计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：APK-CNN 和 Transformer 增强的多域虚假新闻检测模型
作者：李金金，桑国明，张益嘉
收稿日期：2023-10-09
网络首发日期：2024-03-19
引用格式：李金金，桑国明，张益嘉．APK-CNN 和 Transformer 增强的多域虚假新闻检测模型[J/OL]．计算机应用．
https://link.cnki.net/urlid/51.1307.TP.20240315.1430.010

Journal of Computer Applications ISSN 1001-9081
计算机应用 CODEN JYIIDU http://www.joca.cn

doi:10.11772/j.issn.1001-9081.2023091359

APK-CNN 和 Transformer 增强的多域虚假新闻检测模型

李金金，桑国明*，张益嘉

(大连海事大学 信息科学技术学院，辽宁 大连 116026)

(*通信作者电子邮箱 sangguoming@dlmu.edu.cn)

摘 要: 为解决社交媒体新闻中的领域转移、领域标签不完整问题，以及探索更高效的多域新闻文本特征提取和融合网络，提出一种基于 APK-CNN(Adaptive Pooling Kernel Convolutional Neural Network)和 Transformer 增强的多域虚假新闻检测模型 Transm3。首先，设计三通道网络对文本的语义、情感和风格信息进行特征提取和表示，并利用多粒度跨域交互器对这些特征进行视图组合；其次，通过优化的软共享内存网络和域适配器来完善新闻领域标签；接着，将 Transformer 与多粒度跨域交互器结合，使用更先进的融合网络动态加权聚合不同领域的交互特征；最后，将融合特征输入到分类器中用于真假新闻判别。实验结果表明，Transm3 模型与 M^{3}FEND 和 EANN 相比，在中文数据集上宏 F1 值分别提高了 3.68%和 6.46%，在英文数据集上分别提高了 6.75%和 11.93%，在 9 个分领域上 F1 值也得到了明显的提高，充分验证了 Transm3 模型在多域虚假新闻检测工作上的有效性，为社交媒体中的虚假新闻检测提供了有力支持。

关键词: 虚假新闻检测；领域转移；软共享内存网络；Transformer；APK-CNN

中图分类号: TP391.1; TP183; G210.7 **文献标志码**：**A**

APK-CNN and Transformer-enhanced multi-domain fake news detection model

LI Jinjin, SANG Guoming*, ZHANG Yijia

(College of Information Science and Technology, Dalian Maritime University, Dalian Liaoning 116026, China)

Abstract: In order to solve the problems of domain shifting and incomplete domain labeling in social media news, as well as to explore more efficient multi-domain news feature extraction and fusion networks, a multi-domain fake news detection model based on APK-CNN (Adaptive Pooling Kernel Convolutional Neural Network) and Transformer enhancement was proposed, namely Transm3. Firstly, a three-channel network was designed for feature extraction and representation of semantic, emotional, and stylistic information of the text and view combination of these features using a multi-granularity cross-domain interactor. Secondly, the news domain labels were refined by optimized Soft-shared Memory Networking and domain adapters, Then, Transformer was combined with a multi-granularity cross-domain interactor to dynamically weight the aggregation of different domain interaction features. Finally, the fused features were fed into the classifier for true/false news discrimination. Experimental results show that compared with M^{3}FEND and EANN, Transm3 improves the macro F1 values by 3.68% and 6.46% on Chinese dataset, and 6.75% and 11.93% on English dataset. And the F1 values on the nine sub-domains are also significantly improved. The effectiveness of Transm3 model for multi-domain fake news detection work is fully validated, providing strong support for fake news detection in social media.

Keywords: fake news detection; domain shift; soft-shared memory networking; Transformer; APK-CNN（Adaptive Pooling Kernel Convolutional Neural Network）

0 引言

虚假新闻作为一种出于个人或集体利益追求而被故意创作的虚假信息，具有博人眼球、传播速度快的特点，容易引发公众误导。这种形式下，探索更智能和高效的虚假新闻自动检测方法对于新闻平台的持续健康发展具有重要意义。

对于多域虚假新闻检测，Silva 等[1]提出了一种保留 3 个领域特定知识和共享知识的方法。然而，在实际应用中，新闻领域的数量远远超过 3 个，他们忽略了领域转移问题，如

收稿日期: 2023-10-09; 修回日期:2023-12-08; 录用日期: 2023-12-11。
基金项目：国家自然科学基金资助项目(62072070)；中央高校基本科研业务费项目(3132019207)。
作者简介: 李金金(2000—)，女，河南漯河人，硕士研究生，CCF 会员（会员号：P6500G），主要研究方向：自然语言处理、谣言检测；桑国明(1971—)，男，辽宁大连人，副教授，硕士，主要研究方向：自然语言处理、人工智能；张益嘉(1979—)，男，辽宁大连人，教授，博士，主要研究方向：自然语言处理、社会媒体计算。

(a)九域高频词汇图 （b）虚假新闻的风格特征热图 （c）虚假新闻的情感特征热图

（d）真实新闻的风格特征热图 （e）真实新闻的情感特征热图

图1 多域特征分析效果

Fig. 1 Multi-domain feature analysis effectiveness

图 1 所示，不同领域的新闻在高频词汇、情感表达以及写作风格上存在明显差异。

Nan 等[2]提出了 MDFEND 模型，使用 9 个新闻领域数据进行训练，显著提升了模型的准确性和领域适应性；然而，一篇新闻往往同时涉及多个主题，他们忽略了领域标签不完整的问题。Zhu 等[3]对中英文数据集进行了整合，同时引入了文本情感信息和风格信息，分别命名为 Ch-9 和 En-3。为了应对领域标签不完整和领域转移的挑战，Zhu 等[3]提出了软共享内存网络的方法；然而，这种方法面临两个主要困难：一方面，软共享内存网络中的信息具有过时性，无法及时反映当前情况，降低了模型在多域虚假新闻检测中的准确性；另一方面，存储领域相关信息需要大量的存储空间，增加了模型的资源需求，导致运行效率下降。另外，Zhu 等[3]仅使用传统的文本卷积模型(Text Convolutional Neural Network, TextCNN)和向量拼接的方法来进行特征提取和融合，这明显存在较大缺陷：一方面，他们采用传统的 TextCNN 模型进行文本语义特征提取，在训练过程中丢失了大量有效信息；另一方面，他们使用向量拼接的方式来融合特征，这种方法未能充分挖掘特征之间的内在相关性，引发了噪声，并生成了冗余信息，从而使多特征虚假新闻的检测效果难以提升。

针对以上问题，本文提出了一种基于 APK-CNN 和 Transformer 增强的多域虚假新闻检测模型 Transm3。首先，设计了三个通道网络，从语义、情感和风格角度对 9 个领域的新闻片段进行建模，在获取文本语义信息时，提出了一种新的 APK-CNN 模型，有助于捕捉文本中的关键语义特征，增强了模型对于多域新闻的语义表示；其次，通过三步融合网络，对语义、情感和风格特征进行视图组合，并将得到的视图组合与软共享内存网络提供的领域标签信息进行特征点积融合，生成更有效用的多源、多角度视图信息；最后，利用 Transformer 对视图信息进行加权融合，从而生成一个信息丰富、噪声较少、高层抽象以及领域敏感性强的融合特征。

本文的主要工作如下：

1)提出了一种新的 APK-CNN 和 Transformer 增强的多域多特征虚假新闻检测模型 Transm3，深入挖掘软共享内存网络和视图组合方法进行多域虚假新闻检测，并利用 APK-CNN 和 Transformer 实现了深层次的特征提取和融合，实现了高性能的虚假新闻检测。

2)设计了一个自适应优化的多粒度跨域交互器，该模块能够驱动多个领域特征进行视图组合，灵活地学习特定任务的特定注意力分布。此外，本文探索了多粒度跨域交互器与 Transformer 的结合方式，利用端到端的神经网络对多特征向量进行平滑的交互和协作，生成新闻文本特征的全局表示关系。这种深度协同的方式不仅有助于提升特征表示的一致性和丰富性，而且有效地捕捉了新闻文本特征之间的全局关系。

3)提出了一种新的 APK-CNN 模型。通过引入自适应池化层、动态可调卷积神经网络以及扩张卷积等来更好地处理不同长度和特征相异的输入序列，有效地捕获文本中的长期依赖关系，深入保留语义线索的完整性。

4) 通过在模型中引入软共享内存网络来存储并提供新闻领域标签信息，以缓解领域差异性导致的模型性能下降问题。另外，对软共享内存网络中存储和更新特征向量的方式进行了优化调整，通过学习率（取 0.1）精确地判断新闻所属类别并保留重要信息，从而实现更加稳定和准确的内存更新。

1 相关工作

虚假新闻自动化检测技术的发展与新闻主题的多样性密切相关。本章将从单一领域和多领域的虚假新闻检测两个方面进行研究。

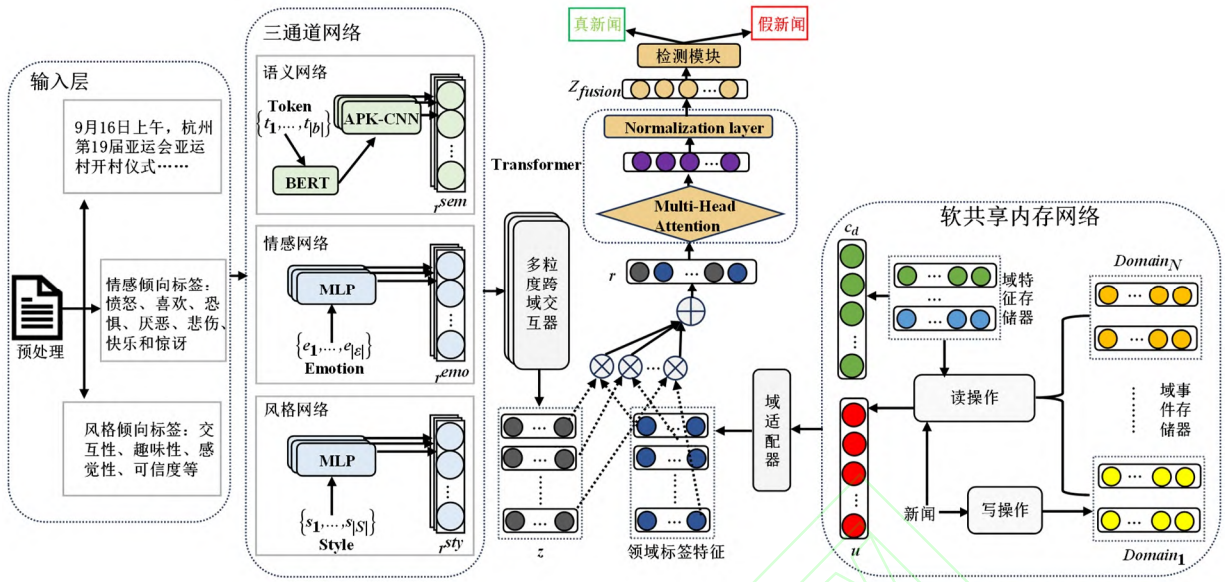


图2 Transm3 模型架构

Fig. 2 Architecture of Transm3 model

1.1 基于单域的虚假新闻检测

借助机器学习技术对特定领域的新闻进行分析并判断其真伪已成为多媒体时代的一个重要研究方向。Singhal 等[4]采用了 BERT(Bidirectional Encoder Representation from Transformers)预训练模型学习特定领域文本的语义特征。其中的优势在于，BERT 模型能够同时捕捉新闻文本的语义和上下文信息，检测出潜在的虚假内容，但由于无法深入理解特征的内在特点和交互性，因此存在一些局限和检测盲区。Ma 等[5]采用了对抗性生成模型，以有效地检测 Twitter 上的谣言。具体来说，他们利用真实和虚假信息的生成网络相互对抗这一特性对特定领域的新闻进行分析并判断其真假，从而提高模型的性能。然而，来源于不同领域的社交上下文信息庞杂且质量参差不齐，单一领域检测模型可能无法全面涵盖不同主题的新闻，使得模型检测结果的可信度降低。

1.2 基于多域的虚假新闻检测

为了更好地识别和防御多领域、多主题、多语言的跨域虚假新闻。Ganin 等[6]提出了学习域不变表示方法，确保特征在不同领域、主题和语言下的一致性，有助于模型识别多域场景下的虚假信息。Ma 等[7]侧重于对域关系进行建模，深入理解和捕捉不同新闻之间的相似性和差异性。上述两种方法都依赖于硬共享机制，通过最大化跨域文本特征交互性来增强模型在不同领域中的适应性。然而，不同领域的语言表达、事件类型和信息传播渠道存在差异。硬共享机制难以应对多种域特征变化，模型在跨领域情境下无法准确捕捉和区分虚假信息，降低了模型的可靠性和鲁棒性。

Zhu 等[8]通过对齐领域特定分布来解决跨领域分类问题。Zadeh 等[9]提出了一种内存储网络，有效提升了多域检测模型的性能。Zhu 等[3]提出了 M^{3}FEND 模型，采用软共享机制聚合不同领域的知识，有效解决了领域转移和领域标签问题。然而，这些模型过于关注领域共享性，忽视了特征提取和融合任务，导致这些模型在实际应用中存在瓶颈。

为了克服上述问题，本文研究了多域多特征虚假新闻检测的新方法，聚焦于领域差异性、领域标签不完整、特征提取和融合问题，利用优化的软共享内存网络有效地存储领域标签信息，同时探索了多通道提取网络和特征融合网络的组合应用。这些创新方法为探索多域虚假新闻检测的研究提供了一个新的方向。

2 Transm3 模型

本章将详细介绍 Transm3 模型，整体结构如图 2 所示，检测模型主要由输入层、三通道网络、软共享内存网络、特征融合网络与检测模块 5 部分组成。首先，通过输入层将文本转换成模型理解的张量；其次，将处理好的文本信息送入三通道网络进行语义、情感和风格特征提取，并利用多粒度跨域交互器对这些特征进行视图组合，捕捉多重维度信息。软共享内存网络负责存储和提供领域标签信息，内部结构如图 2 所示。域适配器将领域标签信息进行转换和适配，以生成区别性领域标签特征。然后，将视图组合和领域标签特征进行点积融合，过滤冗余信息，并将它馈送至 Transformer 模型，从而获得更为精准和全面的聚合特征。最后，利用预测器进行新闻真假分类工作。

2.1 输入层

文本数据预处理过程包括：文本标准化、文本清洗、繁体转换和拼写纠错等步骤，获得最终的新闻文本。设 P 为社交媒体上的一条新闻，由 T 个单词组成，即 $P=\{p_1,p_2,\mathbf{K},p_T\}$，其中，$p_i$ 代表分词后句子的第 i 个单词。域标签 $d\in\{Domain_1,Domain_2,\mathbf{K},Domain_N\}$，$N$ 为新闻领域的个数。我们需要将 P 转换为计算机可以理解的形式。具体地，语义特征的提取过程遵循 Zhu 等[3]的方法，首先对输入的 P 添加[CLS]和[SEP]标记，获得文本语义 ***Token*** 编码为 $\boldsymbol{B}=\{t_1,t_2,\mathbf{K},t_{|b|}\}$。其中，$\boldsymbol{B}$ 代表文本语义信息编码，$t\in R^O$ 表示嵌入，O 表示嵌入维度，$t_{|b|}$ 为第 b 个单词的 ***Token***。情感特征的提取过程遵循 Zhang 等[10]的方法，包括情感类别、情感词典、情感强度以及情感得分等，情感信息编码记为 $\boldsymbol{\omega}=\{e_1,e_2,\mathbf{K},e_{|\varepsilon|}\}$，$e_{|\varepsilon|}$ 表示第e个单词的 ***Token***。同样地，受 Yang 等[11]的启发，提取文本风格信息编码并将其记为 $\boldsymbol{S}=\{s_1,s_2,\mathbf{K},s_{|s|}\}$，$s_{|s|}$ 表示第 s 个单词的 ***Token***。在进行虚假新闻检测时，每条新闻都有一个 ground-truth 标签 $y\in\{0,1\}$，0 和 1 分别代表真实新闻和虚假新闻。

2.2 三通道网络

2.2.1 语义网络

语义网络(Semantic Network, SemNet)由 BERT 预训练模型和 APK-CNN 提取器组成。BERT 模型的作用是将输入的文本内容编码成一个具有语义信息和上下文信息的特征向量，实现通用性跨领域的迁移学习。APK-CNN 是一种进化型文本卷积神经网络(TextCNN)模型，通过引入自适应池化层、动态可调卷积神经网络和扩张卷积网络，实现了对不同长度和特征文本序列灵活、精确的处理，增强了模型对文本长期依赖关系的建模能力，同时提高了模型对文本局部和全局信息的感知，减少了训练过程中的信息损失。此外，为了加速训练并提高泛化性能，APK-CNN 采用了参数共享、字典化和 ReLU(Rectified Linear Unit)激活函数等技术，有效地解决了梯度消失问题，增强了模型的非线性建模能力，使其在文本处理任务中表现卓越。本文使用嵌入编码 $\{t_1,t_2,\mathbf{K},t_{|b|}\}$ 作为语义网络的输入，得到表示情感特征的向量 $\boldsymbol{r}^{sem}$，语义网络的提取过程可通过式（1）表示为：

$$\boldsymbol{r}^{sem}=\mathrm{SemNet}(\{t_1,t_2,\mathbf{K},t_{|b|}\}) \tag{1}$$

2.2.2 情感网络

根据 Castillo 等[12]的方法，本文将新闻片段中的情感信息处理为数值特征，帮助模型更好地理解新闻片段的情感内容，多层感知机(Multilayer Perceptron, MLP)能够处理文本中复杂的非线性关系，学习高级抽象特征，实现端到端学习，能够更全面、精确地捕捉不同领域新闻文本的情感和风格差异。

本文采用多层感知(MLP)作为情感网络(Emotional Network, EmoNet)来学习数据集中的高层次抽象情感特征，MLP 通过多层神经网络的前向传递和反向传播训练，能够从文本中学习并捕捉到有效情感特征，以表达数据之间的复杂关系。具体地，将情感信号 $\{e_1,e_2,\mathbf{K},e_{|\varepsilon|}\}$ 作为输入，通过 MLP 模型进行处理并得到表示情感特征的向量 $\boldsymbol{r}^{emo}$，情感网络的提取过程可通过式（2）表示为：

$$\boldsymbol{r}^{emo}=\mathrm{EmoNet}(\{e_1,e_2,\mathbf{K},e_{|\varepsilon|}\}) \tag{2}$$

2.2.3 风格网络

本文采用多层感知机(MLP)作为风格网络(Stylistic Network, StyNet)来提取新闻片段的风格特征。具体地，将风格信号 $\{s_1,s_2,\mathbf{K},s_{|s|}\}$ 作为输入，通过 MLP 模型进行处理并得到表示风格特征的向量 $\boldsymbol{r}^{sty}$。风格特征已被处理为数值特征，用于后续的建模和分析[13]。在提取过程中，MLP 可以通过反向传播算法不断调整网络参数，以最小化训练数据和目标输出之间的误差，风格网络提取过程可通过式(3)表示为：

$$\boldsymbol{r}^{sty}=\mathrm{StyNet}(\{s_1,s_2,\mathbf{K},s_{|s|}\}) \tag{3}$$

多域虚假新闻的特点复杂多样，仅从单一角度进行建模难以达到高准确率。本文采取多任务学习方法，使用多个训练好的模型作为多通道提取器，分别记为 $\{\boldsymbol{r}_i^{sem}\}_{i=1}^{k_{sem}}$、$\{\boldsymbol{r}_i^{emo}\}_{i=1}^{k_{emo}}$、$\{\boldsymbol{r}_i^{sty}\}_{i=1}^{k_{sty}}$。其中，$k_{sem}$、$k_{emo}$、$k_{sty}$ 分别代表语义网络、情感网络和风格网络的通道号。通过三通道网络，模型能够同时关注不同子空间的表示信息，产生更多样化的交互特征，从而更全面地表征输入数据。

2.3 软共享内存网络

社交媒体数据通常呈现明显的时序性。为了深度挖掘新闻文本中的时序特征，本文提出了软共享内存网络(Soft-shared Memory Networking)，主要用于存储、完善和提供新闻领域标签信息。本文将软共享机制与双向长短期记忆（Bi-directional Long Short-Term Memory, BiLSTM）网络相结合，构建更高性能的内存储网络。通过引入软共享机制，增强了模型灵活适应动态关系的能力，而 BiLSTM 模块的引入，增强了模型对社交媒体文本时序性的建模性，从而能够更精准地捕捉文本主题在时间上的演变模式。

BiLSTM 模块的独特之处在于其对序列数据的敏感性，它可以有效地捕捉文本中的时序变化、情感演变和话题发展。在社交媒体这种动态变化迅速的环境下，引入 BiLSTM 有助于模型更好地理解文本背后的时序结构，能够保留和传递长期的信息。这一融合方法旨在提高模型对社交媒体文本时序

特性的理解能力，进而更精准地分析文本内容中的情感、风格趋势以及主题演变，有助于完善新闻的领域标签信息以及缓解领域差异性对模型的性能影响。软共享内存网络的引入不仅增强了模型的适应性，同时深化了对社交媒体文本时序关联性的理解，使得对文本数据内在结构的抽象和建模水平更为高效，为社交媒体文本的深度挖掘提供了更强有力的工具。这一方法不仅有助于提升模型性能，还为社交媒体文本的深入研究提供了有力的支持。

软共享内存网络主要由一个域特征存储器和多个域事件存储器组成，具体结构如图 2 所示。为了实现稳定且准确的内存更新，本文引入学习率（设为 0.1）来优化特征向量的存储和更新方式，有效地利用了历史信息，通过加权平均混合新旧样本，能更好地反映样本的变化趋势，实现对内存的有效管理。

2.3.1 域特征存储器

域特征存储器用于自动捕获和存储域特征，旨在将每个域的特征表示存储在域内存单元中并将其用作后续任务的输入。首先，通过 BiLSTM 学习每个输入域特征表示序列的时序关系，包括捕获序列中的短时和长时依赖关系。随后，对所学到的时序信息进行编码，并将其与域特征表示一同存储在相应的域内存单元中，这个步骤的目的是将时序信息与域特征表示有机地结合在一起，形成更为丰富和抽象的域内存储结构，以便后续任务能够更有效地利用这些编码的时序关系。c_i 表示第 i 个域的内存单元，N 表示新闻领域的数量，域特征存储器可通过式（4）表示为：

$$\boldsymbol{C}=\{c_i\}_{i=1}^{N} \tag{4}$$

在模型训练中，随机初始化域特征存储器的参数，通过学习每个域的训练数据来自动学习该领域的特征表示，并将其存储在相应的内存单元中。对于每个输入样本，模型可以根据其对应的域内存单元来获取该域的特征表示，这种方式使得模型能够针对不同领域的特征进行个性化学习[14]。

2.3.2 域事件存储器

域事件存储器利用域事件内存矩阵记录某域的新闻，并评估给定新闻片段与所有域事件内存矩阵的相似性，确定其潜在的域标签分布。每个域都有一个域事件内存矩阵，表示该域新闻片段的聚类结果，通过聚类可以发现并表示出潜在的领域标签。具体地，第 j 个域的域事件内存矩阵记为 $\boldsymbol{M}_j=\{\boldsymbol{m}_i\}_{i=1}^{N}$。一个内存单元 $\boldsymbol{m}$ 代表一组相似的新闻片段。域事件存储器的工作过程主要分为三步：初始化、读操作和写操作。

初始化。利用 *K*-means 算法来聚类相似新闻，以初始化内存单元 $\boldsymbol{m}$。一条新闻用 $\boldsymbol{n}$ 表示，即：$\boldsymbol{n}=[G(\{t_1,t_2,\mathbf{K},t_{|b|}\});\{e_1,e_2,\mathbf{K},e_{|e|}\};\{s_1,s_2,\mathbf{K},s_{|s|}\}]\hat{\imath}\ R^{I}$ 表示。$G(\times)$ 为可学习的注意力层，$\{t_1,t_2,\mathbf{K},t_{|b|}\}$ 为语义特征，$\{e_1,e_2,\mathbf{K},e_{|e|}\}$ 为情感特征，$\{s_1,s_2,\mathbf{K},s_{|s|}\}$ 为风格特征。

读操作。读操作是评估给定新闻片段与所有域事件内存矩阵之间的相似性[15]。域事件内存矩阵集合表示为 $\{\boldsymbol{M}_1,\boldsymbol{M}_2,\mathbf{K},\boldsymbol{M}_j\}$，其中 j 为域内存矩阵编号。对于给定的新闻片段，首先在某个域事件内存矩阵 $\boldsymbol{M}_j$ 中找到所有相似的内存单元 $\boldsymbol{m}$，然后通过“聚合”技术，将相似的内存单元聚合成一个域表示，聚合过程表述如式（5）所示：

$$\boldsymbol{O}_j=\mathrm{softmax}(n\boldsymbol{W}g(\boldsymbol{M}_j)/\mathrm{t})\boldsymbol{M}_j \tag{5}$$

此处设定 $\tau=0.01$ 以找到最相似的新闻集群，$\boldsymbol{W}\hat{\imath}\ R^{I'I}$ 为可学习参数矩阵。将聚合后的所有域表示集中到一个矩阵 $\boldsymbol{D}=[\boldsymbol{O}_1,\boldsymbol{O}_2,¼,\boldsymbol{O}_N]\hat{\imath}\ R^{I'I}$ 中，N 表示新闻领域的个数，$V\hat{\imath}\ R^{I'I}$ 表示相似度分布[16]。相似度分布评估过程如式（6）所示：

$$v=\mathrm{softmax}(nVg(\boldsymbol{D})) \tag{6}$$

在处理某条新闻时，首先根据域标签 d 查找域事件内存单元 $\boldsymbol{c}$，得到显式域表示 $\boldsymbol{c}_d$。然后将显式域表示 $\boldsymbol{c}_d$ 与隐式域表示 $\boldsymbol{u}$ 合并为 $[\boldsymbol{c}_d,\boldsymbol{u}]$，其包含了全面的领域标签信息。$\boldsymbol{c}_i$ 是第 i 个域的显示域表示，v_i 表示相似度分布向量的第 i 个元素[17]。隐式域表示 $\boldsymbol{u}$ 计算如式（7）所示：

$$\boldsymbol{u}=\sum_{i=1}^{g} v_i\boldsymbol{c}_i \tag{7}$$

写操作。给定域标签 d 后，确定新闻片段所属领域，并将其存储在对应的域事件内存矩阵 $\boldsymbol{M}_d$ 中。在向 $\boldsymbol{M}_d$ 写入信息时，先将域内存单元 $\boldsymbol{m}_d$ 中原有信息擦除，再添加新信息，使内存中的信息随时间不断演化和更新，更好地反映当前情况和动态变化[18]。相似度得分计算如式（8）所示：

$$sim=\mathrm{softmax}(n\boldsymbol{W}g(\boldsymbol{M}_d)/\mathrm{t}) \tag{8}$$

其中，$g(\times)$ 表示转置函数，$\boldsymbol{W}\hat{\imath}\ R^{I'I}$ 表示可学习的参数矩阵，t=0.01。通过相似度得分 sim_i 计算每个内存单元的加法向量 $\boldsymbol{add}_i$，即 $\boldsymbol{add}_i=sim_i\times n$。参数j (取 0.1)用于控制内存擦除和添加的比例，更新后的内存单元 $\boldsymbol{m}_i$ 表述如式（9）所示：

$$\boldsymbol{m}_i=(1-\mathrm{j})\boldsymbol{m}_i+\mathrm{j}\ \boldsymbol{add}_i \tag{9}$$

2.3.3 域适配器

域适配器的主要作用是对软共享内存网络提供的领域标签信息进行有效筛选，保留更为重要的领域标签信息并对其进行向量映射。此外，本文利用域适配器对视图组合差异进行建模，得到更准确、更有效用的视图组合。具体的工

作流程如下：域表示$[\boldsymbol{c}_d,\boldsymbol{u}]$作为输入传递给域适配器，通过聚合有用的领域标签特征来辅助筛选重要的视图组合。本文使用前馈网络$f(\cdot)$聚合多个领域的特征交互向量，用$\boldsymbol{r}$代表聚合后的交叉视图表示，h为聚合的数量，$\boldsymbol{Z}$表示交互特征向量，并使用权重向量$\boldsymbol{W}\in R^h$来表示各个域交互特征向量的重要性[19]。域适配器的工作原理表述如式（10）所示：

$$\boldsymbol{r}=\sum_{i=1}^{h}\boldsymbol{W}_i\boldsymbol{Z}_i \quad \boldsymbol{w}=\mathrm{softmax}(f([\boldsymbol{c}_d,\boldsymbol{u}])) \tag{10}$$

2.4 特征融合网络

特征融合网络主要分为两个结构：多粒度跨域交互器和Transformer模型。

不加区分的视图可能导致噪声视图组合，进而影响模型的性能。本研究提出了一种多粒度跨域交互器，通过采用基于注意力的子网络[20]，将多个领域的语义、情感和风格特征进行有效组合。该方法能够灵活地调整注意力分布，以适应不同输入样本的任务需求。

大多学者在研究中采用了直接拼接的方法聚合特征，这种融合方式较为简单粗暴，未能充分考虑不同领域中数据的异质性。本文将多粒度跨域交互器与Transformer结合起来，通过更先进的聚合模式来改进融合网络，灵活学习特征之间的内在结构和依赖。首先，多粒度跨域交互器选择性地交互多个新闻领域的语义、情感和风格特征，然后将得到的视图组合与域适配器输出的领域标签特征进行点积聚合，并利用Transformer的多头自注意力机制对其进行进一步交互，得到更全面的融合特征。这种设计弥补了跨域融合网络存在的信息丢失和过拟合问题，模型的性能和领域适应性大大提升。

2.4.1 多粒度跨域交互器

多粒度跨域交互器具有H个头，每个头可以自适应地学习一种领域特征表示并建模它们的重要性，避免了噪声视图的干扰。在多粒度跨域交互器中，引入了三个可学习参数：a^{sem}、a^{emo}、a^{sty}，分别表征语义、情感和风格特征的重要性。$\boldsymbol{Z}$表示交互特征向量，多域特征交互过程表述如式（11）所示：

$$\boldsymbol{Z}=\exp\Big[\sum_{i=1}^{K_{sem}}a_i^{sem}\ln\boldsymbol{r}_i^{sem}+\sum_{j=1}^{K_{emo}}a_j^{emo}\ln\boldsymbol{r}_j^{emo}+\sum_{q=1}^{K_{sty}}a_q^{sty}\ln\boldsymbol{r}_q^{sty}\Big] \tag{11}$$

2.4.2 Transformer融合层

本研究旨在利用Transformer实现多个不同特征的精确匹配和融合。首先，对输入向量$\boldsymbol{r}$进行线性映射得到**Query(*Q*)**、**Key(*K*)**、**Value(*V*)**，对应的权重矩阵分别为$\boldsymbol{W}^q$、$\boldsymbol{W}^k$、$\boldsymbol{W}^v$。接着，将$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$分别传入4个独立的自注意力头，得到4个注意力矩阵，每个矩阵的向量维度为$\boldsymbol{d}_k$，$\boldsymbol{A}_q$表示第q个注意力矩阵，$q\in\{1,2,3,4\}$，$\boldsymbol{K}^{\mathrm{T}}$表示矩阵转置。然后，将4个注意力矩阵合并后与Value矩阵相乘，得到拼接特征$\boldsymbol{U}$。最后，对$\boldsymbol{U}$进行归一化处理得到融合后的共享特征$\boldsymbol{Z}_{fusion}$。Transformer中的权重矩阵由模型自学习得到，输出权重矩阵表示为$\boldsymbol{W}^O$。综上所述，Transformer模型为不同领域的特征融合提供了一种高效的方式，其工作原理表述如式（12）~式（15）所示：

$$\boldsymbol{Q}=\boldsymbol{W}^q\boldsymbol{r} \quad \boldsymbol{K}=\boldsymbol{W}^k\boldsymbol{r} \quad \boldsymbol{V}=\boldsymbol{W}^v\boldsymbol{r} \tag{12}$$

$$\boldsymbol{A}_q=\mathrm{softmax}\Big(\frac{\boldsymbol{Q}_q\boldsymbol{K}_q^T}{\sqrt{d_k}}\Big) \tag{13}$$

$$\boldsymbol{U}=\mathrm{Concat}(\boldsymbol{A}_1\boldsymbol{V}_1,\boldsymbol{A}_2\boldsymbol{V}_2,\boldsymbol{A}_3\boldsymbol{V}_4,\boldsymbol{A}_4\boldsymbol{V}_4)\boldsymbol{W}^O \tag{14}$$

$$\boldsymbol{Z}_{fusion}=\mathrm{Activation}(\mathrm{LayerNorm}(\boldsymbol{U})) \tag{15}$$

2.5 检测模块

检测模块由MLP构成，主要包括两组线性变换层、批归一化层以及sigmoid激活函数。具体而言，线性变换层将输入向量$\boldsymbol{Z}_{fusion}$与相应的权重进行加权，映射为分类器的输入特征，并经过批归一化操作，提高训练的稳定性和收敛速度。然后，在输出层应用sigmoid激活函数，将MLP预测的连续值映射到(0,1)之间的概率，用于对新闻片段P进行真假分类预测，得到一个二元分类结果[21]。预测新闻P为假新闻的概率计算如式（16）所示：

$$\hat{p}=\mathrm{sigmoid}(\mathrm{MLP}(\boldsymbol{Z}_{fusion})) \tag{16}$$

在预测过程中，所有的参数都是可学习的，通过最小化反向传播的交叉熵损失函数来进行优化，量化预测值与真实标签之间的差距，提高模型的预测能力[22]。y代表真实值，该过程可以表述如式（17）所示：

$$V=-y\log\hat{p}-(1-y)\log(1-\hat{p}) \tag{17}$$

3 实验与结果分析

本章实验在中英文数据集上评估了Transm3模型的性能，并回答以下问题：

1）RQ1。Transm3模型对于多域多特征虚假新闻检测是否具有性能提升？

2）RQ2。Transm3在不同语种的数据集上是否均优于其他方法？

3）RQ3。本文提出的AKP-CNN和特征融合网络是否有助于增强模型效果？

4）RQ4。自注意力机制的数量和Encode的层数是否影响模型性能？

3.1 数据集与评价指标

3.1.1 数据集

本文使用的数据集来自于 Zhu 等[错误!未找到引用源。]编制的公共中英文数据集，Ch-9 与 En-3 的数据分布可参见表 1 和表 2。

1)中文数据集：Ch-9 数据集收集自新浪微博的 9 个领域，涵盖科技、军事、教育、灾难、政治、健康、金融、娱乐和社会。

2)英文数据集：En-3 数据集包含了 Gossipcop、Politifact 和 COVID 三个领域的真假新闻，提供了 28764 条样本数据，这些数据有助于提升 Transm3 模型的泛化能力。

表1 中文数据集 Ch-9

Tab. 1 Chinese dataset Ch-9

领域	样本数		领域	样本数	
	真新闻	假新闻		真新闻	假新闻
科技	143	93	健康	485	515
军事	121	222	金融	959	362
教育	243	248	娱乐	1000	440
灾难	185	591	社会	1198	1471
政治	306	546	总计	4640	4488

表2 英文数据集 En-3

Tab. 2 English dataset En-3

领域	真新闻样本数	假新闻样本数
Gossipcop	16804	5067
Politifact	447	379
COVID	4750	1317
总计	22001	6763

3.1.2 评价指标

为了评估 Transm3 模型的可靠性，本文选取准确率(Acc)、精确率(Precision)、召回率(Recall)、F1 值(F1-score)和 ROC 曲线下的面积(Area Under Curve, AUC)作为评价指标。各指标的计算方式可表述如式（18）~式（21）所示：

$$V_{\text{Precision}} = \frac{TP}{TP + FP} \tag{18}$$

$$V_{\text{Recall}} = \frac{TP}{TP + FN} \tag{19}$$

$$V_{\text{AUC}} = \int_0^1 TPR(FPR)\mathrm{d}FPR \tag{20}$$

$$V_{\text{F1}} = \frac{2 \times V_{\text{Precision}} \times V_{\text{Recall}}}{V_{\text{Precision}} + V_{\text{Recall}}} \tag{21}$$

其中：*TP*(True Positive)代表被模型预测为正类的正样本数量；*FP*(False Negative)代表被模型预测为正类的负样本数量；*FN*(False Positive)代表被模型预测为负类的正样本数量；*TN*(True Negative)代表被模型预测为负类的负样本数量；*TPR*(True Positive Rate)代表准确预测为正的样本数占实际正样本总数的比例；*FPR*(False Positive Rate)代表错误预测为正的负样本数占实际负样本总数的比例。

3.2 实验设置

所有实验均在 NVIDIA V100-SXM2 上进行，使用 Python3.8 编程语言和 PyTorch 1.13.0 深度学习框架进行开发。中英文数据集按照 6:2:2 的比例随机划分为训练集、验证集和测试集。在训练过程中，批大小 Mini-batch 设为 64，迭代轮次 Epoch 设为 50。模型使用 Adam 进行损失优化，初始学习率设置为 0.0001，丢弃率 Dropout 设为 0.2，MLP 隐藏层维度为 384，激活函数包括 ReLU 和 sigmoid。英文和中文数据集的最大序列长度分别为 300 和 170，超出部分进行截断处理，不足部分用 0 填充。

3.3 基线模型

为了验证 Transm3 模型的性能，本节对基线模型进行了评估(RQ1)，并将它们分为 3 类：单域方法、混合域基线和精心设计的多域方法[23]。

单域方法即在单个领域中训练模型，单域方法包括三个模型：

1)双向门控循环单元（Bidirectional Gated Recurrent Unit，BiGRU）[14]：通过使用双向门控循环单元建模新闻文本，并将它们合并以形成综合表示，经过全连接层和激活函数进行分类，以判断新闻是否虚假。

2)TextCNN[24]：使用不同大小的卷积核在文本上进行卷积操作，以提取新闻的文本语义特征，并利用全连接层和激活函数进行分类。

3)RoBERTa[25]-[26]：即 Robustly optimized BERT approach，通过大规模的自监督学习从文本数据中学习语言表示，传入全连接层将其映射为二元分类结果。

第二组是混合域基线，将所有领域的数据混合在一起，即将所有领域组合成一个域。除了前述的 BiGRU、TextCNN 和 RoBERTa 模型，还对比了另外两个基线模型：

1)StyleLSTM[27]：采用双向长短期记忆神经网络(BiLSTM)提取文本风格特征并利用全连接层进行真假预测。

2)DualEmo[10]：利用 BiGRU 模型中的双向门控循环单元模型来捕获新闻文本中的情感特征，然后经过多层感知机(MLP) 将情感特征转化为更高级别的抽象表示，最终生成预测结果。

第三组是精心设计的多域方法，包括以下基线模型：

1)EANN[28]：引入事件判别器作为对抗网络的一部分，学习虚假新闻和真实新闻之间的共享特征，并减少模型对于事件特定性的依赖性。

表3 不同模型在 En-3 数据集上的实验结果

Tab. 3 Experimental results of different models on En-3 dataset

模型	不同子数据集上的 F1			overall		
	Gossipcop	Politifact	COVID	F1	Acc	AUC
BIGRU	0.7666	0.7722	0.8885	0.7958	0.8668	0.8840
TextCNN	0.7786	0.8011	0.9040	0.8079	0.8692	0.9023
RoBERTa	0.7810	0.8583	0.9288	0.8184	0.8802	0.9108
BIGRU	0.7479	0.7339	0.7448	0.7501	0.8321	0.8504
TextCNN	0.7519	0.7040	0.8322	0.7679	0.8362	0.8674
RoBERTa	0.7823	0.7967	0.9014	0.8101	0.8744	0.9058
StyleLSTM	0.8007	0.7937	0.9252	0.8285	0.8826	0.9250
DualEmo	0.8056	0.7868	0.9019	0.8270	0.8818	0.9251
EANN	0.7937	0.7558	0.8836	0.8123	0.8743	0.9053
MMoE	0.8022	0.8477	0.9379	0.8361	0.8920	0.9265
MoSE	0.7981	0.8576	0.9326	0.8318	0.8885	0.9252
EDDFN	0.8067	0.8505	0.9306	0.8378	0.8912	0.9263
MDFEND	0.8080	0.8473	0.9331	0.8390	0.8936	0.9237
M^{3}FEND	0.8237	0.8478	0.9392	0.8517	0.8977	0.9342
Transm3	**0.8467**	**0.8982**	**0.9486**	**0.9092**	**0.9212**	**0.9654**

表4 不同模型在 Ch-9 数据集上的实验结果

Tab. 4 Experimental results of different models on Ch-9 dataset

模型	不同领域的 F1									overall		
	科技	军事	教育	灾难	政治	健康	金融	娱乐	社会	F1	Acc	AUC
BIGRU	0.5175	0.3365	0.7416	0.7293	0.8588	0.8373	0.8137	0.7992	0.7918	0.8103	0.8103	0.8902
TextCNN	0.4074	0.3365	0.8059	0.4388	0.8482	0.8819	0.8215	0.7973	0.8615	0.8369	0.8370	0.9094
RoBERTa	0.7463	0.7369	0.8146	0.7547	0.8044	0.8873	0.8361	0.8513	0.8300	0.8477	0.8477	0.9226
BIGRU	0.7269	0.8724	0.8138	0.7935	0.8356	0.8868	0.8291	0.8629	0.8485	0.8595	0.8598	0.9309
TextCNN	0.7254	0.8839	0.8362	0.8222	0.8561	0.8768	0.8638	0.8456	0.8540	0.8686	0.8687	0.9381
RoBERTa	0.7777	0.9072	0.8331	0.8512	0.8366	0.9090	0.8735	0.8769	0.8577	0.8795	0.8797	0.9451
StyleLSTM	0.7729	0.9187	0.8341	0.8532	0.8487	0.9084	0.8802	0.8846	0.8552	0.8820	0.8821	0.9471
DualEmo	0.8323	0.9026	0.8362	0.8396	0.8455	0.8905	0.9053	0.8944	0.8569	0.8846	0.8846	0.9541
EANN	0.8225	0.9274	0.8624	0.8666	0.8705	0.9105	0.8710	0.8957	0.8877	0.8975	0.8977	0.9610
MMoE	0.8755	0.9112	0.8706	0.8770	0.8620	0.9364	0.8567	0.8886	0.8750	0.8947	0.8948	0.9547
MoSE	0.8502	0.8858	0.8815	0.8672	0.8808	0.9179	0.8672	0.8913	0.8729	0.8939	0.8940	0.9543
EDDFN	0.8186	0.9137	0.8676	0.8786	0.8478	0.9379	0.8636	0.8832	0.8689	0.8919	0.8919	0.9528
MDFEND	0.8301	0.9389	0.8917	0.9003	0.8865	0.9400	0.8951	0.9066	0.8980	0.9137	0.9138	0.9708
M^{3}FEND	0.8292	0.9506	0.8998	0.8896	0.8825	0.9460	0.9009	0.9315	0.9089	0.9216	0.9216	0.9750
Transm3	**0.8943**	**0.9807**	**0.9111**	**0.9236**	**0.9074**	**0.9690**	**0.9256**	**0.9490**	**0.9195**	**0.9555**	**0.9557**	**0.9895**

2)MMoE[7]：使用文本卷积神经网络 TextCNN 提取新闻的文本特征，该模型共享多个混合专家(MoE)，每个专家对应一个领域的特定头部，最后利用新闻分类器进行二分类。在 MMoE 中，专家是指多层感知机(MLP)。

3)MoSE[29]：采用了与 MMoE 相似的多专家共享架构，但使用长短期记忆网络(LSTM)作为专家。通过 LSTM 网络来捕捉各个领域的文本信息特征，并将多个领域的文本特征进行拼接后送入全连接层进行分类。

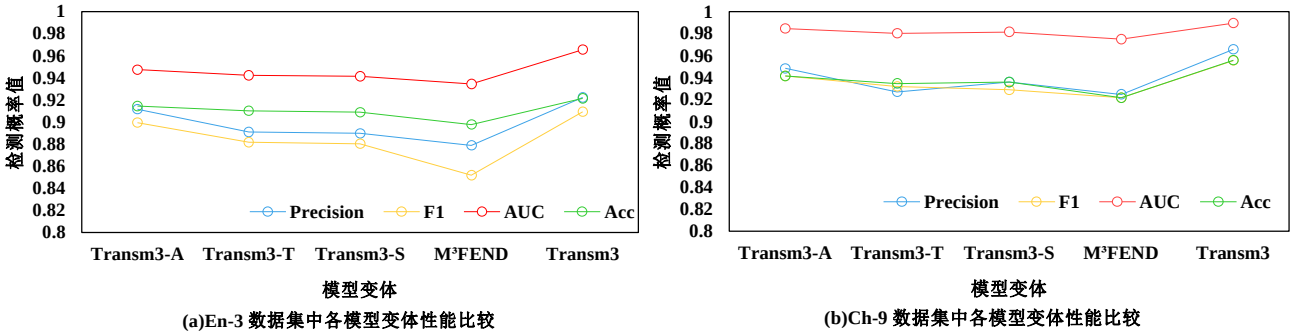
4)EDDFN[1]：提出了一种同时保留领域特定知识和领域共享知识的方法。首先，模型使用无监督的方法学习领域特定的表示。然后，通过 Louvan 算法获得每个新闻的跨领域表示。最后，模型利用这两种表示来重构新闻的真假标签，实现虚假新闻的检测。

5)MDFEND[2]：使用 BERT 对新闻文本进行编码，引入不同专家网络(TextCNN)进行特征提取，通过软共享机制和领域门控制协同工作，利用全连接层进行新闻真假分类。

6)M^{3}FEND[3]：是最新的多域虚假新闻检测模型，使用 TextCNN 提取新闻文本语义信息，使用 MLP 模型提取新闻文本情感信息和风格信息，同时通过显式聚合不同域的交互特征来缓解域差异问题，在大多数任务上表现优异。

3.4 实验结果分析

本节对 Transm3 与基准模型进行实验对比(RQ2)。表 3 和表 4 分别展示了模型在 En-3 和 Ch-9 数据集上的实验结果，包括各个领域的 F1 值、综合性能的 F1 值、Acc 和 AUC，粗体和下划线分别表示最佳和次优结果。实验分析可得：



(a)En-3 数据集中各模型变体性能比较　(b)Ch-9 数据集中各模型变体性能比较

图3　消融实验

Fig. 3 Model ablation experiments

1)M³FEND 模型在中英文数据集上取得了良好的效果，该模型提出了“域内存库”的概念，充分缓解了领域转移问题。与 M³FEND 模型相比，本文提出的软共享内存网络不仅保留了域内存库的软共享机制架构，而且通过双向长短期记忆网络深度学习文本中的时序关系，同时引入了学习率机制以弥补信息过时性和有限存储空间问题。此外，本文提出的 APK-CNN 模型在很大程度上提高了模型的性能和语义表达能力。通过融合多粒度跨域交互器和 Transformer 的优势，构建了更高效的特征融合网络。值得注意的是，在所有实验结果的度量指标上，Transm3 模型都显著超越了 M³FEND 模型，这进一步验证了 Transm3 模型的有效性。

2)混合域方法与多域模型在大多数任务上表现于单域方法，这表明使用多个领域联合训练数据可以实现领域间的知识迁移和共享。然而，在表 3 中，第二组的 BiGRU、TextCNN 和 RoBERTa 在混合域中表现不佳，经分析得出如下解释。首先，领域之间的差异性是性能下降的主要原因，因为这些模型难以有效处理不同领域特征的数据，领域差异性问题导致模型难以泛化到不同领域的任务上。其次，在实验过程中，不同新闻领域的训练数据比重略有不同，数据不平衡问题导致模型在数据量较少的领域上性能下降，解决这一问题需要更加平衡的数据采样策略。

3)在多领域虚假新闻检测中，软共享机制被证明比硬共享机制更有效。实验结果显示，MDFEND、M³FEND 和 Transm3 在大多数任务上均表现优于其他模型，比如 EANN、MMoE、MoSE 和 EDDFN。Transm3 模型与 M3FEND 和 EANN 相比，在中文数据集上宏 F1 值分别提高了 3.68%和 6.46%；在英文数据集上分别提高了 6.75%和 11.93%，在 9 个分领域上 F1 值也得到了明显的提高。分析可得，软共享机制赋予了模型更大的灵活性，使其能够根据不同任务的需求自动调整共享程度。相比之下，硬共享机制往往受到任务之间差异性的制约，导致模型稳健性低。

4)M³FEND 模型和 Transm3 模型在各个领域的实验效果均显著优于其他模型，这表明相较于仅基于文本语义的单一特征检测方法，文本的情感特征和风格特征能够为虚假新闻检测提供互补的语义线索，采用基于文本的多特征检测方法在虚假新闻检测方面具有明显的优势。

3.5　消融实验

图 3 的结果表明，本文所提出的 Transm3 模型具有最好的性能。在实验中，“Transm3-A”是指去除了 APK-CNN 中的自适应池化层、动态可调卷积神经网络和扩张卷积等创新技术；“Transm3-T”是指去除了 Transformer 模块；“Transm3-S”代表去除软共享内存网络；“M³FEND”代表最新的基线模型；“Transm3”代表完整的 Transm3 模型。分析模型在中英文数据集上的实验结果可得如下结论：

1) 移除模型的任何重要部分或重要方法均导致模型性能出现不同程度的下降，强有力地证明了本文模型中的各模块在多域虚假新闻检测任务中的显著有效性。

2) 在 En-3 和 Ch-9 实验中，相较于完整的 Transm3 模型，Transm3-A 变体性能下降，这表明学习灵活处理不同文本特征以及加强长期依赖建模的重要性。该变体去除了 APK-CNN，导致仅依靠传统的 TextCNN 难以对不同长度和特征的文本序列进行灵活而精确的处理，进而造成大量有效信息的丢失，对虚假新闻的全局性检测产生了不良影响。实验结果表明，模型在处理不同类型的新闻时，需要具备对复杂文本结构的敏感性。APK-CNN 的创新性贡献使得 Transm3 模型更有效地捕捉并利用文本中的关键信息，从而提高了虚假新闻检测的全面性和准确性。这一实验证明了 APK-CNN 在模型框架中的重要性，为虚假新闻检测任务中的文本处理提供了有益的方法和洞见。

3) 在 En-3 和 Ch-9 实验中，通过对比移除各模块后模型性能的下降程度，明显观察到 Transm3-T 变体的性能相较于其他变体更为不理想，这表明利用 Transformer 模块进行多特征融合有助于提升模型的鲁棒性和准确性，同时也增强了模型对不同语言和领域信息的处理能力。具体来说，Transformer 模块能够在输入数据中捕获远距离的依赖性，对大规模数据的高效处理使其在模型中具备出色的性能。这一模块的引入加强了模型对不同关系的敏感性，从而更为精准地捕捉输入信息的内在结构和关系。实验结果进一步验证了

模型在中英文上的普适性，显示了该模块在不同语言的虚假新闻检测任务中都能取得显著的性能提升。

4) 在 En-3 和 Ch-9 实验中，根据准确率的下降程度，去除软共享内存网络对于模型性能的影响最为明显，这表明本文所使用的软共享内存网络在完善新闻领域标签和缓解领域差异性方面发挥着重要作用。通过对软共享内存网络进行消融实验，可以发现，捕捉新闻内部的时序关系和主题特征能够帮助模型有效挖掘新闻中隐含的领域标签。这一发现不仅提升了模型的性能，同时也增强了检测过程的可解释性。实验结果表明，在构建具有高度解释性的多域虚假新闻检测模型时，深入挖掘新闻文本中的时序关系、主题特征以及领域标签信息是至关重要的。

3.6 参数敏感性实验

本节在中英文数据集上研究 Transformer 模型中多头自注意力机制的头数 *Head* 和 Encoder 层数 *layer* 对模型的影响(RQ4)，并对实验的 F1 值与 Acc 进行了统计。在实验过程中，通过对这些参数进行调整来找寻最佳的模型配置。

表5 注意力机制头数对模型性能的影响结果

Tab. 5 Results of impact of the number of attention mechanism heads on model performance

数据集	*Head*	F1	Acc
En-3	1	0.8745	0.8801
	2	0.8966	0.8897
	4	0.9092	0.9212
	8	0.8812	0.8963
Ch-9	1	0.9114	0.9115
	2	0.9320	0.9319
	4	0.9555	0.9557
	8	0.9235	0.9235

在实验中，头数 $Head \in \{1,2,4,8\}$，实验结果详见表 5。在中英文数据集上的实验结果均表明，Transm3 模型中的单头注意力机制无法很好地捕捉文本中的多样性特征，增加头数可以提高模型的性能。具体而言，当 *Head*=2 时，模型表现有所提升；当 *Head*=4 时模型表现最佳，此时模型可以同时关注不同的文本特征，并通过参数共享减少模型参数量；当 *Head*=8 时，模型性能下降，因为它过于关注一般特征，而丢失了一些特定信息。因此，在选择头数 *Head* 时，需要平衡模型对一般性和特定性特征的关注程度，并根据具体任务和数据集来选择合适的头数以获得最佳模型效果。

综上所述，在中英文数据集上，适当增加参数 *Head* 的数量均有助于提高模型性能，具体表现为更高的 F1 和 Accuracy 值。这是因为增加 *Head* 个数能够提高模型对复杂语义和特征的捕捉能力，从而提升任务性能。然而，需要注意的是，随着 *Head* 数量的进一步增加，模型的性能并非一味地持续提升。研究表明，在某一点之后，增加 *Head* 反而导致模型的 F1 和 Accuracy 开始下降。这一现象可能源于过多的参数导致模型过拟合，使其在训练数据上表现出色，但在测试数据上的泛化能力下降。因此，在调整参数 *Head* 的数量时，需要综合考虑模型的复杂性和数据集的特点，在性能提升和过拟合之间取得平衡。

表6 Encoder 层数对模型性能的影响结果

Tab. 6 Results of impact of the number of Encoder layers on model performance

数据集	*layer*	F1	Acc
En-3	1	0.9092	0.9212
	2	0.8912	0.9113
	4	0.8798	0.8998
	8	0.8394	0.8691
Ch-9	1	0.9555	0.9557
	2	0.9212	0.9213
	4	0.9098	0.9098
	8	0.8994	0.8999

在中英文数据集实验中，Encoder 层数 $layer \in \{1,2,4,8\}$，实验结果详见表 6。实验结果显示，模型性能与 Encoder 层数的增加呈现出一种“倒 U 型曲线”。当 *layer*=1 时，模型性能最佳，这表明相对较少的 Encoder 层数已经足够学习数据的不同特征。这一发现与传统认知有所不同，强调了在某些情境下，采用简化模型结构的策略可能更有利于提高模型性能；当 *layer* 增加到 2、4、8 时，模型性能逐渐变差，这表明过多的 Encoder 层数会使模型过于复杂，从而在中英文数据集上表现出过拟合的倾向，进而影响模型的泛化能力。这种趋势强调了在多域虚假新闻检测模型设计中，需要在模型复杂性和泛化能力之间进行有效平衡。因此，在调整 Encoder 层数时，应仔细权衡模型性能和过拟合的风险，以确保模型在各个方面都能够取得最佳表现。

综上所述，在中英文数据集上，自注意力机制最佳个数为 4，Encoder 最佳层数为 1。因此，在实际应用中需要根据不同语种和数据集特点进行超参数调整，以提高模型对数据集的拟合能力和泛化性能。

4 结论

针对现有多域虚假新闻检测模型存在的领域转移、领域标签不完整以及语义特征提取和融合网络不佳等问题，本文提出了 Transm3 模型，构建了基于语义、情感和风格特征的端到端多域虚假新闻检测系统。通过优化的软共享内存网络以及其他模块的相互作用，极大解决了域标签不完整和领域转移问题。同时，本文提出了 APK-CNN 模型，能够有效地捕捉新闻文本中的全局语义特征。此外，我们采用 Transformer 来构建特征融合网络，实现了多层次特征交互和长距离建模。经过一系列验证实验，包括模型对比、消融实

验和参数选择，以及在中英文数据集上的测试证明了 Transm3 模型在多域虚假新闻检测中的可行性和有效性。

Transm3 模型提供了对新闻真实性更精细的量化度量，通过概率值为相关人员提供了有力工具，以应对虚假信息在社交媒体新闻中的传播。例如：新闻编辑和记者可以通过评估模型输出的高概率值来审查他们发布的新闻，有助于及时发现并核实可能存在虚假信息的内容。平台管理员能够利用模型对虚假信息进行筛选，有效降低社交媒体平台上虚假信息传播的风险，提升整体信息质量。对于公众而言，通过了解新闻的虚假程度，可以增强对信息真实性的辨别力，从而避免受到虚假信息的误导。这一综合的应用有望在社交媒体新闻领域中提升信息传播的可信度和透明度，为用户提供更为可靠和准确的信息环境。

本文提出的 Transm3 模型还存在一些局限性：Transm3 模型仅关注新闻文本内容，未考虑其他类型信息，例如新闻图片，评论和视频等，因此可以尝试引入其他多模态信息以更全面地理解新闻特征。同时，在未来的研究中需要对新闻进行更细粒度的分类，例如虚假新闻多类别检测任务。

参考文献

[1] SILVA A, LUO L, KARUNASEKERA S, et al. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data[C]// Proceedings of the 2021 AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2021, 35(1): 557-565.

[2] NAN Q, CAO J, ZHU Y, et al. MDFEND: Multi-domain fake news detection[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM, 2021: 3343-3347.

[3] ZHU Y, SHENG Q, CAO J, et al. Memory-guided multi-view multi-domain fake news detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2022.

[4] SINGHAL S, SHAH R R, CHAKRABORTY T, et al. Spotfake: A multi-modal framework for fake news detection[C]// Proceedings of the 2019 IEEE International Conference on Multimedia Big Data. Piscataway: IEEE, 2019: 39-47.

[5] MA J, GAO W, WONG K F. Detect rumors on twitter by promoting information campaigns with generative adversarial learning[C]// Proceedings of the 2019 World Wide Web Conference. New York: ACM, 2019: 3049-3055.

[6] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. The journal of machine learning research, 2016, 17(1): 2096-2030.

[7] MA J, ZHAO Z, YI X, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1930-1939.

[8] ZHU Y, ZHUANG F, WANG D. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources[C]// Proceedings of the 2019 AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019, 33(1): 5989-5996.

[9] ZADEH A, LIANG P P, MAZUMDER N, et al. Memory fusion network for multi-view sequential learning[C]// Proceedings of the 2018 AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018, 32: No.1.

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

[11] YANG Y, CAO J, LU M, et al. How to write high-quality news on social network? Predicting news quality by mining writing style [EB/OL]. [2022-08-17]. https://arxiv.org/pdf/1902.00750.pdf.

[12] 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.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 2017 Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2017: No.30.

[14] 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. New York: ACM, 2016: 3818-3824.

[15] XIA H, WANG Y, ZHANG J Z, et al. COVID-19 fake news detection: A hybrid CNN-BiLSTM-AM model[J]. Technological Forecasting and Social Change, 2023, 195: 122746.

[16] SHAFIQ M, GU Z. Deep residual learning for image recognition: A survey[J]. Applied Sciences, 2022, 12(18): No.8972.

[17] ZENG G, CHI J, MA R, et al. ADAPT: Adversarial domain adaptation with purifier training for cross-domain credit risk forecasting[C]// Proceedings of the 2022 International Conference on Database Systems for Advanced Applications. Cham: Springer, 2022: 353-369.

[18] RAZA S, DING C. Fake news detection based on news content and social contexts: a transformer-based approach[J]. International Journal of Data Science and Analytics, 2022, 13(4): 335-362.

[19] DAVOUDI M, MOOSAVI M R, SADREDDINI M H. DSS: A hybrid deep model for fake news detection using propagation tree and stance network[J]. Expert Systems with Applications, 2022, 198: 116635.

[20] SHAHID W, JAMSHIDI B, HAKAK S, et al. Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities[J]. IEEE Transactions on Computational Social Systems, 2022.

[21] HUANG K H, MCKEOWN K, NAKOV P, et al. Faking fake news for real fake news detection: Propaganda-loaded training data generation [EB/OL]. [2023-03-13]. https://arxiv.org/pdf/2203.05386.pdf.

[22] MOHAPATRA A, THOTA N, PRAKASAM P. Fake news detection and classification using hybrid BiLSTM and self-attention model[J]. Multimedia Tools and Applications, 2022, 81(13): 18503-18519.

[23] DAVOUDI M, MOOSAVI M R, SADREDDINI M H. DSS: A hybrid deep model for fake news detection using propagation tree and stance network[J]. Expert Systems with Applications, 2022, 198: 116635.

[24] KIM Y. Convolutional neural networks for sentence classification [EB/OL]. [2022-12-02]. https://arxiv.org/pdf/1408.5882.pdf.

[25] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach [EB/OL]. [2023-02-12]. https://arxiv.org/pdf/1907.11692.pdf.

[26] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for Chinese BERT[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3504-3514.

[27] PRZYBYLA P. Capturing the style of fake news[C]// Proceedings of the 2020 AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020, 34(1): 490-497.

[28] WANG Y, MA F, JIN Z, 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.

[29] QIN Z, CHENG Y, ZHAO Z, et al. Multitask mixture of sequential experts for user activity streams[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 3083-3091.

This work is partially supported by National Natural Science Foundation of China (62072070), Fundamental Research Funds for the Central University (3132019207).

LI Jinjin, born in 2000, M. S. candidate. Her research interests include natural language processing, rumour detection.

SANG Guoming, born in 1971, M.S., professor. His research interests include natural language processing, artificial intelligence.

ZHANG Yijia, born in 1979, Ph. D., professor. His research interests include natural language processing, social media computing.