



云南大学学报(自然科学版)

Journal of Yunnan University(Natural Sciences Edition)

ISSN 0258-7971,CN 53-1045/N

《云南大学学报(自然科学版)》网络首发论文

题目：基于联合情感的多任务谣言检测方法
作者：马儀，邵玉斌，杜庆治，龙华，马迪南
收稿日期：2023-02-14
网络首发日期：2024-01-04
引用格式：马儀，邵玉斌，杜庆治，龙华，马迪南. 基于联合情感的多任务谣言检测方法[J/OL]. 云南大学学报(自然科学版).
<https://link.cnki.net/urlid/53.1045.N.20240103.1615.012>



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

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



基于联合情感的多任务谣言检测方法

马 儀¹, 邵玉斌^{1**}, 杜庆治¹, 龙 华^{1,2}, 马迪南²
(1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650500;
2. 云南省媒体融合重点实验室, 云南 昆明 650032)

摘要:情感分析在社交媒体谣言检测中有重要作用, 现有的谣言检测方法侧重于使用文章的情感特征, 忽略了用户评论与文章的联合情感, 单任务深度学习谣言检测方法缺少足够的标签数据导致准确率难以提升. 为解决上述问题, 基于谣言检测与联合情感检测两个任务的相关性, 提出了多任务联合学习的谣言检测方法, 在同一模型中实现联合情感检测以及谣言检测两个任务. 首先, 构建编码器提取文章及对应评论的语义特征并映射到同一语义空间, 并通过注意力机制加权融合语义特征; 其次, 通过基于公共情感分类器与情感词典共同构建的情感提取网络提取文章与对应评论的情感特征; 最后, 将语义特征与联合情感特征融合后输入到两个共享损失函数的分类器中, 分别得到谣言检测和联合情感检测的分类结果. 实验结果表明, 多任务模型的效果强于单任务模型, 在公开的中文 Weibo-16 数据集、英文 Twitter-15 数据集上相较于对比方法中最好的方法, 提出的模型在准确率上分别提升了 3.5 和 2.9 个百分点, F1 值提高了 3.1 和 3.9 个百分点.

关键词: 谣言检测; 用户评论; 联合情感; 多任务学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 0258-7971(2024)00-0001-13

拥有庞大用户量的社交媒体平台已成为谣言信息泛滥的场所, 大量谣言迅速传播不仅危害网络空间安全, 还威胁到现实世界的政治、经济、社会等方面的秩序^[1-2]. 自动谣言检测技术通过提取社交媒体中的文章及相关信息的有效特征进行谣言检测, 可以减轻社交媒体平台的审核压力, 节约人工成本, 提高检测效率, 有效减少谣言传播.

情感特征能反映文本情感的强度和正负, 已广泛应用于谣言检测中. Ajao 等^[3]指出, 情感是谣言检测的重要特征, 谣言检测模型中添加情感特征能提升模型效果. Giachanou 等^[4]基于情感词汇从文章中提取了更丰富的情感特征用于谣言检测. 现有的模型大多利用了文章本身传递的情感信号, 但很少关注文章评论的情感. 然而, 为了在人群中快速传播, 谣言往往引起读者强烈的情感^[5-8], 如图 1 微博平台上的两篇谣言文章与对应的评论所示, 相较于文章而言评论中蕴含了更多的情感信息. Ghanem 等^[9]研究发现, 谣言信息会激发读者恐惧、厌恶

和惊讶的情感, 相反事实能激发读者期待、悲伤、喜悦和信任等正向情感, 谣言和事实对读者产生的不同情感造成了谣言比事实传播更加迅速, 因此有必要探讨用户评论的情感是否能增强谣言检测模型的能力. 联合情感表示为文章与用户评论共同反映的情感, 按照文献 [6] 中将联合情感分为正向或负向, 文献 [5] 中证明了联合情感与文章是否是谣言之间具有强相关性.

基于机器学习的自动谣言检测方法难以获得高维、复杂、抽象的特征数据, 模型容易陷入过拟合, 深度学习有更强的特征学习能力, 但谣言早期传播时缺乏标签数据, 导致单任务深度学习效果不佳. 多任务联合学习框架同时对多个任务进行预测, 减少了所需标签数据的数量以及整体模型参数的规模, 使预测更加高效. 同时学习多个相关任务能获得比单一任务更好的结果^[10-11], 能提高模型泛化能力, 降低过拟合风险, 解决数据标注不足的问题, 通过共享信息平均每个任务的噪声实现隐式数据

收稿日期: 2023-02-14; 接受日期: 2023-08-08

基金项目: 云南省媒体融合重点实验室项目(320225403).

作者简介: 马 儀(1998-), 男, 云南人, 硕士生, 主要研究自然语言处理、舆情分析. E-mail: mayi7131@163.com.

** 通信作者: 邵玉斌(1970-), 男, 云南人, 教授, 主要研究自然语言处理、个人通信系统以及信号处理. E-mail: Shaoyubin999@qq.com.

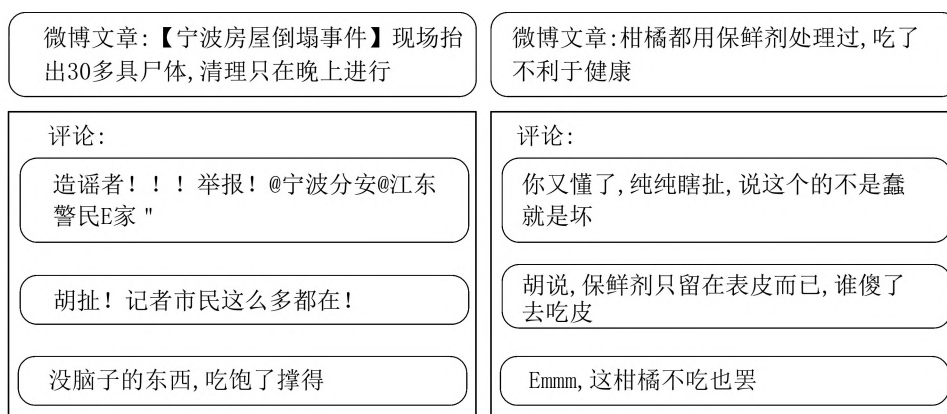


图 1 微博谣言文章与对应评论示意图

Fig. 1 Diagram of Weibo rumor articles and corresponding comments

扩充获得比单任务模型更好的结果. 多任务学习模型的构建方法为融入联合情感到谣言检测模型中提供了解决思路.

选取相关性强的任务能构建效果更好的多任务学习模型^[11-13], 本文基于联合情感检测任务与谣言检测任务之间的强相关性, 提出基于联合情感的多任务学习模型 (Joint Emotion Multitask Learning Model, JEMT), JEMT 可同时进行联合情感检测和谣言检测. 首先 JEMT 利用双向长短期记忆网络 (Bi-directional Long Short-Term Memory, BiLSTM) 分别提取文章和对应评论的语义特征并利用其中隐藏层的信息, 再次使用 BiLSTM 将两者的语义特征映射到同一语义空间中得到融合语义特征, 并通过注意力机制对融合语义特征进行加权和得到语义特征的句向量表示; 其次使用公共情感分类器结合情感字典的方法提取文章与评论的情感特征提升情感检测任务性能, 从而提高 JEMT 模型整体性能和加快整个模型的收敛速度; 最后将情感特征以及语义特征的句向量表示融合后分别送入共享同一个损失函数的两个分类器中, 以多任务学习方式训练模型, 得到谣言检测结果以及联合情感检测结果. 现有的公共谣言数据集缺失情感标签, 先使用情感分类模型为数据打上情感标签后, 再使用 JEMT 进行多任务训练. 本文的贡献如下:

(1) 提出一种新的基于文章与用户评论的语义特征、文章与用户评论的联合情感特征的多任务学习模型. 并使用双层 BiLSTM 网络将文章与用户评论的语义特征映射到同一语义空间, 实现有效地共享多任务参数.

(2) 提出一种以公共情感分类器结合情感词典的情感特征提取方法并融入到提出的谣言检测模

型中, 达到优化模型性能, 加快模型收敛速度的目的.

(3) 构建情感正负分类模型, 为公共数据集打上情感标签, 通过实验验证了提出的谣言检测模型的有效性.

1 相关工作

1.1 情感检测 情绪是生活中至关重要的一部分, 反映了日常活动的快乐、悲伤、喜悦和恐惧等. 社交媒体的文章标题、文章内容等数据被有效地用于设计情感检测模型^[14-15]. Ren 等^[16] 从在线评论中提取情绪信息, 并测试了评论的恐惧、愤怒和悲伤情绪对感知审查的影响. 谣言检测领域中, Giachanou 等^[4] 使用情感特征提升谣言检测模型的性能, 从文章标题中提取情感特征, 以确定文章的可信度, 效果优于仅利用文章本身的情感特征进行谣言检测. 文献^[5] 提出了一种机制, 分别提取文章和对应用户评论的情感感知表示, 通过构建双重情感特征提高谣言检测模型性能. 文献^[7] 使用协同注意力机制解释了情感特征提升谣言检测模型效果的原因.

1.2 多任务学习 多任务学习的目的是通过联合其他相关任务提高目标任务的性能, 通过多个任务共享参数减少单任务模型对标签数据的依赖. 多任务学习方法应用广泛, 在谣言检测中, Kochkina 等^[17] 描述了一个新的多任务框架, 同时实现谣言检测、准确性预测和姿态检测 3 个任务. Wu 等^[18] 提出具有筛选特征功能共享层的多任务谣言检测方法, 该共享层采用注意力和门控机制选择任务之间的共享参数特征. Kumari 等^[19] 结合文章与标题的特征, 构建多任务模型同时实现谣言检测、情感检测、重复性检测 3 个任务.

1.3 谣言检测 谣言检测通常使用机器学习或深度学习的方法. 机器学习谣言检测方法需手工提取或构造大量的特征^[20-21], 如文献 [20] 基于决策树排名的谣言检测模型, 通过搜索数据中的特殊短语和对争议的文章聚类, 并根据统计结果构建特征对聚类结果进行排序以区分谣言和事实; 文献 [21] 基于支持向量机的谣言检测模型, 采用时间序列为谣言信息的社会信息特征(内容、用户和传播等)的变化进行建模以区分谣言和事实. 该类方法突出优点是计算复杂度低、速度快, 但效果弱于使用深度神经网络自动捕获特征的方法. 基于深度学习模型的谣言检测方法避免了机器学习大量的手工设计特征, 深度学习谣言检测可利用文章语义特征^[3, 22-23], 也可以结合传播方式^[24-25]、社会背景^[27-28], 如发布平台、发布者出版商等特征进行谣言检测. 文献 [23] 中利用多层递归神经网络, 以抽取数据深层隐藏信息用于谣言检测. 文献 [25] 考虑到每条微博有许多分支, 使用树状循环神经网络对社交平台文章的分

支现象进行建模, 以增强谣言检测模型性能.

不同于上述方法, 本文设计的重点是考虑用户评论与文章的联合情感对模型的影响, 以及联合情感特征与文章语义特征的提取与融合, 并使用多任务模型提升检测任务性能等方面.

2 基于联合情感的多任务谣言检测方法

2.1 基于联合情感的多任务模型 基于联合情感的多任务谣言检测模型如图 2 所示, 由编码器、联合情感提取网络和任务分类器组成. 编码器提取文章和评论的语义特征表示并嵌入到同一个语义空间中, 得到融合语义特征, 融合语义特征通过注意力机制加权求和后得到文章和评论的句向量表示. 联合情感提取网络使用情感词典和公共情感分类器共同构成, 将句向量表示和联合情感提取网络提取的联合情感特征融合得到融合特征, 将融合特征送入谣言检测任务分类器和情感检测任务分类器中, 以实现联合谣言检测和情感检测.

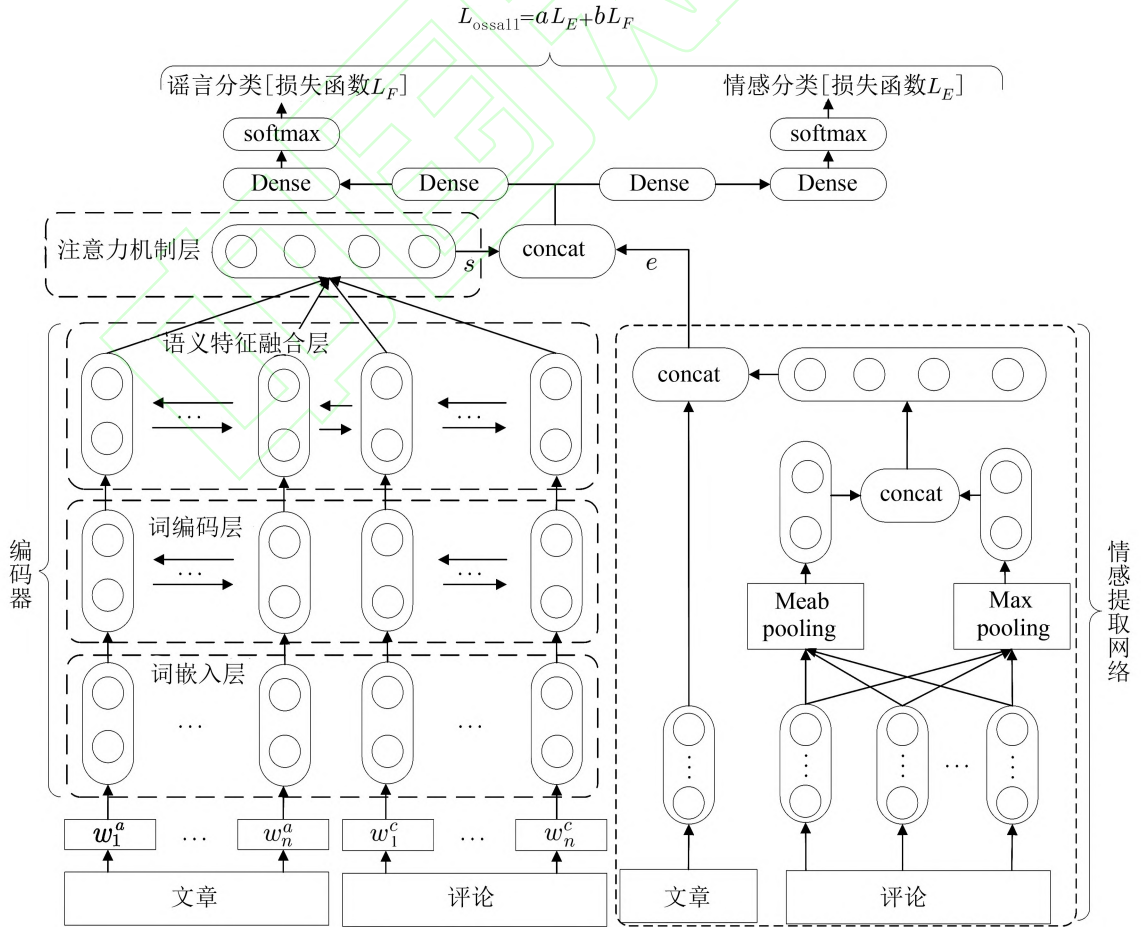


图 2 联合情感的多任务模型图

Fig. 2 Multi-task model for joint emotion analysis

2.2 编码器 编码器由词编码器和句编码器构成,词编码器提取文章与对应评论的词语义特征,句编码器将文章与对应评论词的语义特征嵌入到同一个语义空间中得到融合语义特征,并通过注意力机制层对融合语义特征进行加权求和得到句向量。

2.2.1 词编码器 由于社交媒体(如微博、Twitter等)的文章由于字数限制,长度较短,但对应的多条评论总的文字长度相对较长。因此为了保持输入模型数据形状的一致性,将每篇社交媒体文章视为一个句子,每篇文章的评论则由 M 个句子组成,并将每个句子截断或填充为相同的长度 n ,即每个句子包含 n 个词, $[w_1^a, w_2^a, \dots, w_n^a]$ 表示社交媒体文章和评论的分词结果, w_n^a 代表文章中的第 n 个词,将分词后的每个词经过词嵌入层转化为 d 维词向量,词向量化分词处理时,中文词语没有明显的空格分隔,需借助中文分词工具进行切分。英文语料以空格为单位进行切分成词,词嵌入层将中文分词经过微博语料训练的 300 维预训练词嵌入映射为词向量^[29],英文分词经过使用 Glove 基于 Twitter 语料训练的 300 维预训练词嵌入映射为词向量^[5],分词 $[w_1^a, w_2^a, \dots, w_n^a]$, $[w_1^c, w_2^c, \dots, w_n^c]$ 经过词嵌入层得到词向量矩阵为 $X^A = [x_1^a, x_2^a, \dots, x_n^a] \in \mathbb{R}^{n \times d}$,对应评论的分词结果 $[w_1^c, w_2^c, \dots, w_n^c]$ 经过词嵌入层得到词向量矩阵为 $X^{C_i} = [x_1^{c_i}, x_2^{c_i}, \dots, x_n^{c_i}] \in \mathbb{R}^{n \times d}$,其中 i 的取值为 1 到 M ,表示评论划分的第 i 个句子。

词汇与上下文关联,有较强的双向语义依赖,使用单独的 BiLSTM 层从词的两个方向获取词向量矩阵的隐藏表示, BiLSTM 由前向 $\overrightarrow{\text{LSTM}}$ 和后向 $\overleftarrow{\text{LSTM}}$ 组成,定义获取文章隐藏表示的 BiLSTM 层为 LSTM_A ,设置 LSTM_A 隐藏单元为 d_s 维。如式 (1) 中 $\overrightarrow{\text{LSTM}}_A$ 从左往右处理词向量矩阵获得前向隐藏状态 $\vec{h}^A = [\vec{h}_1^A, \vec{h}_2^A, \dots, \vec{h}_n^A] \in \mathbb{R}^{n \times d_s}$, 式 (2) 中 $\overleftarrow{\text{LSTM}}_A$ 从右往左处理词向量矩阵得到后向隐藏状态输出 $\overleftarrow{h}^A = [\overleftarrow{h}_1^A, \overleftarrow{h}_2^A, \dots, \overleftarrow{h}_n^A] \in \mathbb{R}^{n \times d_s}$,最后连接前向隐藏状态输出和后向隐藏状态输出作为词的上下文表示,如式 (3) 所示, concat 表示拼接操作,则文章中的词上下文表示为 $h^A = [h_1^A, h_2^A, \dots, h_n^A] \in \mathbb{R}^{n \times 2d_s}$ 。

$$\vec{h}_t^A = \overrightarrow{\text{LSTM}}_A(x_t^a, \vec{h}_{t-1}^a), t \in \{1, 2, \dots, n\}, \quad (1)$$

$$\overleftarrow{h}_t^A = \overleftarrow{\text{LSTM}}_A(x_t^a, \overleftarrow{h}_{t-1}^a), t \in \{n, n-1, \dots, 1\}, \quad (2)$$

$$h^A = \text{concat}(\vec{h}_t^A, \overleftarrow{h}_t^A), t \in \{1, 2, \dots, n\}. \quad (3)$$

$\overrightarrow{\text{LSTM}}_A$ 计算 \vec{h}_t^A 的具体内部计算过程如 (4) 所

示, i_t^a 、 f_t^a 、 o_t^a 、 c_t^a 分别表示输入门、遗忘门、输出门、记忆单元, $\sigma(\cdot)$ 表示 sigmoid 函数, W 和 b 是模型参数, \odot 表示逐元素乘法 (Hadamard 乘积), x_t^a 表示 t 时刻的输入 (第 t 个词的词向量)。 $\overrightarrow{\text{LSTM}}_A$ 与 $\overleftarrow{\text{LSTM}}_A$ 运算过程类似,不同之处是 $\overleftarrow{\text{LSTM}}_A$ 反向遍历数据。

$$\begin{aligned} i_t^a &= \sigma(W_{xi}x_t^a + W_{hi}\vec{h}_{t-1}^a + W_{ci}\vec{c}_{t-1}^a + b_i), \\ f_t^a &= \sigma(W_{xf}x_t^a + W_{hf}\vec{h}_{t-1}^a + W_{cf}\vec{c}_{t-1}^a + b_f), \\ c_t^a &= f_t^a \odot \vec{c}_{t-1}^a + i_t^a \odot \tanh(W_{xc}x_t^a + W_{hc}\vec{h}_{t-1}^a + b_c), \\ o_t^a &= \sigma(W_{xo}x_t^a + W_{ho}\vec{h}_{t-1}^a + W_{co}c_t^a + b_o), \\ \vec{h}_t^a &= o_t^a \odot \tanh(c_t^a). \end{aligned} \quad (4)$$

定义提取评论中词的上下文表示的 BiLSTM 层为 LSTM_C , 设置 LSTM_C 隐藏单元维度与 LSTM_A 一致,对于评论中句子的词向量矩阵 X^{C_i} , 提取出的评论的隐藏表示过程如式 (5) ~ (7) 所示,得到评论中的词上下文表示为 $h^{C_i} = [h_1^{C_i}, h_2^{C_i}, \dots, h_n^{C_i}] \in \mathbb{R}^{n \times 2d_s}$ 。

$$\vec{h}_t^{C_i} = \overrightarrow{\text{LSTM}}_C(x_t^{C_i}, \vec{h}_{t-1}^{C_i}), t \in \{1, 2, \dots, n\}, \quad (5)$$

$$\overleftarrow{h}_t^{C_i} = \overleftarrow{\text{LSTM}}_C(x_t^{C_i}, \overleftarrow{h}_{t-1}^{C_i}), t \in \{n, n-1, \dots, 1\}, \quad (6)$$

$$h^{C_i} = \text{concat}(\vec{h}_t^{C_i}, \overleftarrow{h}_t^{C_i}), t \in \{1, 2, \dots, n\}. \quad (7)$$

将评论中 M 个句子的词上下文表示按照词向量的维度拼接得到评论总的表示 h^C , 如下所示:

$$h^C = \text{concat}(h^{C_1}, h^{C_2}, \dots, h^{C_M}) \in \mathbb{R}^{n \times 2Md_s}, \quad (8)$$

拼接 $h^A \in \mathbb{R}^{n \times 2d_s}$ 和 $h^C \in \mathbb{R}^{n \times 2Md_s}$ 得到 $h^S \in \mathbb{R}^{n \times (2+2M)d_s}$, 如式 (9) 所示, h^S 为词编码器的输出。

$$h^S = \text{concat}(h^A, h^C) \in \mathbb{R}^{n \times (2+2M)d_s}. \quad (9)$$

2.2.2 句子编码器 如图 3 所示, 定义语义特征融合层 LSTM_S 映射评论和文章的语义特征到同一语义空间, LSTM_S 使用单个 BiLSTM 构建, 隐藏单

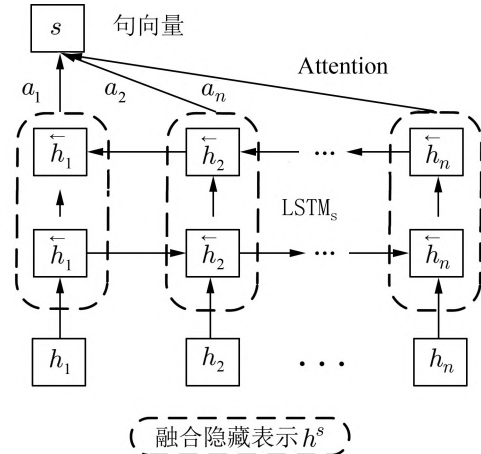


图 3 句向量合成

Fig. 3 Sentence Vector Composition

元维度为 d_c 维, 如式 (10) 所示, 将 h^S 输入到LSTM_s中得到的隐藏表示 $h = [h_1, h_2, \dots, h_n] \in \mathbb{R}^{n \times 2d_c}$ 作为语义融合特征.

$$h = \text{LSTM}_S(h^S) = [h_1, h_2, \dots, h_n] \in \mathbb{R}^{n \times 2d_c}. \quad (10)$$

基于每个词对句子意义的重要性不同的原则, 引入注意力机制提取词对句子的重要性来获得句子向量, 如式 (11) 所示, 将语义融合特征 $h = [h_1, h_2, \dots, h_n] \in \mathbb{R}^{n \times 2d_c}$ 进行注意力加权求和得到句向量 $s \in \mathbb{R}^{2d_c}$, 其中 h_t 为 h 的第 t ($0 < t < n$) 个元素, a_t 为衡量第 t 个词对句子重要性的注意力权重, a_t 由式 (12) 计算得到, 其中 h_w 为随机初始化的可训练的权重矩阵, u_t 为 h_t 经过全连接层后得到, 如式 (13) 所示, 其中 W, b 是全连接层中随机生成的可训练的参数.

$$s = \sum_{t=1}^n a_t h_t, \quad (11)$$

$$a_t = \frac{\exp((u_t)^T h_w)}{\sum_{t=1}^n \exp((u_t)^T h_w)}, \quad (12)$$

$$u_t = \tanh((W h_t) + b). \quad (13)$$

将评论和文章的语义特征映射到同一语义空间得到融合语义特征, 通过对融合语义特征注意力加权求和得到句向量 $s \in \mathbb{R}^{2d_c}$ 作为编码器的输出.

2.3 情感提取网络 如图 2 所示, 情感提取网络通过公共情感分类器以及情感字典提取联合情感向量, 有助于多任务模型更快收敛, 加速学习速度, 从而获得更好的性能.

2.3.1 文章情感特征提取 使用公共情感类别分类器^[5]处理文章, 文章视为一个句子, 情感类别分类器的输出为包含特定情感的概率, 设情感分类器为 f , 输出维度为 d_f , 文章 A 通过分类器的情感概率为 $e_1^A, e_1^A \in \mathbb{R}^{d_f}$, 如式 (14) 所示.

$$e_1^A = f(A). \quad (14)$$

使用情感词典提取特征, 如示例句子: “我看见这个消息非常不开心”, 先对句子进行分词, 该句子分为 7 个词, “开心”表示积极的情感, 在句子中的频率为 1/7, 否定词“不”的权重为-1, 程度词“非常”的权重为 2, 则该句话积极情感基于情感词典的情感分数为 $-1 \times 2 \times 1/7$. 设情感词典的情感类别为 $E = \{e_1^d, e_2^d, \dots, e_l^d\}$, 其中 $e_m^d \in E$, $1 \leq m \leq l$, 对应情感分类词汇表 B_m^d . 对于文章句子分词结果 $[w_1^a,$

$w_2^a, \dots, w_n^a]$, 首先判断每个词是否在情感分类词汇表中, 如式 (15), 其中 $0 \leq j \leq n$.

$$e_m^d(w_j^a) = \begin{cases} 1, & w_j^a \in B_m^d \\ 0, & w_j^a \notin B_m^d \end{cases}, \quad (15)$$

其次使用情感程度词汇词典以及负向词汇词典处理词语 w_j^a , $\deg(w_j^a)$ 表示 w_j^a 对应的情感强烈程度权重, $\text{neg}(w_j^a)$ 表示 w_j^a 对应的负向情感程度权重, w_j^a 的情感程度与负向程度和相邻词有关, 设窗口长度为 L , 表示 w_j^a 相邻的 L 个词, 如式 (16)、(17), 根据窗口检索词语 w_j^a 的情感的强烈程度和是否表达了负向情感.

$$f_d(w_j^a, L) = \prod_{i=j-L}^{j-1} \deg(w_i^a), \quad (16)$$

$$f_n(w_j^a, L) = \prod_{i=j-L}^{j-1} \text{neg}(w_i^a), \quad (17)$$

w_j^a 在情感类别 e_m^d 下情感得分计算如式 (18) 所示, 其中 n 为句子中词的总数.

$$e_2^{am}(w_j^a) = \frac{e_m^d(w_j^a) \cdot \deg(w_j^a, L) \cdot \text{neg}(w_j^a, L)}{n}, \quad (18)$$

计算文章 A 中所有词在情感类别 e_m^d 下的情感得分, 如式 (19).

$$e_2(A, e_m^d) = \sum_{j=1}^n e_2^{am}(w_j^a). \quad (19)$$

将 l 种情感类别下的情感得分拼接后得到基于情感词典提取的文章情感特征 e_2^A , 如式 (20) 所示, $e_2^A \in \mathbb{R}^l$.

$$e_2^A = \text{concat}(e_2(A, e_1^d), e_2(A, e_2^d), \dots, e_2(A, e_l^d)), \quad (20)$$

将 e_1^A 与 e_2^A 拼接得到文章情感向量 e^A , 如式 (21), $e^A \in \mathbb{R}^{l+d_f}$.

$$e^A = \text{concat}(e_1^A, e_2^A). \quad (21)$$

2.3.2 评论情感特征提取 对评论中的 M 个句子进行与文章 A 相同的情感计算, 如评论中第 i ($1 \leq i \leq M$) 个句子对应情感特征为 $e^{ci} \in \mathbb{R}^{l+d_f}$. 将 M 个句子中的情感特征转置后拼接, 得到评论情感矩阵, 如式 (22) 所示, 其中 $e^{cM} \in \mathbb{R}^{M \times (l+d_f)}$.

$$e^{cM} = \text{concat}[(e^{c1})^T, (e^{c2})^T, \dots, (e^{cM})^T], \quad (22)$$

为了让评论情感矩阵与文章情感向量相连, 对评论情感矩阵按列进行平均池化Meanpooling和最

大池化Maxpooling, 目的是提取出平均情感和极端情感. 如式 (23)、(24) 所示: $e_{\text{mean}}^c, e_{\text{max}}^c \in \mathbb{R}^{l+d_f}$.

$$e_{\text{mean}}^c = \text{Meanpooling}(e^{cM}), \quad (23)$$

$$e_{\text{max}}^c = \text{Maxpooling}(e^{cM}). \quad (24)$$

拼接 $e_{\text{mean}}^c, e_{\text{max}}^c$ 作为评论情感向量 e^c , 如式 (25) 所示, $e^c \in \mathbb{R}^{2(l+d_f)}$.

$$e^c = \text{concat}(e_{\text{mean}}^c, e_{\text{max}}^c). \quad (25)$$

将文章情感向量 e^a 与评论情感向量 e^c 拼接后得到联合情感向量 e , 如式 (26) 所示, $e \in \mathbb{R}^{3(l+d_f)}$.

$$e = \text{concat}(e^a, e^c). \quad (26)$$

最后连接句子向量 s 以及联合情感向量 e 作为编码器的输出, 如式 (27) 所示. $s_{\text{out}} \in \mathbb{R}^{3(l+d_f)+2d_c}$

$$s_{\text{out}} = \text{concat}(s, e). \quad (27)$$

2.4 任务分类层 谣言检测和情感检测分类层由两个全连接 Dense 层和使用 softmax 作为分类器函数的输出层组成, 分类层的输入为融合特征 s_{out} , 公式 (28) 表示谣言检测任务的输出结果, 公式 (29) 表示情感检测任务的输出结果. 其中 mlp 表示全连接 Dense 层.

$$\text{softmax}(\text{mlp}_f(s_{\text{out}})), \quad (28)$$

$$\text{softmax}(\text{mlp}_e(s_{\text{out}})). \quad (29)$$

多任务模型中对于每个任务定义一个单独的损失函数 L_1, L_2, \dots, L_n , 如式 (30) 所示, 通过加权求和和单个任务损失函数的方式得到多任务模型总的损失函数 L_{all} , 目的是避免多任务模型被某个任务所主导, 使用加权求和的方式将每个任务的损失函数的尺度统一^[10].

$$L_{\text{all}} = \sum_{i=1}^n a_i L_i, \quad (30)$$

根据任务类型 (分类、回归) 定义单个任务损失函数, 谣言检测以及情感检测任务都为二分类问题, 分类问题中, 通常使用交叉熵或其变种函数作为损失函数, 交叉熵对比于平方和等损失函数在分类问题中能更好地处理模型训练中的梯度爆炸或梯度消失的问题, 以及能更好地处理类间不平衡问题, 并且输出层的分类器函数为 softmax, softmax 会将分类预测结果映射为一个概率分布, 数据真实标签可以使用 one-hot 向量 (对应类别的位置上为 1, 其余位置上为 0 的向量) 表示概率分布, 而从 KL (Kullback Leibler) 散度中得到的交叉熵可以用来衡

量两个概率分布的距离, 两者结合能加快模型训练, 且训练时数值更加稳定. 则使用二元交叉熵损失函数作为谣言检测以及情感检测任务的损失函数.

如式 (31) 所示, 单个样本情况下情感检测的损失函数为 L_E , 其中 y_e 是数据真实的情感标签, p_e 为情感预测标签概率, 谣言检测的损失函数为 L_F , 如式 (32) 所示, 其中 y_f 是数据谣言或事实的真实标签, p_f 为谣言预测标签概率.

$$L_E = -y_e \log p_e - (1 - y_e) \log(1 - p_e), \quad (31)$$

$$L_F = -y_f \log p_f - (1 - y_f) \log(1 - p_f), \quad (32)$$

根据式 (30), 使用加权求和的方式统一谣言检测和情感检测两个任务损失函数的尺度, 将 L_E 和 L_F 的加权求和作为模型总损失函数 L_{all} , 如式 (33) 所示, 其中 a, b 为权重, $a + b = 1$, 在实验部分讨论其取值.

$$L_{\text{all}} = aL_E + bL_F. \quad (33)$$

3 实验结果与分析

3.1 数据集介绍与处理 如表 1 所示, Weibo-16^[23] 数据集是经过聚类算法^[5] 去重后的中文公共谣言检测数据集, 数据集中包含文章内容以及对应的用户评论, 每篇文章平均有 505 条评论. Twitter-15 数据集^[25] 通过收集 Twitter 社交平台上的数据得到, 每篇文章平均有 13 条评论. 数据集中每一篇文章都被标注为事实或谣言. 实验中以 3 : 1 : 1 的比例划分训练集、验证集、测试集, 预处理时对数据集进行删除文本标点, 多余字符清理等操作.

表 1 Weibo-16, Twitter-15 数据集分布

Tab. 1 Distribution table of Weibo-16 and Twitter-15 datasets

数据集	标签	Weibo-16		Twitter-15	
		文章数	评论数	文章数	评论数
训练集	真	1 410	482 226	222	2 484
	假	801	649 673	221	3 175
验证集	真	470	146 948	74	861
	假	268	222 149	73	1 050
测试集	真	471	179 942	76	958
	假	286	193 740	77	1 031

以上介绍的公共数据集缺乏情感标签, 如图 4 所示, 建立联合情感正负分类模型, 该模型输出是文章联合情感的正向或负向情感分类, 为数据集打上情感标签.

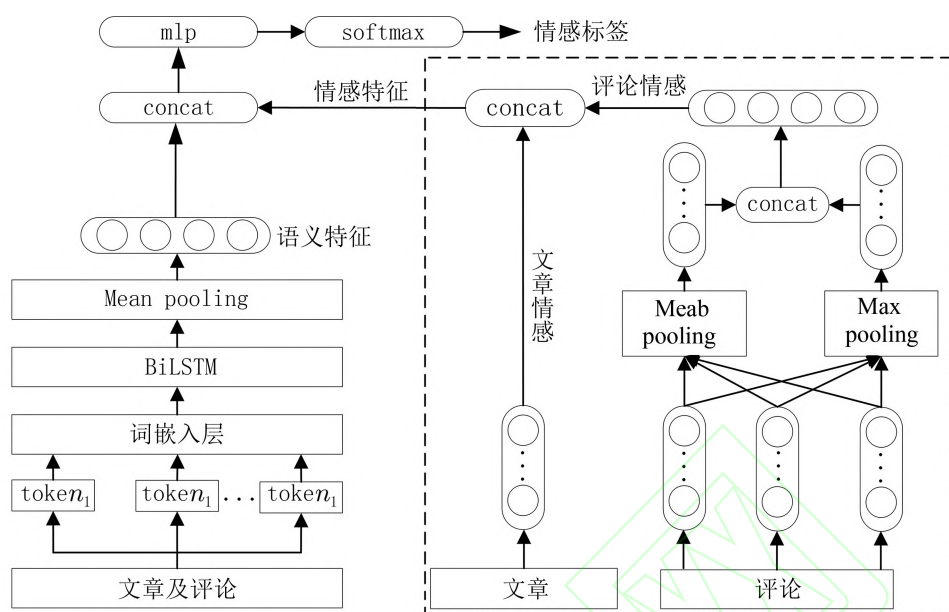


图 4 基于 BiLSTM 情感检测模型图

Fig. 4 Model diagram of emotion detection based on BiLSTM

如图 4 所示构建的模型分为语义特征提取部分和情感特征提取部分. 使用 BiLSTM 网络提取的特征经过平均池化后得到语义特征, 情感特征由 2.3 节介绍的情感提取网络提取, 情感特征提供了丰富的情感信息增强模型分类的准确性, 情感特征与语义特征拼接后输入到分类网络中得到联合情感分类输出. 针对中文使用带情感标签的 Weibo_senti_100k 数据集进行训练; 并将 Weibo-16 数据集作为测试集打上情感标签. 针对英文使用带情感标签的斯坦福 Twitter 情感数据集进行训练; 并将 Twitter-15 数据集作为测试集打上情感标签.

3.2 相关性分析实验 使用卡方统计显著性检验方法检验联合情感与文章属性(谣言或事实)的统计相关性, 参照文献 [5] 的方法, 定义文章真假与文章的情感正负两个变量, 两个变量的卡方自由度为 1, 通过计算两个变量的卡方统计量后对比卡方统计表中自由度为 1 的临界值, 判断两个变量之间是否具有相关性 [5, 30], 卡方统计量大于 95% 概率的临界值时变量相关, 卡方统计量大于 99% 概率的临界值时变量高度相关. 表 2 为数据集样本中两个变量取值的频数列联表, 表 2 的卡方统计量计算如式 (34) 所示 [30], 据式 (34) 计算 Twitter-15 和 Weibo-16 数据集的卡方统计量, Weibo-16 数据集中 $a=540$, $b=1353$, $c=795$, $d=998$ 计算出卡方统计量为 99.67 高于概率为 99% 的临界值 6.64, Twitter-15 数据集 $a=126$, $b=246$, $c=246$, $d=133$ 卡方统计量

为 68.15 高于概率为 99% 的临界值 6.64, 说明中文数据集和英文数据集中联合情感与文章是否是谣言在数量统计上高度相关.

表 2 频数列联表

Tab. 2 Frequency contingency table

情感	谣言	事实	总计
情感正向	a	b	$a+b$
情感负向	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

$$\chi^2 = \frac{(a+b+c+d) \cdot (ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}. \quad (34)$$

3.3 实验设置及评价指标 论文实验的软件集成环境为 Windows10 操作系统, Anaconda3 管理工具, IDE 为 Pycharm, 编程语言为 Python, 使用 GPU 版本的 tensorflow 2.9.1 深度学习框架. 硬件环境基于内存 64 GiB, CPU 处理器为 Intel(R)Core(TM)i7-10700FCPU@2.90 GHz, GPU 显卡为 NVIDIA GeForce RTX 3080.

中文预训练词嵌入为基于微博语料的 300 维词嵌入 [29], 英文预训练词嵌入为基于 Twitter 情感分析语料训练的 300 维词嵌入 [5], 编码器的 BiLSTM 的隐藏单元为 150, softmax 函数为分类器函数, 交叉熵损失作为单个任务损失函数, 使用 Adam 优化器, 学习率为 0.001, $L_{all} = aL_E + bL_F$ 中损失函数权

重 a 与 b 的值设置为 0.5, 提取情感特征时窗口长度为 3, Weibo-16 数据集中每篇文章选取对应的 300 条评论参与训练, Twitter-15 数据集中每篇文章选取对应的 13 条评论参与训练. 使用精确率(P)、召回率(R)、F1 值(F_1)、准确率(A_{cc})作为模型分类评价指标.

$$A_{cc} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \quad (35)$$

$$R = \frac{T_P}{T_P + F_N}, \quad (36)$$

$$P = \frac{T_P}{T_P + F_P}, \quad (37)$$

$$F_1 = \frac{2 \times P \times R}{P + R}, \quad (38)$$

式中: T_P 为谣言文章的正确分类数量, T_N 为非谣言文章的正确分类数量, F_P 为非谣言文章的错误分类数量, F_N 为谣言文章的错误分类数量.

3.4 对比模型介绍 为突出本文模型的优势, 选取目前谣言检测领域流行的机器学习, 单任务深度学习, 多任务深度学习谣言检测模型作为对比模型.

3.4.1 机器学习谣言检测模型 DT-Rank: Zhao 等^[20] 提出的基于决策树排名的谣言检测模型, 提出一些查询短语, 通过搜索查询短语和聚类有争议的声明, 并基于统计特征对聚类结果进行排名, 从排名结果中区分文章是谣言或是事实.

SVM-TS: Ma 等^[21] 提出的基于支持向量机的谣言检测模型, 使用 RFB 作为核函数, 采用时间序列为谣言信息的特征变化进行建模.

3.4.2 单任务深度学习谣言检测模型 Multi-layer GRU: Ma 等^[23] 提出的单任务谣言检测模型, 利用多层递归神经网络, 抽取文章深层语义特征用于谣言检测.

Dual Emotion GRU: Zhang 等^[5] 提出的单任务谣言检测模型, 将文章情感和用户评论情感, 以及两种情感之间的关系作为特征, 添加到谣言检测模型中提升模型性能.

3.4.3 多任务深度学习谣言检测模型 MT-trans-G-A: Wu 等^[18] 提出的多任务谣言检测模型, 使用立场检测任务用提升虚假谣言检测任务的性能.

PMTM: Kumari 等^[19] 提出的多任务谣言检测模型, 通过提取文章语义特征, 构建多任务模型, 同时实现谣言检测、情感检测、重复性检测等 3 个任务.

为获得对比模型较好的结果, 对比模型中参数与其原文保持一致. 基于模型提取特征的不同, 作以下相关说明: (1) 基于统计的手工构建特征的模型: DT-Rank、SVM-TS; (2) 仅提取文章内容特征的模型: Multi-layer GRU、PMTM、MT-trans-G-A; (3) 提取文章内容以及评论特征的模型: Dual Emotion GRU, JEMT.

3.5 实验结果及分析

3.5.1 对比实验 图 5 和图 6 的实验结果表明, 提出的 JEMT 模型在 Weibo-16 与 Twitter-15 数据集上各个指标上都优于其它模型. 机器学习谣言检测模型效果弱于提取特征能力更强的深度学习模型, JEMT 模型在 Weibo-16 数据集中相较于 DT-Rank 和 SVM-TS 模型的准确率分别提高了 20.5 和 11.9 个百分点, 提升效果较为明显. 在实验的数据样本较少情况下, 提出的 JEMT 模型以及 MT-trans-G-A、PMTM 等多任务模型表现优于单任务模型, JEMT 模型在 Weibo-16 数据集中相较于 Multi-layer GRU 和 Dual Emotion GRU 模型的准确率分别提高了 8.5 和 6.6 个百分点, Twitter-15 数据集中分别提高了 10.7 和 5.5 个百分点, F1 值也有明显提升, 这表明多任务模型在标签数据较少的情况下通过共享参数平均噪声能获得更好的表现. JEMT 模型性能优于其余两个对比的多任务模型, 主要原因是对文章语义特征和联合情感特征的深层次提取以及对辅助任务的选取, 选取情感检测作为辅助任务的 JEMT 模型效果强于选取立场检测作为辅助任务的 MT-trans-G-A 模型, JEMT 模型在 Weibo-16 数据集中相较于 MT-trans-G-A 模型准确率提高了 5.1 个百分点, F1 值提高了 4.4 个百分点, 这说明选取更相关的任务是构建多任务谣言检测模型的关键, 相较于立场检测任务, 情感检测任务提供了用户对于信息的态度和行为, 有助于更准确地判断是否存在谣言. 而 PMTM 模型虽然使用了 3 个任务同时构建模型, 但只对文章内容的特征进行提取, 忽略了用户评论情感信息和语义信息, 损失了大量的有用特征, 导致效果弱于 JEMT 模型, JEMT 模型在 Weibo-16 和 Twitter-15 数据集中相较于 PMTM 模型准确率提高了 3.5 和 2.9 个百分点, F1 值提高了 3.1 和 3.9 个百分点, 实验结果表明了 JEMT 模型在实验数据集上是有效的.

3.5.2 消融实验 为验证模块有效性, 将 JEMT 分解为下面 3 个简化模型:

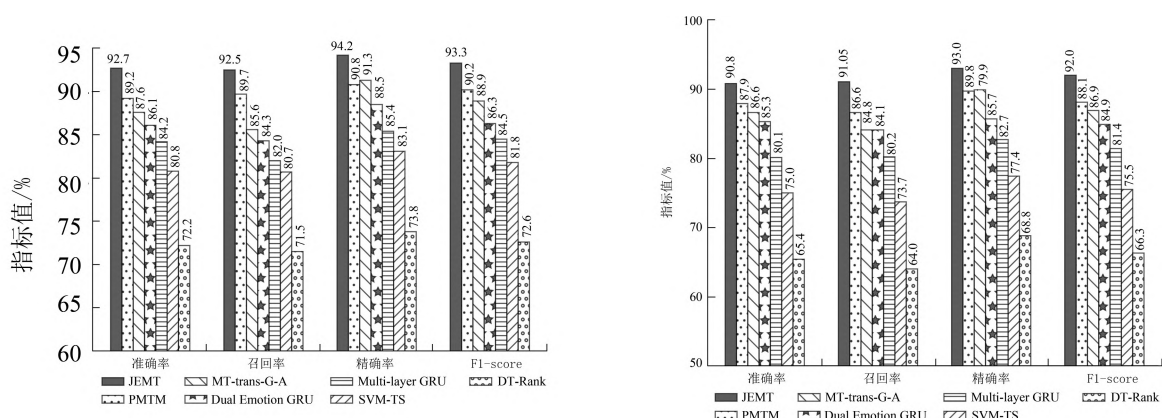


图 5 Weibo-16 数据集模型性能比较图

Fig. 5 Comparison chart of model performance on Weibo-16 dataset

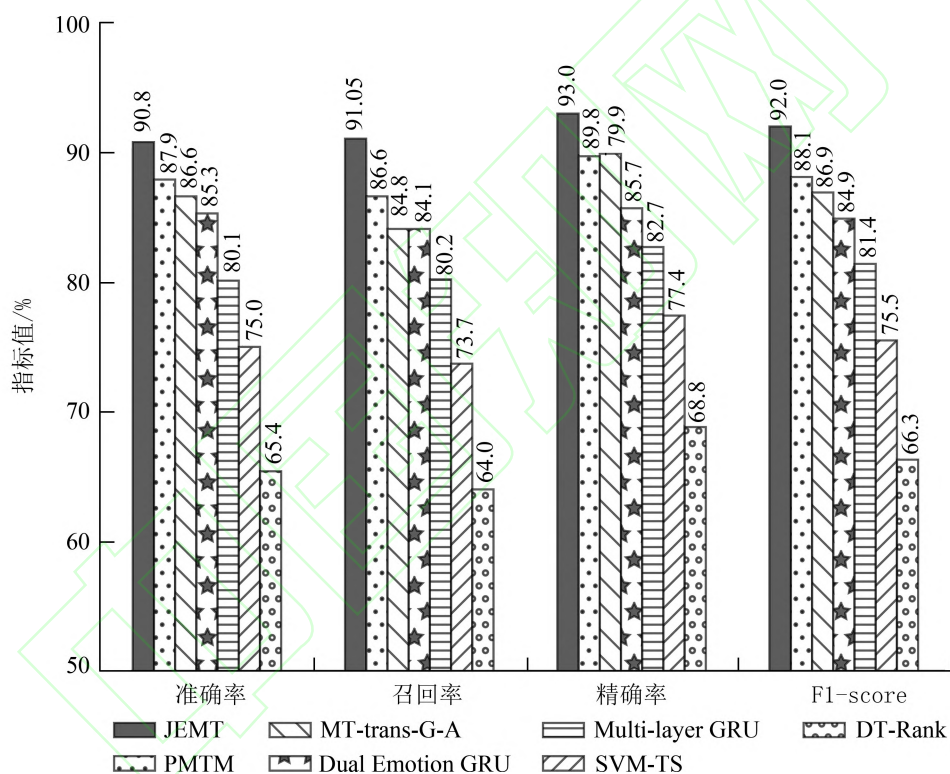


图 6 Twitter-15 数据集模型性能比较图

Fig. 6 Comparison chart of model performance on Twitter-15

(1) JEMT_T: 仅对文章提取语义特征, 进行单任务训练.

(2) JEMT_S: 对文章提取情感特征和语义特征, 进行单任务训练.

(3) JEMT_E: 对文章及其用户评论提取情感特征和语义特征, 进行单任务训练.

(4) JEMT: 本文提出的模型.

实验结果如图 7 和图 8 所示, JEMT_S 的性能在准确率、召回率、精确率, F1 值都优于 JEMT_T, 说明情感特征是进行谣言检测的重要特征, 而 JEMT_

E 的性能强于 JEMT_S, 体现出谣言检测中用户评论带有大量的有用信息, 不可忽视. JEMT_E 通过将单任务训练变为多任务训练就可得到本文提出的模型, 模型的性能也得到了提升, 说明了多任务在标签数据较少情况下的优越性.

3.6 实验主要参数分析

3.6.1 评论数量分析 模型训练时需要提取文章和评论的语义特征以及情感特征, Weibo-16 数据集中, 文章的平均长度为 116 个字符, 每条评论的平均长度为 16 个字符, 每篇文章对应的评论平

均为 505 条; Twitter-15 数据集中, 文章的平均长度为 31 个字符, 每条评论的平均长度为 12 个字符, 每篇文章对应的评论平均为 13 条. 文本总体长度更长的评论比社交媒体文章携带的信息更为丰富, 参与模型训练的评论的数量成为关键因素, 如果评论数量太少, 不能很好地反映出用户的情感和观点, 难以提升谣言检测的准确度和可靠性, 而过多的评论数量也会影响谣言检测的效果, 因为一些评论可能是不相关或者是重复的, 并且过多的评论数量会让模型计算量大大增加.

在实验中, 对于 Weibo-16 数据集中的每篇文章, 抽取了不同数的评论参与到模型训练, 评论数

为 50、100、150、200、250、300; 对于 Twitter-15 数据集中的每篇文章, 抽取不同数量的评论参与模型训练, 评论数为 3、5、7、9、11、13. 实验过程中, 如果某篇文章的评论数未达到实验设定数量, 将使用文章的所有评论进行实验. 如图 9(a) 所示, Weibo-16 数据集评论基数较大, 对模型改进的效果相对较小. 如图 9(b) 所示, Twitter-15 数据集由于其评论基数较少, 改进相对较为明显. 实验结果说明评论数量增加到一定量时对模型的改进效果会逐渐下降. 增加评论数量会带来巨大的计算量. 因此在合理范围内选取评论数量以及选取最相关和有质量的评论才能保持模型的最佳效果. 如何构建合适的

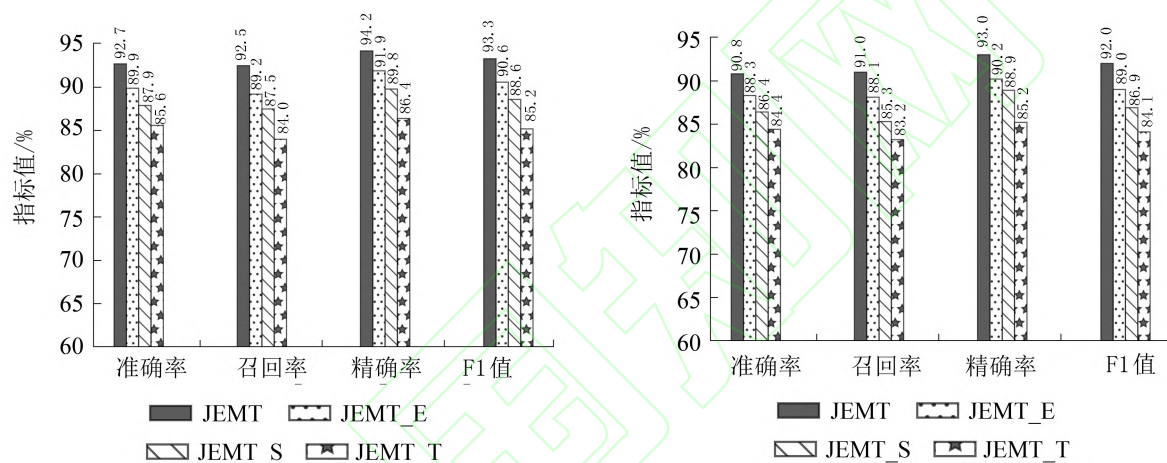


图 7 Weibo-16 数据集消融实验结果图

Fig. 7 Ablation experiment results for Weibo-16 dataset

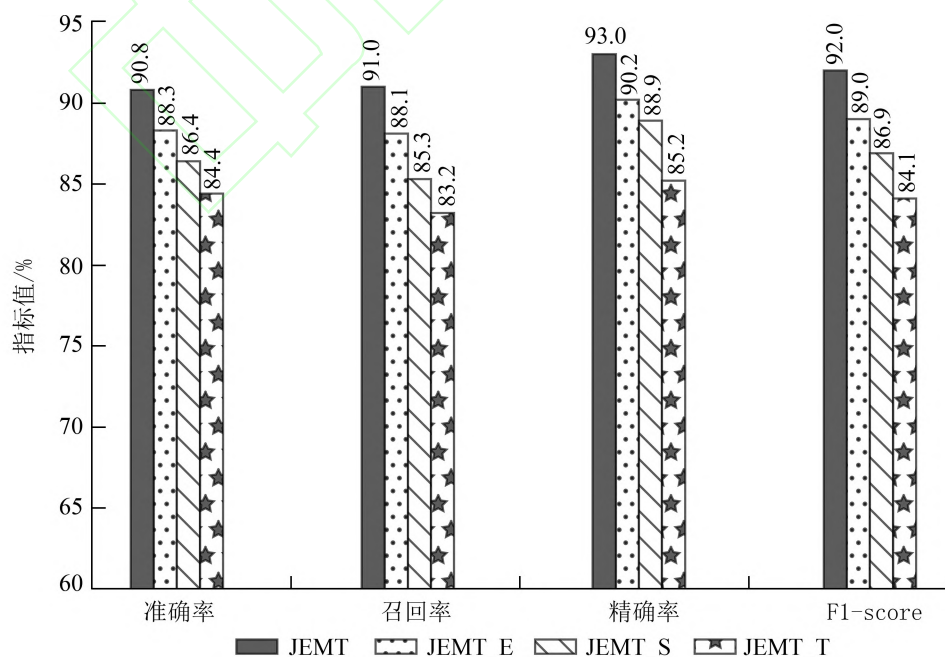


图 8 Twitter-15 数据集消融实验结果图

Fig. 8 Ablation experiment results for Twitter-15 dataset

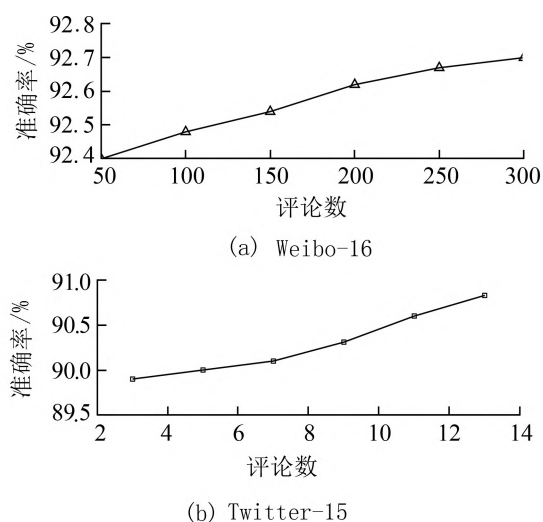


图 9 评论数量与准确率关系图

Fig. 9 Relationship between number of comments and accuracy

算法筛选评论放到未来的工作中. 随着评论数量的增加, 模型的效果有细微的改进.

3.6.2 损失函数权重 由公式 (33) $L_{oss} = aL_E + bL_F$ 可知, 联合情感检测任务以及谣言检测任务的损失函数以权重系数 a 和 b 加权和为总损失函数, 联合情感检测任务损失函数权重系数 a 取值范围为 0.1 到 1.0, 以 0.1 为间隔, 对应权重系数 b 为 $1-a$; 得到的结果如图 10 所示, 当权重为 1:1 时在英文和中文数据集上都能取得最好的结果; 任何一个任务赋予较少的权重都会降低整体性能. 这是由于模型在设计时谣言检测任务和联合情感检测任务的参数是同时更新导致.

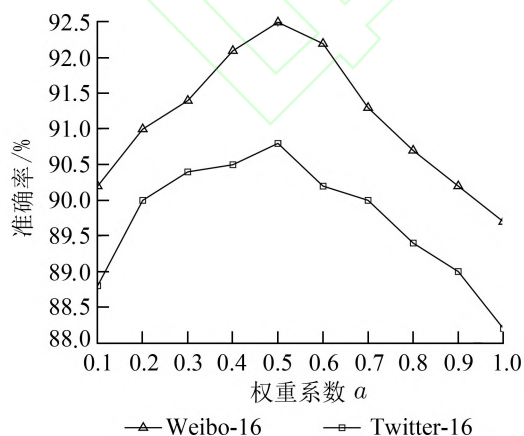


图 10 权重系数与准确率关系图

Fig. 10 Relationship between weight coefficient and accuracy

4 结论

现有的谣言检测方法使用情感辅助谣言检测

时存在两个局限性, 一是只考虑文章本身情感, 忽略了用户评论情感; 二是难以提升标签数据较少的谣言检测领域的准确率. 针对上述问题, 本文提出一个基于联合情感的多任务谣言检测模型, 首先, 使用 BiLSTM 网络处理文章与评论的语义特征, 并将网络隐藏状态变量送入下一层 BiLSTM 网络中将文章与评论的语义特征映射到同一语义空间中, 得到融合语义特征, 并使用注意力机制得到融合语义特征的词向量表示; 其次, 使用情感词典与公共情感分类器构成的情感提取网络提取文章和评论情感特征; 最后, 将词向量表示与情感特征融合后输入到两个分类网络实现谣言检测与联合情感检测. 实验结果表明, 与流行的谣言检测模型相比, 该模型表现得更好, 通过消融实验验证了模型各个模块的有效性. 在未来的工作中, 可以优化多任务谣言检测模型的信息共享层来优化任务之间的参数共享程度, 保留有效特征, 让模型更有效. 其次可以构建算法以筛选出更具代表性的评论参与模型训练, 以提升模型性能并减少计算复杂度.

参考文献:

- [1] Scheufele D A, Krause N M. Science audiences mis-information and fake news[J]. *Proceedings of the National Academy of Sciences*, 2019, 116(16): 7 662-7 669. DOI: [10.1073/pnas.1805871115](https://doi.org/10.1073/pnas.1805871115).
- [2] Sivandzadeh G R, Askari H, Safarpour A R, et al. COVID-19 infection and liver injury: Clinical features, biomarkers, potential mechanisms, treatment, and management challenges[J]. *World Journal of Clinical Cases*, 2021, 9(22): 6 178-6 200. DOI: [10.12998/wjcc.v9.i22.6178](https://doi.org/10.12998/wjcc.v9.i22.6178).
- [3] Ajao O, Bhowmik D, Zargari S. Sentiment aware fake news detection on online social networks[C]// I-CAS-SP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 2019: 2 507-2 511.
- [4] Giachanou A, Rosso P, Crestani F. Leveraging emotional signals for credibility detection[C]//Proceedings of the 42nd International ACM Sigir Conference on Research and Development in Information Retrieval, Paris, France, 2019: 877-880.
- [5] Zhang X, Cao J, Li X, et al. Mining dual emotion for fake news detection[C]//Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 2021: 3 465-3 476.
- [6] Wang W Y. "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection[C]//Proceedings

- of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017: 422-426.
- [7] 葛晓义, 张明书, 魏彬, 等. 基于双重情感感知的可解释谣言检测[J]. 中文信息学报, 2022, 36(9): 129-138.
Ge X Y, Zhang M S, Wei B, et al. Dual emotionaware method for interpretable rumor detection[J]. Journal of Chinese Information Processing, 2022, 36(9): 129-138.
- [8] 谭艳霞, 化存才. 网络舆情反转问题的模糊聚类分析[J]. 云南大学学报(自然科学版), 2019, 41(S1): 16-20. DOI: [10.7540/j.ynu.20170800](https://doi.org/10.7540/j.ynu.20170800).
Tan Y X, Hua C C. Analysis of cluster on the inversion problem of network public opinion events[J]. Journal of Yunnan University (Natural Sciences Edition), 2019, 41(S1): 16-20.
- [9] Ghanem B, Rosso P, Rangel F. An emotional analysis of false information in social media and news articles[J]. ACM Transactions on Internet Technology, 2020, 20(2): 1-18.
- [10] Zhang Y, Yang Q. A survey on multi-task learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 32(12): 5 586-5 609.
- [11] 张钰, 刘建伟, 左信. 多任务学习[J]. 计算机学报, 2020, 43(7): 1 340-1 378.
Zhang Y, Liu J W, Zuo X. Survey of multi-task learning[J]. Chinese Journal of Computers, 2020, 43(7): 1 340-1 378.
- [12] Singh A, Saha S, Hasanuzzaman M, et al. Multitask learning for complaint identification and sentiment analysis[J]. *Cognitive Computation*, 2022, 14(1): 212-227. DOI: [10.1007/s12559-021-09844-7](https://doi.org/10.1007/s12559-021-09844-7).
- [13] Abhadiomhen S E, Nzeh R C, Ganaa E D, et al. Supervised shallow multi-task learning: analysis of methods[J]. *Neural Processing Letters*, 2022, 54(3): 2 491-2 508. DOI: [10.1007/s11063-021-10703-7](https://doi.org/10.1007/s11063-021-10703-7).
- [14] Abdul M M, Ungar L. Emonet: Fine-grained emotion detection with gated recurrent neural networks[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017: 718-728.
- [15] Liew J S Y, Turtle H R. Exploring fine-grained emotion detection in tweets[C]// Proceedings of the NAACL Student Research Workshop, San Diego, California, USA, 2016: 73-80.
- [16] Ren G, Hong T. Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews[J]. Information Processing & Management, 2019, 56(4): 1 425-1 438.
- [17] Kochkina E, Liakata M, Zubiaga A. All-in-one: Multi-task learning for rumour verification[C]// Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, 2018: 3 402-3 413.
- [18] Wu L W, Yuan R, Jin H L, et al. Different absorption from the same sharing: Sifted multi-task learning for fake news detection[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hongkong, China, 2019: 4 644-4 653.
- [19] Kumari R, Ashok N, Ghosal T, et al. Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition[J]. Information Processing & Management, 2021, 58(5): 102631-102 646.
- [20] Zhao Z, Resnick P, Mei Q. Enquiring minds: Early detection of rumors in social media from enquiry posts[C]// Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 2015: 1 395-1 405.
- [21] Ma J, Gao W, Wei Z, 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, Australia, 2015: 1 751-1 754.
- [22] Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for fake news detection [C]// The World Wide Web Conference, San Francisco, USA, 2019: 2 915-2 921.
- [23] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C]// International Joint Conference on Artificial Intelligence, New York, USA, 2016: 3 818-3 824.
- [24] 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 AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018: 354-361.
- [25] Ma J, Gao W, Wong K F. Rumor detection on twitter with tree-structured recursive neural networks[C]// Association for Computational Linguistics, Melbourne, Australia, 2018: 1 980-1 989.
- [26] Guo H, Cao J, Zhang Y, et al. Rumor detection with hierarchical social attention network[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 2018: 943-951.
- [27] Li Q, Zhang Q, Si L. Rumor detection by exploiting user credibility information, attention and multitask

- learning[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019: 1 173-1 179.
- [28] Shu K, Wang S, Liu H. Beyond news contents: The role of social context for fake news detection[C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 2019: 312-320.
- [29] Li S, Zhao Z, Hu R F, et al. Analogical reasoning on Chinese morphological and semantic relations[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018: 138-143.
- [30] 陆运清. 相关四格表数据检验需要注意的几个问题[J]. 统计与决策, 2022, 38(7): 31-34.
- Lu Y Q. Some problems needing attention in data inspection of relevant fourfold table[J]. Statistics & Decision, 2022, 38(7): 31-34.

Multi-task microblog rumor detection based on joint emotion analysis

MA Yi¹, SHAO Yu-bin^{1**}, DU Qing-zhi¹, LONG Hua^{1,2}, MA Di-nan²

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;

2. Yunnan Provincial Key Laboratory of Media Integration, Kunming 650032, Yunnan, China)

Abstract: Sentiment analysis plays a pivotal role in the detection of rumors on social media. Existing rumor detection methods predominantly rely on the emotional features of articles, often neglecting the combined sentiment emanating from user comments and the articles themselves. Single-task deep learning models for rumor detection face challenges due to a shortage of labeled data, hindering substantial improvements in accuracy. To address these challenges, this paper introduces a multi-task joint learning approach for rumor detection, leveraging the inherent correlation between rumor detection and joint sentiment detection tasks. This innovative method integrates joint sentiment detection and rumor detection tasks within the same model. Initially, an encoder is constructed to extract semantic features from articles and their corresponding comments, mapping them into a unified semantic space. An attention mechanism is employed to meticulously blend and weigh these semantic features. Subsequently, an emotion extraction network, founded on a common emotion classifier and emotion lexicon, extracts emotional features from both articles and corresponding comments. Finally, the semantic features and joint emotional features are fused and fed into two shared loss function classifiers, yielding classification results for both rumor detection and joint sentiment detection. Experimental results demonstrate the superior performance of the multi-task model over its single-task counterpart. When evaluated on the publicly available Chinese Weibo-16 dataset and English Twitter-15 dataset, the proposed model exhibits a notable improvement of 3.5 and 2.9 percentage points in accuracy, along with a commendable increase of 3.1 and 3.9 percentage points in F1 score, respectively.

Key words: rumor detection; user comments; joint emotion; multi-task learning