韩晓鸿, 郭恒, 杨港. 基于 XLNet-BiLSTM-Attention 模型的假新闻检测研究[J]. 智能计算机与应用,2025,15(9):96-100.
DOI:10.20169/j.issn.2095-2163.250915

基于 XLNet-BiLSTM-Attention 模型的假新闻检测研究

韩晓鸿, 郭 恒, 杨 港
(河北工程大学 信息与电气工程学院, 河北 邯郸 056038)

摘 要: 随着社交多元化和网络技术的发展,网络上开始出现虚假新闻,给个人和社会造成了不利的影响。针对此现象,本文提出基于 XLNet-BiLSTM-Attention 神经模型的检测方法。首先使用 XLNet 获取具有上下文依赖的词向量,然后通过 BiLSTM 双向门控单元获取深层次语义信息,最后利用 Attention 机制根据特征的重要性赋予不同的特征权重,并进行文本真实性检测。本文模型与 4 种常用神经模型进行对比,准确率达到 94%,均高于其他 4 种模型,从而验证了该模型的有效性。

关键词: 假新闻检测; 神经模型; XLNet; BiLSTM; 特征权重
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2025)09-0096-05

Research on fake news detection based on XLNet-BiLSTM-Attention modeling

HAN Xiaohong, GUO Heng, YANG Gang
(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, Hebei, China)

Abstract: With social diversification and the development of online technology, fake news began to appear on the network, causing adverse effects on individuals and society. Faced with this phenomenon, this paper proposes a detection method based on the XLNet-BiLSTM-Attention neural model. Firstly, XLNet is employed to obtain word embeddings with contextual dependencies. Then, the BiLSTM bidirectional gated units are utilized to capture deep semantic information. Finally, the Attention mechanism is utilized to assign different feature weights according to the importance of the features and perform text authenticity detection. The proposed model is compared with four commonly used neural models, achieving an accuracy of 94%, surpassing the other four models. This reults confirms the effectiveness of the proposed model.

Key words: fake news detection; neural model; XLNet; BiLSTM; feature weights

0 引 言

近年来,随着互联网的迅速发展普及,人们不时就会面临由认知偏见、误导性言论和虚假信息带来的困扰[1]。虚假新闻的传播对社会发展也造成了不良的影响。因此,对虚假新闻检测展开研究既可以维护公共利益和社会稳定,保护好个人权益,也能避免做出错误的决策,从而导致个人或集体遭受不必要的损失。

为此,本文提出了基于 XLNet-BiLSTM-Attention 的自动假新闻检测模型。其中,XLNet 通过排列组合获得包含句子上下文信息的词向量,有效地改善了 Bert 在 Fine-tune 阶段带来的信息误差,然后再利用 BiLSTM 提取文本的全局特征,接着通过 Attention 机制,赋予不同的权重,最后通过 Softmax 激活函数进行判断真假。

1 相关工作

文献[2]详细描述了 WoS 核心库中关于假新闻检测领域的论文发表趋势。研究指出,随着时间的推移,假新闻研究已然取得了多项可观成果。

国内外研究学者针对社交平台上的假新闻检测已开展了丰富的研究工作,大部分研究将其视作有监督的分类问题,通过带标签的数据进行训练,从不同的角度对假新闻检测任务进行建模,取得了不错的成效。早期传统机器学习方法[3-4]和目前更为主流的深度学习方法在假新闻检测研究中发挥着重要作用。在 2021 年,Shifath 等学者[5]提出了一种基于 Transformer 的方法来检测 COVID-19 假新闻。实验准确率最高达到了 97.9%。

作者简介: 韩晓鸿(1972—),女,副教授,主要研究方向:人工智能,数据库。Email:1281999898@qq.com; 郭 恒(1997—),男,硕士研究生,主要研究方向:自然语言处理; 杨 港(1998—),男,硕士研究生,主要研究方向:自然语言处理。

收稿日期: 2023-12-25

特征有普通特征(字词频率[6]、符号、情绪词[7]等)和聚合特征(即 2 个或 2 个以上的普通特征之间的融合)之分[5,8]。Zhang 等学者[9]分别提取了微博发文和评论的情感特征,并通过实验证明这些情感特征可以和多个模型结合,作为一种增强模型表现的手段,提高模型检测性能。孙王斌[10]构建了包括符号特征、情感特征、有效度特征、敏感度特征和热度特征在内的浅层特征,采用 LSTM 提取深层特征,并使用 SVM 进行特征拟合,显著提高了谣言检测的准确率。文献[11-13]探索了 Transformer 架构的变种在虚假新闻检测任务中的有效作用。还有研究者将注意力(Attention)机制引用到谣言检测的研究中[14-16],利用注意力机制给提取出的特征分配权重,有效提高了谣言识别的精度。韩晓鸿等学者[17]提出一种基于元路径的推文-词-用户异质图卷积注意力框架。HGCAN 通过图卷积网络提取文本内容特征,利用 Attention 机制聚合邻居节点的信息并学习子图重要性,进而有效学习节点的特征表示。

2　XLNet-BiLSTM-Attention 模型研究

2.1　研究框架设计

XLNet-BiLSTM-Attention 模型主要由 3 部分组成。本文模型如图 1 所示。

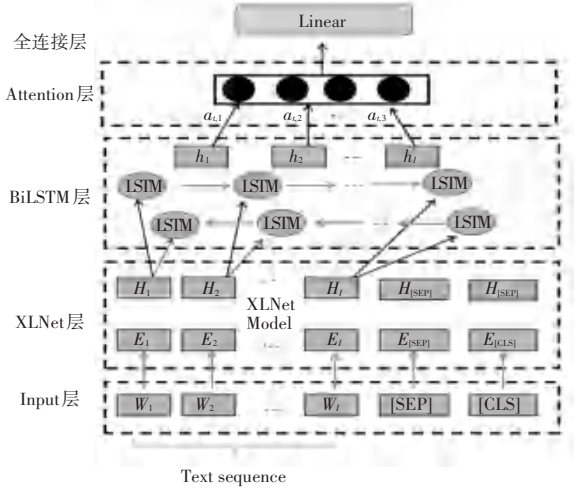

图 1　模型结构图

Fig. 1　Diagram of model structure

2.2　XLNet 提取特征向量

2.2.1　XLNet 模型解析

XLNet[18]模型是由国外研究院所在 2019 年提出。XLNet 不仅改进了 BERT 训练和调试中的依赖性弱点和不一致性问题,而且提出了 PLM 排列语言模型训练方式,并使用 AR 来预测最后几个标记。例如,单词预测如图 2 所示。图 2 中,对于句子“I am a student”,PLM 重新将该句子进行随机排序,生成一个新句子“student I a am”,当预测“a”时就需要用到“student I”。此时并没有用到“am”。当 PLM 产生的一个新序列中“am”是“a”的上文内容的单词时,预测“a”才会用到“am”。

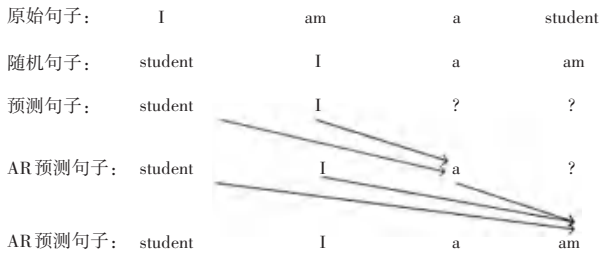

图 2　单词预测

Fig. 2　Word prediction

2.2.2　双流自注意力机制

如果用传统的注意力机制计算 PLM,模型就无法看到被遮掩的词的位置信息。为解决此问题,XLNet 采用双流自注意力机制将位置信息 g^{θ} 加入 AR 模型的目标函数中,推得公式如下:

$$P_{\theta}(X_{z_t} \mid X_{z<t}) = \frac{\exp(e(X)^{\mathrm{T}} g^{\theta}(X_{z<t'} z_t))}{\sum_{x'} \exp(e(x')^{\mathrm{T}} g^{\theta}(X_{z<t'} z_t))} \quad (1)$$

“双流”即 Query stream 和 Content stream。其中,Query stream 可以看到当前词的位置信息、不能看到其内容信息,而 Content stream 既可以看到当前词的内容信息、也可以看到其位置信息。“双流”的更新公式分别如下:

$$g_{zt}^{m} \leftarrow \mathrm{Attn}(Q = g_{zt}^{m-1}, KV = h_{z<t}^{m-1}; \theta) \quad (2)$$

$$h_{zt}^{m} \leftarrow \mathrm{Attn}(Q = h_{zt}^{m-1}, kv = h_{z\leqslant t}^{m-1}; \theta) \quad (3)$$

其中,g 表示查询隐状态;h 表示内容隐状态;m 表示 XLNet 的层数;Q 表示查询向量 Query;K 表示待查向量 Key;V 表示内容向量 Value。Q、K、V 通过 Linear 得到其对应的矩阵。完整的双流自注意力机制实现原理如图 3 所示。

假设上述“I am a student”为 $[x_1, x_2, x_3, x_4]$、当预测 x_3 时,模型能获得 x_1、x_2 和 x_4 的信息。其中,在图 3 的 Content Stream 流中,同时对 X_1 位置信息和内容信息进行了编码,在 Query stream 流中对 X_3 本身位置和 X_3 上下文信息进行了编码,最终 XLNet 模型输出层的词向量为:$H = \{H_1, H_2, \cdots, H_l\} \in R^{l \times d_h}$,这里 d_h 表示维度。取值为 $H = \mathrm{XLNet}(E)$。

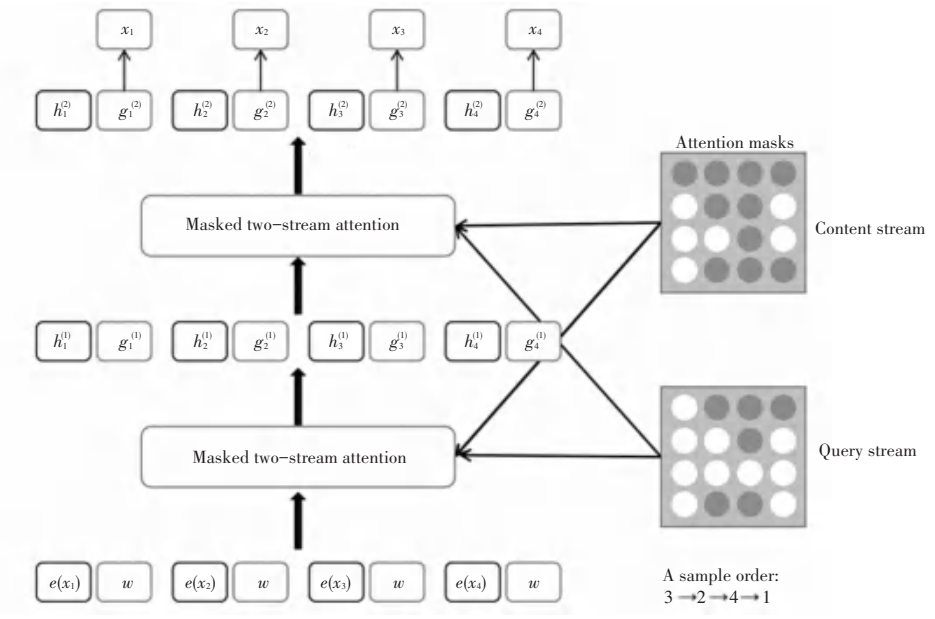

图 3 双流自注意力机制

Fig. 3 Double-stream self-attention mechanism

2.3 BiLSTM 获取上下文信息

通过 BiLSTM 文本特征层提取文本之后,可以更加充分地学习文本上、下文关系。该网络是 RNN 的改进版,其设计结构如图 4 所示。图 4 中,设置了 3 种门控机制:遗忘门、输入门和输出门。分析可知,遗忘某些内容是必要的,比如一条新闻可能涉及到很多主题,此时在建模时遗忘来自主题的部分输入是必要的。调整门的作用是模拟不同节点类别之间的信息变化,例如从创建者到文章之间的信息流变化。选择门的作用是控制输入/状态向量的不同组合。通过对遗忘和信息传递的选择性处理,有效地解决了梯度损失的问题,从而捕捉到取决于文本序列距离的信息。

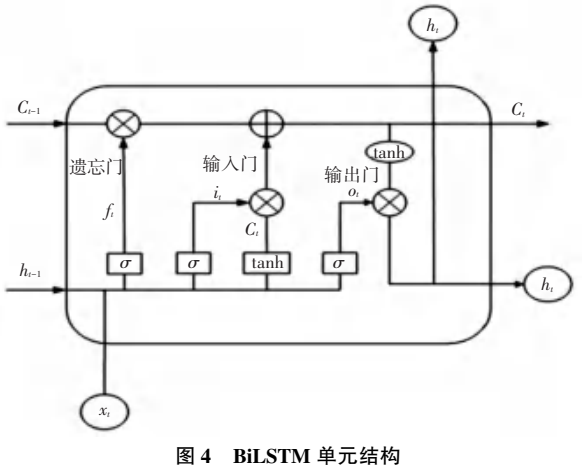

图 4 BiLSTM 单元结构

Fig. 4 Unit structure of BiLSTM

在此基础上,研究给出了输入门、输出门和遗忘门的 3 种阈值的具体计算公式为:

$$f_t = \sigma(W_f \cdot (h_{t-1}, x_t) + b_f) \tag{4}$$

$$i_t = \sigma(W_i \cdot (h_{t-1}, x_t) + b_i) \tag{5}$$

$$o_t = \sigma(W_o \cdot (h_{t-1}, x_t) + b_o) \tag{6}$$

其中,W_f 表示在对 f_t 进行计算时的权重;W_i 表示在计算 i_t 时的权重;W_o 表示计算 o_t 时的权重。

遗忘阈值负责计算要遗忘的信息,输入阈值计算要更新的数据信息,输出阈值负责决定输出信息。对于内部状态信息进行遗忘与更新,对此可表示为:

$$\tilde{C}_t = \tanh(W_c \cdot (h_{t-1}, x_t) + b_c) \tag{7}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{8}$$

其中,W_c 表示计算 $\tilde{C}_t$ 时的权重,是当处在 t 时刻情况下,tanh 所创建的新向量;C_t 表示在时刻更新之后所呈现出的状态信息;h_t 表示最后隐藏层信息输出。

计算最终的隐藏层输出公式为:

$$h_t = o_t \cdot \tanh(C_t) \tag{9}$$

BiLSTM 层的输入 H 是经过 XLNet 层进行 token 之后所获得的句向量,能够针对输入的句向量来获取全局特征,继而对被测新闻上下文信息进行更为全面完整的学习。当处在 t 时刻的情况下,将正向 $\overleftarrow{h}_t$ 与反向 $\overrightarrow{h}_t$ 拼接起来作为 Attention 的输入 $h = \{h_1, h_2, \cdots, h_l\} \in R_{l \times d_h}$,其中 $\overleftarrow{h}_t$ 与 $\overrightarrow{h}_t$ 的数学定义公式可写为:

$$\overrightarrow{h}_t = \overrightarrow{\mathrm{LSTM}}(x_t, \overrightarrow{h}_{t-1}) \tag{10}$$

$$\overleftarrow{h}_t = \overleftarrow{\mathrm{LSTM}}(x_t, \overleftarrow{h}_{t+1}) \tag{11}$$

2.4　Attention 赋予权重

Attention 模块主要是对 BiLSTM 层输出的隐藏层向量进行加权计算。首先将 BiLSTM 层的输出向量输入到 Atention 层，再将输出结果输入到单层感知机中得到隐含表达，然后通过对比与上下文向量的相似性来判定该单词是否为主要单词。最后，利用词向量加权求和计算的方法，实现文本级别的局部特征关注。计算流程如图 5 所示。

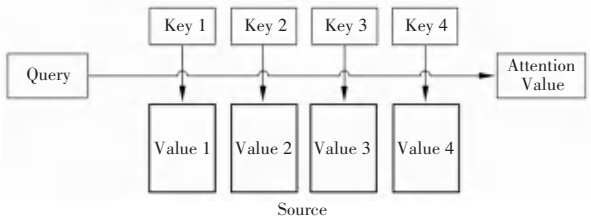

图 5　注意力机制计算流程

Fig. 5　Computation process of attention mechanism

利用注意力机制从全局角度获取新闻文本联系，并关注局部关键文本联系的特点，同时结合优秀的处理冗余信息和信息丢失问题的能力，从而进一步强化文本的语义特征。具体计算公式如下：

$$\boldsymbol{u}_i = \tanh(\boldsymbol{W}\boldsymbol{h}_i + \boldsymbol{b}) \tag{12}$$

$$\alpha_t = \frac{\exp(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_w)}{\sum_{i=1}^{t}\exp(\boldsymbol{u}_i\boldsymbol{u}_w)} \tag{13}$$

$$\boldsymbol{v} = \sum_{i=1}^{t}\alpha_i\boldsymbol{h}_i \tag{14}$$

其中，$\boldsymbol{h}_i$ 表示输入隐向量；tanh 表示双曲正切激活函数；$\boldsymbol{u}_i^{\mathrm{T}}$ 表示训练所得参数向量的转置；t 表示语句序列长度；$\boldsymbol{v}$ 表示利用 tanh 激活函数计算得到的最终句子表示。

3　实验与分析

3.1　数据集

本数据集采用 Ma 等学者[19-20]在其研究中构建的微博谣言数据集 Weibo21。该数据集存在特殊符号与表情等问题，本研究对数据集进行了过滤和清洗，最终得到 1 609 条假新闻、1 076 条真实新闻。

3.2　实验参数

本文模型实验环境和模型参数配置见表 1、表 2。

表 1　实验环境配置

Table 1　Experimental environment configuration

实验环境	具体配置
CPU	Intel Platinum 8255C
内存	128 G
GPU	RTX3080
编程语言	Python3.6

表 2　模型参数配置

Table 2　Model parameter configuration

参数	值
Hidden_dim	384
Activation	ReLU
Loss	CrossEntropy
epoch	10
Batch_size	64
学习率(lr)	0.000 02
Dropout	0.2
Optimizer	Adam

3.3　对比实验分析

表 3 为 XLNet-BiLSTM-Attention 模型与其他 4 种模型通过精准率来评判模型的优劣。

表 3　模型比较

Table 3　Comparison of the models

模型	精准率/%
本文模型	94.0
XLNet	86.2
XLNet-CNN	87.3
XLNet-BiLSTM	91.4
W2V-BiLSTM-Attention	89.6

在对比实验中，本文提出的模型精准率高于其他模型，本文模型与第 4 组的性能比较体现在加入 Attention 机制之后。因为 Attention 机制会给重要特征赋予较大的权重，更加注重突出重要特征的影响，从而可以有效提升模型的准确率。从第 3 组与第 4 组来看，BiLSTM 明显比 CNN 发挥出色，因为 BiLSTM 能够注意到文章的上下文的全部信息，而 CNN 只能注意到文本的局部信息。从本文模型和最后一组的对比来看，由于 XLNet 能获得融合上下文的词向量，并且能表示一词多义，即相同的词在不同的语境能够获得不同的词向量，而 Word2Vec 对相同词语的不同语境产生相同的词向量，所以最后一组模型的精准率要低于本文模型。

4　结束语

本文提出的融合 XLNet-BiLSTM-Attention 的谣言测方法，为假新闻检测领域的发展提供了新的路线和技术支持。

由于如今的新闻分类有许多，单一种类的假新闻检测很难满足多模态应用需求，所以今后将考虑假新闻检测细粒度划分，并对模型进行测试，提高适用性能。

参考文献

[1] RIZOIU M A , GRAHAM T , ZHANG Rui,et al. DebateNight: The role and influence of socialbots on twitter During the 1st 2016 U. S. Presidential[J]. arXiv preprint arXiv,1802. 09808,2018.

[2] 汝绪华. 国外假新闻研究: 缘起、进展与评价[J]. 新闻与传播评论, 2019, 72(5): 58-70.

[3] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675-684.

[4] YANG Fan,LIU Yang,YU Xiaohui,et al. Automatic detection of rumor on Sina Weibo[C]// Proceedings of 2012 ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: 91-97.

[5] SHIFATH S M, KHAN M F, ISLAM M S. A transformer based approach for fighting COVID-19 fake news[J]. arXiv preprint arXiv,2101. 1207,2021.

[6] GUO Chuan, CAO Juan, ZHANG Xueyao, et al. Exploiting emotions for fake news detection on social media[J]. arXiv preprint arXiv,1903. 01728, 2019.

[7] SILVA F C D D, ALVES R V D C, GARCIA A C B. Can machines learn to detect fake news? a survey focused on social media [C]//Proceedings of the 52nd Hawaii International Conference on System Sciences. Hawaii, USA: dblp, 2019: 2763-2770.

[8] WANG Yaqing, MA Fenglong, JIN Zhiwei, et al. EANN: Event adversarial neural networks for multi-modal fake news detection [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 849-857.

[9] ZHANG Xueyao, CAO Juan, LI Xirong, et al. Mining dual emotion for fake news detection [C]//Proceedings of the Web Conference. New York:ACM,2021:3465-3476.

[10] 孙王斌. 多特征融合的可移植谣言早期检测模型[J]. 计算机时代, 2020 (9): 11-16.

[11] SONG Chenguang, NING Nianwen, ZHANG Yunlei, et al. Knowledge augmented transformer for adversarial multidomain multi classification multimodal fake news detection [J]. Neurocomputing,2021,462:88-100.

[12] WU Lianwei, RAO Yuan, ZHANG Cong, et al. Category-controlled encoder-decoder for fake news detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(2): 16.

[13] SADIQ S, WAGNER N, SHYU M, et al. Feaster, high dimensional latent space variational autoEncoders for fake news detection [C]//Proceedings of 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). Piscataway,NJ:IEEE, 2019, 437-442.

[14] FANG Yong, GAO Jian, HUANG Cheng, et al. Self multi-head attention-based convolutional neural networks for fake news detection[J]. PLoS One,2019,14(9): e022271.

[15] 潘德宇,宋玉蓉,宋 波. 一种新的考虑注意力机制的微博谣言检测模型[J]. 小型微型计算机系统,2021,42(2): 348-353.

[16] CHEN Tong, WU Lin, LI Xue, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection[C]// Proceedings of 2018 Pacific Asia Conference on Knowledge Discovery and Data Mining. Melbourne, Australia: dblp, 2018: 40-52.

[17] 韩晓鸿,赵梦凡,张钰涛. 联合异质图卷积网络和注意力机制的假新闻检测[J]. 小型微型计算机系统, 2024 ,45 (2): 301-308.

[18] YANG Zhilin, DAI Zihang, YANG Yiming, et al. XLNet: Generalized autoregressive pretraining for language understanding [C]//Proceedings of Advances in Neural Information Processing Systems. San Francisco, USA: NIPS Foundation, 2019: 5754-5764.

[19] MA Jing, GAO Wei, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence(IJCAI'16). New York:ACM, 2016: 3818-3824.

[20] AWAN M J, YASIN A, NOBANEE H, et al. Fake news data exploration and analytics[J]. Electronics, 2021, 10(19): 2326.