

基于注意力机制的酒店评论分类模型

周生金, 王 勇, 王 瑛

广东工业大学计算机学院, 广东 广州

收稿日期: 2021年11月27日; 录用日期: 2021年12月21日; 发布日期: 2021年12月28日

摘 要

在线评论能够对用户的决策产生重要的影响, 一些无良商家会利用这一点通过雇佣水军等方式褒扬自己的商品或诋毁竞争对手的商品, 从而影响用户的判断, 将利益最大化。为了维护大众消费者的利益, 将商品最客观真实的评价展现给用户, 为用户提供最可靠的参考, 将商品评论分类就显得尤为重要。本文将评论文本和评论发布者特征结合, 分别利用融入注意力机制的卷积神经网络模型(ACNN)提取商品评论文本特征和评论发布者特征, 综合挖掘其中的信息, 从而提高分类的准确率。通过在真实数据集上的多次实验表明, 这种方式在评论有效性分类上的准确率达到87.2%, 相比只提取评论者特征和只在评论文本中融入注意力机制的分类效果均有提高。

关键词

评论分类, 注意力机制, 神经网络, 准确率

Attention-Based for Hotel Review Classification Model

Shengjin Zhou, Yong Wang, Ying Wang

School of Computers, Guangdong University of Technology, Guangzhou Guangdong

Received: Nov. 27th, 2021; accepted: Dec. 21st, 2021; published: Dec. 28th, 2021

Abstract

Online comments have an important impact on the user's decision, some unscrupulous merchants will take advantage of this point through hiring people and other ways to praise their own goods or slander competitors' goods to affect the user's judgment and maximize the interests. In order to ensure the interests of the public consumers and present the most objective and true evaluation of products to users and provide users with the most reliable reference, it is particularly important to classify product reviews. In view of the above the problem, the author combines the text of com-

ments and the features of comment publishers, respectively uses the convolutional neural networks model (ACNN) with attention mechanism to extract the features of product reviews, so as to improve the accuracy of classification. Several experiments on real dataset show that the accuracy of this method in the classification of comment validity reaches 87.2%, which is better than only extracting the features of reviews and only incorporating the attention mechanism into the comment text.

Keywords

Comment Classification, Attention, Neural Networks, Accuracy

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

互联网的发展与普及很大程度上改变了消费者的消费观以及消费意见反馈的途径，越来越多的消费者青睐于网上购物与消费。同时，大部分电商平台允许用户对所购买的产品，消费过的酒店，景点等发表评论，来表达消费者对产品的看法。这些评论能够对以后的消费者的决策起到一定的作用，潜在消费者可以以此作为是否消费的重要参考依据。部分无良商家发现利用这一点获取更大的利润，雇佣水军，学术界也称其为垃圾评论发布者，即发布垃圾评论的人，在本店铺下的商品下进行赞赏，发布积极的评论，来达到诱导潜在客户的目的，或在该商家的竞争对手下，发布一些恶意诋毁的评论，误导该商店的潜在客户，导致客户失去兴趣，从而在一定程度上可以提升自己商店的销量。所以，将商品评论正确分类，将最客观真实的评价展现给用户的研究应运而生[1]。

Jindal 和 Liu [2]在 2008 年最早提出虚假评论这一研究问题，把垃圾评论可以大体分为三类：1) 上文所提到的对产品或服务进行不符合实际的鼓吹或诽谤，从而达到影响用户的观点或消费行为的目的的评论；2) 评论的对象仅仅是品牌、生产商、销售商等与产品本身无关的评论；3) 广告，读者的问题和回复等非评论信息。Abernethy 等人[3]提出的检测垃圾文本的方法可以检测垃圾网页和垃圾邮件，大都基于关键词词频统计的方法，一般不能用于检测第一类产品垃圾评论。

该课题早期的研究是基于文本本身，Li 和 Ott 等人[4]在心理学角度进行分析，通过词袋及词性特征对评论文本的虚假性进行判断，词袋特征对观点挖掘、情感分析等研究方向是一个十分有效的特征，但是它的单独运用对虚假评论的识别效果不理想，语法分析属于该任务的早期研究方法，同时研究表明虚假评论比真实评论包含更多的情感词，他们认为稀疏相加生成模型在该课题中运用情感极性特征可以发现反常评论信息。

目前神经网络在自然语言处理方面表现较好，现有研究方法一般将评论文本本身和评论发布者的特征相结合，从而有效提升检测效果。在对虚假评论文本的检测研究中，现有研究对文本及评论发布者的综合分析和利用，主要通过特征级别的融合实现。Mukherjee 等人[5]在 Yelp 数据集上采用支持向量机分类器，运用评论文本特征获得 65.6%~67.8%的准确度。在加入评论发布者的特征之后，检测准确度提升至 84.8%~86.1%，该研究说明评论发布者特征有助于提升对虚假评论文本的检测能力，同时也说明了 Yelp 垃圾评论过滤算法的合理性。

Lin 等人[6]根据评论内容和评论发布者特征，设计了基于监督的方法和基于阈值的方法来识别虚假

评论, 研究表明基于监督的方法能够在较少的训练样本的情况下对虚假评论检测获得较高的精确率和召回率, 而基于阈值的方法在不需要训练样本的情况下也能取得较好的效果。Heydari 等人[7]有效地利用了评级偏差、评论内容和评论发布者的活跃度, 采用模式识别技术, 对从评论时间序列中获取的可疑时间间隔进行综合研究, 在垃圾评论识别检测中获得更高的准确度。

原福永等人通过建立注意力机制的神经网络模型, 结合评论文本, 评论发布者特征, 商品特征, 在 Yelp 数据集中进行垃圾评论检测, 能够较好地提取特征信息, 准确率高达 86.81%, 但是还可以结合更多的外部特征进行垃圾评论检测, 进一步提升模型对评论发布者特征的提取能力[8]。

产品评论垃圾检测尚有许多值得深入探索的问题, 将文本特征与评论发布者行为等特征相结合也是一类直观、有效的研究角度, 注意力机制能够获得不同的特征在检测中不同的权重[9], 实现更好的检测效果。针对忽略评论者特征影响导致准确率低的问题, 本文首先在美国最大的点评网站 Yelp 中爬取酒店评论数据作为实验的数据集, 然后通过融入注意力机制的卷积神经网络模型进行特征提取完成分类任务, 在数据集上进行对比实验, 最后通过对比实验结果证明本文模型的有效性, 并提出未来可以进一步提升准确率的方向。

2. 相关工作

2.1. 卷积神经网络

卷积神经网络是目前比较流行的深度学习模型, 由于其强大的特征提取能力, 在自然语言处理领域越来越受欢迎。卷积神经网络模型包括卷积层, 池化层和全连接层, 每一层有特定的任务, 如图 1 所示。

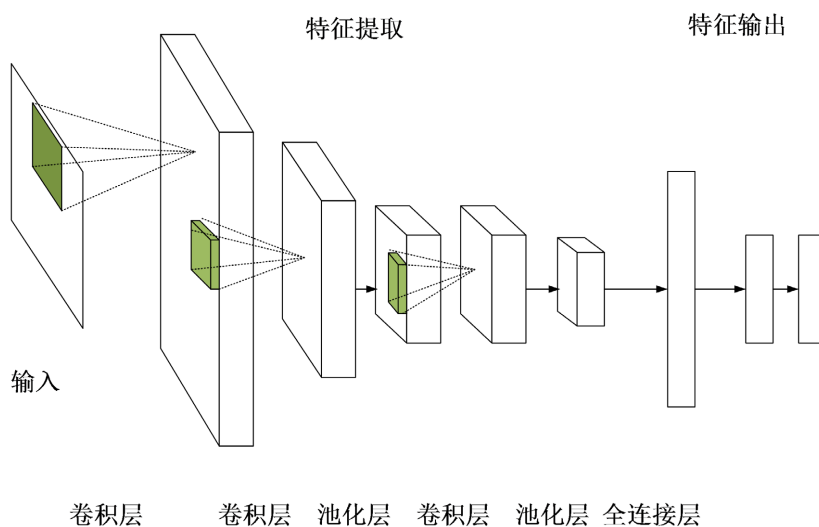


Figure 1. Convolutional neural networks model

图 1. 卷积神经网络模型

其中, 卷积层负责特征提取, 在卷积层进行特征提取后, 池化层负责特征的选择和信息过滤, 一般常用的策略有最大池化和平均池化, 如图 2 所示。

最大池化是指在一定区域内取最大值, 平均池化指一定区域内取平均值, 通过这种操作, 可以达到压缩矩阵的目的。全连接的核心操作是矩阵向量乘积, 把一个特征空间转化到另一个特征空间, 也可以起到分类器的作用。

卷积网络本质就是一种输入到输出的映射, 只要用已知的模式对卷积网络加以训练, 就能够学习大

量的输入与输出之间的映射关系。卷积神经网络有出色的特征提取能力，是当下最流行的深度学习模型之一。

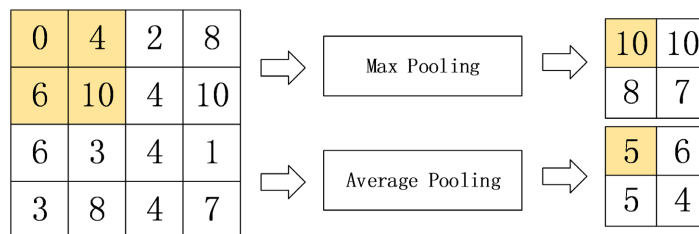


Figure 2. Pooling strategy

图 2. 池化策略

2.2. 注意力机制

在人类视觉中，会选择性关注所有信息的一部分，同时忽略其他信息，这种机制叫做注意力机制。注意力模型最初被用于机器翻译，现在已成为神经网络领域的一个重要概念。在人工智能领域，注意力已成为神经网络结构的重要组成部分，并在自然语言处理、统计学习、语音和计算机等领域有着大量的应用。使用注意力机制，能够获取各种特征在判定中不同的权重，从而达到更好的识别效果[10]。

大多数注意力模型附着在 Encoder-Decoder 框架下，Encoder-Decoder 框架是一种深度学习领域的研究模式，原理就是对于句子对<Source,Target>，将输入句子 Source，通过该框架生成目标句子 Target，在框架中，先将 Source 编码成中间语义 C，再将 C 解码成 Target。

一般在注意力机制的计算过程中，会涉及到 Query (查询)，Key (键)和 Value (值)三个要素，三者构成注意力机制的本质，如图 3 所示。

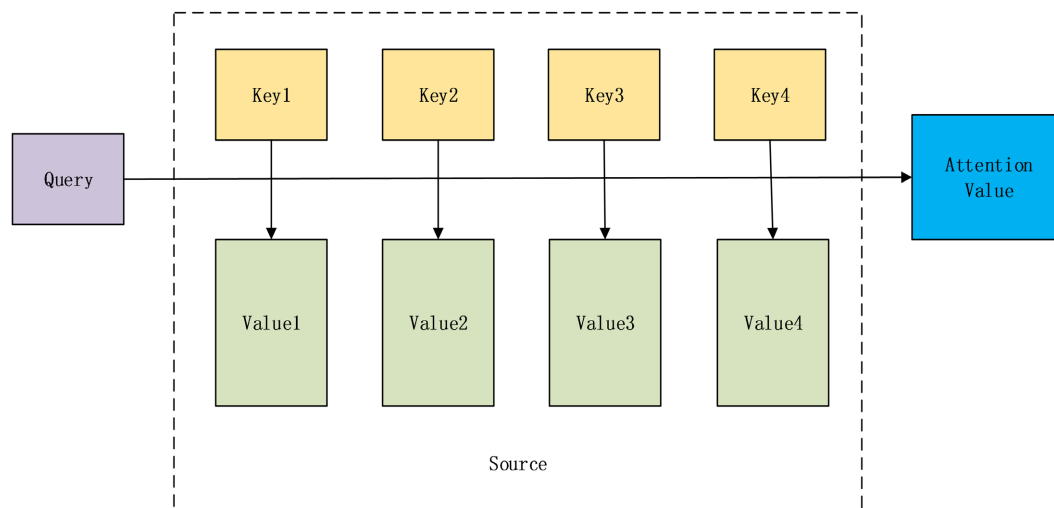


Figure 3. The principle of attention

图 3. 注意力机制原理

Attention 的计算过程一般分为以下三步：

步骤一：计算 Query 和 Key 的相似性，通常使用点积运算(如公式 1 所示)、余弦相似度计算(如公式 2 所示)和神经网络计算(如公式 3 所示)。

$$\text{Similarity}(\text{Query}, \text{Key}_i) = \text{Query} \cdot \text{Key}_i \quad (1)$$

$$\text{Similarity}(Query, Key_i) = \frac{Query \cdot Key_i}{\|Query\| \cdot \|Key_i\|} \quad (2)$$

$$\text{Similarity}(Query, Key_i) = \text{MLP}(Query, Key_i) \quad (3)$$

步骤二：引用 Softmax 函数对步骤一中产生的分值进行数值转换，一方面归一化处理将分值整理成权重之和为 1 的概率分布，另一方面通过 Softmax 的内在机制增加重要元素的权重，一般计算如公式(4)所示，其中， L_x 代表输入句子 Source 的长度， Sim_i 和 Sim_j 表示上一阶段求出的相似度。

$$a_i = \text{Softmax}(Sim_i) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}} \quad (4)$$

步骤三：步骤二中的 a_i 就是 $Value_i$ 的权重系数，加权求和得到 Attention 的值，计算如公式(5)所示。

$$\text{Attention}(Query, Source) = \sum_{i=1}^{L_x} a_i \cdot Value_i \quad (5)$$

Attention 的作用就是通过内部的运算，为每个输入序列分配一个权重，通过计算输入序列的加权和，将关键部分突出，将不重要部分忽略，达到优化模型的目的。

3. 本文方法

本文利用融入注意力机制的卷积神经网络将评论文本和评论发布者特征综合考虑，作为检测依据，从而提升检测准确率[11]，本文的模型(Attention-CNN)架构如图 4 所示。

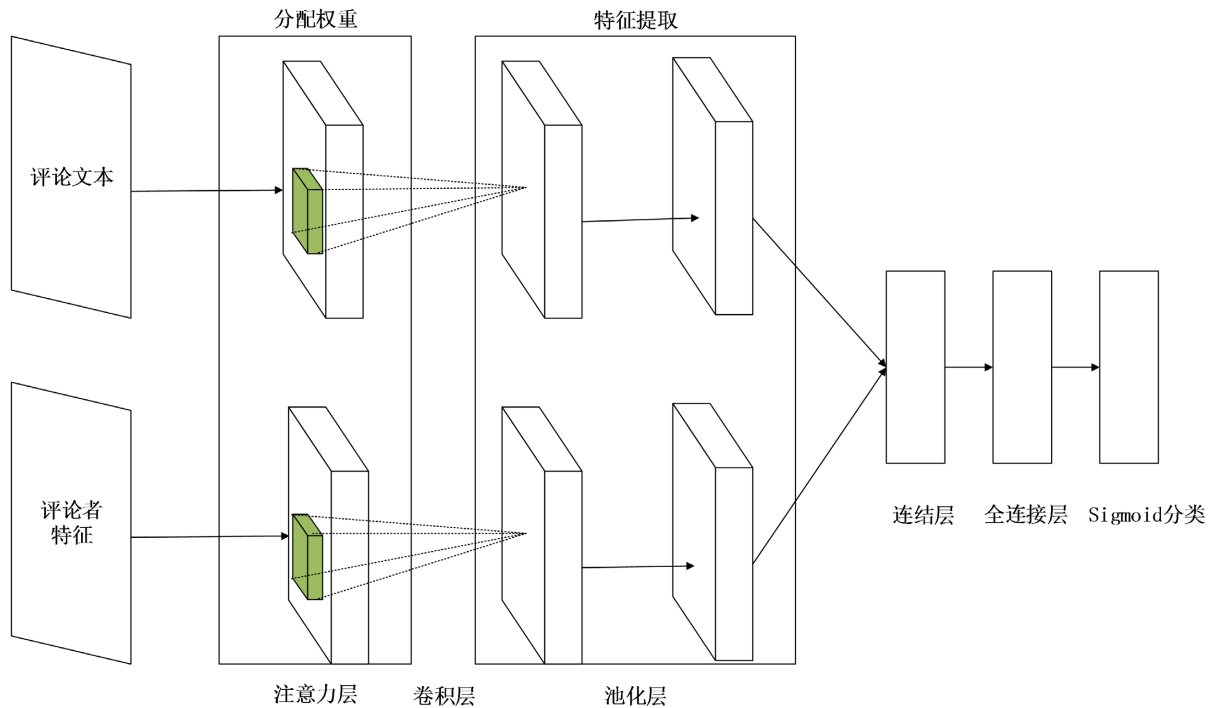


Figure 4. Attention-CNN

图 4. 融入注意力机制的卷积神经网络

对于每一条酒店评论，首先使用 wiki-news-300d 词向量库将评论文本和评论者特征表示，然后在注意力层中将评论文本中的词和评论者特征中的特征分配权重，再到卷积层中提取特征，池化层压缩数据，

这里使用的是最大池化策略，将二者连结，进入全连接层，最后使用 Sigmoid 完成二分类任务。本文模型综合考虑评论文本和评论发布者特征进行检测，同时考虑到不同的特征对结果的影响大小不同，所以该模型的检测效果有一定的提升[12]。

4. 实验

4.1. 实验环境

本实验采用 Pytorch 深度学习框架，操作系统为 64 位的 Linux 系统，内存为 32 GB，CPU 为锐龙 R53600X，GPU 为 RTX 2060 super。

4.2. 实验数据

现有的公共数据集大多只是基于文本本身的内容增加标注，增加评论发布者特征之后如果人工标注则缺少一定的可靠性，本次实验数据集获取方法采用的是在美国最大的点评网站 Yelp 中爬取。Yelp 系统将有效的可推荐的正面评论显示在每一个商家介绍信息下方，而将系统认为无效的，不可推荐的垃圾评论隐藏起来，用户点击显示之后就会显示。本次实验爬取 Yelp 网站中酒店评论正向与反向各 10,000 条，共 20,000 条数据，数据集基本信息如表 1。

Table 1. Basic information about the dataset

表 1. 数据集的基本信息

序号	平均评级	朋友数	评论总数	照片总数	星级评分	发布时间	评论文本
1	4	778	969	3355	5	5/23/2021	I planned a brunch date.....

上述表格列出了数据集中所有字段以及一条正向样本的具体内容，平均评级指该酒店的平均星级评分，总共有五级，朋友数、评论总数和照片总数分别指该评论发布者的好友的数量，注册以来发布评论的总数量和发布的所有评论中的照片总数量，星级评分指该评论者在该评论中对该酒店的评分，总共五级，发布时间指发布该评论的时间，评论文本是评论的具体内容。

4.3. 实验评估指标

本次实验选择准确率(Accuracy)作为实验的评估指标，准确率能够反映模型分类的准确性，计算如公式(6)所示，其中分子表示分类正确的正向和负向样本的总和，分母表示所有分类样本的总和。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (6)$$

4.4. 实验设置

首先提取评论文本的特征，实验选用 wiki-news-300d 的词向量模型表示文本，将每个单词表示为词嵌入向量 $x_i = (a_1, a_2, \dots, a_d)$ ， $x_i \in R^d$ ，文本矩阵表示为 $X_j = (x_1, x_2, \dots, x_T)$ ， $X_j \in R^{d \times T}$ ， T 代表文本的长度。在评论文本的注意力层中，为每个单词分配注意力权重，在评论者特征的注意力层中，为每个特征分配注意力权重，然后输出分配权重之后的向量。在文本中，对于每一个词向量 x_i 分别乘三个权重矩阵 W^Q ， W^K 和 W^V ，得到矩阵 Q_i ， K_i 和 V_i ， W 矩阵开始要初始化，后期随着模型训练更新，然后计算 x_i 与包含本身在内的所有词 x_j 的相关性 $\alpha_{i,j}$ ，如公式(7)所示。

$$\alpha_{i,j} = Q_i \cdot K_j \quad (7)$$

将 $\alpha_{i,j}$ 经过 Softmax 得到 $\alpha'_{i,j}$ ，计算如公式(8)所示。

$$\alpha'_{i,j} = \frac{\exp(\alpha_{i,j})}{\sum_{j=1}^T \exp(\alpha_{i,j})} \quad (8)$$

将每个 $\alpha'_{i,j}$ 与对应的矩阵 V_i 相乘再求和就得到注意力层的输出 b_i ，计算如公式(9)所示。

$$b_i = \sum_{j=1}^T \alpha'_{i,j} V^j \quad (9)$$

在处理评论者特征中，将特征的数据转换成向量 $y = (y_1, y_2, \dots, y_m)$ ，在注意力层做上述过程同样的处理，即可得到分配权重之后的输出 b'_i 。

在卷积层中，计算如公式(10)和公式(11)所示。

$$Z^{l+1} = f(Z^l \otimes w^{l+1}) + b \quad (10)$$

$$L^{l+1} = \frac{L^l + 2p - f'}{s_0} + 1 \quad (11)$$

公式中， Z^l 和 Z^{l+1} 表示第 $l+1$ 层卷积的输入和输出， L^{l+1} 为 Z^{l+1} 的尺寸， w^{l+1} 代表权重， b 为偏置， f' 、 s_0 和 p 为卷积层参数，对应卷积核大小、卷积步长和填充，三者共同决定了卷积层输出特征图的尺寸，是卷积神经网络的超参数。在卷积层进行特征提取后，输出的特征会被传递至池化层进行特征选择和信息过滤，本课题的模型在池化层采用的是最大池化策略。

在评论发布者特征提取中，评论发布者的特征主要表现了评论发布者的个人信息，将数据以向量化的形式表示。数据处理后输入到神经网络中进行特征的提取，将特征向量融合注意力机制，因为评论发布者的每个信息属性对是否判断为垃圾评论的影响是不一样的，注意力机制能选择关键的并忽略不相关的评论发布者信息，从而更加客观准确地实现最终的检测。

通过以上的特征提取，获得文本和评论者的特征，在连结层将两个特征连结转换成一维向量 Z ，将 Z 输入到全连接层，加入 50% 的 dropout 防止过拟合，使用 Adam 作为训练的优化器。本课题的本质是一个二分类问题，在全连接层直接使用 Sigmoid 函数完成分类任务，计算如公式(12)所示，其中 W' 为参数矩阵， B 为偏置项。

$$Z' = \text{Sigmoid}(W' \cdot Z + B) \quad (12)$$

4.5. 对比实验设置

本课题实现了两个对比实验与本文提出的实验模型进行比较。第一个对比实验是只关注评论文本本身的实验，同样利用融入注意力机制的神经网络进行特征提取。第二个对比实验是融合评论发布者多特征实验，在评论发布者特征中只使用卷积神经网络提取特征，为了结果的可靠性，实验采用十折交叉验证方法进行。

4.6. 实验结果

本课题实验与上述两个对比实验的实验结果如表 2 所示。

Table 2. Result data of experiment

表 2. 实验结果

模型	对比实验一	对比实验二	本文实验
准确率	0.845	0.862	0.872

实验证明, 评论发布者特征能够在一定程度上影响评论的分类, 在评论发布者特征中不同的特征对最终的分类结果的影响程度不同。

5. 总结与展望

针对垃圾评论检测因忽略评论者特征导致准确率低的问题, 本文利用融入注意力机制的卷积神经网络, 融合评论发布者特征与评论文本特征进行垃圾评论检测, 实现评论的有效分类, 在 Yelp 酒店评论数据集上, 通过对比实验, 证明了该模型的有效性。下一步的工作可以考虑优化评论发布者特征, 增加可获取的新的特征, 结合领域内更多的专业知识进行检测, 另外可以考虑将模型引入其他类型的评论分类中, 提高模型通用性。

参考文献

- [1] 尤苡名. 虚假评论检测技术综述[J]. 计算机系统应用, 2019, 28(3): 1-9.
- [2] Jindal, N. and Liu, B. (2008) Opinion Spam and Analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, Palo Alto, California, February 2008, 219-230. <https://doi.org/10.1145/1341531.1341560>
- [3] Abernethy, J., Chapelle, O. and Cactillo, C. (2010) Graph Regularization Methods for Web Spam Detection. *Machine Learning*, **81**, 207-225. <https://doi.org/10.1007/s10994-010-5171-1>
- [4] Li, J., Ott, M., Cardie, C., et al. (2014) Towards a General Rule for Identifying Deceptive Opinion Spam. *Meeting of the Association for Computational Linguistics*, Baltimore, June 2014, 1566-1576. <https://doi.org/10.3115/v1/P14-1147>
- [5] Mukherjee, A., Venkataraman, V., Liu, B. and Glance, N. (2013) What Yelp Fake Review Filter Might Be Doing? *Proceedings of the International AAAI Conference on Web and Social Media*, Washington, June 2013, 409-418.
- [6] Lin, Y., Zhu, T., et al. (2014) Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence. *IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining*, Beijing, 17-20 August 2014, 261-264. <https://doi.org/10.1109/ASONAM.2014.6921594>
- [7] Heydari, A., Tavakoli, M. and Salim, N. (2016) Detection of Fake Opinions Using Time Series. *Expert Systems with Applications*, **58**, 83-92. <https://doi.org/10.1016/j.eswa.2016.03.020>
- [8] 原福永, 刘宏阳, 王领, 冯凯东, 黄国言. 融合多特征的垃圾评论检测模型[J]. 小型微型计算机系统, 2020, 41(3): 539-543.
- [9] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. arXiv:1706.03762.
- [10] 曲强, 于洪涛, 黄瑞阳. 基于注意力机制的社交垃圾文本检测方法[J]. 网络与信息安全学报, 2020, 6(1): 54-61.
- [11] Bhuvaneshwari, P., Rao, A.N. and Robinson, Y.H. (2021) Spam Review Detection Using Self Attention Based CNN and Bi-Directional LSTM. *Multimedia Tools and Applications*, **80**, 1-18. <https://doi.org/10.1007/s11042-021-10602-y>
- [12] 贾红雨, 王宇涵, 丛日晴, 等. 结合自注意力机制的神经网络文本分类算法研究[J]. 计算机应用与软件, 2020, 37(2): 200-206.