

# 基于不确定性损失函数和任务层级注意力机制的多任务谣言检测研究<sup>\*</sup>

杨晗迅 周德群 马 静 罗永聪

(南京航空航天大学经济与管理学院 南京 211100)

**摘要:**【目的】通过引入不确定性损失函数和层级注意力机制,解决多任务谣言检测研究中主观设定主任务和辅助任务问题。【方法】融合谣言勘探、立场检测和谣言检测任务的领域信息,构建改进的任务层级注意力机制模型。同时,首次在多任务谣言检测研究中,引入同方差不确定性损失函数,替代传统损失函数。最后使用 PHEME 数据集,将改进模型与传统多分类模型进行对比。【结果】所提模型相比于目前最优模型,在 PHEME4 数据集中,Macro-F 值提升 4.2 个百分点;在 PHEME5 数据集中,Macro-F 值提升 7.6 个百分点。【局限】只在 PHEME 数据集进行实验测试,对于其他谣言检测数据集未测试。【结论】该模型在不划分主任务和辅助任务的情况下,仍可得到理想解。

**关键词:** 不确定性损失函数 多任务学习 谣言检测 注意力机制

**分类号:** TP393

**DOI:** 10.11925/infotech.2096-3467.2020.1216

**引用本文:** 杨晗迅,周德群,马静等. 基于不确定性损失函数和任务层级注意力机制的多任务谣言检测研究[J]. 数据分析与知识发现, 2021, 5(7): 101-110.(Yang Hanxun, Zhou Dequn, Ma Jing, et al. Detecting Rumors with Uncertain Loss and Task-level Attention Mechanism[J]. Data Analysis and Knowledge Discovery, 2021, 5(7): 101-110.)

## 1 引言

近年来,互联网科技发展迅速,以微博、知乎、论坛等为代表的社交媒体在信息传播方面扮演着越来越重要的角色。人们仅仅需要一台移动通信设备,便可在社交平台上发布和传播信息。每天数以亿计的信息在各个平台自由流动,为谣言的产生和传播提供前所未有的有利条件。而谣言通常被认定为达某种目的而被凭空捏造的信息。这类信息的传播可能会对社会生活、经济、政治等各个方面带来重大影响,因此,谣言识别一直是亟待解决的热点和难点问题。

问题。

Zubiaga 等<sup>[1]</sup>将谣言识别细分为以下 4 个子任务,如图 1 所示。谣言勘探:该任务从海量的文本中将无法辨别真伪的信息筛选出来,这些信息被视为疑似谣言文本,并在接下来的任务中做进一步判别;谣言追踪:一旦疑似谣言信息被确认之后,关联追踪其相关信息,包括但不限于评论信息和用户相关信息;立场检测:通过疑似谣言下的相关评论确定用户的态度;谣言检测:作为谣言识别的最终步骤,用于判断疑似谣言的真假。

通讯作者(Corresponding author): 马静(Ma Jing), ORCID: 0000-0001-8472-2518, E-mail: majing5525@126.com。

<sup>\*</sup>本文系南京航空航天大学前瞻性发展策略研究基金项目(项目编号: NW2020001)、国家社会科学基金重点项目(项目编号: 202ZDA092)和南京航空航天大学研究生创新基地(实验室)开放基金项目(项目编号: kfj20200901)的研究成果之一。

The work is supported by the Forward Development Strategy Research Fund Project (Grant No. NW2020001), the National Social Science Fund of China (Grant No. 202ZDA092), the Open Fund Project of Nanjing University of Aeronautics and Astronautics Graduate Innovation Base (Laboratory) (Grant No. kfj20200901).

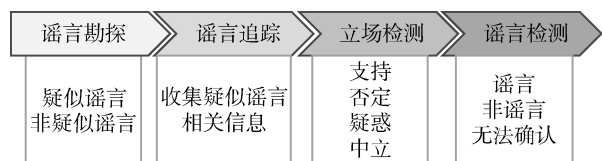


图1 谣言识别流程

Fig.1 Rumor Identification Process

目前,国内外对于谣言识别研究主要聚焦于单一的谣言检测任务,采用基于内容的谣言检测方法<sup>[2]</sup>,通过谣言传播过程的文本特征<sup>[3]</sup>、用户内容特征<sup>[4]</sup>以及随时间变化的趋势<sup>[5]</sup>进行谣言识别。但谣言识别研究中多个任务如何相互作用,如何将各个任务作为一个整体以有效地解决谣言问题是一个十分值得探究且有意义的课题。已有的多任务谣言检测<sup>[6]</sup>研究,对于各个任务的层次关系的界定,都是通过人工设定各个任务的权重,将任务划分为主任务和辅助任务。该方法存在以下不足:首先,各个任务间设定一个合理的权重关系,需要对于数据分布有着深刻的理解和大量的实验基础;其次,当模型更换不同的数据集,各个任务间的权重是否需要发生变化有待考究。因此,人工划分主任务和辅助任务是否具有必要是一个值得讨论的问题。

基于以上问题,本文通过引入同方差不确定性对损失函数加以优化,并使用层级注意力机制对模型加以改进,使模型在训练不同的数据集时自发寻找任务之间的最优权重,避免人工划分主任务和辅助任务,最终得到满意效果。本文以谣言检测为解决谣言问题的关键步骤,联合谣言勘探和立场检测任务,构建基于不确定性损失函数和任务层级注意力机制的多任务神经网络模型。

## 2 研究现状

近年来,国内外学者对于谣言检测已进行大量研究,但目前对于谣言检测相关研究主要以单任务为主。谣言检测早期阶段,研究者的注意力主要集中在特征提取方面。Sejeong等<sup>[7]</sup>对用户、结构、语言和时间等特征进行综合,并计算它们的相对强度,运用统计学习方法区分谣言的真假。Yang等<sup>[8]</sup>使用新浪微博数据集,除了文本的语义特征,还将基于客户端的软件特征和基于发送文本的位置特征纳入模型

考虑范围。Chang等<sup>[9]</sup>强调了发布谣言的用户特征,提出一种基于规则的方法,以识别Twitter上的政治谣言。人工选择的特征虽然可以增加谣言识别的准确率,但是人工标注可能费时费力,不利于实际检测。近年来深度学习的崛起在一定程度上解决了这个问题。Ma等<sup>[10]</sup>使用循环神经网络(Recurrent Neural Network, RNN)和其衍生模型进行谣言检测,得到效果优于手工制作特征的模型。Chen等<sup>[11]</sup>在RNN基础上,引入深度注意力机制,相比于其他模型取得了更好的效果。本文使用三个关联任务进行多任务学习,通过不同的任务学习不同领域的语义信息。该方法既有效学习其他特征,又避免了耗费过多人工成本,实现谣言的自动化检测。

同时,多任务学习在自然语言处理任务中取得了不俗的成绩。在主题-关系推理方面,Collobert等<sup>[12]</sup>使用卷积神经网络,构建出一个以语义角色标注(Semantic Role Labeling, SRL)为主任务,包含词性(Part-of-Speech, POS)标注、浅层句法分析(Chunking)、命名实体识别(Named Entity Recognition, NER)、语言模型(Language Model, LM)的多任务模型。结果表明各种组合的多任务模型效果均优于单模型效果。在文本分类方面,Liu等<sup>[13]</sup>针对文本多分类任务,提出三种不同的多任务贡献分类模型,探索不同任务之间的语义共享,提高分类的准确性。在机器翻译方面,Dong等<sup>[14]</sup>使用多任务学习解决多语言翻译及语料稀疏的问题。在谣言检测方面,Firat等<sup>[15]</sup>使用跨编码器-解码器共享的单一注意力机制,该模型在资源受限的情况下,相比单模型取得了更好效果;Kochkina等<sup>[6]</sup>提出基于长短期记忆网络(Long Short-Term Memory, LSTM)的多任务学习方法,构建分别包含谣言勘探、立场检测和谣言检测任务组合的模型,该实验证明所包含的任务越丰富,模型准确率越高;Ma等<sup>[16]</sup>在此基础上,针对立场检测和谣言检测子任务,提出ES-RNN模型,该模型对于每个子任务附加单独门控循环单元(Gated Recurrent Unit, GRU)层,增强提取的语义信息。Li等<sup>[17]</sup>在多任务模型中引入用户信用信息和注意力(Attention)机制,用于提高谣言检测准确性。

在以往的多任务谣言检测研究中,研究者主观为每个任务赋予固定权重,划分主任务和辅助任务。

然而,主任务和辅助任务划分是否合理有待研究。针对此问题,Sener等<sup>[18]</sup>将多任务模型转化为多目标优化问题,寻找帕累托最优解。Kendall等<sup>[19]</sup>在计算机视觉领域,基于贝叶斯深度学习的不确定性<sup>[20]</sup>,在联合语义分割、实例分割和深度回归的多任务模型上取得较好效果。本文对同方差不确定性损失函数<sup>[19]</sup>进行改进,使之适用于多任务谣言检测研究,并引入层级注意力机制<sup>[17]</sup>,构建基于谣言勘探、立场检测和谣言检测的多任务模型框架。

### 3 基于任务依赖不确定性和注意力机制的多任务谣言检测模型

目前已有的多任务谣言检测方法对于各个任务的层次划分都是通过模型损失函数表示,如公式(1)所示<sup>[19]</sup>。

$$Loss_{total} = \sum w_i Loss_i \quad (1)$$

其中, $Loss_{total}$ 为模型的损失函数,由各个任务损失函数加权求和确定; $w_i$ 为各个任务的权重; $Loss_i$ 为第*i*个任务的损失函数。

如何设定 $w_i$ ,更换不同数据集 $w_i$ 是否会发生变化,多任务谣言检测主观设定主任务和辅助任务是否合理,是一个值得探讨的问题。

本文首次在多任务谣言检测领域引入同方差不确定性,将简单加权求和的损失函数转化为不确定性损失函数。同时,模型在联合谣言勘探、立场检测和谣言检测的多任务检测模型中引入注意力机制,加强模型的语义理解能力。根据每个任务对谣言检测任务的重要度提供不同的权重,增强带有支持、反对等标签的立场信息和疑似谣言文本信息对于谣言检测任务的影响。

#### 3.1 同方差不确定性和损失函数优化

不确定性是深度学习中一个重要的研究方向<sup>[21]</sup>。不确定性可以划分为认知的不确定性和偶然的不确定性。其中,认知的不确定性(Epistemic Uncertainty, EU)用于描述模型本身带有的不确定性,由模型的参数决定;偶然的不确定(Aleatoric Uncertainty, AU)用于描述观测中固有噪声带来的影响。偶然的不确定性可以进一步划分为异方差不确定性和同方差不确定性。在多任务学习中,异方差不确定性根据不同的输入具有不同的差别。同方

差不确定性不依赖于输入数据,但是在模型中与不同任务有关。

在多任务学习领域,同方差不确定性表明了任务的相对置信度,反映了分类和回归问题的不确定性。因此,本文将同方差不确定性作为误差项对多任务学习中各个任务的权重进行优化。基于同方差不确定性中模型依赖于不同任务的特点,推导多任务谣言检测的损失函数。假设模型的语义输入为 $x$ ,参数矩阵为 $W$ ,任务间的误差项为 $\sigma^2$ ,模型输出为 $f^w(x)$ ,对于分类任务的概率估计,如公式(2)所示。

$$p(y|f^w(x), \sigma) = \text{Softmax}(\frac{1}{\sigma^2} f^w(x)) \quad (2)$$

其中, $y$ 代表多任务谣言检测中的任务; $p$ 为概率估计值。

进一步,对于公式(2)求极大似然估计,如公式(3)所示。

$$\log p(y = c|f^w(x), \sigma) = \frac{1}{\sigma^2} f_c^w(x) - \log(\sum_{c'} \exp(\frac{1}{\sigma^2} f_{c'}^w(x))) \quad (3)$$

其中, $c$ 为模型的真实标签; $f_c^w(x)$ 为 $f_c^w(x)$ 中第 $c'$ 元素的向量表示。

将公式(3)扩展到谣言检测多分类任务损失函数 $L(W, \sigma_1, \sigma_2, \sigma_3)$ ,如公式(4)所示。

$$\begin{aligned} L(W, \sigma_1, \sigma_2, \sigma_3) &= -\log p(y_1, y_2, y_3 = c|f^w(x)) \\ &= \text{Softmax}(y_1 = c_1; f^w(x), \sigma_1) \cdot \text{Softmax}(y_2 = c_2; f^w(x), \sigma_2) \cdot \text{Softmax}(y_3 = c_3; f^w(x), \sigma_3) \\ &= \sum_{i=1}^{K=3} \frac{1}{\sigma_i^2} f_c^w(x) - \log(\sum_{c'} \exp(\frac{1}{\sigma_i^2} f_{c'}^w(x))) \\ &= \sum_{i=1}^{K=3} (p(y_i|f^w(x), \sigma) - \frac{\log(\sum_{c'} \exp(\frac{1}{\sigma_i^2} f_{c'}^w(x)))}{(\log(\sum_{c'} \exp(f_{c'}^w(x))))^{\frac{1}{\sigma_i^2}}}) \\ &\approx \sum_{i=1}^{K=3} (L_i(W) + \log(\sigma_i)) \end{aligned} \quad (4)$$

其中, $\sigma_1, \sigma_2, \sigma_3$ 分别代表谣言勘探、立场检测、谣言检测任务的噪声; $L_i(W) = -\log(\text{softmax}(y_i, f^w(x)))$ 为分类任务*i*的交叉熵损失。为了简化计算,粗略地假设



$$\frac{1}{\sigma^2} \sum_{c_i} \exp\left(\frac{1}{\sigma^2} f_{c_i}^W(x)\right) \approx \left(\sum_{c_i} \exp\left(\frac{1}{\sigma^2} f_{c_i}^W(x)\right)\right)^{\frac{1}{\sigma_i^2}}。$$

最终,传统的线性加权求和损失函数转化为通过学习参数  $\log^{\sigma_i}$  的不确定性损失函数。模型摒弃了主观设定权重过程,因此不需要假定主任务和辅助任务。在训练过程中,模型会自动寻找最优的  $\log^{\sigma_i}$ ,求得最终的满意解。

### 3.2 基于注意力机制的多任务谣言检测模型

基于任务层级注意力机制的多任务谣言检测模型结构如图2所示。整个模型具有三个并行分支,依次为谣言勘探任务、谣言检测任务和立场检测任务。各个任务通过浅层的参数共享,补充学习各个领域的相关知识,提升模型泛化效果。

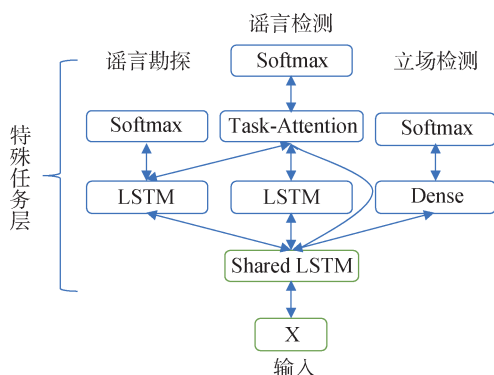


图2 多任务模型

Fig.2 Multitask Model

在模型训练时,首先每个任务接收相同的向量表示  $x$ ,再将  $x$  向量序列输入 LSTM<sup>[22]</sup> 共享层,每个任务根据本身的领域特征,通过 LSTM 层对共享层的输出向量进行进一步的语义编码,最后分别通过 Softmax 函数输出最大标签序列。其中,谣言检测作为目标任务,为接收更加丰富的语义信息,额外接入任务层级注意力机制,对重要信息给予更多关注。

#### (1) 基于LSTM的谣言勘探层

在文本分类任务中,循环神经网络(RNN)<sup>[23]</sup> 由于能接受文本序列的历史信息,成为解决文本分类任务的首选。相较于 RNN 模型,长短期记忆网络(LSTM)在隐含层引入了门机制,通过门状态控制语义传输状态,自动记忆重要信息,过滤次要信息,对于解决长序列训练任务产生的梯度弥散和梯度消

失问题有着重要的贡献。因此,本文谣言勘探任务采用 LSTM 模型,LSTM 网络结构内部转移方程如公式(5)-公式(10)所示。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (8)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (9)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (10)$$

其中,  $x_t$  为  $t$  时刻的输入信息;  $h_{t-1}$  为  $t-1$  时刻输出;  $\sigma$  表示 sigmoid 激活函数;  $\tanh$  表示双曲正切激活函数;  $\otimes$  表示点乘运算;  $f_t$ 、 $i_t$  和  $o_t$  分别是  $t$  时刻的遗忘门、输入门和输出门;  $W$  为权重矩阵,  $b$  为偏置矩阵;  $\tilde{C}_t$  为  $t$  时刻候选的输入信息;  $C_t$  表示  $t$  时刻单元状态;  $h_t$  表示  $t$  时刻的输出。

#### (2) 基于Branch-LSTM的立场检测层

Branch-LSTM 是 Kochkina 等<sup>[24]</sup> 为解决立场检测问题而设计的一种基于 LSTM 的改进算法。传统 LSTM 模型在进行立场检测时,通常只获取输入的文本信息,每个 LSTM 单元对词向量进行编码,LSTM 层学习单个评论文本的语义信息,最后通过分类函数得到概率标签。而 Branch-LSTM 可以有序地学习多个评论的语义信息,同时纳入各个评论的结构信息。由于模型同时接收多个评论的语义信息,立场检测任务从对单条评论的预测任务转化为同时对多个评论共同进行概率预测的分类任务。

Branch-LSTM 中每个 LSTM 单元接收的为整个句子的语义编码,而非单个词的语义编码。若干个 LSTM 单元串连,得到从谣言文本到用户评论的丰富语义信息,如图3所示。Branch-LSTM 在时间步  $t_i$  得到第  $i$  个句子的向量表示,模型同时记录第  $i$  个 LSTM 单元的输出概率,该输出通过若干个带有 ReLU 激活函数的全连接层得到更深层次的语义信息,如公式(11)所示。将更深层次的语义信息输入 50% 的 Dropout 层,增加模型的泛化效果,如公式(12)所示。最终通过 Softmax 函数进行分类,得到最大的标签序列,如公式(13)所示。

$$f_r = \max(0, W_r \cdot h_t + b_r) \quad (11)$$

$$f_d = f_r \times p \quad (12)$$

$$f_{out} = \frac{e^{f(w_d \cdot f_{di} + b_d)}}{\sum_{k=1}^K e^{f(w_d \cdot f_{dk} + b_d)}} \quad (13)$$

其中,  $W$  为权重矩阵;  $b$  为全连接层的偏置矩阵;  $h_t$  为 LSTM 第  $t$  个时间步的输出概率;  $p$  为 Dropout 的删除神经元的比例;  $e$  为无理常数;  $K$  为标签类别数;  $f_r$  为经过 ReLU 激活函数的输出;  $f_d$  为经过 Dropout 层的输出;  $f_{out}$  为经过 Softmax 激活函数的输出概率。

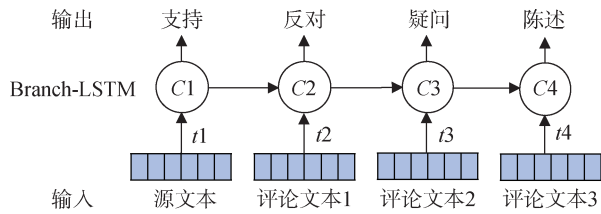


图3 Branch\_LSTM模型

Fig.3 Branch\_LSTM Model

### (3) 基于任务层级注意力机制的谣言检测层

在谣言识别任务中,不同的评论信息对于谣言的判定会有不同的影响。例如,若用户评论的立场大部分被识别为支持标签,疑似谣言最终可能会被确定为谣言;若大部分为反对标签,疑似谣言最终可能会被确定为非谣言;若大部分为疑问或无法确定,疑似谣言可能最终仍然无法判定。除此之外,谣言勘探的结果也会直接影响谣言检测结果,从海量信息中筛选出疑似谣言文本,疑似谣言文本确认为谣言的概率大于非疑似谣言文本确认为谣言的概率。

为更好地利用立场检测的领域知识和谣言勘探的领域知识,本文提出基于任务层级注意力机制的多任务谣言检测模型,如图4所示。一般的注意力机制是通过计算单个词对于输入文本语义表示的贡献度。然而,基于任务层级的注意力机制是通过融合谣言勘探任务、立场检测任务和谣言检测任务的领域信息,计算各个任务的语义表示对于谣言检测任务检测判定的贡献度。

在基于任务层级注意力机制中,每个时间步  $t$ , 谣言勘探任务、立场检测任务和谣言检测任务通过 LSTM 单元得到带有谣言勘探领域知识、立场检测领域知识和谣言检测领域知识的语义输出分别为

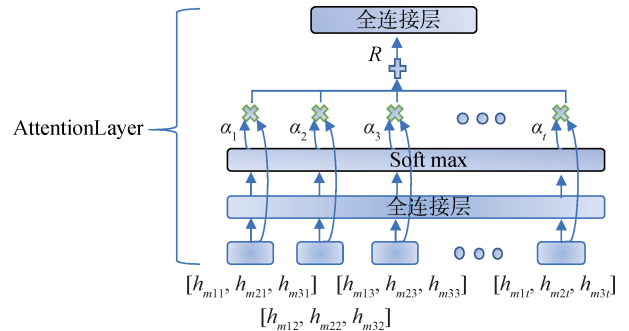


图4 基于任务层次注意力机制图

Fig.4 Attention Mechanisms Based on Task Hierarchy

$h_{m1t}$ 、 $h_{m2t}$  和  $h_{m3t}$ 。将该语义信息进行拼接,得到包含三个领域知识的语义编码  $(h_{m1t}, h_{m2t}, h_{m3t})$ 。然后,将所有时间步的语义编码输入 Softmax 激活函数的全连接层,得到权重向量  $\alpha$ ,如公式(14)所示。最后,权重向量与包含各个任务的语义信息  $(h_{m1t}, h_{m2t}, h_{m3t})$  相乘,得到带有各个任务贡献度的语义信息  $C$ ,如公式(15)所示。最终将带有权重的语义信息  $C$ ,通过 Softmax 函数,得到概率最大的标签序列。

$$\alpha = \frac{e^{W_a \cdot (h_{m1t}, h_{m2t}, h_{m3t}) + b_a}}{\sum_{j=1}^J e^{W_a \cdot (h_{m1t}, h_{m2t}, h_{m3t}) + b_a}} \quad (14)$$

$$C = \sum_{i=1}^J \alpha_i (h_{m1t}, h_{m2t}, h_{m3t}) \quad (15)$$

其中,  $W_a$  为权重矩阵;  $b_a$  为偏置矩阵;  $e$  为无理常数;  $J$  为标签类别数;  $(h_{m1t}, h_{m2t}, h_{m3t})$  为谣言勘探任务、立场检测任务、谣言检测任务在  $t$  时间步产生的语义信息的拼接值;  $\alpha$  为权重向量,  $\alpha_t$  为  $t$  时刻的权重;  $C$  为语义信息。

## 4 实验设置与结果分析

### 4.1 数据集

本文使用扩展版本的 PHEME 数据集<sup>①</sup>,该数据集围绕 Charlie Hebdo、Sydney Siege、Ferguson、Ottawa Shooting、Germanwings-crash、Putin Missing、Prince Toronto、Gurlitt、Elbola Essien 等 9 个事件,总共包含 6 425 条事件信息,105 354 条回复信息,如表 1 所示。

此数据集包含三种类型标签,第一类为谣言勘

①[https://figshare.com/articles/PHEME\\_dataset\\_for\\_Rumour\\_Detection\\_and\\_Veracity\\_Classification/6392078](https://figshare.com/articles/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078).

表1 PHEME数据集

Table 1 PHEME Dataset

事件名称	事件文本量	回复文本量	疑似谣言	非疑似谣言	谣言	非谣言	无法判定
Charlie Hebdo	2 079	38 268	458	1621	193	116	149
Sydney Siege	1 221	23 996	522	699	382	86	54
Ferguson	1 143	24 175	284	859	10	8	266
Ottawa Shooting	890	12 284	470	420	329	72	69
Germanwings-crash	469	4 489	238	231	94	111	33
Putin Missing	238	835	126	112	0	9	117
Prince Toronto	233	902	229	4	0	222	7
Gurlitt	138	179	61	77	59	0	2
Elbola Essien	14	226	14	0	0	14	0
合计	6 425	105 354	2 402	4 023	1 067	638	697

探标签,合计包含2 402个疑似谣言标签,4 023个非疑似谣言标签;第二类为立场检测标签,由于原数据集只包含部分的立场检测标签,剩余标签使用Branch-LSTM模型进行自动标注;第三类为谣言检测标签,合计包含1 067个谣言标签,638个非谣言标签,697个无法判定标签。

传统的交叉验证方法将数据按照一定比例划分训练集和验证集,保证训练集和验证集分布一致。考虑到现实中的谣言大部分围绕特定事件爆发,但在不同事件中谣言标签、非谣言标签、无法判定标签可能因事件性质不同而具有不同的分布。目前模型存在难以及时获取谣言爆发时相关事件的语料信息问题,只能依靠其他事件的语料进行训练。因此,采用新的交叉验证方法,同时输入若干组事件,依次剔除某个事件作为验证集,并将剩余事件作为训练集,模拟现实情况进行谣言预测。虽然该方法可能导致训练集和验证集产生分布差异,但更符合实际同时更具有挑战性。

#### 4.2 数据预处理

在统一的多任务模型框架内,为使各个任务共享相同的语义输入,本文参考Branch-LSTM输入格式。谣言文本和评论信息共同构成树状结构,如图5所示,选取Charlie Hebdo事件中具有三个评论分支的树状结构。由于多任务模型每次都并行输出谣言勘探标签、立场检测标签和谣言检测标签,模型需要同时输入关于评论文本的语义信息和关于谣言文本的语义信息。因此,需要对数据集进行特殊的数据预处理,以获得完整的语义信息。具体操作如下。

首先,对于输入的文本信息,利用正则表达式删

除非字母字符,并把所有单词转化为小写字母。然后,将只含单词的文本进行分词操作,得到长度为 $x$ 的输入。分词之后,通过Google Data News<sup>[25]</sup>预训练的词向量,获得句子中每个单词的向量表示,最终获得的句子向量维度为 $[x, 300]$ 。将句子向量在第一维做均值化处理,得到包含整个句子信息的压缩向量,维度为 $[1, 300]$ 。最后对同一分支下各个文本做相同处理,按第一维度拼接,得到带有整个分支信息的向量表示,维度为 $[X, 300]$ ( $X$ 为分支的长度)。对于不同长度的树状分支,本实验使用零填充和Masking填充,得到具有树状结构信息的向量编码。

#### 4.3 实验系统环境与参数设置

模型运行的实验环境如表2所示。为检验不同算法的性能优劣,使用正确率( $acc$ )和宏平均F值( $Macro-F$ )作为多任务效果的度量指标,如公式(16)~公式(18)所示。

$$acc = \frac{TP + TN}{P + N} \quad (16)$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (17)$$

$$Macro - F = \frac{1}{n} \sum_{i=1}^n F_{1i} \quad (18)$$

其中, $P$ 表示正例标签; $F$ 表示负例标签; $TP$ 表示正例中实际预测为正例的数量; $FP$ 表示正例中实际预测为负例的数量; $n$ 为预测类别个数; $F_{1i}$ 为每种类别的 $F_1$ 值。

#### 4.4 实验设计

##### (1) 实验划分

① 使用 Charile Hebdo、Sydney Siege、Ottawa

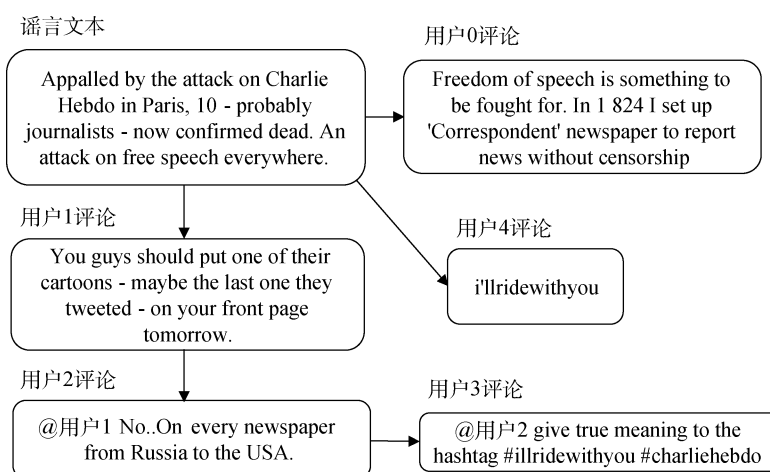


图 5 具有三个分支的谣言文本与评论的树状结构

Fig.5 A Tree Structure with Three Branches of Rumor Text and Comments

表 2 实验环境

Table 2 Experimental Environment

操作系统配置	参数或版本
CPU	Xeon(R) Gold 5218 CPU
GPU	NVIDIA T4(16GB)
Python	3.6
TensorFlow	1.1.31
Keras	2.3.1
内存	1TB

Shooting、Gemanwings-crash 这 4 个标签分布相似的事件作为一组实验。由表 1 可知,前 5 个数据集数据量占总数据量的 90.3%,除第三个事件 Ferguson 外,其他 4 个事件谣言检测标签主要为谣言和非谣言,无法判定标签所占比例较少;而在 Ferguson 事件中,无法判定的文本标签占比高达 93.7%。可能由于过大的数据分布差异,在交叉验证时,数据影响模型检测效果。

② 使用 Charile Hebdo、Sydney Siege、Ottawa Shooting、Gemanwings-crash、Ferguson 这 5 个最大的数据集作为一组实验,不考虑 Ferguson 事件标签分布与其他事件分布不同。

③ 使用所有的数据集作为一组实验。

(2) 对比实验

① 多数投票法 (Majority): 如果选择的某个元素的数量大于总数的一半,则返回它;否则返回 -1。在 PHEME 数据集谣言检测任务中,由于类别的不均衡性,该方法是一个较好的基线对比方法。

② NileTMRG\*: Kochkina 等<sup>[6]</sup>提出基于 NileTMR<sup>[26]</sup>的改进方法 NileTMRG\*,在 RumourEval 数据集中,取得更优效果。

③ Branch-LSTM: 该方法基于 LSTM 模型,结合社交媒体文本信息与位置信息对谣言进行判断。

④ MTL3: Kochkina 等<sup>[6]</sup>提出的基于谣言勘探、立场检测和谣言检测的多任务模型。

⑤ ES-ATT-MTL3: Li 等<sup>[17]</sup>提出结合谣言检测和立场检测的多任务模型,每个任务具有独立的任务处理层,且谣言检测任务输出最后接入注意力层。

⑥ Task-ATT-MTL3: 本文提出的基于层级任务注意力机制的多任务谣言检测模型,在 MTL3 进行优化和改进,增加任务层级注意力机制。

⑦ Un-Task-ATT-MTL3: 本文提出的模型方法,基于同方差不确定性和层级注意力机制的多任务谣言检测模型。在任务层级注意力机制基础上,首次在多任务谣言检测研究中,引入基于分类任务的同方差不确定性损失函数。

#### 4.5 实验结果与分析

(1) 总体结果分析

由表 3 谣言检测任务实验结果发现,ES-ATT-MTL3 模型效果在三组实验都低于 MTL3 模型,因此,选择 MTL3 为基准模型。Un-Task-ATT-MTL3 相较于 MTL3,在实验一和实验二中,Macro-F 分别提升了 4.2 个百分点和 7.6 个百分点;实验三中,本文模型结果与 MTL3 结果相近。考虑到谣言检测是一个



不平衡任务, *Macro-F* 可以更好地代表模型检测效果。结果说明, 同方差不确定性损失函数优化和任务层级注意力机制, 对于多任务谣言检测任务, 可在不指定主任务和辅助任务的情况下, 达到更好的检测效果, 主观划分主任务和辅助任务并非必要步骤。

表3 谣言检测任务实验结果  
Table 3 Results of Rumor Detection Task

实验	算法	acc	Macro-F
实验一	Majority( True)	0.591	0.247
	NileTMRG*	0.444	0.205
	Branch-LSTM	0.466	0.362
	MTL3	0.462	0.322
	ES-ATT-MTL3	0.395	0.263
	Task-ATT-MTL3	0.494	0.333
	Un-Task-ATT-MTL3	0.425	<b>0.364</b>
实验二	Majority( True)	0.511	0.226
	NileTMRG*	0.438	0.339
	Branch-LSTM	0.454	0.336
	MTL3	0.492	0.396
	ES-ATT-MTL3	0.459	0.280
	Task-ATT-MTL3	0.505	0.372
	Un-Task-ATT-MTL3	0.467	<b>0.472</b>
实验三	Majority( True)	0.444	0.205
	NileTMRG*	0.360	0.297
	Branch-LSTM	0.314	0.259
	MTL3	0.405	0.405
	ES-ATT-MTL3	0.356	0.240
	Task-ATT-MTL3	0.418	0.347
	Un-Task-ATT-MTL3	0.385	0.393

表3实验结果中值得注意的是, 由于数据集存在一定的类偏移, 故多数投票法正确率较高而 *Macro-F* 值偏低。在单任务模型中, 随着数据集的增加, 虽然 Branch-LSTM 的分类效果逐渐降低, 但相比其他单分类模型, Branch-LSTM 效果依然可观, 这也说明 Branch-LSTM 结构对于多任务模型具有较好的贡献度。在多任务模型中, ES-ATT-MTL3 虽然作为 MTL3 的改进模型, 但在联合谣言勘探、立场检测任务和谣言检测任务的多任务模型上效果表现不佳, 每个任务单独的任务处理层可能并不能很好地处理语义信息, 单纯地增加模型复杂度, 对于模型效果的提升影响并不大。Task-ATT-MTL3 作为 ES-ATT-MTL3 的改进模型, 舍弃了任务处理层, 在实验一、实验二和实验三中, 正确率较 MTL3 分别提升 3.2 个百分点、1.3 个百分点、1.3 个百分点, 说明基于

层级注意力机制的多任务谣言检测模型的有效性。Un-Task-ATT-MTL3 在数据类偏斜较小的情况下(实验一和实验二), 虽然准确率下降, 但是 *Macro-F* 提升显著。随着数据类偏斜严重(实验三), Un-Task-ATT-MTL3 效果最终与 MTL3 模型效果接近。该结果说明不确定性损失函数和任务层级注意力机制在不指定各个任务权重的情况下, 模型仍可达到理想效果。

纵向对比表3结果可以发现, 随着数据集的增加, 各个模型的分类效果总体而言存在先上升后下降的趋势。实验二与实验一相比, 引入 Ferguson 事件后, 数据总量上升, 在一定程度上弥补 Ferguson 事件的类别标签与其他 4 个事件类别标签相异的问题; 而实验三相比于实验二, 需额外引入剩余的 4 个数据集, 新引入的事件数据的标签分布与其他 5 类数据的标签分布也不同, 但数据的增加不能弥补标签分布的差异。这也体现了每次将一个事件数据分开, 作为测试集对谣言进行检测的难度和高质量数据对于模型训练的重要性。

## (2) 单事件结果分析

为进一步理解 Un-Task-ATT-MTL3 模型在多任务谣言识别任务中的作用, 选取数据量最多的 5 个事件(实验二)做进一步分析, 统计单个事件在谣言检测任务的 *Macro-F* 和单个事件中谣言类、非谣言类、无法确定类的平均  $F_1$  值, 如表4所示。

表4 单事件结果  
Table 4 Results of Single Event

事件	acc	Macro-F	谣言类 平均 $F_1$ 值	非谣言 类平均 $F_1$ 值	无法确 定类平 均 $F_1$ 值
Charlie Hebdo	0.292	0.213	0.147	0.131	0.362
Sydney Siege	0.339	0.204	0.400	0.162	0.501
Ferguson	0.697	0.268	0	0.004	0.801
Ottawa Shooting	0.675	0.306	0.806	0.107	0.010
Germanwings-crash	0.356	0.245	0.310	0.355	0.091

由表4可知, 各个事件的 *Macro-F* 值。其中 Ferguson 事件的标签主要集中在无法确定类, *acc* 为 0.697; Ottawa Shooting 事件标签主要集中在谣言类, *acc* 为 0.675, 该结果说明模型训练过程中, 预测结果并未偏向某一类, 而是根据事件本身的语义信息进行预测, 不确定性损失函数和层级注意力机制在谣



言的语义提取上效果更好。

总体来说,本文方法效果优于目前多任务谣言检测最好的方法。Un-Task-ATT-MTL3 模型在避免主观划分主任务和辅助任务的同时,检测效果优于目前已有的模型。其中,Un-Task-ATT-MTL3 模型对 ES-ATT-MTL3 模型结构进行优化,舍弃了单独的任务处理层,合理地减少了模型的复杂度,提升了模型识别谣言的准确率。本文首次通过引入基于分类任务的同方差损失函数替代同方差损失函数,帮助模型自发寻找各个任务的最优权重,且理论上适用于不同数据集。多任务谣言检测不需要主观划分主任务和辅助任务也可以达到理想的效果。

## 5 结 语

现有关于谣言检测、立场检测、谣言勘探的研究工作大部分都被视为单一任务。本文尝试使用一个统一的多任务框架,联合优化上述三个任务,用于提升谣言检测的准确率。通过大量的对比实验表明,Un-Task-ATT-MTL3 模型在避免人工划分主任务和辅助任务的同时,保证了模型训练效果。Un-Task-ATT-MTL3 模型作为 MTL3 和 ES-ATT-MTL3 的改进模型,任务层级注意力机制的引入和不确定性损失函数的使用,使模型对于谣言检测的准确率有进一步的提升。该方法为解决谣言问题提供了一个新的解决方法与思路。

除了谣言勘探任务和立场检测任务,用户信誉度对于谣言检测也有着重要影响。可以考虑引入用户信誉度评估任务,以提升谣言检测任务的准确率。此外,多任务谣言检测模型中各个任务之间领域知识能否找到更好的融合方式有待探究。未来工作将以这两个问题为切入点,对多任务谣言检测模型进行更深入研究。

## 参考文献:

[1] Zubiaga A, Aker A, Bontcheva K, et al. Detection and Resolution of Rumours in Social Media: A Survey[J]. ACM Computing Surveys (CSUR), 2018, 51(2): 1-36.

[2] 陈燕方,李志宇,梁循,等. 在线社会网络谣言检测综述[J]. 计算机学报, 2018, 41(7): 1648-1676. (Chen Yanfang, Li Zhiyu, Liang Xun, et al. Review on Rumor Detection of Online Social Networks[J]. Chinese Journal of Computers, 2018, 41(7): 1648-

1676.)

[3] Qazvinian V, Rosengren E, Radev D, 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] Liang G, He W B, Xu C, et al. Rumor Identification in Microblogging Systems Based on Users' Behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.

[5] Kwon S, Cha M, Jung K, et al. Prominent Features of Rumor Propagation in Online Social Media[C]//Proceedings of the 13th International Conference on Data Mining. IEEE, 2013: 1103-1108.

[6] Kochkina E, Liakata M, Zubiaga A. All-in-One: Multi-task Learning for Rumour Verification[OL]. arXiv Preprint, arXiv: 1806.03713.

[7] Sejeong K, Meeyoung C, Kyomin J, et al. Rumor Detection over Varying Time Windows[J]. PLoS One, 2017, 12(1): e0168344.

[8] Yang F, Liu Y, Yu X H, et al. Automatic Detection of Rumor on Sina Weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012:13.

[9] Chang C, Zhang Y H, Szabo C, et al. Extreme User and Political Rumor Detection on Twitter[C]//Proceedings of International Conference on Advanced Data Mining and Applications. 2016: 751-763.

[10] 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.

[11] Chen T, Li X, Yin H Z, et al. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection[C]//Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2018: 40-52.

[12] Collobert R, Weston J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning[C]//Proceedings of the 25th International Conference on Machine Learning. 2008: 160-167.

[13] Liu P F, Qiu X P, Huang X J. Recurrent Neural Network for Text Classification with Multi-task Learning[OL]. arXiv Preprint, arXiv: 1605.05101.

[14] Dong D X, Wu H, He W, et al. Multi-task Learning for Multiple Language Translation[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015: 1723-1732.

[15] Firat O, Cho K, Bengio Y. Multi-way, Multilingual Neural Machine Translation with a Shared Attention Mechanism[OL]. arXiv Preprint, arXiv: 1601.01073.

[16] Ma J, Gao W, Wong K F. Detect Rumor and Stance Jointly by

- Neural Multi-task Learning[C]//Companion Proceedings of the Web Conference 2018. 2018: 585-593.
- [17] Li Q Z, Zhang Q, Si L. Rumor Detection by Exploiting User Credibility Information, Attention and Multi-task Learning[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 1173-1179.
- [18] Sener O, Koltun V. Multi-task Learning as Multi-objective Optimization[C]//Proceedings of the 32nd Conference on Neural Information Processing Systems. 2018: 527-538.
- [19] Kendall A, Gal Y, Cipolla R. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 7482-7491.
- [20] Kendall A, Gal Y. What Uncertainties do We Need in Bayesian Deep Learning for Computer Vision? [C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 5580-5590.
- [21] Gal Y. Uncertainty in Deep Learning[D]. University of Cambridge, 2016.
- [22] Staudemeyer R C, Morris E R. Understanding LSTM - a Tutorial into Long Short-Term Memory Recurrent Neural Networks[OL]. arXiv Preprint, arXiv: 1909.09586.
- [23] Fabbri M, Moro G. Dow Jones Trading with Deep Learning: The Unreasonable Effectiveness of Recurrent Neural Networks[C]//Proceedings of the 7th International Conference on Data Science, Technology and Applications. 2018: 142-153.
- [24] Kochkina E, Liakata M, Augenstein I. Turing at Semeval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM[OL]. arXiv Preprint, arXiv: 1704.07221.
- [25] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[OL]. arXiv Preprint, arXiv: 1301.3781.
- [26] Enayet O, El-Beltagy S R. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter[C]//Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017: 470-474.

#### 作者贡献声明:

杨哈迅:提出研究思路,设计实验方案,进行实验,起草论文;  
周德群:完善实验方案;  
马静:设计研究思路;  
罗永聪:数据清洗与转换,论文审阅与修订。

#### 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:

支撑数据由作者自存储, E-mail: 1370117133@qq.com。  
[1] 杨哈迅, 周德群, 马静, 罗永聪. Multi\_task\_rumor\_detection.zip. 程序源码。  
[2] 杨哈迅, 周德群, 马静, 罗永聪. res.zip. 实验及对比实验结果。

收稿日期:2020-12-06  
收修改稿日期:2021-03-18

## Detecting Rumors with Uncertain Loss and Task-level Attention Mechanism

Yang Hanxun Zhou Dequn Ma Jing Luo Yongcong

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics,  
Nanjing 211100, China)

**Abstract:** [Objective] This paper proposes a new model with the help of uncertainty loss function and task-level attention mechanism, aiming to address the issue of setting main and auxiliary tasks in rumor detection. [Methods] First, we integrated the domain knowledge of rumor exploration, stance classification, and rumor detection. Then, we constructed a modified model with task-level attention mechanism. Third, we used uncertainty loss function to explore the weight relationship of each task and obtain better detection results. Finally, we examined our model's performance with the PHEME4 and PHEME5 datasets. [Results] Compared to the existing models, the *Macro-F* of our model increased by 4.2 and 7.6 percentage points with PHEME4 and PHEME5. [Limitations] We only examined our model with the PHEME dataset. [Conclusions] The proposed method could effectively detect rumors without dividing the main and auxiliary tasks.

**Keywords:** Uncertain Loss Multi-task Learning Rumor Detection Attention Mechanism