

文章编号: 1003-0077(2021)10-0119-09

## 基于增强的双向树表示的推特谣言立场检测模型

杨利君, 滕 冲

(武汉大学 国家网络安全学院 空天信息安全与可信计算教育部重点实验室, 湖北 武汉 430040)

**摘 要:** 谣言立场检测任务是通过分析社交媒体平台上用户发表的评论, 判别他们对谣言所持的立场是支持、反对或其他。谣言立场检测有助于甄别谣言真假。现有的工作将社交对话数据建模为单向树结构, 仅考虑了对话树的局部语义和结构信息。针对这些不足, 该文提出了一种增强的双向树神经网络模型。首先, 设计了一种门控机制, 用于融合自底向上和自顶向下两个传播方向上的表示, 从而更有效地提取对话的全局上下文信息。其次, 在模型中引入了一个局部推理模块, 增强了谣言与回复之间的语义联系。在 RumourEval 2017 Twitter 语料集上的实验证明, 该文提出的模型在多分类评价指标  $\text{macro-}F_1$  上获得了 52.5% 的性能, 相较于基线中最好的模型提升了 1.6%, 尤其在检测最具挑战性的否定立场优势的实验上更为明显。

**关键词:** 谣言立场检测; 双向树神经网络; 深度学习

**中图分类号:** TP391

**文献标识码:** A

## Enhanced Bidirectional Tree Representation Based Rumor Stance Detection from Twitter Conversations

YANG Lijun, TENG Chong

(Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education,  
School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei 430040, China)

**Abstract:** Rumor stance detection is to determine if each user's stance on the rumor is supporting, denying or others by analyzing the posts on social media. Existing works models the conversations as unidirectional trees, focusing only on partial structure and semantic information of them. To address this issue, an enhanced bidirectional tree neural networks model is proposed. Firstly, a gate mechanism is designed to learn the representations jointly from the bottom-up and top-down propagation, which effectively extracted the global context of the conversation. Then, a local inference module is incorporated into the model to strengthen the semantic relations between the rumor and responsive posts. Results on the RumourEval 2017 Twitter dataset demonstrate that the proposed model achieves the best performance of 52.5% in macro-averaged  $F_1$  scores (1.6% improvements), especially good at detecting the denying stance which is the most challenging for stance detection.

**Keywords:** rumor stance detection; bidirectional tree neural networks; deep learning

## 0 引言

随着互联网的飞速发展, 未经证实的信息传播速度非常快, 尤其是在微博、Twitter 这样的社交媒体上, 越来越多的用户喜欢在网上浏览、转发热点信息, 并发表自己的观点, 从而使得谣言和虚假信息得以广泛传播<sup>[1]</sup>。谣言会引起公众恐慌和社会动荡,

而有效检测谣言和扼制谣言的传播有助于社会的安定和健康发展, 具有重要的现实意义<sup>[2]</sup>, 因此大量针对社交网络上的谣言自动检测研究开始出现<sup>[3-6]</sup>。一些学者指出, 从谣言所引发的讨论中挖掘出用户的立场, 有助于确定该谣言的真实性<sup>[7]</sup>。所以, 近年来, 谣言立场检测任务也得到了越来越多的关注。

谣言立场检测旨在利用自然语言处理技术, 分析用户发表的文本, 挖掘出其对谣言的支持、反对或

收稿日期: 2020-12-28 定稿日期: 2021-04-08

基金项目: 国家自然科学基金(61772378); 国家重点研发计划(2017YFC1200500); 教育部研究基金(18JZD015)

者中立的态度。Twitter 等社交媒体上的评论通常以嵌套的对话形式出现,具有时序性和结构性,而且文本也具有简短、用词不规范、表达灵活、常采用反讽和隐喻等修辞手法的特点,往往需要利用更广泛的语境识别有效特征,以便于推断文本立场<sup>[8]</sup>。已有的方法证明,利用对话中的传播信息有助于提升立场检测的效果<sup>[9]</sup>。然而,以往的模型仅考虑对话中的局部上下文信息,例如,交互(评论或回复)的推文、时间线上相邻的推文,而忽略了全局信息,而且传播方向也是单一的。目标推文不仅依赖于所回复的文本内容,其下方回复的评论对目标推文也有一定的影响,例如,支持或评论类型的回复能够增强目标推文的立场,而否定和询问类型的回复则会削弱目标推文的立场。另外,对话进行得越深入,目标推文与谣言,即源推文的距离就越远,很容易在长距离的传播过程中丢失源推文的重要语义信息,从而无法正确判断目标推文针对源推文谣言的立场。

针对以上问题,本文提出了一种增强的双向树表示模型 BiTreeInfer,该模型利用双向的树形网络改进了单方向的传播信息的提取,另外,引进了一个局部推理模块以解决源推文语义信息丢失的问题。本文的贡献如下:

(1) 利用社交媒体上对话的树形结构特征,将单向的传统树形长短时记忆神经网络(TreeLSTM)拓展成双向结构;

(2) 有效融合自底向上和自顶向下的两个传播方向上学习到的上下文信息,丰富了目标推文的语义和结构信息。

(3) 加强了目标推文与源推文之间的语义关系,从而更有效地提取目标推文对于源推文的立场。

## 1 相关工作

随着国内外语义评测比赛 NLPCC-2016 Task 4<sup>[10]</sup>、SemEval-2016 Task 6<sup>[11]</sup>、RumourEval 2017<sup>[12]</sup>、RumourEval 2019<sup>[13]</sup>等的陆续开展,社交媒体上的谣言立场检测研究吸引了越来越多的参与者。目前,国内外的立场检测方法主要分为基于特征工程的机器学习方法和基于神经网络的深度学习方法。

传统机器学习方法对立场分析的研究工作主要集中在特征工程的构造上。Mohammad 等<sup>[14]</sup>从推文中提取了字符级和单词级 N-grams 特征、序列标注特征(POS)、词向量特征和情感特征等,并且使用支持向量机作为分类器。郑海洋等<sup>[15]</sup>利用“情感词

+主题词”组合构建立场检测特征,进行立场分类。Zubiaga 等<sup>[16]</sup>创建了两类分类器——线性条件随机场(Linear-Chain CRF)和树形条件随机场(Tree CRF)来学习对话中的树形结构特征。为了进一步提升结果,部分模型还融合了多种机器学习算法,例如,莫雨洁等<sup>[17]</sup>融合基于词频统计的特征字向量、词向量等特征,使用加权平均的方法融合多个分类器,如支持向量机(SVM)、随机森林(RF)和梯度提升决策树(GBDT)模型分类结果。但上述方法都过于依赖领域知识,需要耗费大量的人力、物力在特征工程的构造上。

相比于过于依赖特征工程的机器学习方法,深度学习由于其具有自动学习文本特征的能力,而受到了广泛的关注和使用。例如,Vijayaraghavan 等<sup>[18]</sup>利用卷积神经网络(CNN)对推特立场检测数据集进行单词级和字符级的训练,最后融合两种模型进行立场分析。Dey 等<sup>[19]</sup>在长短时记忆网络(LSTM)模型基础上引入注意力机制,通过强化文本中关键词的语义表示,提升模型立场分类性能。Kumar 等<sup>[20]</sup>将卷积核引入 TreeLSTM 的记忆单元,从对话树中获取了更丰富的传播信号。王安君等<sup>[21]</sup>采用 Bert<sup>[22]</sup>来获取具有上下文语义信息的句向量,设计出基于 BERT-Condition-CNN 的中文微博立场检测方法。Yang 等<sup>[23]</sup>按对话结构顺序组织文本,输入到强大的预训练语言模型进行微调,以获取全局上下文信息。

已有的工作都未研究对话树中双向的传播信息所带来的影响。在基于方面的情感分析任务中,Luo 等<sup>[24]</sup>证明了利用双向树形网络结构提取方面术语的有效性。该任务从句子中提取了依存树的结构信息,而谣言立场检测任务将句子级别的结构特征提取扩展到了文档级别,需要从对话中提取有用的结构信息。针对这一特性,如何科学合理地设计出可以学习到丰富的语义和结构特征表示的模型是本文研究的重点。

## 2 推特谣言立场检测模型 BiTreeInfer

图 1 展示了 Twitter 上两组树形结构的对话示例,每组对话通常由一条源推文(可能为谣言)以及多条嵌套的回复推文所组成。每条推文都有其回复的对象或获得的评论,可以视作其父节点和子节点。句首的斜体表示目标推文在对话中所处的深度,句尾的粗体表示目标推文基于源推文的立场标

签。假设一组对话被定义为  $T = \{t_1, t_2, \dots, t_N\}$ ,  $N$  为对话中推文的数量,  $t_i$  表示对话中第  $i$  条推

文。立场标签集合为  $Y = \{\text{support}, \text{deny}, \text{query}, \text{comment}\}$ 。

#### Example 1:

[depth=0] A: We understand that there are two gunmen and up to a dozen hostages inside the cafe under siege at Sydney.. ISIS flags remain on display #7News [support]

[depth=1] A1: not ISIS flags [deny]

[depth=1] A2: sorry - how do you know its an ISIS flag? Can you actually confirm that? [query]

[depth=2] A21: no she cant cos its actually not [deny]

[depth=1] A3: More on situation at Martin Place in Sydney, AU LINK [comment]

[depth=1] A4: Have you actually confirmed its an ISIS flag or are you talking shit [query]

#### Example 2:

[depth=0] B: These are not timid colours; soldiers back guarding Tomb of Unknown Soldier after today's shooting #StandforCanada PICTURE [support]

[depth=1] B1: Apparently a hoax. Best to take Tweet down. [deny]

[depth=1] B2: This photo was taken this morning, before the shooting. [deny]

[depth=1] B3: I don't believe there are soldiers guarding this area right now. [deny]

[depth=2] B31: wondered as well. I've reached out to someone who would know just to confirm that. Hopefully get response soon. [comment]

[depth=3] B311: ok, thanks. [comment]

图 1 Twitter 对话示例

本文提出一种基于增强的双向树表示的立场检测模型 BiTreeInfer。模型的总体框架如图 2 所示, 该框架主要由四部分组成: 输入表示、双向树模型、局部推理模块和分类器。首先, 通过组合不同的特征初始化每条推文的输入表示; 接着, 使用双向树模型更新每个节点的表示; 然后, 利用局部推理模块增强节点的语义表示; 最后, 将每个节点的最终表示输入到分类器中, 获得预测的标签。

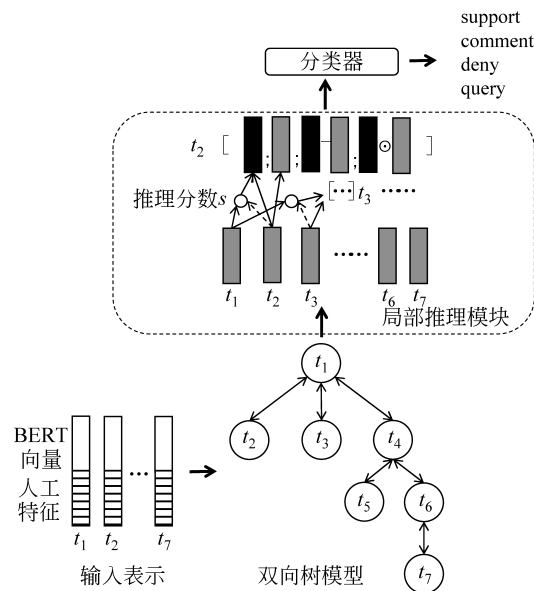


图 2 BiTreeInfer 模型框架

## 2.1 输入表示

文本的输入表示分为两个部分。首先, 对文本进行预处理, 将推文中的 url 链接以及 @ 的用户名

称全部替换成 “\$ url \$” 和 “\$ mention \$” 字符, 同时, 将所有的单词转换成小写形式。然后, 利用 BERT 模型获得预训练的词向量, 取句子中所有词的平均向量作为输入表示的一部分。另一部分, 则参考李峤等<sup>[25]</sup>的工作, 人工提取最相关的文本特征和对话特征, 如表 1 所示, 其中文本之间的相似性特征由余弦相似度 (Cosine Similarity) 计算。最后, 将 BERT 向量与人工提取的特征向量拼接, 获得目标推文  $t_i$  的初始化输入表示  $x_{t_i}$ 。

表 1 人工特征

类别	特征名称	特征说明
文本特征	否定词数量	推文中否定词汇的数量
	负面词数量	推文中负面词汇的数量
	问号	推文中是否存在问号
	感叹号	推文中是否存在感叹号
	url	推文中是否存在链接
	文本长度	文本字符数
对话特征	源推文	是否为源推文
	源推文相似性	目标推文与源推文之间的相似性
	回复推文相似性	目标推文与上一条回复对象之间的相似性
	其他推文相似性	目标推文与上下文(除自己以外的所有推文)之间的相似性
	推文深度	从源推文到目标推文经历的节点数量

## 2.2 双向树结构

双向树模型的整体结构如图 3 所示。树中每个节点的隐层状态由自底向上和自顶向下的两种传播方式分别进行更新。然后使用一种特定的交互方式对两个方向上的信息进行有效融合。以下将先对两种传播过程分别进行介绍,随后介绍所提出的一种双向扩展结构。

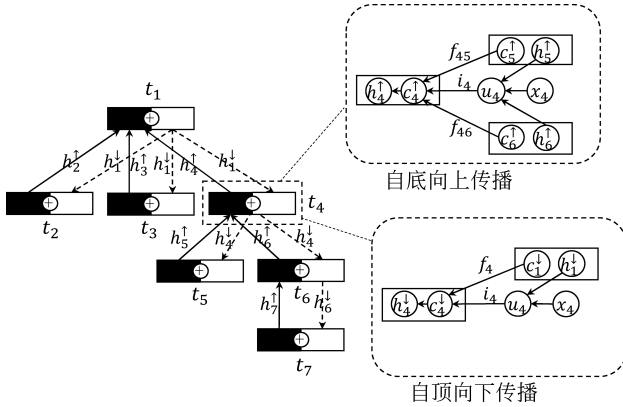


图 3 双向树模型

### 2.2.1 自底向上的树形结构

Tai 等<sup>[26]</sup>证明了树形长短时记忆神经网络(TreeLSTM)能有效捕捉到结构信息和长距离信息。其核心思想是通过递归访问底部的叶子节点到顶部的根节点,为每个子树生成特征向量。实验将它扩展到文档级别的谣言立场检测任务中,树形网络中的每一个节点代表一个句子,而不仅仅是一个单词。图 3 展示了自底上传播模块的细节。令  $C(j)$  表示节点  $j$  的所有子节点的集合, $j$  节点在自底向上树模型中更新的过程如式(1)~式(7)所示。

$$h_C^{\uparrow} = \sum_{k \in C(j)} h_k^{\uparrow} \quad (1)$$

$$i_j = \sigma(W^{\uparrow(i)} x_j + U^{\uparrow(i)} h_C^{\uparrow} + b^{\uparrow(i)}) \quad (2)$$

$$f_{jk} = \sigma(W^{\uparrow(f)} x_j + U^{\uparrow(f)} h_k^{\uparrow} + b^{\uparrow(f)}) \quad (3)$$

$$o_j = \sigma(W^{\uparrow(o)} x_j + U^{\uparrow(o)} h_C^{\uparrow} + b^{\uparrow(o)}) \quad (4)$$

$$u_j = \tanh(W^{\uparrow(u)} x_j + U^{\uparrow(u)} h_C^{\uparrow} + b^{\uparrow(u)}) \quad (5)$$

$$c_j^{\uparrow} = i_j \odot u_j + \sum_{k \in C(j)} f_{jk} \odot c_k^{\uparrow} \quad (6)$$

$$h_j^{\uparrow} = o_j \odot \tanh(c_j^{\uparrow}) \quad (7)$$

其中,箭头“ $\uparrow$ ”表示自底向上的方向, $i, f, o, u$  分别代表输入门、遗忘门、输出门和更新门,  $W^{\uparrow(*)}$ ,  $U^{\uparrow(*)}$  为权重矩阵,  $b^{\uparrow(*)}$  为偏置向量,  $*$   $\in \{i, j, o, u\}$ 。  $x_j$  表示节点  $j$  的初始输入表示,  $c_j^{\uparrow}$  和  $c_k^{\uparrow}$  代表节点  $j$  的细胞状态和其子节点  $k$  的细胞状态,  $h_j^{\uparrow}$

和  $h_k^{\uparrow}$  代表  $j$  的隐层状态和其子节点  $k$  的隐层状态,  $h_C^{\uparrow}$  表示节点  $j$  所有子节点的隐层状态之和。 $\sigma$  表示 sigmoid 激活函数,  $\odot$  表示元素的乘积。与标准 LSTM 单元不同的是, TreeLSTM 单元中门向量和细胞状态的更新依赖于所有与之相关的子单元的状态,另外,相比于标准 LSTM 的单个遗忘门, TreeLSTM 拥有多个遗忘门  $f_{jk}$ , 分别对应当前单元的每个子单元  $k$ , 这使得 TreeLSTM 可以从不同的子节点中获取信息。

### 2.2.2 自顶向下的树形结构

通常目标推文都是基于前一条推文进行表态和回复的,对话间存在着自顶向下的时序关系,所以本文还建立了自顶向下的 TreeLSTM 模块,图 3 中这一模块的细节显示了信息是如何从父节点传播到子节点的。自顶向下的 TreeLSTM 核心思想在于从根节点到子节点递归地聚合特征,直到到达所有的叶子节点。与自底向上的传播方式不同的是,  $j$  节点的隐层状态  $h_j^{\downarrow}$  由其自身的输入表示  $x_j$  以及其父节点  $P(j)$ , 而非由子节点的隐层状态  $h_{P(j)}^{\downarrow}$  计算得到。假设非叶子节点的隐层状态能够同步地传递到子节点中,  $j$  节点在自顶向下树模型中更新的过程如式(8)~式(13)所示:

$$i_j = \sigma(W^{\downarrow(i)} x_j + U^{\downarrow(i)} h_{P(j)}^{\downarrow} + b^{\downarrow(i)}) \quad (8)$$

$$f_j = \sigma(W^{\downarrow(f)} x_j + U^{\downarrow(f)} h_{P(j)}^{\downarrow} + b^{\downarrow(f)}) \quad (9)$$

$$o_j = \sigma(W^{\downarrow(o)} x_j + U^{\downarrow(o)} h_{P(j)}^{\downarrow} + b^{\downarrow(o)}) \quad (10)$$

$$u_j = \tanh(W^{\downarrow(u)} x_j + U^{\downarrow(u)} h_{P(j)}^{\downarrow} + b^{\downarrow(u)}) \quad (11)$$

$$c_j^{\downarrow} = i_j \odot u_j + f_j \odot c_{P(j)}^{\downarrow} \quad (12)$$

$$h_j^{\downarrow} = o_j \odot \tanh(c_j^{\downarrow}) \quad (13)$$

这与标准 LSTM 单元更新类似,只是将上一步的隐层状态替换成了父节点的隐层状态。

### 2.2.3 双向树的扩展结构

由于单向网络结构在捕捉全局上下文时具有局限性,本文提出了一种双向的 TreeLSTM 扩展结构,以此来更好地融合两个传播方向上的信息。推文  $t_i$  最终的隐层状态输出  $h_{t_i}$  由式(14)计算得到:

$$h_{t_i} = g(h_{t_i}^{\uparrow}, h_{t_i}^{\downarrow}) \quad (14)$$

其中,  $g(\cdot)$  表示结合自底向上的隐层状态  $h_{t_i}^{\uparrow}$  和自顶向下的隐层状态  $h_{t_i}^{\downarrow}$  的交互作用函数。为了更好地对特征进行融合,本文尝试了不同的交互方式,如表 2 所示。相加和拼接是不包含任何额外参数的最直接的交互方式,而多层感知机使用了 ReLU 激活函数以及两个方向上单独的权重矩阵  $W^{\uparrow}$  和  $W^{\downarrow}$  来建立交互关系。本文还设计了一个门



控机制,将初始输入表示  $x_{t_i}$  进行线性变换并激活得到控制变量  $T(x_{t_i})$ , 以此从两个方向上选取不同权重的信息。

表 2 交互作用函数列表

函数名称	函数公式
相加 (Addition)	$h_{t_i}^{\uparrow} + h_{t_i}^{\downarrow}$
拼接 (Concatenation)	$h_{t_i}^{\uparrow} \oplus h_{t_i}^{\downarrow}$
多层感知机 (MLP)	$\text{ReLU}(W^{\uparrow} h_{t_i}^{\uparrow} + W^{\downarrow} h_{t_i}^{\downarrow})$
门 (Gated)	$h_{t_i}^{\downarrow} \cdot T(x_{t_i}) + h_{t_i}^{\uparrow} \cdot (1 - T(x_{t_i}))$ $T(x_{t_i}) = \sigma(Wx_{t_i} + b)$

### 2.3 局部推理

尽管双向树模型能够使得目标节点包含丰富的全局上下文信息,但在深度很大的树中,叶子节点离源节点越远,其关于源节点的信息越模糊,而立场检测任务的目标在于判断目标推文基于源推文谣言的立场。所以,本文引入了一个局部推理模块以增强目标推文与源推文之间的语义关系。源推文的隐层状态  $h_{t_s}$  和目标推文的隐藏状态  $h_{t_i}$  之间的推理分数  $s(i)$  由端对端的双线性模型计算得出,如式(15)所示。

$$s(i) = h_{t_s}^T W_s h_{t_i} \quad (15)$$

$W_s$  是训练得到的权重矩阵,  $h_{t_s}^T$  是  $h_{t_s}$  的转置。接着将对话中所有目标—源推文对的推理分数进行 softmax, 得到更新后的隐层状态  $\tilde{h}_{t_i}$ , 如式(16)所示。

$$\tilde{h}_{t_i} = \frac{\exp(s(i))}{\sum_{j=1}^N \exp(s(j))} h_{t_i} \quad (16)$$

参考 Chen 等<sup>[27]</sup> 在自然语言推理任务中的做法,利用  $\langle h_{t_i}, \tilde{h}_{t_i} \rangle$  间的差和积与原向量拼接的方式进一步增强目标向量的语义表示,如式(17)所示。

$$\bar{h}_{t_i} = [h_{t_i}; \tilde{h}_{t_i}; h_{t_i} - \tilde{h}_{t_i}; h_{t_i} \odot \tilde{h}_{t_i}] \quad (17)$$

### 2.4 分类器

目标节点通过双向树模型获得了全局信息,通过局部推理模块获得了局部关系信息,将增强的双向树表示  $\bar{h}$  输入到多层前向神经网络 (Feed Forward Neural Networks, FFNN) 后进行线性变换,得到最终预测的类别概率  $P(\hat{y})$  和标签结果  $\hat{y}$ , 如式(18) ~ 式(19) 所示。

$$P(\hat{y}) = \text{softmax}(W \cdot \text{FFNN}(\bar{h}) + b) \quad (18)$$

$$\hat{y} = \text{argmax}(P(\hat{y})) \quad (19)$$

## 3 实验设置

### 3.1 数据集

本文使用的数据集来自于 RumourEval 2017 评测任务,该数据集从 Twitter 上爬取了 325 组对话集合,为 5 568 条推文标注了立场标签。这些表达舆论的推文围绕 9 个不同的突发新闻事件展开,例如巴黎枪击事件、德国之翼飞机坠毁事件等。数据集被划分成了训练集、开发集、测试集。推文的分布与统计结果如表 3 所示,S 表示支持,D 表示否定,Q 表示询问,C 表示评论。可以看到,这是一个不平衡的数据集,评论类型的标签数量远大于其他标签数量,这对模型的训练来说是一个挑战。

表 3 数据集的统计信息

	S	D	Q	C
训练集	841	333	330	2 734
开发集	69	11	28	173
测试集	94	71	106	778
总计	1 004	415	464	3 685

### 3.2 训练

本文使用多分类交叉熵作为损失函数,定义如式(20)所示。

$$\text{loss} = - \sum_{i=1}^N \sum_{c=1}^{|C|} P(y_i^c) \log P(\hat{y}_i^c) \quad (20)$$

其中,  $|C|$  代表类别集合,  $P(y_i^c)$  和  $P(\hat{y}_i^c)$  分别为第  $i$  条推文在  $c$  类别上的标准概率和预测概率。

实验表明,模型训练迭代次数达到 50 轮时,网络的交叉熵损失值基本保持不变。所以最大迭代次数设置为 50,将一组对话作为一个批次 (batch) 输入模型进行训练。实验使用 Adam<sup>[28]</sup> 更新参数,初始学习率设置为 0.001,每 10 轮下降 0.1% 进行更新。加载的预训练词向量为 768 维的 BERT 模型 (BERT-base-uncased)。自顶向下和自底向上树形神经网络的隐层状态都设置为 64 维,交互作用函数选择了门机制,分类器中的前向神经网络有 2 层,每层 64 维。实验搭建于 PyTorch 框架。

### 3.3 基线

**SVM**<sup>[29]</sup>: 该模型提取单个推文的情感特征和对话特征输入到线性支持向量机中, 得到分类结果。

**BranchLSTM**<sup>[30]</sup>: 该模型将树形对话中处于同一分支的推文组织成语序序列, 输入到 LSTM 网络中, 推文的输入特征由 Google News Word2Vec 词向量和人工特征组合而成。

**Temporal ATT**<sup>[31]</sup>: 该模型根据对话中每条推文发表的时间线, 使用注意力机制学习相邻推文的上下文权重信息, 丰富当前推文的语义特征。

**Conversational-GCN**<sup>[32]</sup>: 该模型利用图卷积神经网络将树形对话中推文之间的关系进行建模, 以更新推文节点的特征。

**Hierachical Transformer**<sup>[33]</sup>: 该模型将一段对话分解成几组子对话, 分别输入到预训练 BERT 模型中, 以获得子对话集合中推文的交互信息和局部上下文信息, 然后将它们输入到 Transformer 层进行融合, 使每条推文获得全局交互信息和上下文信息。

## 4 实验结果

### 4.1 结果比较

数据集标签分布的不平衡使得准确率不足以衡量一个模型的综合表现, 所以本文使用宏观平均  $F_1$  值 (macro-averaged  $F_1$ )<sup>[34]</sup> 作为最终的评价指标, 对不同类别的  $F_1$  值求平均, 以下简称为 macro- $F_1$ 。

表 4 给出了不同模型的实验结果, 可以看到, 本文提出的 BiTreeInfer 模型在所有模型中表现最佳, 获得了 0.525 的 macro- $F_1$  值。表 4 还列出了分类评估结果, 由于数据集中否定类别的推文数量最少, 该类别的立场检测最具挑战性, 大部分模型的识别效果欠佳, 而 BiTreeInfer 模型在否定类别上获得了最高的 macro- $F_1$  值 0.283, 这对于该领域的研究具有很大的意义, 因为否定立场对于谣言的真实性判别提供了至关重要的线索。在模型分析模块中, 会对该类别效果提升的原因进行进一步探究。

表 4 不同模型的结果对比

模型	分类评估				总体评估	
	Support- $F_1$	Deny- $F_1$	Query- $F_1$	Comment- $F_1$	Macro- $F_1$	Accuracy
SVM	0.410	0.000	0.580	<b>0.880</b>	0.470	0.795
BranchLSTM	0.403	0.000	0.462	0.873	0.434	0.784
Temporal ATT	—	—	—	—	0.482	<b>0.820</b>
Conversational-GCN	0.311	0.194	<b>0.646</b>	0.847	0.499	0.751
Hierachical Transformer	<b>0.421</b>	0.255	0.520	0.841	0.509	0.763
BiTreeInfer	0.389	<b>0.283</b>	0.568	0.860	<b>0.525</b>	0.770

### 4.2 模型分析

#### 4.2.1 消融实验

表 5 所示的消融实验证明了不同模块在 Bi-TreeInfer 模型中的作用。将交互作用函数替换成相加、拼接和多层感知机后, macro- $F_1$  值分别下降了 3.5%、1.8%、2.8%, 证明了门机制在筛选、提取和融合两个方向上信息的有效性。去除局部推理模块后, macro- $F_1$  值下降了 3.2%, 这表明该模块对语义信息的增强有明显帮助。如果继续将预训练的 BERT 向量替换成 Google News Word2Vec 向量, macro- $F_1$  值会再次下降 1.8%。

表 5 消融实验

模型	Macro- $F_1$
BiTreeInfer	0.525
$g(\bullet) = \text{Addition}$	0.490
$g(\bullet) = \text{Concatenation}$	0.507
$g(\bullet) = \text{MLP}$	0.497
—局部推理模块	0.493
—BERT 向量	0.475

#### 4.2.2 双向树结构分析

为了进一步研究 BiTreeInfer 模型是否得益于其双向树结构, 实验对不同深度的推文进行了分析。

表 6 统计了测试集中推文深度的分布情况,观察到大部分推文集中在 1~3 的深度,最大的深度可达到 13,为了方便分析,将不低于 6 的深度的推文统计成了一组。

表 6 不同深度的推文统计

深度	总数	S	D	Q	C
0	28	26	2	0	0
1	704	61	60	81	502
2	128	3	6	7	112
3	60	2	1	5	52
4	41	0	0	3	38
5	27	1	0	1	25
6+	61	1	2	9	49

实验做了两个对比模型 BUTreeInfer 和 TDTreeInfer,除了树的传播方向以外,其他设置与 BiTreeInfer 相同,分别表示自底向上(bottom-up)和自顶向下(top-down)的树形结构。图 4 中的对比结果显示,不同模型有其擅长识别的深度,而能识别出更多小数据量类别中推文的模型获得的 macro- $F_1$  值更高,所以在 0 深度上,BUTreeInfer 模型的 macro- $F_1$  值最低,因为其无法检测出否定立场的推文,但在 1 深度上,它的综合识别能力有所提高。TDTreeInfer 在识别 0~3 深度上推文的效果尚佳,但推文位置加深后,它的识别能力稍显逊色。

总体来看,BiTreeInfer 在大部分位置上都取得了最佳的结果,而且在大于 5 的深度的推文集合中,仍然保持较高的检测水平,优于任何单向的网络结构。这说明双向树结构能够在长距离中捕捉到更多的语义和结构信息。

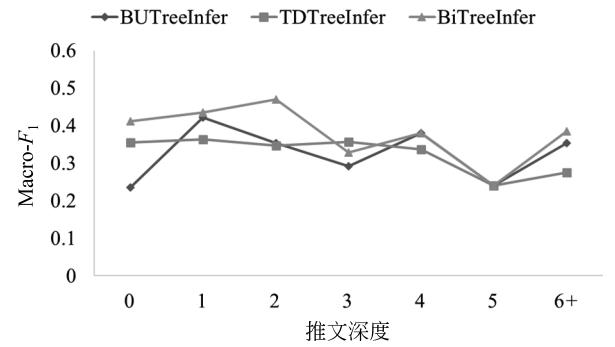


图 4 不同推文深度上的模型结果对比

4.2.3 局部推理模块分析

图 5 展示了对话中具有不同立场标签回复推文的推理分数。背景颜色代表它们与源推文之间的二元关系,颜色越深,说明两者之间关系越紧密。在第一个例子中,否定类型的推文与源推文在语义上相悖,所以它们之间的局部距离较远,推理分数较低。第二个例子中,支持类型的推文通过转发或者重复描述源推文的谣言显示出了与源推文语义关系上的一致性,所以获得了较高的推理分数。分析结果与 4.1 节中的实验结果一致,证明了局部推理模块有助于识别否定这类数据量少但语义关系明显的标签。

**Case 1:**  
Source: Update - AFP: an unnamed police source says the killing of the policewoman in #Paris on Thursday is linked to the #CharlieHebdo attack

Label	Inference score	Reply tweet
deny	6.9658e-06	Wow and you were so sure there was no connection despite having no info. Tell us the news not your view as news. Numpties.
comment	1.0741e-01	Of course it was going to be linked, something needs done before they take over the world !!
deny	3.3956e-05	I d rather wait for official confirmation than your scaremongering speculative social media sourcing news. #givefacts #unnamed
comment	1.0820e-03	At least it hasn t been forgotten... there's been so little coverage, no mention of finding the shooter that escaped!!

**Case 2:**  
Source:: That was very "SLICK" of the Police Chief to create a narrative that #MikeBrown was a "suspect" in a robbery! #Ferguson

Label	Inference score	Reply tweet
support	8.2440e-01	Yeah...I caught that too. "VERY SLICK" Seeing the framing of the lie starting now!
deny	2.9373e-03	We need b careful of jumping 2 conclusions. Waiting to see this videotape.

图 5 两组对话案例中回复推文的推理分数

## 5 结语

本文提出了一种端对端的立场检测框架 Bi-TreeInfer。该框架通过双向的树形神经网络对 Twitter 对话进行建模。节点的初始输入表示由预训练 BERT 模型获得的句向量特征和人工提取的特征组合而成。改进的树形递归神经网络能够有效地提取自底向上和自顶向下两个传播方向上的语义和结构信息,并且通过不同的交互作用函数融合了双向表示。实验证明,双向树结构有助于捕捉长距离信息,并且检测出对话深处的推文的立场。另外,模型还引入了一个局部推理模块以加强源推文和回复推文之间的语义联系,结果证明该方法能够检测出数据量最少、最具挑战性的否定立场。在未来的工作中,我们将进行立场检测和谣言判别的多任务实验,以进一步探究社交平台中谣言的传播规律。

## 参考文献

- [1] Mendoza M, Poblete B. Twitter under crisis: Can we trust what we RT? [C]//Proceedings of the 1st Workshop on Social Media Analytics, 2011: 71-79.
- [2] 李奥,但志平,董方敏,等.基于改进生成对抗网络的谣言检测方法[J].中文信息学报,2020,34(09): 78-88.
- [3] 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(03):198-204.
- [4] 王志宏,过弋.微博谣言事件自动检测研究[J].中文信息学报,2019,33(06):132-140.
- [5] Tacchini E, Ballarin G, Della Vedova M L, et al. Some like it Hoax: Automated fake news detection in social networks[C]//Proceedings of the 2nd Workshop on Data Science for Social Good, SoGood 2017. CEUR-WS, 2017: 1-15.
- [6] Vosoughi S, Roy D, Aral S. The spread of true and false news online [J]. Science, 2018, 359 (6380): 1146-1151.
- [7] Zubiaga A, Liakata M, Procter R, et al. Crowdsourcing the annotation of rumourous conversations in social media[C]//Proceedings of the 24th International Conference on World Wide Web, 2015: 347-353.
- [8] 刘玮,彭鑫,李超,等.立场分析研究综述[J].中文信息学报,2020,34(12):1-8.
- [9] Zubiaga A, Kochkina E, Liakata M, et al. Discourse-aware rumour stance classification in social media using sequential classifiers[J]. Information Processing & Management, 2018, 54(2): 273-290.
- [10] Xu R, Zhou Y, Wu D, et al. Overview of NLPCC shared task 4: Stance detection in Chinese microblogs [M]. Natural Language Understanding and Intelligent Applications. Springer, Cham, 2016: 907-916.
- [11] Mohammad S, Kiritchenko S, Sobhani P, et al. SemEval-2016 Task 6: Detecting stance in tweets[C]//Proceedings of the 10th International Workshop on Semantic Evaluation, 2016: 31-41.
- [12] Derczynski L, Bontcheva K, Liakata M, et al. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours[C]//Proceedings of the 11th International Workshop on Semantic Evaluation, 2017: 69-76.
- [13] Gorrell G, Kochkina E, Liakata M, et al. SemEval-2019 Task 7: RumourEval, determining rumour veracity and support for rumours[C]//Proceedings of the 13th International Workshop on Semantic Evaluation, 2019: 845-854.
- [14] Mohammad S M, Sobhani P, Kiritchenko S. Stance and sentiment in tweets[J]. ACM Transactions on Internet Technology (TOIT), 2017, 17(3): 1-23.
- [15] 郑海洋,高俊波,邱杰,等.基于词向量技术与主题词特征的微博立场检测[J].计算机系统应用,2018,27(09):118-123.
- [16] Zubiaga A, Kochkina E, Liakata M, et al. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations [C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 2438-2448.
- [17] 莫雨洁,金琴,吴慧敏.基于多文本特征融合的中文微博的立场检测[J].计算机工程与应用,2017,53(21): 77-84.
- [18] Vijayaraghavan P, Sysoev I, Vosoughi S, et al. DeepStance at SemEval 2016 Task 6: Detecting stance in tweets using character and word-level CNNs [C]//Proceedings of the 10th International Workshop on Semantic Evaluation, 2016: 413-419.
- [19] Dey K, Shrivastava R, Kaushik S. Topical stance detection for Twitter: A two-phase LSTM model using attention[C]//Proceedings of the European Conference on Information Retrieval. Springer, Cham, 2018: 529-536.
- [20] Kumar S, Carley K M. Tree LSTMs with convolution units to predict stance and rumor veracity in social media conversations [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 5047-5058.
- [21] 王安君,黄凯凯,陆黎明.基于 Bert-Condition-CNN 的中文微博立场检测[J].计算机系统应用,2019,28(11):45-53.
- [22] Devlin J, Chang M W, Lee K, et al. BERT: Pre-



- training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.
- [23] Yang R, Xie W, Liu C, et al. BLCU\_NLP at SemEval 2019 Task 7: An inference chain-based GPT model for rumour evaluation[C]//Proceedings of the 13th International Workshop on Semantic Evaluation, 2019: 1090-1096.
- [24] Luo H, Li T, Liu B, et al. Improving aspect term extraction with bidirectional dependency tree representation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(7): 1201-1212.
- [25] 李峤, 刘宇. 基于机器学习的推特谣言立场分析研究[J]. 电子设计工程, 2019, 27(21): 36-39.
- [26] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015: 1556-1566.
- [27] Chen Q, Zhu X, Ling Z H, et al. Enhanced LSTM for natural language inference[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 1657-1668.
- [28] Kingma D, Ba J. Adam: A method for stochastic optimization[C]//Proceedings of the International Conference on Learning Representations, 2015.
- [29] Pamungkas E W, Basile V, Patti V. Stance classification for rumour analysis in Twitter: Exploiting affective information and conversation structure[C]//Proceedings of the 2nd International Workshop on Rumours and Deception in Social Media. CEUR-WS, 2019, 2482: 1-7.
- [30] Kochkina E, Liakata M, Augenstein I. Turing at SemEval-2017 Task 8: Sequential approach to rumour stance classification with Branch-LSTM[C]//Proceedings of the 11th International Workshop on Semantic Evaluation, 2017: 475-480.
- [31] Veysseh A P B, Ebrahimi J, Dou D, et al. A temporal attentional model for rumor stance classification[C]//Proceedings of the ACM Conference on Information and Knowledge Management, 2017: 2335-2338.
- [32] Wei P, Xu N, Mao W. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 4789-4800.
- [33] Yu J, Jiang J, Khoo L M S, et al. Coupled hierarchical transformer for stance-aware rumor verification in social media conversations[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2020: 1392-1401.
- [34] Lewis D D, Schapire R E, Callan J P, et al. Training algorithms for linear text classifiers[C]//Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1996: 298-306.



杨利君(1996—), 硕士研究生, 主要研究领域为自然语言处理。  
E-mail: yanglijun@whu.edu.cn



滕冲(1974—), 通信作者, 副教授, 硕士生导师, 主要研究领域为大数据和人工智能、自然语言处理、语义分析。  
E-mail: tengchong@whu.edu.cn