

融合元路径学习和胶囊网络的社交媒体 谣言检测方法



刘楠, 张凤荔*, 王瑞锦, 张志扬, 赖金山

(电子科技大学信息与软件工程学院 成都 610054)

【摘要】以源推特文本为研究对象, 深度挖掘推特正文内容的语义信息, 并强调谣言在具有异质性的社交网络传播过程中存在的结构特征, 以达到提升谣言检测效果的目的。采取基于 One-Hot Encoding 的词嵌入方法, 结合 Multi-head attention 机制实现推特正文内容初级语义特征的提取, 并进一步基于胶囊网络 (CapsNet) 构建内容胶囊 (content-capsule) 模块实现对正文内容深度语义特征的提取, 结合图卷积胶囊 (GCN-Capsule) 模块实现谣言在社交网络中传播结构特征的提取, 将两种胶囊向量采用一种动态路由机制进行融合, 进一步丰富输入特征, 之后输出源推特的分类结果, 进而实现源推特的谣言检测。实验结果显示, 该模型对谣言识别的正确率达 93.6%。

关键词 胶囊网络; 数据挖掘; 元路径学习; 谣言检测

中图分类号 TP181 文献标志码 A doi:10.12178/1001-0548.2021219

Rumor Detection Based on Meta-Path Learning and Capsule Network

LIU Nan, ZHANG Fengli*, WANG Ruijin, ZHANG Zhiyang, and LAI Jinshan

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Aiming at taking the source Twitter texts as the research object, this paper deeply explores semantic information of Twitter body content and emphasizes structural features of heterogeneous rumor spreading social networks, so as to improve rumor detection effect. This paper combines one-hot encoding word embedding method and multi-head attention mechanism to extract primary semantic feature of source Twitter text content. Furthermore, the content-capsule module is constructed based on CapsNet to extract the deeper semantic features of text content, and the structure features of rumor propagation in social networks are extracted by combining with GCN-capsule module. In order to further enrich the input, two kinds of capsule vectors are fused with a dynamic routing mechanism. And then the classification results of source tweets are output, and source tweets rumors detection is finished. Experimental results show that the accuracy of the model proposed in this paper reaches 93.6%.

Key words CapsNet; data mining; Meta-path learning; rumor detection

社交媒体的快速发展为人们提供了获取、处理和共享信息的便捷平台, 促进了海量信息的传播和扩散。其中, 谣言的传播会带来不可逆、破坏性强、影响极广的负面影响^[1-2]。谣言的自动识别有助于早期预防、减少损失, 因此, 谣言检测技术^[3]应运而生。

早期的谣言检测方法主要采用监督学习^[4-6], 利用特征工程从文本内容^[7-9]、用户信息^[7]和传播模式^[10-13]中提取可区别的特征, 如传统机器学习模

型^[4]。随着深度学习模型的出现, 基于 RNN (recurrent neural network)、CNN(convolutional neural network) 和 AE(autoencoder) 的方法在特征提取上有所改进, 在情绪分析、机器翻译、文本分类等方面均取得了显著成果。文献 [10] 利用递归神经网络捕捉微博中谣言源帖及其转发帖的语义差异, 从而根据语义的变化进行下一个传播点的预测。这是首个引入深度神经网络捕获谣言在整个传播过程的潜在时序变化的研究。文献 [14] 基于树的递归神

收稿日期: 2021-08-10; 修回日期: 2021-12-21

基金项目: 国家自然科学基金 (61802033, 61472064, 61602096); 四川省区域创新合作项目 (2020YFQ0018); 四川省科技计划重点研发项目 (2021YFG0027, 2020YFG0475, 2018GZ0087, 2019YJ0543); 博士后基金 (2018M643453); 广东省国家重点实验室项目 (2017B03031413); 网络与数据安全四川省重点实验室开放课题 (NDSMS201606)

作者简介: 刘楠 (1995-), 女, 博士生, 主要从事谣言检测、舆论信息传播和图表示学习等方面的研究。

*通信作者: 张凤荔, E-mail: fzhang@uestc.edu.cn

经网络模型以捕获谣言在传播结构中的潜在语义信息特征。文献 [15] 使用一种变分自动编码器 (variational autoencoder, VAE) 获取帖子涵盖的文本特征和图像特征, 以确定该帖子是否为谣言。文献 [16] 将源帖的传播路径建模为一个多元时间序列, 利用 RNN 和 CNN 捕捉相关帖子参与者的用户特征沿传播路径的变化。上述模型多采用单一的文本内容检测模型, 忽略了社会网络结构信息。此外, 部分方法仅从信息个体角度进行考虑, 忽略了社交网络信息之间所存在的结构相关性。如果同一用户发布或转发了多个帖子, 则可以连接这些帖子。这样的关联可在连接的实例之间共享知识, 帮助彼此检测以提高性能。

近年来, GCN(graph convolutional networks) 从信息结构化的度检测谣言和假新闻。如文献 [17] 建立了一个深度扩散网络模型, 学习新闻文章、创建者和主题的融合表示, 挖掘社交网络的结构性特征。此外, 信息在社交网络上的传播过程所构成的图网络具有异质性, 从异构图的构建与分析角度可有效提高虚假信息检测模型的性能。如文献 [18] 通过从社交网络上的帖子、评论和相关用户构建的异构图中, 捕获图结构中的语义信息。虽然目前 GCN 和异构图网络在谣言检测方面性能表现良好, 但仍存在部分问题。首先, GCN 针对图中每个学习到的节点表示采用的是标量式编码, 需要逐一编码节点包含的所有属性, 当数据量过大时, 效率会大大降低。其次, 现有异构图网络着重强调谣言传播过程的文本内容语义变化, 忽略了用户之间的社交关系, 在一定程度上对检测模型的性能进行了限制。此外, 目前已有的谣言检测模型, 对社交网络的异构性研究缺乏用户之间社交关系的考虑, 而在真实的社交网络中, 社交关系是一个较大的影响因素。

针对上述问题, 本文提出了一种融合元路径学习和胶囊网络的社交媒体谣言检测方法 (rumor detection based on meta-path learning and capsule network, CNMLRD), 联合图嵌入和文本内容语义嵌入两方面对谣言在社交网络上的特征学习进行表示, 利用胶囊网络以矢量编码增强学习到的特征。该方法首次将胶囊网络矢量编码模型用于谣言早期检测中, 针对传统神经网络本身特性导致的检测模型编码效率低下的问题提出了一种新的解决思路。此外, 该方法涉及基于元路径学习的异构图分解模型, 实现了对用户潜在社交关系及

图结构的全局语义信息挖掘, 不仅提高了谣言早期检测模型的效率和精度, 并在一定程度上增强了模型的可解释性。

1 问题描述

为了准确描述面向社交网络的谣言检测问题, 对以下概念进行定义。

定义 1 社交媒体关系: 定义为社交传播实体与其对应的传播内容的集合 $S = \{e_1, e_2, \dots, e_s\}$, 其中 e_s 指第 s 个社交传播实体和其所传播的内容。

定义 2 社交传播实体: 在社交网络中参与了发表、转发和评论帖子等行为的用户个体, 用集合 $U = \{u_1, u_2, \dots, u_n\}$ 表示, 其中 u_n 表示第 n 个用户实体。

定义 3 传播内容: 用户所发表的帖子, 并且这些帖子至少会有不少于 1 次的转发和评论, 用集合 $T = \{t_1, t_2, \dots, t_m\}$ 表示, 其中 t_m 表示第 m 个帖子实体。

定义 4 异常传播实体: 以用户是否发起或转发过一条谣言帖子作为评判标准, 将用户分为正常用户和异常用户。

根据以上概念, 可以构建基础社交信息传播网络, 并利用异常传播实体的评判标准将基础社交信息传播网络转化为异构图网络, 然后采用图神经网络模型得到每一个传播实体与传播内容的低维向量特征表示, 谣言的潜在特征可以结合信息在社交媒体网络上的结构特征以及信息内容的文本语义特征得到。

综上, 本文将谣言检测任务看作二分类问题, 目标是训练一个模型 $f(\cdot)$ 以预测一个给定信息的标签 $f(t_i)$, 若 $f(t_i) = 1$, 则 t_i 为非谣言; $f(t_i) = 0$, 则 t_i 为谣言。

2 方法介绍

2.1 总体框架和流程

整个模型框架如图 1 所示, 主要包含 4 个模块: 用户-帖子异构图构建模块, 图节点结构特征胶囊模块、文本内容特征胶囊模块以及特征融合模块。其中, 用户-帖子异构图构建模块主要是完成对原始数据集的清理及预处理, 再依据应用场景的需求构建适当的异构图; 图节点结构特征胶囊嵌入模块主要是利用图卷积胶囊网络将异构图中节点的特征以胶囊形式嵌入得到图节点胶囊的表示, 充分保留节点的属性; 文本内容特征胶囊嵌入模块主要是利用内容胶囊网络将帖子文本内容的语义特征以胶囊形式进行嵌入得到帖子文本内容胶囊的表

示, 充分挖掘文本的语义特征; 特征融合模块主要是将帖子在社交网络中的图节点胶囊表示与其文本

内容胶囊表示进行融合, 并在此基础上实现对谣言的划分。

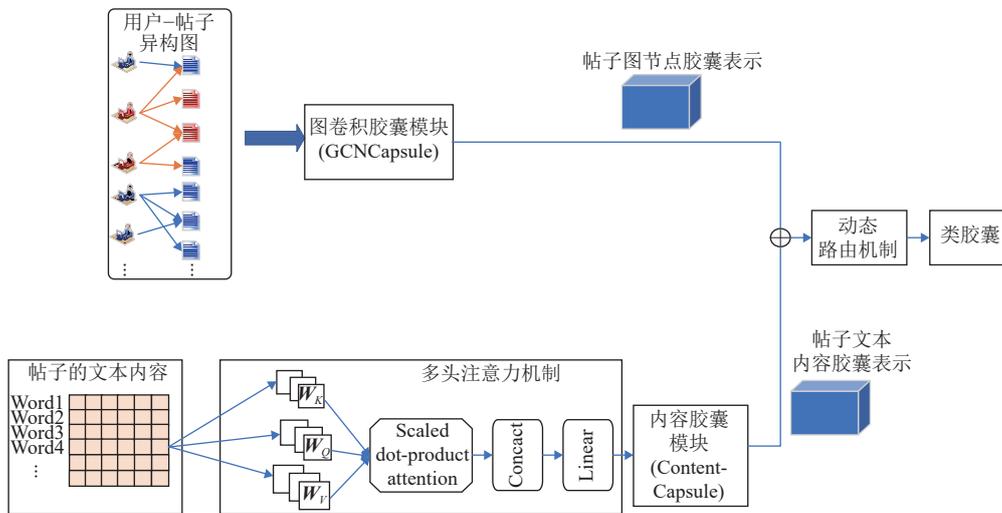


图 1 总体架构和流程

2.2 异构图构建

异构图构建模块的目的是对原始数据集进行预处理, 进一步对给定的数据按照其在实际社交网络中存在的点边关系, 提取对应的点集、边集。针对常见社交网络构建的用户-帖子异构图 G , 图 $G = (V, E)$, V 表示图中节点的集合, 包含所有的用户节点和帖子节点。 E 表示图中边的集合, 凡是用户发起、转发或者评论过的帖子, 该用户与帖子之间存在社交媒体关系, 这两者的节点之间存在一条边。

进一步考虑用户的异常状态, 以用户是否发起或转发过一条谣言帖子作为评判标准, 将用户分为正常用户和潜在威胁用户:

$$V = \{U_B, U_Y, T\} \quad (1)$$

式中, $U_B = \{u_{b_1}, u_{b_2}, \dots, u_{b_n}\}$ 表示社交网络潜在威胁用户的集合, u_{b_i} 表示潜在威胁用户; $U_Y = \{u_{y_1}, u_{y_2}, \dots, u_{y_n}\}$ 表示正常用户的集合, u_{y_i} 表示正常用户; $T = \{t_1, t_2, \dots, t_m\}$ 表示一个话题下的所有帖子的集合, t_i 表示帖子, t_1 表示源帖。节点之间的连接关系可以通过邻接矩阵 $A = \{0, 1\}^{|V| \times |V|}$ 表示获得。

2.3 图节点结构特征胶囊嵌入

此模块在 CapsGNN 网络^[19] 基础上, 根据本文的实际应用场景做出相应的调整, 实现异构图节点的胶囊嵌入。在原有的 CapsGNN 网络中, 只考虑了图结构信息, 而在本文中需要将文本内容信息进行联合, 因此在本文的图嵌入部分的基础上实现对图节点胶囊的嵌入可以简化模型的复杂度。此模

块主要包括 GCN 层、Capsule 层和图节点胶囊构建层 3 部分, 如图 2 所示。

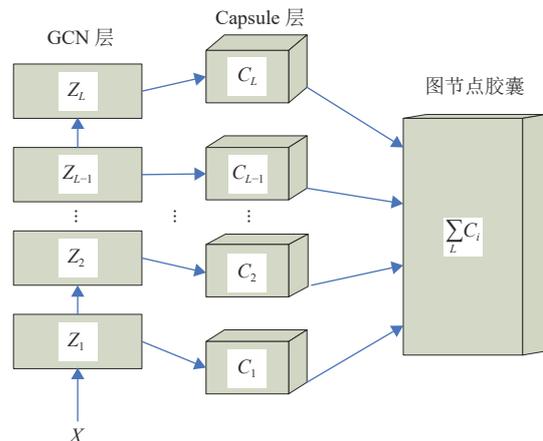


图 2 图卷积胶囊 (GCNCapsule) 模块图

其中, GCN 层的目的是从原始图数据中提取图结构的低级特征, 每一个 Z_i 表示一个 GCN 模块, 可进行 h 次图卷积操作, h 属于超参可设置。GCN 的应用对象是图结构数据, 利用图结构上节点的多阶邻域节点所包含的信息对该节点特征进行学习。假设图 $G = (V, E)$, $V = \{v_1, v_2, \dots, v_N\}$ 表示节点集合, $E = \{e_1, e_2, \dots, e_m\}$ 表示边集合, 节点初始特征表示为 $X \in \mathbb{R}^{N \times d_0}$, 其中 d_0 示节点的特征维度。

GCN 采用的是逐层叠加的学习方式, 所有节点同步更新。因此在 GCN 每层的学习过程中, 卷积运算应用于每个节点及其邻居, 并通过激活函数计算每个节点的新表示。其过程可表示为:

$$Z^l = \sigma(\tilde{D}^{-\frac{1}{2}} \mathbf{A} \tilde{D}^{-\frac{1}{2}} Z^{l-1} \mathbf{W}^{l-1}) \quad (2)$$

式中, $Z^{l-1} \in \mathbb{R}^{d \times d'}$ 表示第 $l-1$ 层提取到的节点特征, d 表节点的维度; $Z^0 = X$ 为输入的初始特征; \mathbf{A} 表示节点的邻接矩阵; \tilde{D} 表示邻接矩阵对应的度矩阵, 可由式 (3) 表示; $\mathbf{W}^{l-1} \in \mathbb{R}^{d \times d'}$ 为第 $l-1$ 层的权重矩阵; σ 是非线性激活函数。

$$\tilde{D} = \begin{pmatrix} \sum_{i=0}^{V-1} A_{0,i} & 0 & \dots & 0 \\ 0 & \sum_{i=0}^{V-1} A_{1,i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sum_{i=0}^{V-1} A_{v-1,i} \end{pmatrix} \quad (3)$$

Capsule 层的主要目的是将图中所包含的节点特征表示转化为向量形式, 有效保留图节点的完整属性。通过将每一个 GCN 模块提取的节点特征进行封装来实现。即原 GCN 层的 GCN 模块之间不再是简单的直接卷积, 而是在原有的卷积上进一步捕获 GCN 模块内部特征通道之间存在的潜在关系, 其过程可表示为:

$$Z_j^l = \sigma\left(\sum_i \tilde{D}^{-\frac{1}{2}} \mathbf{A} \tilde{D}^{-\frac{1}{2}} Z_i^{l-1} \mathbf{W}_{ij}^{l-1}\right) \quad (4)$$

式中, $Z_j^l \in \mathbb{R}^{N \times d'}$, $Z^0 \in X$; $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$; $\tilde{D} = \sum_j \tilde{\mathbf{A}}_{ij}$; $\mathbf{W}_{ij}^{l-1} \in \mathbb{R}^{d \times d'}$ 是一个可训练得到的权重矩阵, 它表示的是从第 $l-1$ 层的第 i 个特征通道在第 l 层的第 j 个特征通道中所占比重; σ 是非线性激活函数, 选用的是 \tanh 函数。

最后, 在图节点胶囊层, 通过累加的方式将每一个 GCN 模块得到的 capsule 叠加起来得到最后的图节点胶囊, 即完成了对图节点胶囊的嵌入表示。

2.4 文本内容特征胶囊嵌入

文本内容特征胶囊嵌入的主要目的是为了将每个帖子的文本表示用胶囊的形式嵌入, 本文采用多头注意力机制和胶囊网络来实现, 具体模型如图 1 下半部分所示。

为了对文本特征进行有效嵌入, 首先利用文本截断的方式对文本进行预处理, 设定处理后的文本为固定长度 L 。若文本长度大于 L , 则以 L 长度截断文本, 若长度小于 L , 则用 0 填充, 帖子 t_i 的文本内容表达式为:

$$\mathbf{X}_{1:L}^i = [x_1^i, x_2^i, \dots, x_L^i] \quad (5)$$

式中, $x_j^i \in \mathbb{R}^d$ 为帖子 t_i 中第 j 个单词的词嵌入表示, 采用热编码方式实现, $\mathbf{X}_{1:L}^i \in \mathbb{R}^{L \times d}$ 。

然后, 使用多头注意力机制来细化词的嵌入, 利用每一个头捕获某一方面的潜在关系, 来捕获词之间的依赖关系。多头注意力机制构建原理的核心是计算三大矩阵集合, 分别是索引矩阵集合 Q 、关键词矩阵集合 K 和值矩阵集合 V 。以本文采用的词嵌入 $\mathbf{X}_{1:L}^i \in \mathbb{R}^{L \times d}$ 为例, 采用的是 h 头注意力, 则 $Q = \{Q_1, Q_2, \dots, Q_h\}$, $K = \{K_1, K_2, \dots, K_h\}$, $V = \{V_1, V_2, \dots, V_h\}$, 其中, $Q_j \in \mathbb{R}^{L \times \frac{d}{h}}$, 表示第 j 个头注意力所对应的索引矩阵, $K_j \in \mathbb{R}^{L \times \frac{d}{h}}$, 表示第 j 个头注意力所对应的关键词矩阵, $V_j \in \mathbb{R}^{L \times \frac{d}{h}}$, 表示第 j 个头注意力所对应的值矩阵, 则 Q_j, K_j, V_j 可以通过式 (6) 计算得到:

$$Q_j = \mathbf{X}_{1:L}^i \mathbf{W}_j^Q \quad K_j = \mathbf{X}_{1:L}^i \mathbf{W}_j^K \quad V_j = \mathbf{X}_{1:L}^i \mathbf{W}_j^V \quad (6)$$

式中, $\mathbf{W}_j^Q, \mathbf{W}_j^K, \mathbf{W}_j^V \in \mathbb{R}^{L \times \frac{d}{h}}$, 分别对应 Q, K, V 的权重矩阵, 可以通过训练学习得到。在此基础上, 结合注意力动态调整方法, 按照式 (7) 计算得到某一头的值 $Z_j \in \mathbb{R}^{L \times \frac{d}{h}}$:

$$Z_j = \text{softmax}\left(\frac{Q_j K_j}{\sqrt{d}}\right) V_j \quad (7)$$

最后将计算得到的 Z_1, Z_2, \dots, Z_h 进行拼接, 得到通过词向量嵌入方式的文本初级嵌入 $Z \in \mathbb{R}^{L \times d}$ 。

在文本初级嵌入的基础上, 利用一个内容胶囊 (ContentCapsule) 模块对文本的语义信息进行嵌入, 模块的具体内容如图 3 所示。

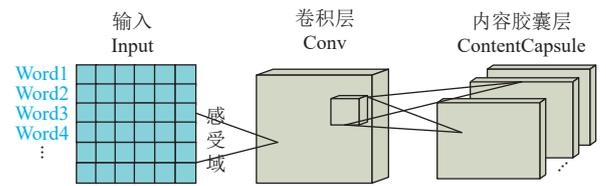


图3 内容胶囊 (ContentCapsule)

在 ContentCapsule 模块的卷积层, 将通过多头注意力机制得到的词向量嵌入作为输入, 通过卷积层对文本初级嵌入进行一次局部特征检测, 抽取低级特征, 这样网络就可以在层次较少的情况下尽可能感受多的信息。随后, 将卷积层得到的结果作为主胶囊层的输入构建相应的张量结构, 在这一层, 对上层输出执行 n 次不同权重的卷积操作, 然后将这 n 个卷积结果单元封装在一起组成 Capsule 神经元, 这样就得到文本内容的低级别特征的向量。

2.5 特征融合

为了提高模型的检测精确度,在分别得到帖子传播图结构和帖子文本内容的特征矢量以后,本模块将这两个表征矢量进行融合,进一步学习更深层次的特征表示,用于后续对谣言的分类及检测。此模块的结构如图 4 所示。

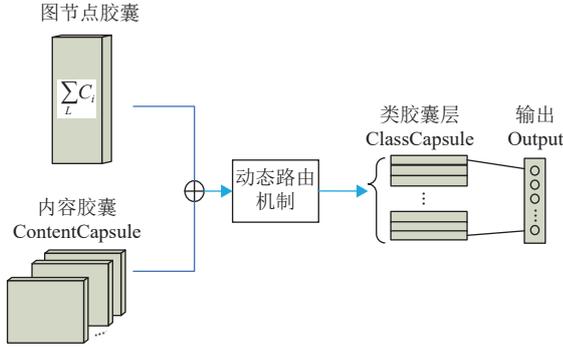


图 4 特征融合

此模块将帖子的图节点胶囊与内容胶囊进行融合,再在此基础上获取高级别特征的向量,由于这不是传统神经网络标量到标量,而是向量到向量,因此,需要用到一个动态路由机制进行传播和更新,得到类胶囊 (ClassCapsule),最终根据输出向量模长得类别概率向量,路由机制实现伪码如算法 1 所示。

算法 1 层级胶囊映射的路由选择

输入: 子级胶囊 S , 初始权重矩阵集 W , 可选迭代次数 t

输出: 父级胶囊 H

路由选择伪码:

1) 针对所有的 $l-1$ 层的胶囊 i , 计算其与对应 l 层的胶囊 j 之间的预测标量积 $u_{ji} = S_i^T W_{ij}$

2) 定义 r_{ij} 是 $l-1$ 层胶囊 i 连接到 l 层胶囊 j 的可能性, 初始值为 $r_{ij} \leftarrow 0$

3) 执行步骤 4)~7) t 次

4) 针对 $l-1$ 层的胶囊 i , 将 r_{ij} 转化为概率

$\tilde{r}_i: \tilde{r}_i \leftarrow \text{softmax}(r_{ij})$

5) 对 l 层的胶囊 j 进行加权求和得到 h_j :

$h_j \leftarrow \sum_i \tilde{r}_i u_{ji}$

6) 对 h_j 进行压缩得到 \tilde{h}_j : $\tilde{h}_j \leftarrow \text{squash}(h_j)$

7) 更新 r_{ij} , $r_{ij} \leftarrow r_{ij} + u_{ji} \tilde{h}_j$

8) 返回 \tilde{h}_j

结束

其中, squash 和 softmax 函数的具体计算如式

(8)~式 (9) 所示:

$$\tilde{h}_j = \text{squash}(h_j) = \frac{\|h_j\|^2}{1 + \|h_j\|^2} \frac{h_j}{\|h_j\|} \quad (8)$$

$$\tilde{r}_j = \text{softmax}(r_{ij}) = \frac{\exp(r_{ij})}{\sum_k \exp(r_{ik})} \quad (9)$$

3 实验分析

3.1 数据集

本文采用测评数据集 Twitter15 和 Twitter16^[12] 进行实验验证, 2 种数据集分别包含 1 490 和 818 条谣言源推文。数据集中的每一条源推特被标记为真实谣言 (true rumor, TR)、虚假谣言 (false rumor, FR)、未经证实的谣言 (unverified rumor, UR) 或非谣言 (non-rumor, NR)。由于原始数据集不包括用户配置文件信息, 调用 Twitter API3 抓取与源推文相关的所有用户的配置文件。数据集的其他细节如表 1 所示。

表 1 实验数据集统计详情

数据集	Twitter15	Twitter16
总谣言源推文数/条	1 490	818
真实谣言数/条	372	207
虚假谣言数/条	370	205
未经证实的谣言数/条	374	201
非谣言数/条	374	205
总源推文数/条	331 612	204 820
用户数/人	276 663	173 487

3.2 基线模型

针对本文所选用的评测数据集, 为验证本文所提出模型的有效性, 与下列 7 种谣言检测模型进行对比。

1) DTR^[20]: 基于决策树的模型, 通过正则表达式对从 Twitter 流中提取的集群进行排序以识别谣言。

2) DTC^[1]: 基于决策树模型, 利用特征工程提取的推文统计特征得到识别谣言的决策树分类器^[1]。

3) BU-RvN^[10]: 基于从叶子到根节点的传播树遍历方向的递归神经网络, 捕获扩散线索和内容语义。

4) TD-RvNN^[10]: 基于从根节点到叶子节点的传播树遍历方向的递归神经网络, 捕获传播线索和内容语义。

5) PPC^[15]: 由递归和卷积网络组成的传播路径分类器建模用户特征序列。

6) GLAN^[9]: 构建整体-局部的注意力网络捕获源帖及相关帖子传播结构的局部语义关联和全局结构关联。

7) HGAN^[17]: 通过构建异构图注意网络框架, 捕获源帖及相关帖子在传播结构中的全局语义关联和结构关联。

3.3 评估指标

本文所研究的谣言检测问题本质上是二分类问题, 本文选用基于分类的评价指标进行谣言检测性能评测。针对本文选用的数据集, 采用各类别判断的准确率 (accuracy, Acc) 和各类别的 F_1 值来评估模型的性能, 计算方式为:

$$\text{Acc} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (10)$$

$$F_1 = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (11)$$

式中, TP(true positive) 表示真实类别为正例, 预测类别也为正例的数量; FP(false positive) 表示真实类别为负例, 预测类别为正例的数量; FN (false negative) 表示真实类别为正例, 预测类别为负例的数量; TN(true negative) 表示真实类别为负例, 预测类别为负例的数量。

3.4 实验结果

实验基于 PyTorch 框架实现, 使用 Adam 优化器, 初始学习率为 0.005, 在模型训练过程中逐渐降低。根据验证集上的性能选择最佳参数设置, 并在测试集中评估方法性能。初始化词向量设置为 300 维。模型训练的批量大小 mini batch 设置为 32。

如表 2 和表 3 所示, 本文方法在两个数据集上的性能优于其他所有基线。具体而言, 本文方法在这两个数据集上分别实现了 92.5% 和 93.6% 的分辨率, 比最佳基线分别提高了 1.4% 和 1.2%。虽然只有一个百分点, 但就数据呈指数级扩增的规模而言, 一个百分点带来的效应也是不可低估的, 这表明本文方法能够有效地捕获谣言文本内容的全局语义关系, 有助于谣言检测。

此外, 基于传统机器学习方法 (DTR 和 DTC) 的对比实验模型表现不佳, 深度学习方法 (如 BU-RvNN、TD-RvNN、PCC 和 GLAN) 比基于传统机器学习的方法有更好的性能, 这表明深度学习方法更容易捕获有效的特征用于谣言检测。此外, GLAN 在所有对比模型中表现最好, 因为它捕捉到谣言传播源推文的局部语义和全局结构信息, 而其

他基线未能捕捉到这部分信息。

表 2 Twitter15 数据集实验结果

模型	Acc	F_1			
		NR	FR	TR	UR
DTR	0.49	0.501	0.331	0.364	0.473
DTC	0.454	0.733	0.355	0.317	0.415
BU-RvNN	0.708	0.695	0.728	0.759	0.653
TD-RvNN	0.723	0.682	0.758	0.821	0.654
PPC	0.842	0.818	0.875	0.811	0.790
GLAN	0.890	0.936	0.908	0.897	0.817
HGAN	0.911	0.953	0.929	0.905	0.854
本文	0.925	0.962	0.936	0.910	0.875

表 3 Twitter16 数据集实验结果

模型	Acc	F_1			
		NR	FR	TR	UR
DTR	0.414	0.394	0.273	0.630	0.344
DTC	0.465	0.643	0.393	0.419	0.403
BU-RvNN	0.718	0.723	0.712	0.779	0.659
TD-RvNN	0.737	0.662	0.743	0.835	0.708
PPC]	0.863	0.843	0.898	0.820	0.837
GLAN	0.902	0.921	0.869	0.847	0.968
HGAN	0.924	0.935	0.913	0.947	0.899
本文	0.936	0.945	0.918	0.952	0.975

4 结束语

本文提出了一种融合元路径学习和胶囊网络的社交媒体谣言检测方法 (CNMLRD), 利用胶囊网络矢量编码的优势特性, 弥补由传统神经网络特性所导致的检测模型编码效率较低的缺陷; 利用元路径学习的异构图分解方法, 实现了对用户潜在在社交关系以及图结构的全局语义信息挖掘, 增强对检测模型的可解释性。实验表明, 在同等数据集的情况下, 该方法相比于其他方法, 检测结果的精确度有所提升, 这样的优势在数据集呈指数级增加的情况下更为凸显。未来研究将考虑加入视频、图像、声音等数据, 利用多模态解决社交媒体谣言检测问题, 并进一步提高方法的泛化性。

参考文献

- [1] ZHANG H L, ABDUL A M, LI X, et al. Misinformation in online social networks: Detect them all with a limited budget[J]. *ACM Trans Inf Syst*, 2016, 34(18): 1-24.
- [2] ALLCOTTLLCOTT H, GENTZKOW M. Social media and fake news in the 2016 election[J]. *Journal of Economic Perspectives*, 2017, 31(2): 211-236.

- [3] SUSHILA S, VAHIDA A. Source detection of rumor in social network: A review[J]. *Online Soc Networks Media*, 2019, 9: 30-42.
- [4] WU K, SONG Y, ZHU K. False rumors detection on Sina Weibo by propagation structures[C]//The 31st IEEE International Conference on Data Engineering. Seoul: IEEE Computer Society, 2015: 651-662.
- [5] MA J, GAO W, WEI Z Y, et al. Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne: ACM, 2015: 1751-1754.
- [6] KWON S, CHA M, JUNG K. Rumor detection over varying time windows[J]. *Plos One*, 2017, 12(1): e0168344.
- [7] QAZVINIAN V, ROSENGREN E, MEI Q Z, et al. Rumor has it: Identifying misinformation in microblogs[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh: ACL, 2011: 1589-1599.
- [8] POPAT K. Assessing the credibility of claims on the Web[C]//Proceedings of the 26th International Conference on World Wide Web Companion. Perth: ACM, 2017: 735-739.
- [9] YUAN C Y, MA Q W, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C]//2019 IEEE International Conference on Data Mining. Beijing: IEEE, 2019: 796-805.
- [10] 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. Melbourne: ACL, 2018: 1980-1989.
- [11] SAMPSON J, MORSTATTER F, WU L, et al. Leveraging the implicit structure within social media for emergent rumor detection[C]//Proceedings of the 25th ACM International Conference on Information and Knowledge Management. Indianapolis: ACM, 2016: 2377-2382.
- [12] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL, 2017: 708-717.
- [13] MA J, GAO W, KWAON S, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: IJCAI/AAAI Press, 2016: 3818-3824.
- [14] KHATTAR D, GOUD J S, GUPTA M, et al. Mvae: Multimodal variational autoencoder for fake news detection[C]//Proceeding of 2019 the World Wide Web Conference. New York: ACM, 2019: 2915-2921.
- [15] LIU Y, WU Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018: 354-361.
- [16] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. [2018-02-04]. <http://arxiv.org/abs/1710.10903>.
- [17] HUANG Q, YU J S, WU J, et al. Heterogeneous graph attention networks for early detection of rumors on Twitter[C]//2020 International Joint Conference on Neural Networks. Glasgow: IEEE, 2020: 1-8.
- [18] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web. Hyderabad: ACM, 2011: 675-684.
- [19] VERMA S, ZHANG Z L. Graph capsule convolutional neural networks[EB/OL]. (2018-05-21). <https://arxiv.org/abs/1805.08090>.
- [20] ZHAO Z, RESNICK P, MEI Q Z. Enquiring minds: Early detection of rumors in social media from enquiry posts[C]//Proceedings of the 24th International Conference on World Wide Web. Florence: ACM, 2015: 1395-1405.

编辑 蒋晓