Research Article

OWGC-HMC: An Online Web Genre Classification Model Based on Hierarchical Multilabel Classification

Guozhong Dong,1 Weizhe Zhang,1,2 Rahul Yadav,1 Xin Mu,3 and Zhili Zhou4

1The Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen 518000, China
2School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
3Frontier Research Centers, Peng Cheng Laboratory, Shenzhen 518000, China
4School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

Correspondence should be addressed to Weizhe Zhang; weizhe.zhang@pcl.ac.cn and Zhili Zhou; zhou_zhili@163.com

Received 9 November 2021; Revised 10 February 2022; Accepted 3 March 2022; Published 29 March 2022

Academic Editor: Mamoun Alazab

Copyright © 2022 Guozhong Dong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Web genre plays an important role in focused crawling, web link analysis, and contextual advertising. In this paper, web genre is defined as the functional purpose and the information type contained in the website. The intelligent classification of web genre can predict the content and functional type of website. However, there are several critical challenges to solve the web genre classification problem: lack of web genre classification dataset and efficient web genre classification mechanism. To improve web genre classification performance, we crawled Chinese websites of different web genres and converted crawled data into a hierarchical multilabel classification dataset. A website knowledge graph is constructed based on the relationship of website and meta tag features. Using entity features extracted from the knowledge graph, we propose an online web genre classification model based on hierarchical multilabel classification (OWGC-HMC) to mine the functional purpose of the corresponding website. Experimental results show that our OWGC-HMC model can mine hierarchical multilabel structure of web genre and outperform other web genre classification methods.

1. Introduction

Web classification [1] is the process of assigning websites to one or more classification labels. Web classification plays an important role in content recommendation [2–4] and contextual search [5–8]. According to different types of classification label, web classification can be divided into different classification problems such as topic classification [9, 10] and genre classification [11, 12]. The intelligent recognition of web genre can predict the content and functional type of website, facilitating web retrieval based on keywords. Web genre can be understood as functional categories of website such as online shopping, news media, government organization, resource download, and specialized information search. When a user is searching with keywords and web genres, the user can easily find search results about web genres [13]. Researchers have proposed different definitions of web genre and investigated various genre classification algorithms [14]. In this paper, web genre is defined as the functional purpose and the information type contained in the website. For example, the functional purpose of Taobao (https://www.taobao.com) is to provide users with an interface for searching products, shopping online, and making payments. The purpose of our work is to classify a website into specific web genres.

Web genre defined in this paper has a hierarchical structure. The hierarchical structure is determined by its functional types. Each web genre is distinguished by its unique functional purpose. A website may have one or more genres from predefined web genres. Some web genre classifiers were proposed to assign a single genre to a website in web genre classification research. However, multiple functional types can be easily combined in a website, which requires multiple web genre labels. Multilabel classification
is more suitable to capture the functional type of website. Multiple web genres of a website may have a hierarchical structure. For example, some “online shopping” websites can share conventions of “online mall.” Besides, some web genres may form a hierarchy relationship. “IT comprehensive” has a hierarchy relationship with “Internet technology.” The above characteristics of the web genre classification can be mapped to the hierarchical multilabel classification (HMC) problem.

In order to classify the functional purpose of a website, web pages within the website need to be crawled and analyzed. However, some researchers regarded meta tags of the homepage as website summary and classified websites only by analyzing the homepage of website. A canonical website contains some meta tags, which contain some useful information and can be used to index web page content. Title, description, and keywords are the most well-known and widely used meta tags. In this way, we do not need to analyze all web pages of the website [15]. Besides, there are several critical challenges to solve web genre classification problem: lack of web genre classification dataset and efficient web genre classification mechanism. In this paper, we first construct large-scale Chinese web genre classification dataset that captures both hierarchical and multilabel characteristics. Collected websites are annotated with several of the 119 genres. We removed web genres with fewer than 400 in our collected websites. Final dataset contains 77 web genres. Secondly, we propose a website knowledge graph based on the relationship of website and meta tag features. An online web genre classification model based on hierarchical multilabel classification (OWGC-HMC) is proposed to mine the functional purpose of corresponding website. Finally, we implement an online web genre classification system and verify the validity of the OWGC-HMC model.

The remainder of this paper is organized as follows. Section 2 presents previous works about web genre classification methods. Section 3 describes the online web genre classification system and OWGC-HMC model. Evaluation of OWGC-HMC is presented in Section 4. Finally, Section 5 concludes our work and proposes future research.

2. Related Works

In this section, we present and summarize some recent research about web genre classification methods.

Lim et al. [16] proposed multiple feature sets to classify genres for web documents. Multiple features were extracted from URL and HTML tags. Based on a dataset of 15 genres, they found that the main body and anchor text information are most effective to identify web genres. Kennedy and Shepherd [17] proposed a neural network classifier to classify homepages as personal homepages, corporate homepages, or organization homepages. A dataset of 321 web pages was evaluated, and F-measures were from 0.78 to 0.85 when evaluations were conducted on personal and corporate home pages. Elizabeth and Howe [18] proposed a logistic regression classifier to classify genres of web documents. Experiments were performed on a dataset of ten genres, and average correctness of 91.5% was obtained. Gams [19] proposed surface, structural, presentation, and other features, which were utilized to design a genre classifier with several machine learning algorithms. Also, the author indicated that the genre classifier could be a useful addition to search engines. Vidulin et al. [20] used standard ML algorithms and web genre classification to better capture the relations between web genres. Abramson and Aha [21] proposed a new genre classification algorithm based on URL, which combined linear interpolation smoothing for classifying URLs in a Naive Bayes classifier. Jebari [12, 22] proposed a new multilabel centroid-based approach with URL, title, headings, and link features, which calculated the similarity between the new page and each genre centroid. Deri et al. [23] defined 13 genres and extracted the list of all registered .it domains from the domain database. They evaluated how to combine probabilistic classifier and SVM classifier to improve the results. Madjarov et al. [24] constructed a hierarchy of web genres in the web genres labels. The HMC method was used to boost the predictive performance. The experimental results showed that the hierarchy structure of web genres could improve the performance of classifiers. Chaker [25] proposed a segment-based weighting approach for web genre classification, which extracted character n-grams from the URL. Experiments conducted on three web genre datasets showed that the approach could achieve encouraging results. Ebubekir and Banu [15] proposed meta tag features and developed an RNN-based deep learning system. The experimental results showed that the accuracy reaches 85%. Gjorgji et al. [13] treated web genre classification as a predictive task. They investigated the structuring of the output space by constructing hierarchies. The experimental evaluation showed that surface and paragraph features offer the best performance.

An overview of web genre classification approaches indicates a lack of large-scale web genre data that capture both hierarchical and multilabel aspects. In this paper, we construct a web genre classification dataset and website knowledge graph that can be used by multilabel web genre classification. More specifically, we provide the OWGC-HMC model to grasp hierarchical and multilabel structures of website based on entity features extracted from the website knowledge graph.

3. Web Genre Classification Based on OWGC-HMC

In this section, we present the online web genre classification system and OWGC-HMC model.

3.1. Overview of Online Web Genre Classification System. The framework of the online web genre classification system is shown in Figure 1, which comprises three functional layers, namely, hierarchical multilabel classifier construction, web genre feature extraction, and web genre prediction. Hierarchical multilabel classifier construction is designed to construct a website knowledge graph and hierarchical multilabel classifier. Each website node in the
website knowledge graph has a subdomain relationship and external hyperlink relationship with other website nodes. Each website node has meta entities, which are extracted from the title and description in meta tags. Web genres and meta tags of website in the website ranking list are crawled according to aizhan.com and Chinaz.com. Entities and entity relationships are extracted from the external link crawler and DNS log system. Web genre features extracted from the website knowledge graph and corresponding web genre labels are utilized to train the OWGC-HMC classifier. Web genre feature extraction crawls meta information and extracts entity features based on the website knowledge graph. Web genre prediction utilizes hierarchical multilabel classifier API combined with website features to predict web genres.

3.2. Online Web Genre Classification Model Based on Hierarchical Multilabel Classification. Let $D = \{E_i, N_i\}_{i=1}^M$ denote the web genre classification dataset where each $E_i$ is entity set of website, $N_i$ is web genre label set, and $M$ is the number of websites. The hierarchy of web genres is a tree structure, which is defined over web genres set by their parent-child relationships. Collected websites were annotated with several of the 119 genres. The hierarchical structure of web genres is shown in Figure 2. For convenience, $\tau(n)$ represents the parent of the web genre $n$ in web genre tree, $y_{in} \in \{-1, 1\}$ denotes whether $E_i$ can be classified into web genre $n$. The problem of hierarchical multilabel web genre classification is to construct a classification model that predicts the target web genre labels of a given website with the smallest possible error.

The hierarchical dependencies between the web genre labels are encoded in a hierarchy used in the learning process. The framework of online web genre classification based on a hierarchical multilabel classification model is composed of four layers.

The input layer determines the entity set of website, which is extracted and processed based on the website knowledge graph. For a website $i$, let $EO_i$, $EH_i$, and $ES_i$ denote its own entity set, external hyperlink entity set, and subdomain entity set. The weight of an entity $e \in E_i$ can be formalized as

$$\omega_e = 0.5 \times C_e(EO_i) + 0.3 \times C_e(ES_i) + 0.2 \times C_e(EH_i),$$

where $C_e(EO_i)$, $C_e(ES_i)$, and $C_e(EH_i)$ represent the number of entities $e$ appearing in the corresponding entity set. Finally, the entities with top $K$ entity weight are selected to $E_i$.

The embedding of input entities will be generated at the embedding layer. Region embedding is used in our OWGC-HMC model, which is a supervised embedding method. The representation of input entities includes the embedding of the entity itself and a weighting matrix to interact with the local entities.

The encoder input entity features $e_1, e_2, \ldots, e_N$ and the embedding of input entities. We use a bag of n-grams as additional features to capture partial information about the local entity order. We maintain a fast and efficient memory mapping of the n-grams by using the hashing trick with the same hashing function as in [26]. The features are embedded and averaged to form the hidden variable. The entity representations are then averaged into a text representation, which is fed to a linear classifier. The text representation is a hidden variable that can be potentially reused. This architecture is similar to the model of Mikolov et al. [27], where the middle word is replaced by a label.

The output layer determines hierarchical web genre classification tasks. For our classification task, we use BCEWithLogitsLoss for hierarchical classification and add a recursive regularization. Using this regularization framework, we incorporate the hierarchical dependencies between web genre labels into the regularization structure of the parameters. In the problem of hierarchical multilabel classification, the prediction function is formalized as a set of parameters:

$$W = \{w_n, n \in N\}.$$ (2)

Each web genre $n$ in the hierarchy is associated with a parameter vector $w_n$. The prediction web genre labels are parameterized by a set of parameters $\hat{w}$ which are then estimated in the learning process.

$$\hat{w} = \arg \min_\lambda (\lambda (\hat{w}) + \alpha \times R_{emp}),$$ (3)

where $R_{emp}$ presents the empirical risk or loss on the web genre training dataset, $\lambda (\hat{w})$ represents the regularization term, and $\alpha$ is a parameter that controls the trade-off between fitting to the given training instances and the complexity of the hierarchical multilabel classification model.
Figure 2: The hierarchical structure of web genres.
The empirical risk in our OWGC-HMC model is defined as the loss incurred by the instances at the leaf nodes of the hierarchy:

\[ R_{\text{emp}} = \sum_{n \in T} \sum_{i=1}^{M} L(y_{in}, E_i, w_n), \tag{4} \]

where \( L \) can be any convex loss function. BCEWithLogitsLoss function is used in our OWGC-HMC model.

We use the hierarchy in the learning process by incorporating a recursive structure into the regularization term for \( W \):

\[ \lambda(W) = \sum_{n \in T} \frac{1}{2} \| w_n - w_{\tau(n)} \|^2. \tag{5} \]

This recursive form of regularization enforces the parameters of web genre node to be similar to the parameters of its parent web genre node under the Euclidean norm. The hierarchical dependencies are considered, which can encourage parameters nearby in the hierarchy to be similar. This can leverage information from nearby web genre labels while estimating model parameters.

4. Experimental Results and Analysis

This section presents some experiments to evaluate our OWGC-HMC model and compares it with other web genre classification methods. This section is organized into three subsections. The first section presents the dataset used in our experiments. The second section describes our performance measures. Finally, the third section discusses the experimental results.

4.1. Dataset. We construct the hierarchical multilabel dataset that contains 1,16,350 websites. Selected websites are of high interest for users, which are collected from a highly ranked list of aizhan.com. Collected websites are annotated with several of the 119 genres. We removed web genres with fewer than 400 in our collected websites. Final dataset contains 77 web genres. Web genres and corresponding numbers in the final dataset are listed in Table 1. For our final dataset, the web genre cardinality is 2.21, which is the average web genre number for each website in the dataset. The genre density is 0.029, which is the average web genre number for each web divided by the total number of web genre.

| Web genres                                  | Number  |
|---------------------------------------------|---------|
| Domestic services                           | 35148   |
| Education culture                           | 17367   |
| Leisure and entertainment                   | 16103   |
| Internet technology                         | 14482   |
| Local portal                                | 13822   |
| Industry enterprise                         | 11887   |
| News media                                  | 10393   |
| School website                              | 8267    |
| Government organization                     | 7135    |
| Online shopping                             | 6704    |
| Government portal                           | 6594    |
| Online store                                | 6379    |
| Forum community                             | 6200    |
| Educational examination                     | 5409    |
| Job recruitment                             | 4467    |
| Travel                                      | 4462    |
| Game website                                | 4283    |
| Video movie                                 | 3825    |
| Software download                           | 3772    |
| Category message                            | 3457    |
| News portal                                 | 3374    |
| Real estate website                         | 3261    |
| Medical health                              | 3049    |
| Novel reading                               | 2754    |
| Financial investment                        | 2538    |
| Training institution                        | 2142    |
| Design material                             | 2141    |
| Technical programming                       | 1980    |
| IT comprehensive                            | 1887    |
| Auto website                                | 1876    |
| Physical fitness                            | 1832    |
| Travel tips                                 | 1682    |
| Travel agency                               | 1680    |
| Photography picture                         | 1661    |
| Music website                               | 1645    |
| Newspapers and magazines                    | 1575    |
| Diet                                        | 1540    |
| Home building                               | 1449    |
| Star fashion                                | 1278    |
| Association organization                    | 1183    |
| Literary papers                             | 1169    |
| Web navigation                              | 1125    |
| Common query                                | 1120    |
| Webmaster resources                         | 1049    |
| Hospital clinic                             | 1032    |
| Mobile digital                              | 1027    |
| Bancassurance                               | 1025    |
| Animated cartoons                           | 924     |
| Early childhood education                   | 899     |
| Dating                                      | 878     |
| Group-buying website                        | 839     |
| Search engine                               | 819     |
| Health care                                 | 793     |
| Broadcasting                                | 770     |
| Rebate discount                             | 767     |
| Ticket booking                              | 763     |
| Foreign language synthesis                  | 752     |
| Technology information                      | 749     |
| Government functions                        | 633     |
| Data analysis                               | 623     |

4.2. Performance Measures. To evaluate our multilabel classification model, we use precision, micro-F1, macro-F1, and Hamming loss evaluation measures. In our experiments, we follow the 10-fold cross-validation procedure. The evaluation result is the average of 10 individual performances. The ratio of the final training set, verification set, and test set is 8:1:1. After training, the test set is used to test the web genre classification model.

Micro-F1 is a conventional evaluation metric used to evaluate classification performance. The micro-averaged Micro-F1 is given by
where $TP_l$, $FP_l$, and $FN_l$ represent the true positives, false positives, and false-negatives for web genre label $l \in L$.

Macro-F1 gives equal weight to each class label, which is given by

$$P' = \frac{TP_l}{TP_l + FP_l},$$

$$R' = \frac{TP_l}{TP_l + FN_l},$$

(7)

$$\text{macro } - F1 = \frac{1}{L} \sum_{l=1}^{L} \frac{2PR'}{P' + R'},$$

4.3. Results and Discussion. In this section, we conduct three experiments. The aim of the first experiment is to compare the accuracy of different encoders [28–30] using word entity and character features in the encoder layer. Experiment results using different encoders are presented in Figure 3, where OWGC-HMC got the best accuracy. We can conclude that the best results are achieved using word entity rather than character. This can be explained by the fact that word entities in title and description contain more web genre specific word entities than characters. The second experiment aims to compare Hamming loss and training time obtained using word entity and character features. Experiment results using different features are presented in Figures 4 and 5, where OWGC-HMC has the lowest loss and training time. Overall, OWGC-HMC outperforms other classifications using word entity and character features concerning all evaluation metrics.
In order to evaluate the classification performance of the OWGC-HMC model for different web genres, we construct the confusion matrix on the test dataset and compute it for the classification accuracy of different web genres. Classification accuracy of different web genres is shown in Table 2.

As shown in Table 2, the F1 score of most web genres in the test set exceeds 0.85. We analyze these web genres of low F1 score and find that their meta tag descriptions are most ambiguous and irregular, such as local portal and entertainment gossip. Most websites have professional SEO managers and meta tag features, which can be applied to web genre classification tasks.

In the third experiment, we compare our OWGC-HMC model with WGC-SOP [13]. WGC-SOP proposes ten different feature representations of the websites. Predictive
clustering trees are used to assess the influence of the different information sources. The hierarchy construction method in WGC-SOP is manual (MAN) for the comparison experiment. Figure 6 shows the comparison results of accuracy. It can be seen from Figure 6 that our method has higher precision, recall, and F1 score compared with WGC-SOP. Compared with WGC-SOP, the improvement of the OWGC-HMC model is obvious. This proves that the OWGC-HMC model based on meta tag features has better classification performance and only fewer features are extracted, which can be used in the online environment.

5. Conclusion

In this paper, we constructed a Chinese web genre classification dataset and converted it into an HMC dataset. Based on the Chinese web genre classification dataset, we proposed the OWGC-HMC model to mine hierarchical and multilabel structures of web genre data. Besides, we developed an online web genre classification system based on hierarchical multilabel classification. In our future research work, we intend to evaluate our model on more hierarchical multilabel classification datasets and provide visualization techniques for hierarchical multilabel classification.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported in part by the Key-Area Research and Development Program of Guangdong Province (2020B0101360001), the Peng Cheng Laboratory Project (grant no. PCL2021A02), and the National Natural Science Foundation of China (62106114).

References

[1] M. Hashemi, “Web page classification: a survey of perspectives, gaps, and future directions,” Multimedia Tools and Applications, vol. 79, no. 17, Article ID 11945, 2020.
[2] K. Yang, S. Long, W. Zhang, J. Yao, and J. Liu, “Personalized news recommendation based on the text and image integration,” Computers, Materials & Continua, vol. 64, no. 1, pp. 557–570, 2020.
[3] B. Janakiraman and S. Arumugam, “Personalized nutrition recommendation for diabetic patients using optimization techniques,” Intelligent Automation & Soft Computing, vol. 26, no. 2, pp. 269–280, 2020.
[4] W. Gong, W. Zhang, M. Bilal, Y. Chen, X. Xu, and W. Wang, “Efficient web APIs recommendation with privacy-preservation for mobile app development in industry 4.0,” IEEE Transactions on Industrial Informatics, p. 1, 2021.
[5] Z. Zhou, Q. M. J. Wu, Y. Yang, and X. Sun, “Region-level visual consistency verification for large-scale partial-duplicate image search,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 16, no. 2, pp. 1–25, 2020.
[6] Z. Fang, J. Wang, B. Wang, J. Zhang, and Y. Shi, “Fuzzy search for multiple Chinese keywords in cloud environment,” Computers, Materials & Continua, vol. 60, no. 1, pp. 351–363, 2019.
[7] Y. Xue, T. Tang, and W. Pang, “Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifiers,” Applied Soft Computing, vol. 88, Article ID 106031, 2020.
[8] T. Wang, Y. Quan, X. S. Shen, T. R. Gadekallu, W. Wang, and K. Dev, “A privacy-enhanced retrieval technology for the cloud-assisted Internet of things,” IEEE Transactions on Industrial Informatics, p. 1, 2021.
[9] P. Kalaiavani and D. R. Thangarajan, “Enhancing the classification accuracy in sentiment analysis with computational intelligence using joint sentiment topic detection with medlda,” Intelligent Automation & Soft Computing, vol. 26, no. 1, pp. 71–79, 2020.
[10] J. Ma, J. Cheng, L. Zhang, L. Zhou, and B. Chen, “A phrase topic model based on distributed representation,” Computers, Materials & Continua, vol. 64, no. 1, pp. 455–469, 2020.
[11] Y. Kim and S. Ross, “Examining variations of prominent features in genre classification,” in Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), pp. 132–142, Waikoloa, HI, USA, January 2008.
[12] C. Jebari, “Multi-label genre classification of web pages using an adaptive centroid-based classifier,” Journal of Information and Knowledge Management, vol. 15, no. 1, Article ID 16500088, 2016.
[13] G. Madjarov, V. Vidulin, I. Dimitrovski, and D. Kocev, “Web genre classification with methods for structured output prediction,” Information Sciences, vol. 503, no. 1, pp. 551–573, 2019.
[14] G. Chen and B. Choi, “Web page genre classification,” in Proceedings of the 2008 ACM symposium on Applied computing, pp. 2353–2357, Fortaleza, Ceara, Brazil, March 2008.
[15] B. Ebubekir and D. Banu, “Web page classification using rnn,” Procedia Computer Science, vol. 154, no. 1, pp. 62–72, 2019.
[16] C. S. Lim, K. J. Lee, and G. C. Kim, “Multiple sets of features for automatic genre classification of web documents,” Information Processing & Management, vol. 41, no. 5, pp. 1263–1276, 2005.

[17] A. Kennedy and M. Shepherd, “Automatic identification of home pages on the web,” in Proceedings of the 38th Annual Hawaii International Conference on System Sciences, pp. 236–251, Big Island, HI, USA, January 2005.

[18] S. Elizabeth and A. Howe, “Genre classification of web documents,” in Proceedings of the 20th national conference on Artificial intelligence, pp. 1596-1597, Pittsburgh, PA, USA, July 2005.

[19] V. Vidulin, M. Luštrek, and M. Gams, “Using genres to improve search engines,” in Proceedings of the International Workshop Towards Genre-Enabled Search Engines (RANLP 2007), pp. 45–51, Borovets, Bulgaria, September 2007.

[20] V. Vidulin, M. Lustrek, and M. Gams, “Multi-label approaches to web genre identification,” JLCL, vol. 24, no. 1, pp. 97–114, 2009.

[21] M. Abramson and D. Aha, “What’s in a url? genre classification from urls,” in Proceedings of the AAAI Workshop-Intelligent Techniques for Web Personalization and Recommender Systems, pp. 2–9, Toronto, Ontario, Canada, July 2012.

[22] C. Jebari and M. Wani, “A multi-label and adaptive genre classification of web pages,” in Proceedings of the International Conference on Machine Learning & Applications, pp. 578–581, Boca Raton, FL, USA, December 2012.

[23] L. Deri, M. Martinelli, D. Sartiano, M. Serrecchia, L. Sideri, and S. Prignoli, “Implementing web classification for tlds,” in Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 85–88, Singapore, December 2015.

[24] G. Madjarov, V. Vidulin, I. Dimitrovski, and D. Kocev, “Web genre classification via hierarchical multi-label classification,” in Proceedings of the 16th International Conference on Intelligent Data Engineering and Automated Learning, pp. 9–17, Wroclaw, Poland, October 2015.

[25] J. Chaker, “A segment-based weighting technique for url-based genre classification of web pages,” Polibits, vol. 53, no. 1, pp. 43–48, 2016.

[26] A. Joulin, E. Grave, and P. Bojanowski, “Bag of tricks for efficient text classification,” in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 427–431, Valencia, Spain, April 2017.

[27] T. Mikolov, K. Chen, and G. Corrado, “Efficient estimation of word representations in vector space,” in Proceedings of the International Conference on Learning Representations, pp. 1–12, Scottsdale, AR, USA, May 2013.

[28] Y. Kim, “Convolutional neural networks for sentence classification,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1746–1751, Doha, Qatar, October 2014.

[29] R. Wang, Z. Li, and J. Cao, “Convolutional recurrent neural networks for text classification,” in Proceedings of the Proceedings of 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–6, Budapest, Hungary, July 2019.

[30] A. Vaswani, N. Shazeer, and N. Parmar, “Attention is all you need,” in Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 5998–6008, CA, USA, December 2017.