基于多模态异质图的社交媒体谣言检测模型*

强子珊　顾益军
（中国人民公安大学信息网络安全学院 北京 100038）

摘要：【**目的**】验证谣言不同模态之间存在关联性，以提高谣言检测准确率，进而提出一种基于多模态异质图的社交媒体谣言检测模型。【**方法**】以社交平台上多模态的帖子为研究对象，首先通过预处理提取文本、图片两种模态信息及用户属性信息的特征表示，按照文本、图片、用户三者间的关联关系构建异质图，然后按照指定的元路径提取文本类型节点的嵌入表示，最后将其输入分类器中，判断其是否是谣言。【**结果**】在公开的数据集上进行实验，结果表明，所提模型在两个数据集上的准确率分别达到91.3%和93.8%，其他评价指标也高于基线模型。【**局限**】由于共享多模态谣言的三类节点会使所构建的异质图存在较大的稀疏性，所提模型更适用于小型的话题社区。【**结论**】谣言的不同模态之间存在关联性，所提模型利用该特征在多模态谣言检测中表现出良好的效果。

关键词：谣言检测　节点嵌入　多模态　异质图　注意力机制
分类号：TP391　G350
DOI：10.11925/infotech.2096-3467.2022.0905
引用本文：强子珊，顾益军. 基于多模态异质图的社交媒体谣言检测模型[J]. 数据分析与知识发现，2023，7（11）：68-78.(Qiang Zishan, Gu Yijun. Detecting Social Media Rumors Based on Multimodal Heterogeneous Graph[J]. Data Analysis and Knowledge Discovery, 2023, 7(11): 68-78.)

1　引　言

随着互联网的发展，虚假消息可以借助社交平台快速传播，这些虚假消息被称为谣言，它们不仅会给人民群众带来恐慌，还可能引导错误的舆论，破坏社会秩序。因此，谣言检测对于维护社会稳定具有重要作用。早期的谣言传播形式主要是文本这类单模态的形式，随着社交平台功能的不断完善，如今文本、图片相结合的多模态形式的谣言更为常见，此类谣言更容易得到用户的关注，传播范围也更广泛。因此，开展多模态的谣言检测研究变得尤为重要。

谣言检测是通过提取相关消息的特征表示，然后对其进行分类，判断真假。现有的单模态谣言检测方法是利用特征提取模型提取文本或者图片一种模态的特征作为谣言表示，而多模态谣言检测方法除特征提取外还需要对所提取的不同模态特征进行融合作为谣言表示，此类先提取特征后融合的思路没有充分利用不同模态特征之间的多种关联关系，可能导致获得的谣言表示损失部分重要信息。因此，如何有效利用模态间的多种关联关系从而使得获得的嵌入表示包含的信息量更丰富成为多模态谣言检测面临的关键挑战。

本文提出一种基于多模态异质图的社交媒体谣言检测模型（Rumor Detection Model Based on Multimodal Heterogeneous Graph，Het_MMRD），利用异质图嵌入提取消息的特征表示，即同时开展特征提取和融合两项工作：首先对数据进行预处理，获

通讯作者（Corresponding author）：顾益军（Gu Yijun），E-mail：guyijun@ppsuc.edu.cn。

*本文系公安部科技强警基础工作专项（项目编号：2020GABJC02）和中央高校基本科研业务费专项（项目编号：2022JKF02039）的研究成果之一。

The work was supported by the Ministry of Public Security Science and Technology to Strengthen the Basic Work of the Police Project（Grant No.2020GABJC02），the Fundamental Research Funds for the Central Universities（Grant No.2022JKF02039）.

取文本、图片、用户三类节点特征，并按照上述三类节点及其之间的连接关系构建异质图。相较于其他两类信息，文本信息对公众的影响最大，因此模型融合异质图中指定的多条元路径信息得到文本类型节点的嵌入表示作为消息的特征表示，最后对其进行分类，判断真假。

2　相关研究

2.1　单模态的谣言检测

谣言检测的本质是提取谣言的相关特征，如内容信息、情感倾向等，然后采用相关分类方法对其进行分类。单模态谣言检测方法的主要工作在于特征提取，即采用人工选择、神经网络等方法提取谣言的相关特征，获得相应的嵌入表示。

早期的单模态谣言检测方法是先人工提取特征，然后使用决策树、支持向量机、贝叶斯网络[1-3]等方法对其进行分类，其中人工选择提取的特征包括符号特征、关键词分布特征、相似度特征等，之后也有学者对其进行补充，提出基于情感倾向特征的观点差异的判断方法[4]，该方法取得了不错的效果且对复杂网络研究产生了一定的影响[5-6]。

随着神经网络的不断发展，一些研究采用基于神经网络的方法解决了人工提取特征耗时耗力的问题。Ma 等[7]提出一种基于循环神经网络谣言的检测模型，利用信息的时序特征进行检测；刘政等[8]使用卷积神经网络学习谣言事件帖子之间的关系以挖掘文本深层的特征；Ma 等[9]利用信息的传播结构提出一种基于树结构的递归神经网络模型；Bian 等[10]同样针对信息的传播结构使用图卷积神经网络提取信息。

上述方法仅使用文本信息，而现在的很多谣言是文本、图片等内容结合的多模态形式，而且有研究表明，不同模态的信息之间存在一定的关联关系[11]，单模态的方法没有充分利用相关信息，限制了模型的性能。

2.2　多模态的谣言检测

多模态的谣言检测方法是在单模态方法的基础上增加了融合谣言不同模态特征的工作，笔者对当前多模态信息融合方法及多模态谣言检测特征融合工作中采用的方法进行论述。

（1）多模态信息融合方法

多模态的信息凭借其内容丰富、形象生动等特征成为当前社交媒体上信息传播的主要形式，进而出现了很多视觉-语言任务，如多模态情感识别、多模态谣言检测等，上述任务均需要获得多模态信息的嵌入表示。为更好地表达信息内容，融合不同模态的数据[12]成为一项重要工作。

早期的融合方法主要是简单融合，如采用拼接、求和、求平均等操作，也可以按照融合的阶段不同将其分为特征级融合和决策级融合。随着注意力机制和神经网络[13]的不断发展，Lu 等[14]提出协同注意力机制分别计算文本注意力和视觉注意力；Nam 等[15]基于类似的想法提出双重注意力网络；Wöllmer 等[16]在情感分析中使用长短期记忆网络；Hu 等[17]提出一种基于图卷积网络的多模态融合模型，有效提升了情感识别模型的性能；Jin 等[18]结合注意力机制与神经网络设计的模型也表现出良好的效果。

（2）多模态谣言检测方法

现有多模态的谣言检测方法多是基于图片和文字或用户和文字内容，特征提取思路与单模态方法基本一致，此处不再赘述。多模态特征融合主要采用简单融合和注意力融合结合神经网络的方法。

采用简单融合的多模态谣言检测方法主要是对所提取的不同模态数据特征进行拼接。刘金硕等[19]提出一种多模态网络谣言检测方法，在图片、文本内容的基础上提取图片内嵌文本的信息，并对三者进行直接拼接；陈志毅等[20]提出一种集成式多模态谣言检测方法，对文本、图片以及社会特征如用户情绪等进行直接拼接；孟佳娜等[21]提出一种基于对抗神经网络的跨模态谣言检测方法，对所提取的文本特征表示和视觉特征表示进行直接拼接。

采用注意力融合的多模态谣言检测方法分别针对不同模态数据训练相应的注意力权重以减少融合过程中信息的丢失。戚力鑫等[22]提出基于注意力机制的多模态融合方法，将文本特征与图片特征进行信息交互训练注意力权重；张少钦等[23]使用类似的思路，在训练过程中训练注意力权重更新视觉特征和文本特征；唐樾等[24]提出一种基于增强对抗网络和多模态融合的谣言检测方法，在特征融合过程中

应用注意力机制；陶雪等[25]提出一种基于注意力和多模态混合融合的谣言检测方法，在前期的特征融合阶段应用注意力机制。目前应用最多的融合方法包括加型注意力和点积型注意力，两种方法存在注意力权重计算方式上的差异。

但上述研究忽略了谣言不同模态的内容之间可能存在多种关联关系[11, 26]，如何深入挖掘此类关系并将其融入谣言的特征表示中是当前多模态谣言检测的主要任务。

2.3 应用异质图的谣言检测

（1）异质图相关定义

① 异质图

异质图是一种特殊的图，可将其定义为 $G=<V,E>$，其中 V 和 E 分别表示图中节点和边的集合。异质图不同于同质图之处在于其存在节点类型的映射函数 $\varphi: V \to \mathcal{A}$ 和边类型的映射函数 $\phi: E \to \mathcal{R}$，其中 $\mathcal{A}$ 和 $\mathcal{R}$ 分别代表节点和边类型的集合，同时 $|\mathcal{A}|+|\mathcal{R}|>2$，即，异质图中的节点或边的类型最少有两种。

② 元路径和元路径实例

元路径是异质图特有的一种定义，是定义在异质图上的路径形式，可以将其解释为异质图上的节点类型序列，表示为 $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_{n-1}} A_n$。元路径可以连接图中没有直接连接关系的节点，从而进一步提取图中的信息。给定异质图上的一条元路径 P，元路径实例即为该异质图上按照元路径所遍历到的节点序列。

③ 基于元路径的邻居

给定异质图上的一条元路径 P，基于元路径的邻居可以定义为节点 v 的集合 N_v^P，表示节点 v 通过元路径 P 下的元路径实例所连接到的节点集。

④ 谣言的多模态异质图

当前很多谣言都以包括图片和文字在内的多模态形式呈现，同时用户也与这些模态信息存在密切的关联，如一个用户发布一段或多段文字、一张或多张图片，每段文字配有一张或多张图片等，将文本、图片、用户定义为节点类型，结合上述节点间的关联关系可以构建谣言的多模态异质图，图上的元路径蕴含丰富的信息。以“文本-图片-文本”的元路径为例，其可以在共用一张图片的两条文本之间建立关联，如图1所示。以文本1为目标节点，按照该元路径可以得到“1-A-2”和“1-A-3”两条元路径实例，2和3就是1基于该元路径得到的邻居。

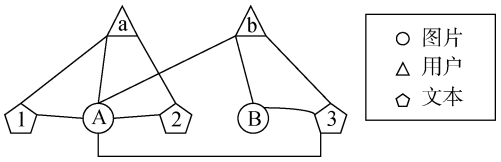

图1 谣言多模态异质图示例

Fig.1 Example of Multimodal Information Heterogeneous Graph

（2）应用异质图的谣言检测方法

应用异质图可以更多地提取图中的信息，Wang等[27]提出异质图注意力网络，通过不同层次的注意力获得最终的节点表示；Zhang等[28]对其进行改进，获得的节点表示捕捉到了网络结构上的异质性；Fu等[29]提出基于元路径的模型，按照元路径获得的节点表示包含更加丰富的语义信息。同时，异质图的应用领域也更加广泛，如推荐系统[30]、恶意软件检测[31]等，也出现了一些通过异质图进行谣言检测的研究。Huang等[32]通过用户、词语和文本构建异质图，划分用户-文本和词语-文本两个子图，分别对子图提取特征再进行融合，但是该模型仅融合了与目标节点有直接连接关系的节点信息；毕蓓等[33]提出MicroBlog-HAN（MicroBlog-Heterogeneous Graph Attention Network）模型，将用户、消息及二者之间的不同关系建模为一个异质图，按照不同的元路径获得最终的微博表示进行谣言检测，但是该模型只针对两种语义进行研究，同时忽略了元路径上的中间节点信息。

3 Het_MMRD模型构建

为解决多模态关联信息利用不足的问题，同时受前人关于异质图研究的启发，本文提出基于多模态异质图的社交媒体谣言检测模型。首先根据多模态信息及用户关联构建异质图，由于文本类型节点包含的内容最为丰富，以文本类型节点为出发点，按照语义关系人为确定异质图的4条元路径：“文本-用户-文本”“文本-图片-文本”“文本-用户-图片-用户-文本”“文本-图片-用户-图片-文本”，上述4条

元路径分别描述了两条文本信息之间4种不同的关系，即同一用户发表的文本信息、附带同一图片的文本信息、使用同一图片的不同用户发表的文本信息、同一用户使用不同图片发表的文本信息，按照元路径可以充分捕捉到文本模态信息与用户及图片之间的关联，获取到融合图片及用户信息的文本类型节点的特征表示，从而进一步判断其是否是谣言。模型整体结构如图2所示，共包括5个部分：预处理、构建异质图、不同模态之间的关联表示、相同模态之间的关联表示、结果预测。

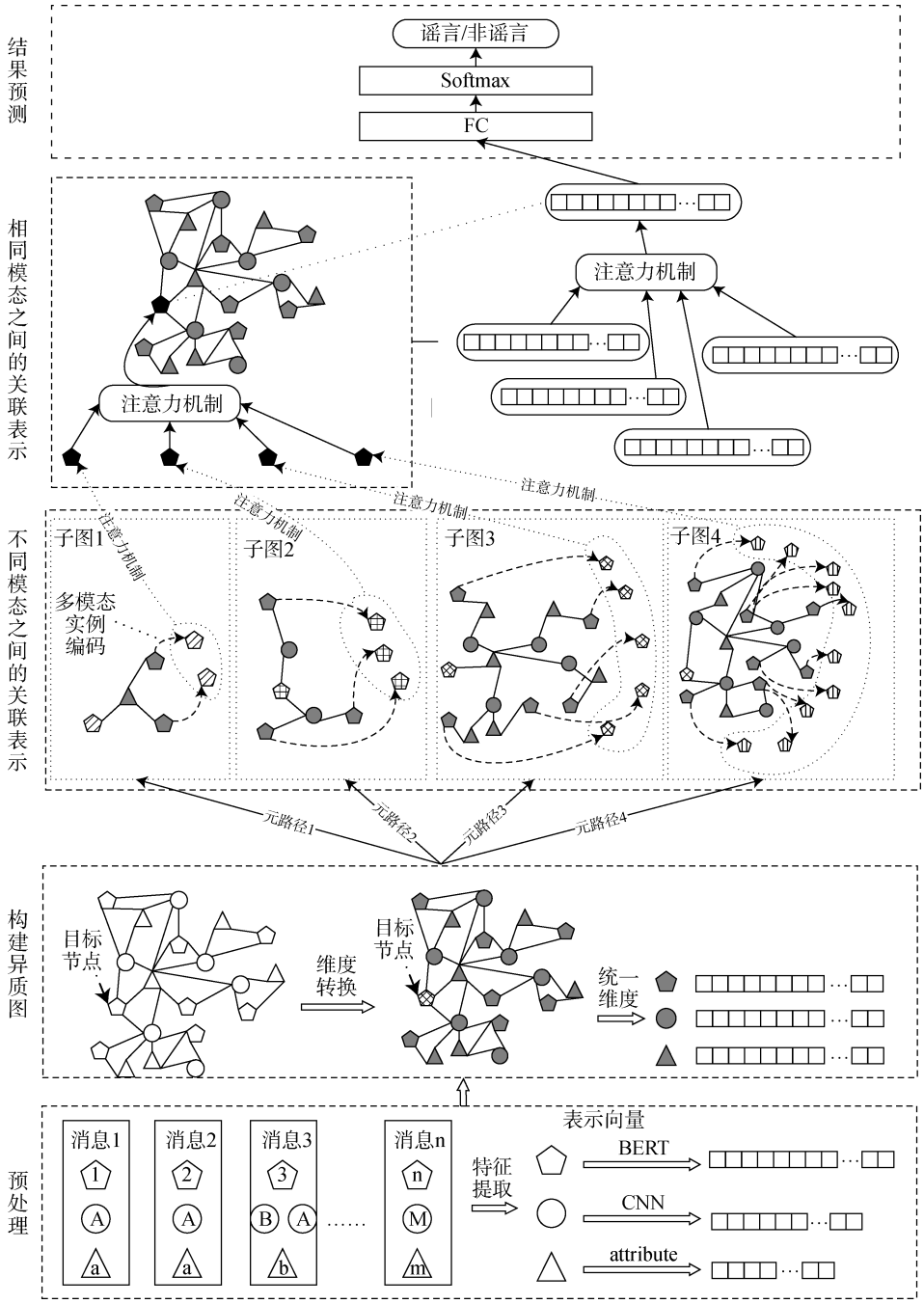

图2　Het_MMRD模型

Fig.2　Het_MMRD Model

3.1 预处理

为便于计算，针对不同模态的数据采用不同的预处理方法。同时，为了更好地检验模型的有效性，在输入模型之前对不同模态特征进行预处理：利用卷积神经网络（Convolutional Neural Network，CNN）提取图片特征作为图片类型节点的特征表示；使用BERT预训练模型提取文本类型节点的特征表示；用户属性作为用户节点的特征表示，其内容包括粉丝数量、关注数量、帖子数量等，相应位置数值对应该属性的数值。

3.2 构建异质图

将多模态消息集定义为 $T=\{T_0, T_1, \cdots, T_n\}$，$T_i$ 表示其中的第 i 条消息，每条消息由三部分组成：图片、文本、用户，定义为 $T_i=\{p_i, t_i, u_i\}$，将每条消息的三种信息类型定义为异质图中的节点类型，同一条消息内的三种信息之间存在直接关联，不同消息的信息之间基于相同的用户或图片也会建立关联，上述节点及其之间的关联关系共同构成了谣言的多模态异质图。

为便于计算，对图片、文本、用户经过预处理后的表示向量进行线性变换，将其投影到同一维度，如公式（1）-公式（3）所示。

$$\boldsymbol{h}_p = \boldsymbol{W}_p \cdot \boldsymbol{x}_p \tag{1}$$

$$\boldsymbol{h}_t = \boldsymbol{W}_t \cdot \boldsymbol{x}_t \tag{2}$$

$$\boldsymbol{h}_u = \boldsymbol{W}_u \cdot \boldsymbol{x}_u \tag{3}$$

其中，$\boldsymbol{x}_p$、$\boldsymbol{x}_t$、$\boldsymbol{x}_u$ 分别代表图片、文本、用户经过预处理之后获得的原始特征向量；$\boldsymbol{W}_p$、$\boldsymbol{W}_t$、$\boldsymbol{W}_u$ 分别代表三者的权重矩阵；$\boldsymbol{h}_p$、$\boldsymbol{h}_t$、$\boldsymbol{h}_u$ 分别代表三者转换后的表示向量。转换后的三种表示向量的维度均为 d。

3.3 不同模态之间的关联表示

为获得文本类型节点与其他类型节点之间的关联表示，确定目标文本类型节点后，以该节点为出发点，在异质图上提取其基于不同类型元路径所形成的多个子图，然后基于各子图对该节点的表示信息进行完善，具体实现过程如下：首先对每个子图上的元路径实例分别进行编码，即多模态实例编码；然后融合同一子图上的编码信息，进而获得该节点不同模态之间的关联表示。

（1）多模态实例编码

多模态实例编码的目的是获得元路径实例的表示，从而对目标节点的表示信息进行更新，其结构如图3所示。不同的元路径实例蕴含不同的信息，以“文本-用户-文本”元路径为例，即同一用户发表的两段文本，当两段文本针对同一个事件时，文本的重要性程度可能更高，因此文本信息应更多地予以保留；当两段文本针对不同事件时，用户的重要性程度可能更高，因此用户信息应更多地予以保留。为更好地实现元路径实例表示，受知识图嵌入启发，模型采用关系旋转编码器[34]，给定一条元路径实例，按照节点之间连接的顺序关系对节点信息进行融合，这样所获得的表示信息可以保留节点序列的结构。

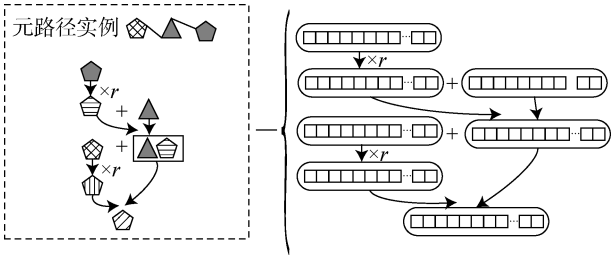

图3 多模态实例编码

Fig.3 Multimodal Instance Coding

定义元路径实例 $P(v,z)=(m_0, m_1, \cdots, m_n)$，其中 $m_0=z$，$m_n=v$；定义参数向量 $\boldsymbol{r}_i$，表示节点 m_{i-1} 和 m_i 之间的关系，具体计算过程如公式（4）-公式（6）所示。

$$\boldsymbol{o}_0 = \boldsymbol{h}_{m_0} = \boldsymbol{h}_z \tag{4}$$

$$\boldsymbol{o}_i = \boldsymbol{h}_{m_i} + \boldsymbol{o}_{i-1} \circ \boldsymbol{r}_i \tag{5}$$

$$\boldsymbol{h}_{P(v,z)} = \frac{\boldsymbol{o}_n}{n+1} \tag{6}$$

其中，$\circ$ 表示两个向量的对应元素相乘，对所获得的 $\boldsymbol{o}_n$ 平均化获得对应元路径实例的表示 $\boldsymbol{h}_{P(v,z)} \in \mathbb{R}^d$。

（2）不同模态信息融合

按照公式（4）-公式（6）的计算可以获得元路径实例的表示 $\boldsymbol{h}_{P(v,z)}$，对目标节点在同一元路径下的元路径实例进行融合可以使其蕴含的信息更丰富。考虑到不同的元路径实例对于目标节点表示的贡献程度不同，在将元路径实例编码为向量表示之后应用注意力机制，为每个元路径实例学习到一个重要性权重 α_{vz}^{p}，具体计算如公式（7）-公式（9）所示。

$$e_{vz}^{P} = LeakyReLU\left(\boldsymbol{a}_{P}^{\mathrm{T}} \cdot [\boldsymbol{h}_{v} \| \boldsymbol{h}_{P(v,z)}]\right) \tag{7}$$

$$\alpha_{vz}^{P} = \frac{\exp(e_{vz}^{P})}{\sum_{s \in N_{v}^{P}} \exp(e_{vs}^{P})} \tag{8}$$

$$\boldsymbol{h}_{v}^{P} = \sigma\left(\sum_{z \in N_{v}^{P}} \alpha_{vz}^{P} \cdot \boldsymbol{h}_{P(v,z)}\right) \tag{9}$$

其中，$\boldsymbol{a}_{P} \in \mathbb{R}^{2d}$ 表示元路径 P 的注意力参数向量；$\|$ 表示拼接操作；e_{vz}^{P} 表示元路径实例 $P(v,z)$ 对节点 v 的重要性，使用 Softmax 函数进行归一化后获得注意力权重 α_{vz}^{P}，最后计算节点 v 基于元路径 P 的表示 $\boldsymbol{h}_{v}^{P}$。

为了保持学习过程的稳定性，使用 K 头注意力机制，然后对输出进行拼接，计算过程如公式（10）所示。

$$\boldsymbol{h}_{v}^{P} = \Big\|_{k=1}^{K} \sigma\left(\sum_{z \in N_{v}^{P}} ([\alpha_{vz}^{P}]_{k} \cdot \boldsymbol{h}_{P(v,z)}\right) \tag{10}$$

其中，$[\alpha_{vz}^{P}]_{k}$ 表示节点 v 的元路径实例 $P(v,z)$ 在第 k 个注意力头归一化的重要性。

因为元路径实例中包含图片、用户的信息，因而可以对文本类型的节点表示起到补充作用。用 A 表示文本类型，元路径集合表示为 $P_{A} = \{P_{1}, P_{2}, P_{3}, P_{4}\}$，目标节点 v 基于不同元路径所获得的表示集合为 $\{\boldsymbol{h}_{v}^{P_1}, \boldsymbol{h}_{v}^{P_2}, \boldsymbol{h}_{v}^{P_3}, \boldsymbol{h}_{v}^{P_4}\}$，按照各元路径，可以在来自相同用户或使用相同图片的文本模态信息之间建立关联关系，从而使文本类型节点所包含的信息更加丰富。

3.4　相同模态之间的关联表示

按照 3.3 节计算可以获得文本类型节点基于不同元路径的表示信息，相同模态之间的关联是通过应用注意力机制对目标节点基于 4 条元路径的表示进行融合。假设构建的异质图中文本类型节点数量为 $|V|$，对于所有 $v \in V$，都存在 $\{\boldsymbol{h}_{v}^{P_1}, \boldsymbol{h}_{v}^{P_2}, \boldsymbol{h}_{v}^{P_3}, \boldsymbol{h}_{v}^{P_4}\}$，不同的元路径在表达节点时所发挥的重要性不同，为便于计算 4 条元路径各自的注意力权重，首先平均化所有特定元路径的表示向量，如公式（11）所示。

$$\boldsymbol{s}_{P_i} = \frac{1}{|V|} \sum_{v \in V} \tanh\left(\boldsymbol{M}_{A} \cdot \boldsymbol{h}_{v}^{P_i} + \boldsymbol{b}_{A}\right) \tag{11}$$

其中，$\boldsymbol{M}_{A} \in \mathbb{R}^{d_m \times d}$ 和 $\boldsymbol{b}_{A} \in \mathbb{R}^{d_m}$ 是可学习的参数。

然后使用注意力机制对元路径的向量表示进行融合，计算过程如公式（12）-公式（14）所示。

$$e_{P_i} = \boldsymbol{q}_{A}^{\mathrm{T}} \cdot \boldsymbol{s}_{P_i} \tag{12}$$

$$\beta_{P_i} = \frac{\exp(e_{P_i})}{\sum_{P \in P_{A}} \exp(e_{P})} \tag{13}$$

$$\boldsymbol{h}_{v}^{P_A} = \sum_{P \in P_{A}} \beta_{P} \cdot \boldsymbol{h}_{v}^{P} \tag{14}$$

其中，$\boldsymbol{q}_{A} \in \mathbb{R}^{d_m}$ 是指定元路径的注意力参数向量；β_{P_i} 是元路径 P_i 的重要性；对基于元路径获得的 4 个表示信息进行求和获得目标节点 v 的表示向量 $\boldsymbol{h}_{v}^{P_A}$。

3.5　结果预测

全连接层实际上是一个线性层，其作用是将所获得向量表示投影到特定的维度，以便对其分类，如公式（15）所示。

$$y = FC(\boldsymbol{h}_{v}^{P_A}) = \sigma(\boldsymbol{W}_{F} \cdot \boldsymbol{h}_{v}^{P_A} + \boldsymbol{b}_{F}) \tag{15}$$

其中，$\sigma(\cdot)$ 是一个线性激活函数；$\boldsymbol{W}_{F} \in \mathbb{R}^{d_F \times d_m}$ 是一个权重参数矩阵；$\boldsymbol{b}_{F} \in \mathbb{R}^{d_F}$ 是一个参数向量。

然后连接 Softmax 层，将每个类别的概率映射到 0-1 之间，获得各类别归一化的概率表示 $\hat{y}$，如公式（16）所示。

$$\hat{y} = Softmax(y) \tag{16}$$

选取输出概率值最大的标签作为该节点即该消息的分类结果 $\hat{y}_i$，如公式（17）所示。

$$\hat{y}_i = \mathrm{MAX}(\hat{y}) \tag{17}$$

模型通过最小化交叉熵损失函数和 Adam 算法进行优化，将获得的标签 $\hat{y}_i$ 与实际标签 y_i 进行比较计算损失值，如公式（18）所示。

$$L = -\sum_{i=1}^{|V|} (y_i (\log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))) \tag{18}$$

其中，L 为损失函数；$|V|$ 为文本类型节点总数。使用反向传播对其进行优化，对参数进行更新。

4　实验与分析

4.1　实验数据集

本文采用公开的多模态数据集 image-verification-corpus，其中包括 MediaEval2015 和 MediaEval2016 两个数据集[35-36]。为便于挖掘同时期用户及信息之间的关联同时检验模型效果，选取两个数据集中的部分数据进行实验，其中包括推文的文本内容及相应的用户信息、图片和标签，每条推文对应一个用户和一张或多张图片，具体统计情况如表 1 所示，每个数据集中用户和图片总数均小于文本总数，即文本和用户之间存在一对一和多对一

的关系，文本和图片之间存在一对一、一对多和多对一的关系，用户和图片之间存在一对一、多对一、一对多和多对多的关系，上述三类节点和边的分布符合异质图结构特征。另外，由于现实场景中谣言的占比并不大，因此本文选取的每个数据集中真假推文的数量比在2∶1左右，训练集和测试集划分比例为4∶1。

表1　数据统计

Table 1　Data Statistics

数据集	总数	谣言	非谣言	文本数量	用户数量	图片数量	节点总数
MediaEval2015	1 144	385	759	1 144	1 100	40	2 284
MediaEval2016	1 206	400	806	1 206	1 157	96	2 459

4.2　对比实验设计

为验证信息模态之间是否存在关联性，选取两类模型进行对比：单模态模型和多模态模型，其中单模态模型将经过预处理所提取的特征表示转换维度直接输入分类器判断；多模态模型按照不同类型的融合方法和异质图的方法进一步提取特征后再输入分类器。

（1）单模态谣言检测模型

由于用户无法判断其真假，而假文本常常与假图片相伴而生，因此单模态模型仅使用文本和图片信息分别进行判断，其中文本和图片的标签均与对应的信息标签一致。

①Text：仅使用文本特征进行谣言检测，即将通过预处理获得的文本信息的特征表示直接输入分类器，经过全连接层和Softmax层输出分类结果。

②Picture：仅使用图片特征进行谣言检测，即将通过预处理获得的图片信息的特征表示直接输入分类器，经过全连接层和Softmax层输出分类结果。

（2）多模态谣言检测模型

①FCC（FeatureConcate）[37]：特征级拼接模型，即将预处理获得的三种信息的特征表示进行直接拼接后输入分类器。

②DCC（DecisionConcate）[37]：决策级拼接模型，即将预处理获得的三种信息的特征表示分别输入分类器，将三种分类结果拼接后再进行最终的分类。

③Add_Att[25]：加型注意力模型，即按照加型注意力计算公式计算各信息的注意力权重，获得融合后的特征表示再对其进行分类。

④Dot_Att[22]：点积型注意力模型，即按照点积注意力计算公式计算各信息的注意力权重，获得融合后的特征表示再对其进行分类。

⑤MicroBlog-HAN模型[33]：为保证与实验数据及原文方法一致从而更好地进行对比，该模型采用与本文相同的元路径连接相应的文本类型节点，然后使用图注意力网络按照相同类型节点之间的关联更新节点特征后对其进行分类。

⑥Het_MMRD模型：本文提出的基于多模态异质图的社交媒体谣言检测模型。

4.3　实验设置

在预处理阶段，使用BERT模型提取文本类型节点的特征，设置两个卷积层提取图片类型节点的特征，为便于计算，将上述两类特征的输出维度均设置为128。用户属性直接使用数据集中的好友数量、粉丝数量、帖子数量等共计7个属性的数值，其中是否有URL和是否被验证两个属性的标签True和False分别用1和0表示，为保证用户特征与其他特征维度一致，在每个用户表示向量后拼接一个全零的向量。不同模态关联阶段应用多头注意力机制，注意力头数量设置为8。模型训练阶段，设置Epoch为100，学习率为0.005，同时应用早停法，当准确率在15个Patience内不再上升时停止训练，通过最小化交叉熵损失函数和Adam优化器对其进行优化。

4.4　实验结果

为验证本文模型的有效性，采用常见的评价指标[38]准确率（Acc）、精确率（P）、召回率（R）和F1值进行评价。两个数据集上的实验结果如表2和表3所示。

表2　MediaEval2015数据集实验结果

Table 2　Experimental Results of MediaEval2015 Dataset

模态	模型	*Acc*	*P*	*R*	*F*1
单模态模型	Text	0.555	0.368	0.405	0.386
	Picture	0.454	0.269	0.439	0.333
多模态模型	FCC	0.655	0.500	0.506	0.503
	DCC	0.620	0.431	0.317	0.365
	Add_Att	0.838	0.728	0.848	0.784
	Dot_Att	0.681	0.568	0.317	0.407
	MicroBlog-HAN	0.887	0.835	0.835	0.835
	Het_MMRD	**0.913**	**0.839**	**0.924**	**0.880**

表3　MediaEval2016数据集实验结果

Table 3　Experimental Results of MediaEval2016 Dataset

模态	模型	*Acc*	*P*	*R*	*F*1
单模态模型	Text	0.558	0.354	0.430	0.389
	Picture	0.482	0.279	0.537	0.367
多模态模型	FCC	0.632	0.432	0.405	0.418
	DCC	0.591	0.384	0.418	0.400
	Add_Att	0.711	0.571	0.456	0.507
	Dot_Att	0.703	0.557	0.430	0.485
	MicroBlog-HAN	0.909	0.880	**0.835**	0.857
	Het_MMRD	**0.938**	**0.985**	0.823	**0.897**

（1）单模态模型和多模态模型对比

FCC、DCC、Add_Att、Dot_att、MicroBlog-HAN、Het_MMRD这6种多模态模型的准确率均比Text和Picture两种单模态模型高，表明图片信息中含有文本信息中所不包含的特征，对该特征进行有效利用可以提高模型效果，同时也验证了多模态模型的有效性。

（2）多模态模型对比

注意力模型总体上比直接拼接模型效果更好。对比单模态模型的实验结果可以发现，文本特征比图片特征在谣言检测工作中发挥的作用更大，因此使用注意力机制赋予不同模态特征不同的注意力权重可以有效提升模型性能。同时，对比直接拼接的两个模型可以发现，相较于FCC模型，DCC模型的效果并不是很稳定，这是因为决策级融合是对各个模态的分类结果进行整合，解决了特征级融合中存在的数据异步性问题，但由于各个模态信息可能存在数据缺失，进而会对分类结果产生影响。另外，对比两种注意力模型的结果，Add_Att模型的效果总体上优于Dot_Att模型，这是因为二者采用不同的计算方法计算注意力权重，进而可能导致结果的差异。

MicroBlog-HAN和Het_MMRD两种基于异质图的模型比上述两类多模态模型效果更佳。这是因为这类模型基于元路径捕捉到文本之间的关联关系，所获得的特征表示中嵌入了文本之间的结构特征，使目标节点所包含的语义更加丰富，在提高模型性能的同时，验证了异质图在谣言检测工作中的有效性。

（3）基于异质图的多模态模型对比

Het_MMRD模型在两个数据集上的准确率分别达到91.3%和93.8%，均高于MicroBlog-HAN模型。这是因为本文针对元路径实例进行编码，有效保留了用户、图片中的有效信息并将其与文本信息进行融合，这种按照信息之间的多种关联关系提取特征的方法，有效验证了谣言不同模态之间是相互关联的，彼此之间可以起到互补的作用，融合三种信息所获得的向量表示内容更加丰富，进而有效提升模型检测效果。

虽然在MediaEval2016数据集中，MicroBlog-HAN模型的召回率最高，考虑到计算公式的差异，出现准确率高、召回率稍低的结果属于正常现象。另外，综合比较模型在两个数据集上的准确率、精确率、召回率和F1值，Het_MMRD模型的效果依然最佳。

4.5　消融实验分析

为验证本文模型的有效性，同时检验人为定义的4条元路径效果，设计了如下4种变体，即分别去除“文本-用户-文本”（T-U-T）、“文本-图片-文本”（T-P-T）、“文本-用户-图片-用户-文本”（T-U-P-U-T）、“文本-图片-用户-图片-文本”（T-P-U-P-T）4条元路径。

（1）去除“文本-用户-文本”元路径（$\text{Het\_MMRD}_{\text{without T-U-T}}$），即按照T-P-T、T-U-P-U-T、T-P-U-P-T三条元路径更新文本类型节点表示。

（2）去除“文本-图片-文本”元路径（$\text{Het\_MMRD}_{\text{without T-P-T}}$），即按照T-U-T、T-U-P-U-T、T-P-U-P-T三条元路径更新文本类型节点表示。

（3）去除“文本-用户-图片-用户-文本”元路径（$\text{Het\_MMRD}_{\text{without T-U-P-U-T}}$），即按照T-P-T、T-U-T、T-P-U-P-T三条元路径更新文本类型节点表示。

（4）去除“文本-图片-用户-图片-文本”元路径（$\text{Het\_MMRD}_{\text{without T-P-U-P-T}}$），即按照T-P-T、T-U-T、T-U-P-U-T三条元路径更新文本类型节点表示。

实验结果如表4所示。首先，本文提出的Het_MMRD模型在两个数据集上的准确率、精确率、召回率和F1值均高于4种变体，由此证明了按照语义人为定义的4条元路径的有效性，利用文本间的关联可以增强语义的丰富性。其次，无论去除哪条元路径，模型的性能均会有不同程度的下降，证明元路径有效性的同时，也说明按照元路径提取的嵌

入表示所发挥的作用大小可能与数据集中数据的分布情况有关，不同时期的数据分布呈现出不同的特征，这会导致每条元路径下的元路径实例数量不同，影响所获得的表示信息，从而影响实验结果。另外，由于实验中使用的数据规模并不是很大，至于元路径在模型中所发挥的作用大小是否与数据分布有关，有待使用其他数据集开展实验进一步探讨。

表4　消融实验结果对比

Table 4　Comparison of Ablation Results

数据集	模型	*Acc*	*P*	*R*	*F*1
MediaEval2015	Het_MMRD$_{\text{without T-U-T}}$	0.873	0.798	0.848	0.822
	Het_MMRD$_{\text{without T-P-T}}$	0.900	0.833	0.886	0.859
	Het_MMRD$_{\text{without T-U-P-U-T}}$	0.887	0.791	0.911	0.847
	Het_MMRD$_{\text{without T-P-U-P-T}}$	0.908	0.837	0.911	0.873
	Het_MMRD	0.913	0.839	0.924	0.880
MediaEval2016	Het_MMRD$_{\text{without T-U-T}}$	0.930	0.984	0.798	0.881
	Het_MMRD$_{\text{without T-P-T}}$	0.913	0.953	0.772	0.853
	Het_MMRD$_{\text{without T-U-P-U-T}}$	0.901	0.983	0.709	0.824
	Het_MMRD$_{\text{without T-P-U-P-T}}$	0.926	0.984	0.785	0.873
	Het_MMRD	0.938	0.985	0.823	0.897

5　结　语

现有的多模态谣言检测模型忽略了模态之间存在的多种关联关系，因此，为提升多模态谣言检测模型的性能，本文构建了一种基于多模态异质图的社交媒体谣言检测模型，并使用公开的数据集进行实验，通过与单模态谣言检测模型及单独开展特征融合的多模态谣言检测模型进行比较，验证了不同模态信息之间是相互关联的，同时也验证了本文模型在多模态谣言检测任务上的有效性。

但模型仍存在一定的不足，一是消息中的图片经过多次转发、保存、修改，其像素、尺寸等会发生变化，如何识别这些肉眼相同的图片进而构建异质图中的边；二是对社交媒体上的所有节点进行共享可能会使网络存在较大的稀疏性。因此，在实际应用中，还需要使用pHash等算法先对肉眼相同的图片进行识别，然后构建异质图；另外，帖子量虽然很大，但其中很多帖子针对的是同一事件，而这些帖子并不是很多，将其定义为该事件的讨论社区，针对一个社区中的多模态帖子构建异质图可以有效降低网络的稀疏性。未来将针对社交媒体上的话题社区发现工作开展研究，同时探索其他有效的元路径减小模型对数据的依赖。

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. ACM, 2011: 675-684.

[2] Wu K, Yang S, Zhu K Q. False Rumors Detection on Sina Weibo by Propagation Structures[C]//Proceedings of 2015 IEEE 31st International Conference on Data Engineering. 2015: 651-662.

[3] Qazvinian V, Rosengren E, Radev D R, et al. Rumor Has It: Identifying Misinformation in Microblogs[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011: 1589-1599.

[4] 祖坤琳, 赵铭伟, 郭凯, 等. 新浪微博谣言检测研究[J]. 中文信息学报, 2017, 31(3): 198-204.(Zu Kunlin, Zhao Mingwei, Guo Kai, et al. Research on the Detection of Rumor on Sina Weibo[J]. Journal of Chinese Information Processing, 2017, 31(3): 198-204.)

[5] 李文政, 顾益军, 闫红丽. 基于网络贝叶斯信息准则算法的社区数量预测研究[J]. 数据分析与知识发现, 2020, 4(4): 72-82.(Li Wenzheng, Gu Yijun, Yan Hongli. Predicting Community Numbers with Network Bayesian Information Criterion[J]. Data Analysis and Knowledge Discovery, 2020, 4(4): 72-82.)

[6] 王本钰, 顾益军, 彭舒凡. 基于粒子竞争机制的半监督社区发现算法[J]. 计算机科学与探索, 2023, 17(3): 608-619.(Wang Benyu, Gu Yijun, Peng Shufan. Semi-supervised Community Detection Algorithm Based on Particle Competition[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(3): 608-619.)

[7] 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. 2016: 3818-3824.

[8] 刘政, 卫志华, 张韧弦. 基于卷积神经网络的谣言检测[J]. 计算机应用, 2017, 37(11): 3053-3056, 3100.(Liu Zheng, Wei Zhihua, Zhang Renxian. Rumor Detection Based on Convolutional Neural Network[J]. Journal of Computer Applications, 2017, 37(11): 3053-3056, 3100.)

[9] 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(Volume 1: Long Papers). 2018: 1980-1989.

[10] 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. 2020: 549-556.

[11] 王雷全. 基于图模型的多模态社会媒体分析[D]. 北京: 北京邮电大学, 2016. (Wang Leiquan. Analysis of Multi-modal Social Media Based on Graph Model[D]. Beijing: Beijing University of Posts and Telecommunications, 2016.)

[12] 吴友政, 李浩然, 姚霆, 等. 多模态信息处理前沿综述:应用、融合和预训练[J]. 中文信息学报, 2022, 36(5): 1-20. (Wu Youzheng, Li Haoran, Yao Ting, et al. A Survey of Multimodal Information Processing Frontiers: Application, Fusion and Pre-training[J]. Journal of Chinese Information Processing, 2022, 36 (5): 1-20.)

[13] Ngiam J, Khosla A, Kim M, et al. Multimodal Deep Learning[C]// Proceedings of the 28th International Conference on International Conference on Machine Learning. 2011: 689-696.

[14] Lu J S, Yang J W, Batra D, et al. Hierarchical Question-Image Co-attention for Visual Question Answering [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems.2016: 289-297.

[15] Nam H, Ha J W, Kim J. Dual Attention Networks for Multimodal Reasoning and Matching [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2156-2164.

[16] Wöllmer M, Metallinou A, Eyben F, et al. Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression Using Bidirectional LSTM Modeling[C]// Proceedings of the 11th Annual Conference of the International Speech Communication Association. 2010: 2362-2365.

[17] Hu J W, Liu Y C, Zhao J M, et al. MMGCN: Multimodal Fusion via Deep Graph Convolution Network for Emotion Recognition in Conversation[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 5666-5675.

[18] 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. 2017: 795-816.

[19] 刘金硕, 冯阔, Jeff Z. Pan , 等. MSRD: 多模态网络谣言检测方法[J]. 计算机研究与发展, 2020, 57(11): 2328-2336. (Liu Jinshuo, Feng Kuo, Jeff Z. Pan , et al. MSRD: Multi-modal Web Rumor Detection Method[J]. Journal of Computer Research and Development, 2020, 57(11): 2328-2336.)

[20] 陈志毅, 隋杰. 基于DeepFM和卷积神经网络的集成式多模态谣言检测方法[J]. 计算机科学, 2022, 49(1): 101-107. (Chen Zhiyi, Sui Jie. DeepFM and Convolutional Neural Networks Ensembles for Multimodal Rumor Detection[J]. Computer Science, 2022, 49(1): 101-107.)

[21] 孟佳娜, 王晓培, 李婷, 等. 基于对抗神经网络的跨模态谣言检测[J]. 数据分析与知识发现, 2022, 6(12): 32-42. (Meng Jiana, Wang Xiaopei, Li Ting, et al. Cross-Modal Rumor Detection Based on Adversarial Neural Networks[J]. Data Analysis and Knowledge Discovery, 2022, 6(12): 32-42.)

[22] 戚力鑫, 万书振, 唐斌, 等. 基于注意力机制的多模态融合谣言检测方法[J]. 计算机工程与应用, 2022, 58(19): 209-217. (Qi Lixin, Wan Shuzhen, Tang Bin, et al. Multimodal Fusion Rumor Detection Method Based on Attention Mechanism[J]. Computer Engineering and Applications, 2022, 58(19): 209-217.)

[23] 张少钦, 杜圣东, 张晓博, 等. 融合多模态信息的社交网络谣言检测方法[J]. 计算机科学, 2021, 48(5): 117-123. (Zhang Shaoqin, Du Shengdong, Zhang Xiaobo, et al. Social Rumor Detection Method Based on Multimodal Fusion[J]. Computer Science, 2021, 48(5): 117-123.)

[24] 唐樾, 马静. 基于增强对抗网络和多模态融合的谣言检测方法[J]. 情报科学, 2022, 40(6): 108-114. (Tang Yue, Ma Jing. A Rumor Detection Method Based on Enhance Adversarial Network and Multimodal Fusion[J]. Information Science, 2022, 40(6): 108-114.)

[25] 陶霄, 朱焱, 李春平. 基于注意力与多模态混合融合的谣言检测方法[J]. 计算机工程, 2021, 47(12): 71-77. (Tao Xiao, Zhu Yan, Li Chunping. Rumor Detection Method Based on Attention and Multi-modal Hybrid Fusion[J]. Computer Engineering, 2021, 47(12): 71-77.)

[26] 何俊, 张彩庆, 李小珍, 等. 面向深度学习的多模态融合技术研究综述[J]. 计算机工程, 2020, 46(5): 1-11. (He Jun, Zhang Caiqing, Li Xiaozhen, et al. Survey of Research on Multimodal Fusion Technology for Deep Learning[J]. Computer Engineering, 2020, 46(5): 1-11.)

[27] Wang X, Ji H Y, Shi C, et al. Heterogeneous Graph Attention Network[C]//Proceedings of WWW'19. 2019: 2022-2032.

[28] Zhang C X, Song D J, Huang C, et al. Heterogeneous Graph Neural Network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 793-803.

[29] Fu X Y, Zhang J N, Meng Z Q, et al. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding[C]//Proceedings of the WWW 2020 . 2020: 2331-2341.

[30] 李昊杰. 异质图神经网络研究及其在推荐系统中的应用[D]. 成都: 电子科技大学, 2022.(Li Haojie. Research on Heterogeneous Graph Neural Network and Its Application in Recommendation System[D]. Chengdu: University of Electronic Science and Technology of China, 2022.)

[31] 李良训. 基于异质图嵌入的Android恶意软件检测的研究与实现[D]. 北京: 北京邮电大学,2021.(Li Liangxun. Research and Implementation of Heterogeneous Graph Embedding for Android Malware Detection[D]. Beijing: Beijing University of Posts and Telecommunications, 2021.)

[32] Huang Q, Yu J S, Wu J, et al. Heterogeneous Graph Attention Networks for Early Detection of Rumors on Twitter[C]// Proceedings of 2020 International Joint Conference on Neural Networks. 2020: 1-8.
[33] 毕蓓, 潘慧瑶, 陈峰, 等. 基于异构图注意力网络的微博谣言监测模型[J]. 计算机应用, 2021, 41(12): 3546-3550.(Bi Bei, Pan Huiyao, Chen Feng, et al. Microblog Rumor Detection Model Based on Heterogeneous Graph Attention Network[J]. Journal of Computer Applications, 2021,41(12): 3546-3550.)
[34] Sun Z, Deng Z H, Nie J Y, et al. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space[OL]. arXiv Preprint, arXiv:1902.10197.
[35] Boididou C, Andreadou K, Papadopoulos S, et al. Verifying Multimedia Use at MediaEval 2015[C]//Proceedings of MediaEval 2015 Workshop. 2015.
[36] Maigrot C, Claveau V, Kijak E, et al. MediaEval 2016: A Multimodal System for the Verifying Multimedia Use Task[C]// Proceedings of MediaEval 2016 Workshop. 2016.
[37] 王壮, 隋杰. 基于多级融合的多模态谣言检测模型[J]. 计算机工程与设计, 2022, 43(6): 1756-1761.(Wang Zhuang, Sui Jie. Multimodal Rumor Detection Model Based on Multilevel Fusion [J]. Computer Engineering and Design, 2022, 43(6): 1756-1761.)
[38] Godbole S, Sarawagi S. Discriminative Methods for Multi-labeled Classification[C]//Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2004: 22-30.

作者贡献声明：

强子珊：提出研究思路，设计研究方案，进行实验，撰写论文；
顾益军：提出研究思路，设计研究方案，论文最终版本修订。

利益冲突声明：

所有作者声明不存在利益冲突关系。

支撑数据

[1] 强子珊. 多模态谣言检测数据集. DOI: 10.57760/sciencedb.j00133.00365.

收稿日期：2022-08-28
收修改稿日期：2022-12-17

Detecting Social Media Rumors Based on Multimodal Heterogeneous Graph

Qiang Zishan　Gu Yijun
(College of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China)

Abstract: [**Objective**] This paper proposes a social media rumor detection model based on the multimodal heterogeneous graph, aiming to verify the correlation between different rumor modalities and improve the accuracy of rumor detection. [**Methods**] First, we retrieved multimodal posts from social platforms. Then, we extracted feature representations of texts, pictures, and user attributes through preprocessing. Third, we constructed a heterogeneous graph based on the correlation between texts, pictures, and users. Fourth, we extracted the embeddings of text-type nodes according to their specified meta path. Finally, we input the embedding into the classifier to determine whether or not it is a rumor. [**Results**] We examined the proposed model with two open data sets. The accuracy of our model reached 91.3% and 93.8%, which were also higher than the baseline models. [**Limitations**] The three types of nodes from the sharing multimodal rumors will make the heterogeneous graph sparse. The proposed model is more suitable for small topic communities. [**Conclusions**] There is a correlation between different modalities of rumors, which helps the proposed model effectively detect multimodal rumors.

Keywords: Rumor Detection　Node Embedding　Multimodal　Heterogeneous Graph　Attention Mechanism