

基于 ASEC 的多模态虚假新闻检测的研究

甘甜甜¹ 王 亮¹ 黄世奇¹
GAN Tiantian WANG Liang HUANG Shiqi

摘 要

当前的新闻检测模型尚存在模态特征融合困难、不够准确等问题,针对这些问题融入用户社交特征,设计了一种 ASEC 融合模型。首先采用 ALBERT 和 Swin transformer 模型分别提取文本和图片特征,加入归一化处理的模态特征;然后通过 co-attention 注意力机制和 ECANet 注意力机制组成的 ASCE 将这三种模态进行融合,合理分配这三个特征的权重,经过全连接层实现检测;最后在 MediaEval2015 数据集上检测出虚假性的准确率为 93.92%,精确率为 94.07%,比起单一的机制融合多个模态,ASEC 模型的准确率提升了 3.35%。实验结果表明,使用所提出的检测虚假新闻算法模型,检测的准确率、精确度较高,能避免模态信息的损失,也能够更好地识别虚假新闻。

关键词

多模态虚假新闻检测; ALBERT; Swin transformer; ECANet 注意力机制; co-attention 注意力机制

doi: 10.3969/j.issn.1672-9528.2024.01.042

0 引言

近年来,信息技术的飞跃式发展使得社交媒体在人们的社会生活中愈发不可或缺。人们可以通过社交媒体更便捷、更高效地获取各类信息,随时随地都能够通晓天下事。然而新闻报道一味地追求吸引人的眼球,虚假新闻利用人工智能、大数据等相关技术鱼目混珠,导致了很多悲剧,例如夸大造谣胡鑫宇失踪事件博大众眼球、造谣北京协和医院 2022 级硕士复试存在“黑幕”、疫情期间散播涉及疫情的虚假信息引起民众的恐慌等。信息的不确定性使得网民们陷入惶恐、迷

茫,损害了媒体公信力和社会秩序,所以对虚假新闻的监控和修正尤为重要。因此,综合分析新闻的图片、文字、用户社交等特征是虚假新闻检测的重中之重。

近些年,多模态虚假新闻检测大致有两个方向,一个是基于虚假信息的传播进行检测,另一个是基于信息模态检测。基于传播信息来检测新闻代表的有 DAVOUDI 等人^[1]通过对传播树和立场进行动态、静态、结构综合研究分析构建特征来检测虚假新闻。Hu 等人^[2]提出了含有句子、实体和主题的传播图,加入图注意力网络和实体对比网络,更好地捕捉实体之间的关联。SONG 等人^[3]提出了 TGAT(时序图注意力)模型,把过往的信息形成传播图的记忆向量,来实时学

1. 沈阳化工大学计算机科学与技术学院 辽宁沈阳 110142

[3] RATTANATAMORONG P,BOONPALIT Y,SUWANJINDA S,et al.Overhead study of telegraf as a real-time monitoring agent[C]//International Joint Conference on Computer Science and Software Engineering.Piscataway:IEEE,2020:42-46.

[4] 张世贤,张少春,谢晓东.基于 InfluxDB 的监控设备通用运维管理平台[J].计算机系统应用,2021,30(12):123-127.

[5] 李宇.基于一致性 Hash 算法的分布式缓存数据冗余[J].软件导刊,2016(1):47-50.

[6] 陈茂棠,郑圣安,游理通,等.一种基于 RDMA 多播机制的分布式持久性内存文件系统[J].计算机研究与发展,2021,58(2):384-396.

【作者简介】

孙辽东(1988—),男,山东济宁人,学士,中级工程师,

研究方向:人工智能平台。

王超(1988—),男,黑龙江齐齐哈尔人,博士,中级工程师,通信作者,研究方向:深度学习算法、人工智能平台。

陈培(1982—),男,河南郑州人,博士,高级工程师,研究方向:深度学习算法、人工智能平台。

王德奎(1985—),男,山东德州人,硕士,中级工程师,研究方向:人工智能平台。

李世刚(1987—),男,山东德州人,学士,中级工程师,研究方向:人工智能平台。

张书博(1996—),男,辽宁本溪人,学士,中级工程师,研究方向:人工智能平台。

王文潇(1991—),男,山东泰安人,硕士,中级工程师,研究方向:人工智能平台。

(收稿日期:2023-10-13)

习不同时刻的结构特征。考虑到事件随着时间变化的因素，一些学者把研究角度放在了元路径，加入了 GRU 聚合函数^[4]，防止了新闻信息的损失。

基于信息模态来对虚假新闻检测是近些年的热点，把虚假新闻的图片、文本、视频等信息结合起来，高效提取相关的模态，并对多种模态进行相互权重之间的融合进行检测，在研究中举足轻重。Long Y^[5] 提出了 MTMN 模型，它是一种端到端的多模态记忆网络，利用语义、实体的多模态信息补充并增强每个模态。王婕等人^[6] 提出了 BPMF 模型，基于 BERT 和 VGG-19 分别来捕获文本、图片特征，使用矩阵分解双线性池化方法进行特征融合。Ying G^[7] 提出了 MANN 模型，加入相互关注融合模块，学习每种不同模态之间的联系。Zhiping L^[8] 采用 XLNet 和 VGG-19 分别来捕获文本、图片特征，加入 BLIP，解除了信息遗漏给模型带来的干扰。

同时，许多研究者会结合外部信息进行特征提取进行信息的补充，Cui X 与 Li Y^[9] 在处理图像、文本特征的同时加入了外部证据特征并进行交互并建模，证实了外部证据在检测虚假新闻中有效。刘帅等人^[10] 基于知识图谱加入了外部知识特征，强调用户之间的关系和交互，实现了多维度信息融合。这些学者的研究足以证明用户信息特征在虚假新闻检测方面的重要性。

基于以上的研究，本文利用 ALBERT、Swin transformer 分别提取视觉特征向量和文本特征向量，加入归一化处理的用户的特征向量，减少了信息的遗漏。采用 ECANet 注意力机制和 co-attention 机制进行特征向量融合，合理分配特征权重，使得融合更佳，从而实现检测准确率的提升。

1 ASEC 模型

本文使用的 ASEC 主要有模态输入、特征向量提取、向量交融、分类四个结构模块，ASEC 模型结构如图 1 所示。首先采用 ALBERT 和 Swin transformer 模型分别提取文本和

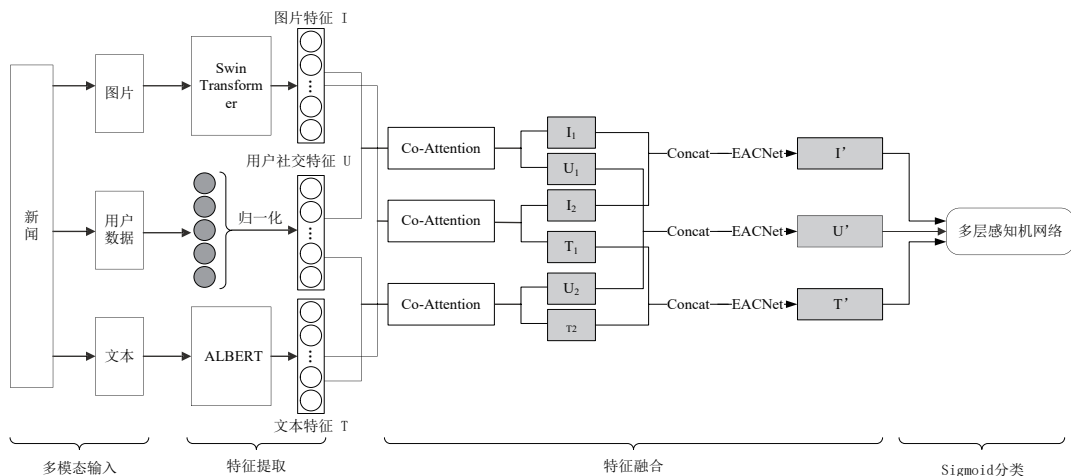


图 1 ALBERT 模型结构

图片模态特征，加入归一化处理的模态特征。其次加入了 co-attention 联合注意力机制和 ECANet 注意力机制的组合，合理分配这三个特征的权重，实现新闻的三个模态有效融合。最后特征向量经过全连接层实现虚假新闻的检测。

1.1 ALBERT 模型结构

ALBERT 是谷歌公司的 Lan 等人提出的一种轻量化的预训练文本模型，该模型在 BERT 原有的基础上对参数量和运行速度进行了改进，如图 2 所示。

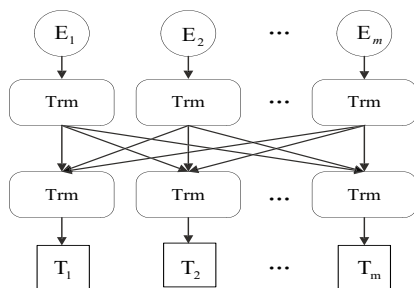


图 2 ALBERT 模型结构

(1) Embedding 层低秩分解。降低了该层向量的因式分解，添加变换矩阵，使得 Embedding 层的输出向量由维度为 E 变成维度为 H ，参数规模从 BERT 的 $V \times H$ 下降为 ALBERT 的 $V \times E + E \times H$ ，明显减少了参数量和空间。

(2) 跨层参数共享。attention 和 FFN 模块在 transformer Block 层实现了参数共享，尽可能减少模型的存储参数量。

(3) SPO 句子次序预测。ALBERT 模型中 SPO 与 BERT 模型中的 NSP 相比，选取负样本有不同之处。NSP 选取负样本是随机选取语义关系不大的句子，而 SPO 选取负样本是交换正样本中两个句子的位置。

(4) Dropout 移出。ALBERT 模型移出了 Dropout，大幅度减少了模型训练时使用的内存。

1.2 Swin transformer 模型结构

Swin transformer 模型预训练精准且高效。Swin trans-

former 模型有四个阶段，这些阶段包含了碎片分割层 (patch partition)、线性嵌入层 (linear embedding)、碎片合成层 (patch merging) 和 Swin transformer Block。

如图 3 所示，Swin transformer 模型把尺寸为 $H \times W \times 3$ 的图片放入 patch parti-

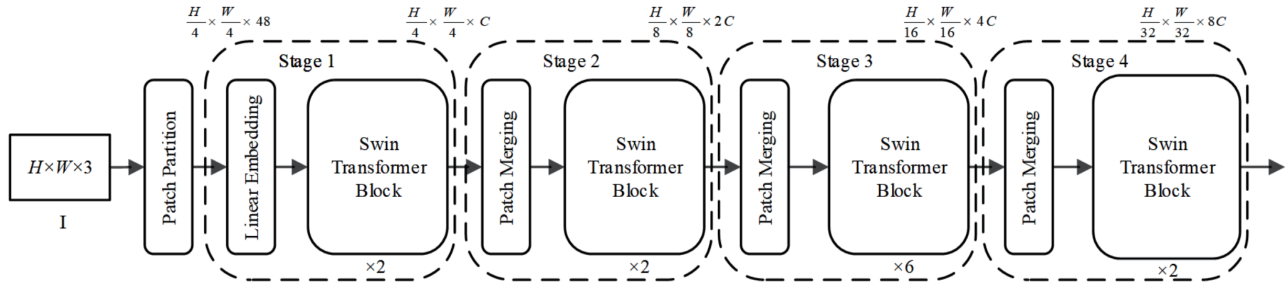


图 3 Swin transformer 模型

tion 碎片分割模块中，分割成 $H/4 \times W/4$ 个图片碎片，图片碎片是维度为 48×1 的向量。首先，通过 linear embedding 线性嵌入层对 48 的碎片维度线性变换成 C ，得到维度为 $H/4 \times W/4 \times C$ 的特征图。其次，通过 patch Merging 碎片合并层把 4 个相邻的碎片进行合成，碎片维度由 C 线性变换成 $2C$ ，得到维度为 $H/8 \times W/8 \times 2C$ 的特征图。最后，通过两个 patch Merging 碎片合并层依次得到特征图，它们的维度分别是 $H/16 \times W/16 \times 4C$ 和 $H/32 \times W/32 \times 8C$ ，输出的特征图的维度层层递进。碎片分割层把图片分成一个个碎片，每一个碎片输入 transformer 中进行操作。线性嵌入层将碎片维度线性变换成 C 。碎片合成层下采样降低图片分辨率，减半通道数，并形成层次化碎片。

1.3 归一化处理模型

本文采用了 MediaEval2015 数据集，分析其中的用户信息，该数据集包括的用户相关数据有：用户被关注数、发布博文数、与其他用户互动数、认证情况。本文对比了虚假新闻和真实性新闻的用户模态的账号的差别，对数据集中的用户特征数据进行归一化处理，把数据转换成 $0 \sim 1$ 之间的小数，从而把不同一量级的特征值转成同一量级的特征值，以便后续高效地进行特征融合。其中，本文使用的是归一化方法整理用户数据：

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中： x 是待处理的用户特征值， $x_{\min}$ 和 $x_{\max}$ 分别是特征值的最小、最大值，通过处理后得到 x^* ，从而可以得到用户特征向量： $U = [x_1^*, x_3^*, \dots, x_5^*]$ 。

1.4 co-attention 和 ECANet 机制结构

联合注意力机制经过对比模态特征当中的相似程度，合理调整模态特征的权重，不同模态间的特征相互指引，把图像、用户、文本三个特征向量输入联合注意力机制中的 Parallel-attention 机制，从而实现了特征向量的两两融合。ECANet 注意力机制主要是在特征维度不缩减的情况下增强特征的聚合，进而各个通道完成全局平均池化层，去除全连接层，在每个通道及其 k 个近邻的一维卷积来实现跨通道交互， k 与通道维度成比例。关于权重分配，对模态局部

特征实行加权求和，捕获关键特征，随之而来的特征数量增多、模型复杂度增高、计算量增大和计算成本高昂。然而 ECANet 能结合通道和空间，使用较少的参数量来高效地学习不同特征通道权重。结合这两个机制的优点，本文将 co-attention 联合注意力机制和 ECANet 注意力机制联结，不但能捕获各个模态间的联系，还能清晰表达出各个模态中具有虚假性的重要性。

co-attention 机制有两种实现方式：Parallel-attention 和 Alternating-attention 机制。本文使用的 Parallel-attention 机制结构如图 4 所示，Parallel-attention 机制选取重要的信息输入神经网络进行计算，经过对文本、图片、用户两两之间进行注意力机制融合，计算出各特征基于各模态的相似程度，完成不同模态之间的交互。

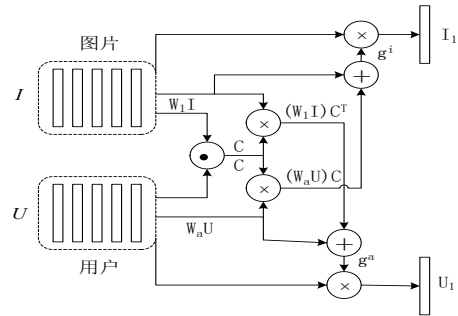


图 4 Parallel-attention 机制结构

令 I 、 U 分别为原始的图像特征向量、原始的用户特征向量，而 I 、 U 分别为通过 co-attention 机制融合后生成的图像向量和用户向量。其公式为：

$$\begin{aligned} C &= \tanh(U^T W_b I) \\ H^i &= \tanh(W_i I + (W_a U) C) \\ H^a &= \tanh(W_a U + (W_i I) C^T) \\ g^i &= \text{softmax}(w_{hi}^T H^i) \\ g^a &= \text{softmax}(w_{ha}^T H^a) \end{aligned} \quad (2)$$

式中： w_{hi} 、 w_{ha} 、 W_a 、 W_b 、 W_i 是权重参数， g^i 和 g^a 是每个图片区域 i_n 和用户区域 a_i 的注意力概率。根据上面的计算可以得到图片特征和用户特征的加权求和为：

$$I_l = \sum_{n=1}^N g_n^i i_n \quad (3)$$

$$U_l = \sum_{t=1}^T g_t^a a_t \quad (4)$$

ECANet 结构图如图 5 所示, 该机制首先对 co-attention 机制新生成特征进行权重分配计算, 然后能得到每个模态最后的特征向量。其中, 输入的特征 $I \in R^{H \times W \times C}$ 首先经过平均池化层压缩得到 $I_{avg} \in R^{1 \times 1 \times C}$, 其次经过一维卷积得到特征图 $I_{conv} \in R^{1 \times 1 \times C}$, 实现了跨通道交互, 然后对 I_{conv} 激活得到权重系数 $M_c(I) \in R^{1 \times 1 \times C}$, 最后将 $M_c(I)$ 与 I 相乘从而得到 I' , 其公式为:

$$M_c(I) = \sigma(i^k(AvgPool(I))) = \sigma(i^k(I_{avg}^c)) \quad (5)$$

$$I' = M_c(I) \otimes I$$

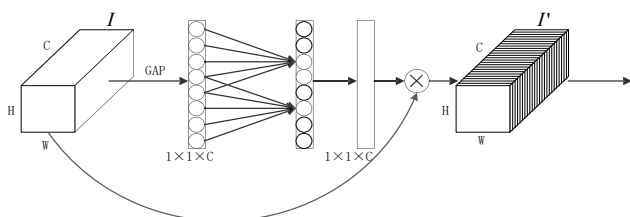


图 5 ECANet 结构

1.5 Sigmoid 函数

把新闻的文本、图片、用户信息预处理, 然后通过 co-attention 机制和 EACNet 注意力机制进行三个模态的特征融合, 输入多层感知机网络从而进行新闻是否虚假的分类。虚假新闻分类计算公式为:

$$\sigma(z_i) = \frac{1}{1 + e^{-z_i}} \quad (6)$$

式中: z_i 代表多模态虚假新闻集第 i 个文本、图片、用户特征向量通过 co-attention 机制和 EACNet 注意力机制融合后准备输入全连接层的模态特征向量; $\sigma(z_i)$ 代表第 i 个向量经过 Sigmoid 函数融合后的虚假新闻概率。

2 实验

2.1 数据集

本文使用 Media Eval2015 数据集, 该数据信息包含文本内容、相关的图片内容以及用户社交信息。其中训练集包括 17 个事件, 事件中虚假新闻 9000 条, 而真实新闻 6000 条, 测试集包括 35 个事件, 相关新闻数据 2000 条。其中用户社交模态中有 nums_followers (粉丝数)、nums_friends (关注数)、times_listed (被 @ 次数)、nums_tweets (发布推文数)、verified_ratio (账户认证比) 这五个部分, 它们虚假性占真实性的占比情况约为: 1:7、1:2、1:8、1:3、1:2。

2.2 数据预处理

首先对导入符号列表, 根据列表对数据集中文字进行分词操作, 接着把冗余的停用词 (例如 “the” “to” “and” “a”

等) 进行删除, 其次对数据集中的文本里的各个单词的词形归并还原, 然后对文本的大小写进行统一转换, 最后清洗对模型没有实际意义的数据。

2.3 实验设置

本文的实验是在 Windows 10 系统中运行的, CPU 为 Intel i7, GPU 为 NVIDIA GTX950 M, 内存为 32 GB, 使用 Python 在 PyTorch 框架下编写。本实验使用官方提供的 ALBERT、Swin transformer 提取文本、图片。本实验把文本的词嵌入维度和字长分别设为 64、200, 超出 200 就拦截, 不足补零, 把图片输入设置为 $224 \times 224 \times 3$, 其中激活函数有: tanh、Sigmoid。本实验选择 Adam 函数去优化模型, 选择 CrossEntropy Loss 作为损失函数。同时本实验的初始学习率为 0.000 1, 批处理为 32, dropout 为 0.15。

2.4 评价指标

本文把准确率 (accuracy) 作为新闻检测的整体评价指标, 同时还有精确率 (precision)、召回率 (recall) 和 F_1 值来评估该模型的性能指标。其中准确率表示预测正确的样本所占总全部样本的比例, 精确率表示样本能被正确检测出来的比例, 召回率表示被正确检测出的正样本占全部被检测为虚假新闻的样本比例, F_1 值对 precision 和 recall 进行协调均等。其公式为:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (10)$$

2.5 结果对比与分析

考虑到模态的个数可能会影响实验, 分别对文本模态、文本和用户信息模态、文本和图片模态以及文本、图片、用户三个模态相结合来研究, 如表 1 所示实验各个性能都有一定的上升。表 1 表示不同模态对检测的影响。

表 1 不同模态对检测的影响 (%)

模态	准确度	精确度	召回率	F_1 值
T	87.74	88.62	89.32	88.96
T+U	89.52	90.24	91.01	90.62
T+I	89.23	90.48	92.09	91.27
T+U+I	93.92	94.07	93.18	93.62

如表 2 所示,对注意力融合多模态的两种方法进行了对比,分别对 co-attention 注意力机制、co-attention 注意力机制 + ACENet 注意力进行探究,能够发现两种注意力结合、合理分配模态权重的实验效果更佳。同时,如表 3 所示,将本文使用的 ASEC 与现有的 CARMN 和 Spotfake 进行比较,准确程度有 2.41% 的提高。

表 2 不同融合方法准确率比较 (%)

融合方法	准确度	精确度	召回率	F_1 值
联合注意力机制	90.57	91.72	90.56	91.13
联合 +ACENet 注意力	93.92	94.07	93.18	93.62

表 3 与现有模型对比结果 (%)

模型	准确度	精确度	召回率	F_1 值
CARMN ^[12]	89.17	90.68	90.64	90.65
Spotfake ^[11]	91.51	91.97	91.06	91.51
ASEC	93.92	94.07	93.18	93.62

3 结语

针对目前多模态虚假新闻检测算法的不足,本文使用了一种基于 ASEC 的模型检测虚假新闻,各个模态的特征向量经过联合注意力机制和 ECANet 注意力机制的结合,捕捉模态之间的关联,进行引导融合,完成权重计算,然后输入全连接层,生成新闻真假性标签。本文模型在公开数据集中进行实验,并记录了相关实验的对比与分析,ASEC 模型对检测性能有一定的提升。

参考文献:

[1] 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: 116-125.

[2] HU L, YANG T, ZHANG L, et al. Compare to the knowledge:graph neural fake news detection with external knowledge[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2021: 754-763.

[3] SONG C, SHU K, WU B. Temporally evolving graph neural network for fake news detection[J]. Information processing & management, 2021, 58(6): 102-117.

[4] CUI J, KIM K, NA S H, et al. Meta-path-based fake news detection leveraging multi-level social context information[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta. New York: ACM, 2022: 325-334.

[5] LONG Y,HUI Y,WANG J, et al. Fake news detection via multi-modal topic memory network[J]. IEEE access, 2021(9): 818-829.

[6] 王婕,刘芸,纪淑娟.基于矩阵分解双线性池化的多模态融合虚假新闻检测[J]. 计算机应用研究, 2022,39(10):2968-2973+2978.

[7] YING G. A mutual attention based multimodal fusion for fake news detection on social network[J]. Applied intelligence, 2022, 53(12): 15311-15320.

[8] LIANG Z P. Fake news detection based on multimodal Inputs[J]. University of melbourne, 2023,75(2):4519-4534.

[9] CUI X,LI Y. Fake news detection in social media based on multi-modal multi-task learning[J]. International journal of advanced computer science and applications (IJACSA), 2022, 13(7): 912-917.

[10] 刘帅,傅丽芳.融合外部知识和用户交互特征的虚假新闻检测[J]. 数据分析与知识发现, 2023,7(11):79-87.

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

[12] SONG C, NING N, ZHANG Y, et al. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks[J]. Information processing & management, 2021, 58(1): 102-116.

【作者简介】

甘甜甜(1996—),女,河南信阳人,硕士研究生,研究方向:多模态虚假新闻检测。

王亮(1971—),男,辽宁沈阳人,博士,副教授,硕士研究生导师,研究方向:自然语言处理、推荐系统和多模态。

黄世奇(1998—),女,四川达州人,硕士研究生,研究方向:深度学习、个性化推荐系统。

(收稿日期:2023-10-13)