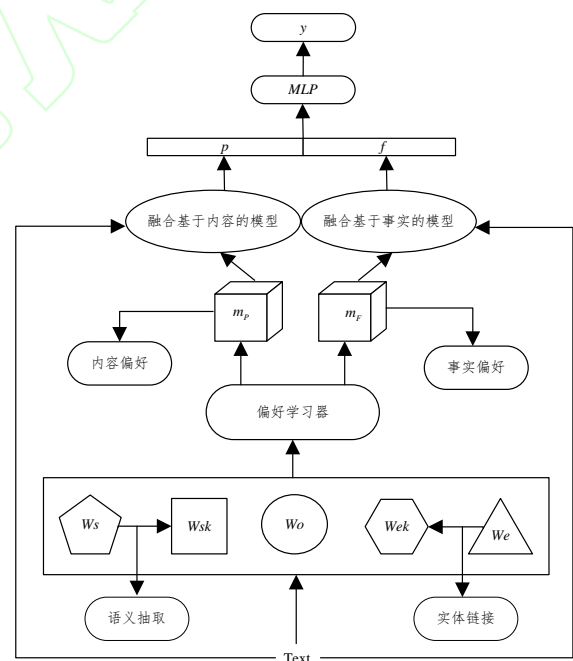


图 1 Pfnd 模型

Fig.1 Pfnd model

向量, 利用 MLP 感知机进行最终预测。

2.1 语义抽取和实体链接

Pref-FEND 模型从文本中提取风格词、实体词和普通词, 共 3 种词语作为输入向量。为了更好的表示文本内容和外部事实

对应的偏好，本文提出的 Pfnd 模型分别扩充了风格词和实体词，共 5 种不同类型的词语作为输入特征。

基于新闻文本内容的方法通常聚焦于新闻文章的行文风格，如语义风格、情感风格等。因此，对于给定的新闻文本，Pfnd 首先提取新闻中可以代表此新闻内容风格的词语。根据张^[21]等人的研究，可以利用情绪相关特征以及其他语言特征来表示新闻文本的行文风格，例如表示否定、愤怒、赞赏等相关的情绪词语。因此，根据预定义的包含多种不同情感类型的情感词典，利用 BERT 提取数据集中对应的词语作为情感标签，用 $W_s = \{s_1, s_2, \dots, s_n\}$ 表示新闻文本中的内容风格词语。

基于外部事实的方法通常关注可验证的客观事实。根据 James^[22]等人的研究，这些事实中包含一个或多个实体。为了识别并提取出新闻中包含的实体，使用 LAC^[23]和 TextSmart^[24]两种工具。其中，LAC 用于从中文数据集中提取实体，TextSmart 用于从英文数据集中提取实体。用 $W_e = \{e_1, e_2, \dots, e_n\}$ 来表示提取到的实体词语。

同时，为了更好融合新闻中文本的不同风格特征，考虑提取其他语义特征。根据相关研究，层次化的语言知识结构能够丰富文本中原词语的表示，通过一个特定词语，可以获得相关的语义知识，如上下文词、同义词、义原等，能够有效挖掘词语潜在的语义信息。因此，根据上文的风格词语 W_s 得到其对应的语义知识词。即利用 HowNet^[25]提取中文风格词语的相关语义知识词，利用 WordNet^[26]提取英文风格词语的相关语义知识词。用 $W_{sk} = \{w_{sk1}, w_{sk2}, \dots, w_{skn}\}$ 表示文本中风格词对应的相关语义知识词。

相关研究表明，通过实体链接可以将新闻文本中已识别的实体词正确地指向外部知识源中的目标实体，有助于扩充新闻实体中的信息量，为新闻判别做信息补充。因此，通过上文的实体词语 W_e 得到对应的语义知识词。对于中文数据集，利用外部知识库 CN_DBpedia 提供的实体链接 API，将 W_e 中的中文实体词和知识库中存在的实体词对应起来；对于英文数据集，利用实体链接工具 TAGME，将 W_e 中的英文实体词与 WiKi 知识库中的实体词对应。得到的实体词表示为 $W_{ek} = \{w_{ek1}, w_{ek2}, \dots, w_{ekn}\}$ 。

最后，文本中其他词表示为 $W_o = \{w_{o1}, w_{o2}, \dots, w_{on}\}$ 。

综上，分别得到新闻文本中五种词语表示，即风格词 W_s 和其对应的语义知识词 W_{sk} ，实体词 W_e 和其对应的实体链接词 W_{ek} ，以及其他词 W_o 。这五种词语在图 1 中分别用五种不同的形状图来表示。

2.2 偏好学习器的构建

不同类型的节点描述了新闻文本的不同方面，为了区分这些方面并得到对应的特征表示，同时为了使不同类型节点的信息能够动态、充分地交互，本文引入了一种深度异构图卷积网络 (HDGCN)，相比 Pref-FEND 中引入的 HGNCN^[27]，HDGCN 的深度策略和多层注意力机制能让模型更加聚焦于自身的偏好感知并减少冗余信息。

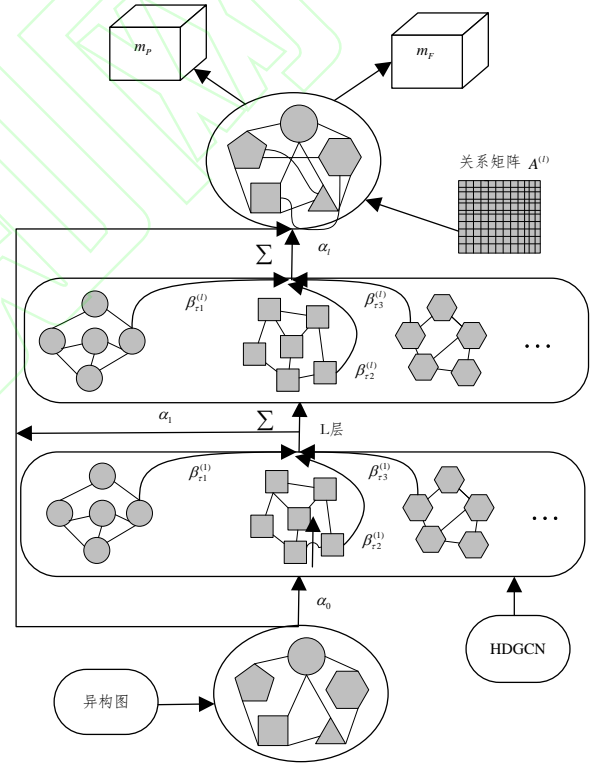


图 2 偏好学习器

Fig.2 Preference learner

相比传统的图卷积网络(Graph Convolutional Network, GCN)和异构图卷积网络(HGCN),HDGCN 在不同特征的提取和融合方面效果提升明显。一方面，为了在异构图中获得更多有用的信息，HDGCN 引入了层次注意力机制，该机制可以捕捉不同关系、不同节点类型和不同层对节点嵌入的对应贡献；另一方面，随着图中节点数量和对应关系的急剧增加，会导致“邻域爆炸”和过度平滑的问题。为此，HDGCN 引入相邻域采样策略来

控制相邻聚合节点的数量,同时,在每一层中重复采样,这不仅保留了图中的节点信息,而且缓解了“邻域爆炸”和过度平滑问题。

偏好学习器的构建流程如图2所示。

2.2.1 异构图的构建

对2.1节得到的五种词语,利用预训练语言模型BERT^[28]进行特征提取并进行初始化,得到对应的向量表示 $H^0 \in R^{n \times d}$ (d 表示每个词语特征向量的维度), $H^0 = [H_s, H_e, H_{sk}, H_{ek}, H_o]$ 。

节点初始化的目的是获得每种类型节点的合理初始表示。

异构图中的每条边表示所连接节点之间的相关性。通过计算实体对的余弦相似性来初始化相邻节点间的边权重:

$$A^{(0)}(i, j) = \frac{h_i^{(0)} \cdot h_j^{(0)}}{2 \|h_i^{(0)}\| \|h_j^{(0)}\|} + 0.5 \quad (1)$$

其中, $h_i^{(0)}$ 和 $h_j^{(0)}$ 是初始化的节点特征, $A^{(0)}(i, j) \in (0, 1)$ 是连接第*i*个和第*j*个节点的边权重。定义第*l*层的归一化关系矩阵

$$\tilde{A}^{(l)} = (D^{(l)})^{-0.5} A^{(l)} (D^{(l)})^{-0.5} \quad . D^{(l)} \text{ 是第 } l \text{ 层的度矩阵,}$$

$$D^{(l)}(i, j) = \sum_j A^{(l)}(i, j) .$$

2.2.2 深度异构图卷积网络(HDGCN)的构建

HDGCN 采用深度策略来缓解“邻域爆炸”和过度平滑的问题。一方面,将节点的邻域拓展到 d_{\max} 并将其中所有的节点作为候选节点;另一方面,引入参数*p*来表示节点被保留的概率,每个候选节点都有概率为*p*的可能性作为邻域节点,邻域节点的最大数量为 n_{\max} ,超过这个值的节点将被忽略。对于每一层,重复采样过程,然后将所选节点聚合到中心节点。通过这种方式,每个层都有一个由不同采样节点聚合的输出表示。

相比传统的HGCN,HDGCN更加关注图节点间的相关性,HDGCN引入了两种注意力机制,具体过程如下:

首先,分别处理不同类型的节点,并聚合交互的特征:

$$H^{(l+1)} = \sigma \left(\sum_{\tau \in T} \beta_{\tau}^{(l)} * \tilde{A}_{\tau} \cdot H_{\tau}^{(l)} \cdot W_{\tau}^{(l)} \right) \quad (2)$$

其中, $T = \{W_s, W_{sk}, W_e, W_{ek}, W_o\}$ 。 $\tilde{A}_{\tau} \in R^{|\mathcal{V}| \times |\mathcal{V}|}$ 是 \tilde{A} 的子矩阵, \tilde{A} 表示对称归一化矩阵,并且 $\tilde{A} = M^{-0.5} \cdot A \cdot M^{-0.5}$, A 为加权矩阵。

\tilde{A}_{τ} 的行代表图中的节点,列代表节点的类型。 $\beta_{\tau}^{(l)}$ 代表第*l*层中

τ 类型的注意力权重。 σ 为Relu激活函数。 $W_{\tau}^{(l)}$ 为可学习的转移

矩阵。 $H_{\tau}^{(0)} = X_{\tau}$, X_{τ} 是 τ 型节点的初始化向量。

其次,添加层注意力机制到网络中,即连接图卷积网络中具有不同注意力得分的所有层。由于每一层的输出都是由不同的采样节点聚合而成的,并且每一层特征提取的深度不同,因此实现了对不同层给予不同的关注权重。卷积网络中的每一层都利用其前面所有层的输出作为输入,即 $H^{(l+1)}$ 由来自所有先前层的图卷积转换组成:

$$H^{(l+1)} = T(\alpha_{l+1} * f(H^{(l)}, W^{(l)}), \alpha_l * H^{(l)}, \dots, \alpha_0 * H^{(0)}) \quad (3)$$

其中*T*是一个连接函数,它可以是加法函数或串联函数。 α_i 是该层的注意力得分。

最后,更新关系矩阵:

$$\Delta A^{(l+1)} = \sigma(H^{(l+1)} W_A^{(l+1)} H^{(l+1)T}) \quad (4)$$

$$A^{(l+1)} = \alpha A^{(l)} + (1 - \alpha) \Delta A^{(l+1)} \quad (5)$$

其中 $W_A^{(l+1)}$ 是用于更新相关性的可学习权重矩阵, α 为激活函数, $\alpha \in (0, 1)$ 为激活因子。

2.2.3 偏好感知 m_p 和 m_F 的生成

经过*L*层HDGCN,得到关系矩阵 $A^{(l)}$ 。同时,需要得到每个词特征向量对基于内容的模型和基于事实的模型的偏好感知分数。对于第*i*个节点,基于内容的偏好得分表示为:

$$m_{p_i} = \sum_{j=1}^n A^{(L)}(i, j) - \sum_{k=1}^{n_e} A_E^{(L)}(i, k) - \sum_{k=1}^{n_{sk}} A_{EK}^{(L)}(i, k) \quad (6)$$

同样,基于事实的偏好得分表示为:

$$m_{F_i} = \sum_{j=1}^n A^{(L)}(i, j) - \sum_{k=1}^{n_s} A_S^{(L)}(i, k) - \sum_{k=1}^{n_{sk}} A_{SK}^{(L)}(i, k) \quad (7)$$

通过归一化每个词特征向量的相关性来获得最终的偏好分数:

$$m_p = \left| \frac{m_{p_i}}{\sum_{j=1}^n m_{p_j}} \right|_{i=1}^n \quad (8)$$

$$m_F = \frac{m_{F_i}}{\sum_{j=1}^n m_{F_j}} \quad (9)$$

2.3 偏好感知与对应模型的融合

通过 2.2 节，得到新闻文本基于两种模型的不同偏好:内容偏好、事实偏好。本节利用自适应融合机制将两种偏好分别与对应模型融合，以达到检测目的。本节选取的基于内容和基于外部事实的模型均为通用的常见模型，可直接集成到本文的框架中。

2.3.1 将 m_p 与基于内容的模型融合

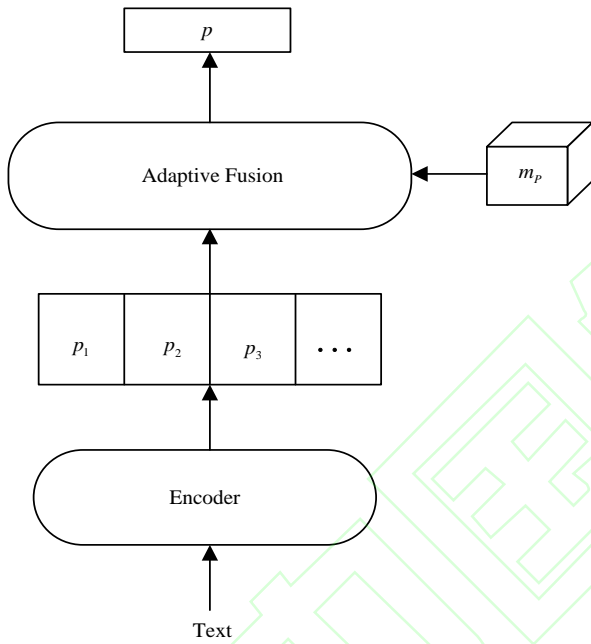


图 3 融合 m_p 与基于内容的模型

Fig.3 Fusion of m_p and content-based model

基于内容的模型通常使用文本特征提取器以获得最终预测向量，如 LSTM，BERT 等。新闻文本经过 BERT 等特征提取器后，得到特征向量 $[p_1, p_2, \dots, p_n]$ 。受 Wang^[29]等人工作的启发，本文采用自适应融合机制实现模型与对应偏好的融合。相比于直接融合，自适应融合可以更有效的结合向量间的特征，并且不会损害模型的原始表示能力。融合过程包括如下步骤：首先，通过注意力引导，灵活地对齐偏好向量 m_p 与特征向量 p_i ；其次，采用多个门控模块选择性地提取交互语义信息，最终得到融合向量 p 。经自适应融合后的最终特征向量可表示为：

$$p = \sum_{i=1}^n m_{p_i} p_i \quad (10)$$

融合过程如图 3 所示。

2.3.2 将 m_f 与基于外部事实的模型融合

基于外部事实的模型利用外部事实来源进行检索，以获得相关文章或证据。本节同样使用自适应机制融合基于外部事实的偏好感知与对应模型。设事实文章向量表示为 $[d_1, d_2, \dots, d_n]$ ，首先通过注意力引导，将偏好向量 m_p 与特征向量 d_i 对齐；接着利用门控模块选择性地提取向量间的语义信息，然后将其与事实一起送入推理模块，该模块通常用于捕捉文本和事实证据中复杂的交互作用，如一致性和冲突性。推理模块的输出向量作为最终的融合向量：

$$f = \sum_{i=1}^n m_{f_i} f_i \quad (11)$$

融合过程如图 4 所示。

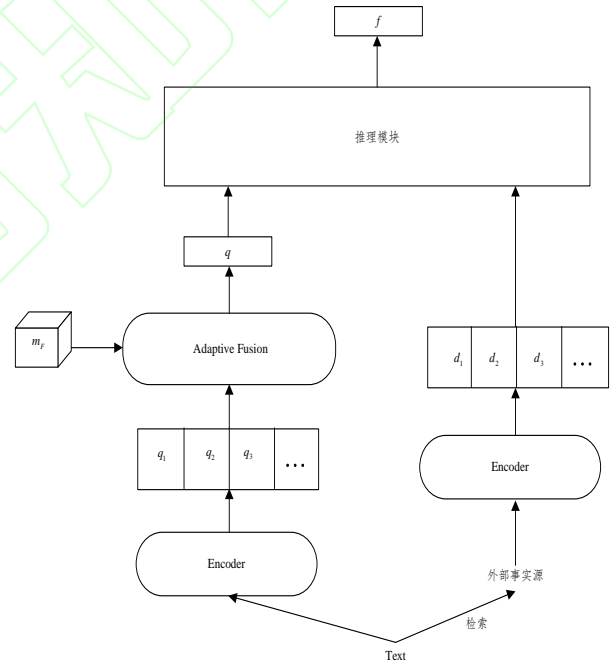


图 4 融合 m_f 与基于事实的模型

Fig.4 Fusion of m_f and fact-based model

2.4 最终预测

将 2.3 节得到的输出向量 p 和 f 连接起来，并输入到多层感知器 MLP 中，以获得最终预测：

$$\hat{y} = MLP([p; f]) \quad (12)$$

$$f = InferenceModule(q, [d_1, d_2, \dots, d_n]) \quad (13)$$

其中 $InferenceModule$ 为基于外部事实模型中的推理模块。

本文提出的模型共使用三种损失函数进行训练：

首先，使用交叉熵损失函数作为标签与真实值的损失：

$$L_{cls}(y, \hat{y}) = CELoss(y, \hat{y}) \quad (14)$$

其次，最小化内容偏好和事实偏好之间的余弦相似性作为第二层损失：

$$L_{cos} = \frac{m_p \cdot m_F}{\|m_p\| \|m_F\|} \quad (15)$$

以及交换两种模型的输入偏好后得到的损失：

$$L_{cls}(y_{rev}, \hat{y}') = CELoss(y_{rev}, \hat{y}') \quad (16)$$

其中， $y_{rev} = |1 - y|$ ，预测值 $\hat{y}' = MLP([p'; f'])$ ， p' ， f' 分别是交换两种偏好图作为输入之后的模型输出。

最后，总的损失表示为：

$$L = \alpha_1 L_{cls}(y, \hat{y}) + \alpha_2 L_{cos} + \alpha_3 L_{cls}(y_{rev}, \hat{y}') \quad (17)$$

其中 $\alpha_1, \alpha_2, \alpha_3 \in (0, 1)$ 。

3 实验与分析

3.1 实验数据集

本文使用的数据集为 Qiang 等人构建的公开数据集 Weibo 和 Twitter，该数据集同时提供新闻文本和外部事实文章，包括中文数据和英文数据。本文对其进行了预处理。原始数据集中新闻文本包含的内容为{文本，标签，分词，词类型}，其中词类型为 (W_s, W_e, W_o) 。本文利用 2.1 节的方法，扩充了词类型为 $(W_s, W_e, W_{sk}, W_{ek}, W_o)$ ，即现在的分词中不仅包括风格词 W_s 、实体词 W_e ，还包括其对应的风格语义词 W_{sk} 、实体链接词 W_{ek} 。

详细的数据集内容如表 1 所示。

Weibo 数据集包含 6362 条中文新闻文本和 17849 篇中文事实文章，其中新闻文本包括 3161 条假新闻数据和 3202 条真新闻数据。对于训练、验证和测试集，以 6:2:2 的比例保持其原始的分割。

Twitter 数据集包含 14709 条英文新闻文本和 12419 篇英文事实文章，其中新闻文本包括 5699 条假新闻数据和 9010 条真新闻数据。对于训练、验证和测试集，以 6:2:2 的比例保持其原始的分割。

3.2 用于实验的基础模型

本文选用三种基于内容的通用模型和三种基于外部事实的通用模型作为基础模型。

基于内容的模型：

- Bi-LSTM^[30] 为一种广泛使用的基于 RNN 的传统模型，它从新闻内容中获取特征编码并进行预测。为了便于比较，本文实现了一个最大序列长度为 100、隐藏大小为 128 的单层 Bi-LSTM。

- BERT-Emo^[31] 为一种典型的关注新闻内容的模型，它利用 BERT 对新闻文本进行编码并聚焦于捕捉新闻情感。本文设置其最大序列长度为 150，嵌入向量大小为 768。

- LDAVAE^[32] 专注于使用可解释的特征和方法检测文本内容中的假信息，该模型利用变分自动编码器和双向长短期记忆网

表 1 数据集统计

Table 1 Dataset Statistics

| | Weibo | | | Twitter | | |
|-----------|-------|------|------|---------|------|------|
| | Train | Val | Test | Train | Val | Test |
| Fake News | 1896 | 632 | 633 | 3419 | 1140 | 1140 |
| Real News | 1920 | 640 | 641 | 5406 | 1802 | 1802 |
| Total | 3816 | 1272 | 1274 | 8825 | 2942 | 2942 |
| Relevant | 17849 | | | 12419 | | |

络，以集成新闻文本的密集表示。本文同样设置其最大序列长度为 150，嵌入向量大小为 768。

基于外部事实的模型：

- DeClar 为一种使用特定关注度来关注新闻文章中事实词语的模型，其编码器是一个隐藏大小为 128 的单层 Bi-LSTM。

- MAC^[33] 为一种分层的多头注意力网络，其结合单词和文章级别的注意力，包含两个多头注意力模块。

- GET^[34] 提出了一种统一的基于图的语义挖掘框架，专注于基于外部事实的假新闻检测。为了将其集成到本文提出的框架中，设置最大序列长度为 150，嵌入向量大小为 768。

3.3 实验设置

表 2 单一模型对比

Table 2 Single model comparison

| | Weibo | | | | | | Twitter | | | | | |
|-----------|--------------|--------------|------------|------------|------------|------------|--------------|--------------|------------|------------|------------|------------|
| | A_{cc} | F_1 | P_{fake} | R_{fake} | P_{real} | R_{real} | A_{cc} | F_1 | P_{fake} | R_{fake} | P_{real} | R_{real} |
| 基于内容的模型 | | | | | | | | | | | | |
| Bi-LSTM | 0.667 | 0.660 | 0.626 | 0.820 | 0.744 | 0.516 | 0.767 | 0.732 | 0.753 | 0.923 | 0.811 | 0.635 |
| Pref-FEND | 0.709 | 0.709 | 0.696 | 0.735 | 0.723 | 0.683 | 0.793 | 0.788 | 0.870 | 0.779 | 0.700 | 0.754 |
| Pfnd/s | 0.721 | 0.719 | 0.708 | 0.757 | 0.724 | 0.693 | 0.812 | 0.805 | 0.886 | 0.789 | 0.708 | 0.771 |
| BERT-Emo | 0.712 | 0.708 | 0.667 | 0.839 | 0.787 | 0.587 | 0.794 | 0.762 | 0.769 | 0.950 | 0.873 | 0.675 |
| Pref-FEND | 0.746 | 0.744 | 0.703 | 0.847 | 0.811 | 0.647 | 0.804 | 0.776 | 0.781 | 0.945 | 0.870 | 0.697 |
| Pfnd/s | 0.759 | 0.753 | 0.712 | 0.838 | 0.807 | 0.709 | 0.811 | 0.787 | 0.782 | 0.936 | 0.879 | 0.701 |
| LDAVAE | 0.731 | 0.727 | 0.687 | 0.855 | 0.788 | 0.602 | 0.803 | 0.775 | 0.781 | 0.951 | 0.879 | 0.677 |
| Pref-FEND | 0.749 | 0.744 | 0.712 | 0.855 | 0.832 | 0.654 | 0.807 | 0.779 | 0.788 | 0.956 | 0.875 | 0.703 |
| Pfnd/s | 0.761 | 0.760 | 0.723 | 0.870 | 0.838 | 0.704 | 0.817 | 0.799 | 0.794 | 0.958 | 0.881 | 0.732 |
| 基于外部事实的模型 | | | | | | | | | | | | |
| DeClarE | 0.684 | 0.678 | 0.642 | 0.820 | 0.755 | 0.549 | 0.786 | 0.753 | 0.765 | 0.941 | 0.853 | 0.543 |
| Pref-FEND | 0.706 | 0.701 | 0.661 | 0.840 | 0.785 | 0.574 | 0.798 | 0.785 | 0.823 | 0.854 | 0.754 | 0.710 |
| Pfnd/s | 0.724 | 0.722 | 0.683 | 0.857 | 0.801 | 0.586 | 0.808 | 0.794 | 0.844 | 0.874 | 0.748 | 0.717 |
| MAC | 0.724 | 0.723 | 0.695 | 0.793 | 0.763 | 0.657 | 0.791 | 0.764 | 0.777 | 0.924 | 0.829 | 0.581 |
| Pref-FEND | 0.749 | 0.748 | 0.728 | 0.790 | 0.773 | 0.708 | 0.804 | 0.784 | 0.800 | 0.907 | 0.815 | 0.642 |
| Pfnd/s | 0.753 | 0.758 | 0.729 | 0.804 | 0.778 | 0.709 | 0.817 | 0.795 | 0.799 | 0.924 | 0.819 | 0.682 |
| GET | 0.737 | 0.738 | 0.702 | 0.799 | 0.749 | 0.664 | 0.799 | 0.776 | 0.775 | 0.938 | 0.830 | 0.592 |
| Pref-FEND | 0.748 | 0.747 | 0.712 | 0.814 | 0.757 | 0.684 | 0.805 | 0.783 | 0.769 | 0.935 | 0.836 | 0.598 |
| Pfnd/s | 0.767 | 0.764 | 0.734 | 0.819 | 0.771 | 0.698 | 0.820 | 0.801 | 0.778 | 0.967 | 0.870 | 0.556 |

实验时，设置 HDGCN 的层数为 2，激活因子 $\alpha = 0.5$ ，

(Precision) 和召回率 (recall) 作为对比指标。

$\alpha_1 = 0.5, \alpha_2 = 0.5, \alpha_3 = 1$ 。预训练的语言模型是从

3.5 实验结果分析

HuggingFace^[34]获得，中文模型为 bert-base-chinese，英文模型为

3.5.1 单一模型分析

bert-base-uncased，词向量维度为 768。

本文分别在 Weibo 和 Twitter 数据集上进行实验。为了与基

实验环境为：Ubuntu 16.04 LTS 64 位操作系统，64G 内存，

于内容或基于外部事实的单一模型进行公平比较，本文将所提出

CPU i7-8700，显卡 NVIDIA GeForce GTX1080Ti。采用编程语言

框架简化为单一的模型版本 Pfnd/s。即与基于内容的模型进行对

为：Python 3.7，深度学习框架为：Pytorch 1.8。

比时，删除 Pfnd 框架中基于外部事实的模型，但保留偏好感知

3.4 评价指标

模块；与基于外部事实的模型进行对比时，删除 Pfnd 框架中基

本文选取精准度(Accuracy)和 F1 值作为评价指标。同时，对

于内容的模型，保留偏好感知模块。实验对比结果如表 2 所示。

于真(Real)假(Fake)两种类型的新闻，分别测试其精确率

如表 2 所示，对于单一模型，Pfnd 在选取的六种基础模型上

都取得了一定提升。在 Weibo 数据集上, Pfnd/s 框架下的 LDAVAE 比单一的 LDAVAE 模型的 F1 值高出 3.3%; 在 Twitter 数据集上高出 2.4%。对比结果表明, 偏好感知模块可以较好的增强模型对特定内容的关注, 排除非偏好信息的干扰。在 Pfnd 框架下, 基于内容的模型可以更好地关注文本内容, 基于外部事实的模型则更好的聚焦于外部事实信息。同时, LDAVAE 的性能明显优于 Bi-LSTM 和 BERT-Emo, 在 Weibo 数据集上比 Bi-LSTM 的 F1 值提升 6.7%, 比 BERT-Emo 提升 1.9%, 这是因为 LDAVAE 使用了变分自动编码器, 能从多种文本信息中精确找到对模型偏好更有利的信息; 而 GET 的性能优于 MAC 和 DeClarE, 这是因为 GET 对于事实文章中的证据利用的更加充分, 它能整合文章中分散的证据并删掉冗余的部分, 进而提升模型的性能。

同时, 本文的框架相比 Pref-FEND 也有提升, 在三种基于内容的模型和三种基于事实的模型上, 检测效果均优于 Pref-FEND。在 Weibo 数据集上, 相比于 Pref-FEND 框架, Pfnd/s 框架下

LDAVAE 和 GET 模型的 F1 值分别提高 1.6%, 1.7%; 在 Twitter 数据集上, F1 值分别提高 2.4%和 2.5%; 其他基础模型在 Pfnd/s 框架下的实验结果同样优于 Pref-FEND, 如 Pfnd/s 框架下, GET 模型的 F1 值在 Weibo 数据集上提高 1.7%, 在 Twitter 数据集上提高 1.8%。对比结果充分表明本文所扩充词语类型的有效性, 即可以让两种模型更好的学习对应偏好。具体的, 相比于 Pref-FEND 框架的三种词语, 通过挖掘风格与词对应的相关语义词, 以及实体词对应的外部链接词, 扩充后的五种语义词极大地丰富了图网络中节点的有效表征, 可以帮助两种模型更好地理解文本中的对应的语义信息。同时, Pfnd 框架引入的 HDGCN 可以更好地融合异构图中的节点表示, 从而提升模型的偏好感知能力, 实验结果证实了引入的深度异构图卷积网络对学习不同模型偏好的能力的提升, 在消融实验中会做进一步对比。

表 3 融合模型对比
Table 3 Fusion model comparison

| | Weibo | | | | | | Twitter | | | | | |
|-----------------|--------------|--------------|------------|------------|------------|------------|--------------|--------------|------------|------------|------------|------------|
| | A_{cc} | F_1 | P_{fake} | R_{fake} | P_{real} | R_{real} | A_{cc} | F_1 | P_{fake} | R_{fake} | P_{real} | R_{real} |
| Bi-LSTM+DeClarE | | | | | | | | | | | | |
| Pref-FEND-last | 0.697 | 0.696 | 0.721 | 0.637 | 0.678 | 0.757 | 0.798 | 0.768 | 0.775 | 0.945 | 0.866 | 0.566 |
| Pfnd-last | 0.703 | 0.708 | 0.731 | 0.644 | 0.674 | 0.759 | 0.805 | 0.774 | 0.778 | 0.965 | 0.875 | 0.576 |
| Pref-FEND | 0.714 | 0.712 | 0.684 | 0.788 | 0.754 | 0.640 | 0.812 | 0.792 | 0.803 | 0.917 | 0.832 | 0.645 |
| Pfnd | 0.727 | 0.728 | 0.709 | 0.798 | 0.747 | 0.658 | 0.829 | 0.811 | 0.810 | 0.921 | 0.856 | 0.534 |
| LDAVAE+GET | | | | | | | | | | | | |
| Pref-FEND-last | 0.730 | 0.734 | 0.726 | 0.697 | 0.724 | 0.719 | 0.809 | 0.811 | 0.787 | 0.955 | 0.887 | 0.621 |
| Pfnd-last | 0.738 | 0.744 | 0.735 | 0.701 | 0.764 | 0.725 | 0.817 | 0.810 | 0.787 | 0.941 | 0.853 | 0.663 |
| Pref-FEND | 0.753 | 0.754 | 0.755 | 0.840 | 0.785 | 0.574 | 0.827 | 0.806 | 0.823 | 0.854 | 0.754 | 0.710 |
| Pfnd | 0.767 | 0.771 | 0.771 | 0.855 | 0.791 | 0.627 | 0.839 | 0.824 | 0.819 | 0.834 | 0.748 | 0.711 |

3.5.2 融合模型分析

实验选取 Bi-LSTM+DeClarE、LDAVAE+GET 两种组合作为

融合模型。同时为了公平比较, 分别将两种模型的最后一层特征连接起来作为最终预测, 即 Pfnd-last。实验结果如表 3 所示。

实验结果表明, 首先, Pfnd 优于上述基于内容和基于外部事实的方法。与单一方法相比, 在 Weibo 数据集上, 使用 LDAVAE+GET 融合模型比单独的 LDAVAE 模型的 F1 值高出 4.4%, 比单独的 GET 高出 3.3%; 在 Twitter 数据集上的 F1 值分别高出 4.9%, 4.8%, 这表明 Pfnd 在融合两种模型对应偏好方面具有良好效果; 其次, 相比 Pref-FEND 模型, 本文提出的模型 Pfnd 也取得了较大提升。在 Weibo 数据集上, 对于融合模型 Bi-LSTM+DeClarE, 本文的模型比 Pref-FEND 的 F1 值高出 1.7%, 在 Twitter 数据集上高出 1.8%。这表明引入的 HDGCN 以及扩充词语类型的有效性; 最后, 实验表明在两种框架下, 利用最后一层特征融合得到的检测性能均较低, 即这种融合方式不足以对齐两种模型方法的特征空间, 从而导致负面的融合效果。

3.6 消融实验

为验证本文提出的模型中各个模块的有效性, 现分别从两个方面进行消融对比实验。

表 4 消融实验 1

| Table 4 Ablation experiment 1 | | | | |
|-------------------------------|--------------|--------------|--------------|--------------|
| | Weibo | | Twitter | |
| | A_{cc} | F_1 | A_{cc} | F_1 |
| Bi-LSTM+DeClarE | | | | |
| Pfnd/se0 | 0.714 | 0.716 | 0.820 | 0.799 |
| Pfnd | 0.728 | 0.729 | 0.824 | 0.811 |
| LDAVAE+GET | | | | |
| Pfnd/se0 | 0.756 | 0.760 | 0.829 | 0.808 |
| Pfnd | 0.763 | 0.769 | 0.837 | 0.819 |

3.6.1 扩充词语表示的有效性

如 2.1 节所示, 为了验证扩充的语义特征词以及实体链接词对于模型的有效性, 本文删除了这两种词表示, 即只用 作为图卷积网络的节点表示, 其他模块保持不变。将此模型命名为 Pfnd/se0, 并在 Weibo 数据集和 Twitter 数据集上进行实验, 结果

如表 4 所示。结果表明, 在 Weibo 数据集上, 在 Bi-LSTM+DeClarE、LDAVAE+GET 两种融合模型情况下, Pfnd/se0 分别比 Pfnd 的 F1 值降低 1.3%和 1.2%, 在 Twitter 数据集上的结果同样低于 Pfnd。对比结果充分证明了扩充词语类型的有效性, 即挖掘更多潜在的语义信息可以帮助模型更好的学习不同偏好, 并进行分类预测。

3.6.2 深度异构图卷积网络(HDGCN)的有效性

表 5 消融实验 2

Table 5 Ablation experiment 2

| | Weibo | | Twitter | |
|-----------------|--------------|--------------|--------------|--------------|
| | A_{cc} | F_1 | A_{cc} | F_1 |
| Bi-LSTM+DeClarE | | | | |
| Pfnd/rand | 0.689 | 0.688 | 0.791 | 0.793 |
| Pfnd/gcn | 0.709 | 0.710 | 0.803 | 0.790 |
| Pfnd/hetgcn | 0.714 | 0.713 | 0.815 | 0.797 |
| Pfnd | 0.727 | 0.726 | 0.827 | 0.814 |
| LDAVAE+GET | | | | |
| Pfnd/rand | 0.722 | 0.719 | 0.798 | 0.782 |
| Pfnd/gcn | 0.746 | 0.742 | 0.814 | 0.793 |
| Pfnd/hetgcn | 0.754 | 0.750 | 0.829 | 0.806 |
| Pfnd | 0.769 | 0.763 | 0.837 | 0.819 |

如 2.2 节所示, 本文引入了 HDGCN 作为模型的偏好感知模块。为验证此模块的有效性, 分别删除了此模块,即随机初始化偏好映射 (Pfnd/rand); 将此模块替换为传统图卷积网络 GCN(Pfnd/gcn); 将此模块替换为异构图卷积网络 HetGCN(Pfnd/hetgcn)。其他模块保持不变。选取 LDAVAE+GET 作为融合模型, 实验结果如表 5 所示。在 Weibo 数据集和 Twitter 数据集上的实验表明, HDGCN 对模型的性能提升较为明显。这证明了此模块对偏好学习的有效性, 即其引入的多种注意力机制可以充分利用图节点的特征表示, 使不同类型节点的信息能够动态、充分地交互。

4 结束语

为了更好的提升虚假新闻检测的准确度,本文集成了基于内容和外部事实的联合检测模型 Pfnd。具体来说,引入深度异构图卷积网络(HDGCN)作为偏好学习器,它可以为两种模型生成对应的偏好感知并减少冗余信息;同时,为了更有效的建模图神经网络的节点表征,利用语义挖掘扩充了新闻中的风格词,利用实体链接扩充了新闻中的实体词,共同作为图网络的节点表示。在两个公开数据集 Weibo 和 Twitter 上的实验结果都表明了 Pfnd 模型的有效性。

然而,本文的模型仍有一些不足和可以改进之处,主要有:

(1)本模型仅对新闻的文本内容进行了特征建模,而社交媒体的新闻中通常包含评论等附加的社交上下文信息,如把这些信息同样集成到模型中,对模型的偏好感知和检测能力会有较大提升,这是未来关注的方向。

(2)如何进一步增强两种模型间的特征表示与联系仍是值得探究的地方,未来将会继续尝试加强模型间的交互能力,并使用更先进的融合策略,进而提升检测性能。

参考文献

[1] HUNT ALLCOTT, MATTHEW GENTZKOW. Social Media and Fake News in the 2016 Election [J]. Journal of Economic Perspectives, 2017, 31(2): 211-236.

[2] NAEEM S B, BHATTI R. The Covid 'infodemic': a new front for information professionals [J]. Health Information & Libraries Journal, 2020.

[3] YINGTONG DOU, KAI SHU, CONGYING XIA, PHILIP S, LICHAO SUN. User Preference-aware Fake News Detection [C]. SIGIR2021.

[4] YI-JU LU, CHENG-TE LI. Attention Networks for Explainable Fake News Detection on Social Media [C]. ACL: 505-514, 2020.

[5] SHENGVI JIANG, XIAOTING CHEN, LIMING ZHANG, SUTONG CHEN, HAONAN LIU. User Characteristic

Enhanced Model for Fake News Detection in Social Media [C].

NLPCC: 634-646, 2019.

- [6] CARLOS CASTILLO, MARCELO MENDOZA, BARBARA POBLETE. Information Credibility on Twitter [C] // In Proceedings of the 20th International Conference on World Wide Web (Hyderabad, India) (WWW '11). Association for Computing Machinery, New York, NY, USA, 2011: 675-684.
- [7] VOLKOVA S, SHAFFER K, JANG J Y, et al. Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2017: 647-653.
- [8] PIOTR PRZYBYLA. Capturing the Style of Fake News [C] // In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. AAAI Press, 2020: 490-497.
- [9] NAN Q, CAO J, ZHU Y, et al. MDFEND: Multi-domain Fake News Detection [J]. 2022.
- [10] KASHYAP POPAT, SUBHABRATA MUKHERJEE, ANDREW YATES, GERHARD WEIKUM. Declare: Debunking fake news and false claims using evidence-aware deep learning [J]. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 22-32.
- [11] NGUYEN VO, KYUMIN LEE. Learning from fact-checkers: Analysis and generation of fact-checking language [J]. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 335-344.
- [12] POTTHAST, M, KIESEL, J, REINARTZ, K, BEVENDORFF, J, STEIN, B. A stylometric inquiry into hyperpartisan and fake news [J]. arXiv preprint arXiv:1702.05638, 2017.
- [13] LIMENG CUI, HAESEUNG SEO, MARYAM TABAR, FENGLONG MA, SUHANG WANG, DONGWON LEE.

- DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation [C] // KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.ACM, 2020:492–502.
- [14] JAMES THORNE, ANDREAS VLACHOS, OANA COCARASCU, CHRISTOS CHRISTODOULOPOULOS, ARPIT MITTAL. The Fact Extraction and VERification (FEVER) Shared Task [J]. In Proceedings of the First Workshop on Fact Extraction and VERification(FEVER) (Brussels, Belgium). Association for Computational Linguistics, 2018:1–9.
- [15] KASHYAP POPAT, SUBHABRATA MUKHERJEE, ANDREW Y ATES, GERHARD WEIKUM. Declare: Debunking fake news and false claims using evidence-aware deep learning [J]. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018:22–32.
- [16] JING MA, WEI GAO, SHAFIQ JOTY, AND KAM-FAI WONG. Sentence-Level Evidence Embedding for Claim Verification with Hierarchical Attention Networks [J]. ACL, 2019:2561–2571.
- [17] WU L, RAO Y, YANG X, et al. Evidence-Aware Hierarchical Interactive Attention Networks for Explainable Claim Verification [C] // Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20, 2020: 1388–1394.
- [18] NGUYEN VO AND KYUMIN LEE. Hierarchical Multi-head Attentive Network for Evidence-aware Fake News Detection [J]. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 2021:965–975.
- [19] QIANG SHENG, XUEYAO ZHANG, JUAN CAO, LEI ZHONG. Integrating Pattern-and Fact-based Fake News Detection via Model Preference Learning [J]. arXiv e-prints, 2021.
- [20] ZHEZHOU KANG, YANAN CAO, YANMIN SHANG. Fake News Detection with Heterogenous Deep Graph Convolutional Network [J]. PAKDD 2021, LNAI 12712, 2021:408–420.
- [21] XUEYAO ZHANG, JUAN CAO, XIRONG LI, QIANG SHENG, LEI ZHONG, AND KAI SHU. Mining Dual Emotion for Fake News Detection. In Proceedings of the Web Conference 2021 [C] // Association for Computing Machinery, New York, NY, USA, 2021:3465–3476.
- [22] DI Y OU, NGUYEN V O, KYUMIN LEE, AND QIANG LIU. Attributed multi-relational attention network for fact-checking url recommendation [C] // Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019:1471–1480.
- [23] ZHENYU JIAO, SHUQI SUN, AND KE SUN. Chinese Lexical Analysis with Deep Bi-GRU-CRF Network [J]. arXiv.1807.01882, 2018.
- [24] YANG, DONG YU, FENG ZHANG, ZHANHUI KANG, SHUMING SHI. TexSmart: A System for Enhanced Natural Language Understanding [C] // In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations (Online). Association for Computational Linguistics, 2021: 1–10.
- [25] MILLER G A. WordNet: a lexical database for English [J]. Commun ACM, 1995, 38:39–41.
- [26] DONG Z D, DONG Q. HowNet—a hybrid language and knowledge resource [C] // Proceeding of the International Conference on Natural Language Processing and Knowledge Engineering, 2003.
- [27] LINMEI, H., YANG, T., SHI, C., JI, H., LI, X.: Heterogeneous

- graph attention net-works for semi-supervised short text classification [C] // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019.
- [28] JACOB DEVLIN, MING-WEI CHANG, KENTON LEE, AND KRISTINA TOUTANOVA. BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding [J].In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 Long and Short Papers, 2018.
- [29] Sirui Wang, Di Liang, Jian Song, Yuntao Li, Wei Wu. DABERT: Dual Attention Enhanced BERT for Semantic Matching [C]. COLING 2022.
- [30] ALEX GRAVES AND JÜRGEN SCHMIDHUBER. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural networks 18, 5-6 ,2005, 602–610.
- [31] XUEYAO ZHANG, JUAN CAO, XIRONG LI, QIANG SHENG, LEI ZHONG, AND KAI SHU. Mining Dual Emotion for Fake News Detection [J]. In Proceedings of the Web Conference 2021 (Ljubljana, Slovenia) (WWW '21). Association for Computing Machinery, New York, NY, USA,2019, 3465–3476.
- [32] MARJAN HOSSEINI, ALIREZA JAVADIAN SABET. Interpretable Fake News Detection with Topic and Deep Variational Models [J]. arXiv:2209.01536, 2022.
- [33] NGUYEN VO AND KYUMIN LEE. Hierarchical Multi-head Attentive Network for Evidence-aware Fake News Detection [J]. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (Online). Association for Computational Linguistics, 2021,965–975.
- [34] WEIZHI XU, JUNFEI WU. Evidence-aware Fake News Detection with Graph Neural Networks [J]. WWW '22, April ,2022 25–29.
- [35] WOLF T , DEBUT L , SANH V ,et al.Transformers: State-of-the-Art Natural Language Processing [C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, 2020:38–45.

张明道，出生于 1998 年，硕士，研究方向为自然语言处理。

周欣，出生于 1985 年，博士，毕业于四川大学，研究方向为自然语言处理。

吴晓红，出生于 1970 年，博士，副教授，硕士生导师。主要研究领域包括图像处理和模式识别。

卿粼波，出生于 1982 年，博士，教授，博士生导师，四川大学电子信息学院副院长。研究方向为图像处理、图像通信、嵌入式系统设计等

何小海，出生于 1964 年，博士，教授，博士生导师，四川大学电子信息学院副院长，无线电系主任，图像信息研究所所长。研究方向为通信与信息处理、图像处理与信息系统、机器视觉与智能系统。



ZHANG Ming-dao, born in 1998, postgraduate.His main research intersets include Natural Language Processing and so on.



WU Xiao-hong, born in 1970, PhD, Associate Professor, postgraduate supervisor.Her main research intersets include Image Processing and Pattern Recognition.