Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 3, March 2019, pg.97 – 111

Assessing the Quality of Scientific Articles Using Artificial Neural Networks

Iman KHALEEL[1]; Dr. Sefer KURNAZ[2]
ALTINBAŞ University, Engineering and Naturel Science, Mahmutbey, Istanbul, Turkey
[1] menh_saad@yahoo.com, [2] sefer.kurnaz@altinbas.edu.tr

Abstract: Assessing the quality of scientific articles being submitted for publishing through digital libraries is an essential step that ensures that the published articles meet the qualifications required by the journal, to maintain the reputations of these journals. Normally, these articles are peer-reviewed by experts in the field the article investigates, which has dramatically increased the time required to assess these articles before publishing. The quality of an article is measured mainly by the quality of the writing and the significance of the field it investigates. In this study, a quality assessment technique is proposed, which uses artificial neural networks to predict the number of citations the article is expected to gain as a measure of its quality. The evaluation is based on three main components of the article, the title, keywords and abstract, as these components can reflect the overall quality of the article, without the need of excessive processing of the entire article. Two approaches are evaluated, the first attempt to measure the quality per each of the components, separately, and fuse the measures into a single overall quality measure. The second approach uses a single hybrid neural network that combines the three networks together, so that, these components are processed simultaneously. Per each approach, three types of neural networks are evaluated, which are the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Long- Short-Term Memory (LSTM). The results show that the use of the CNN with the hybrid network has achieved the best predictions, with 4.52 Mean Squared Error (MSE).

Keywords: Artificial Neural Networks; Convolutional Neural Networks; Recurrent Neural Networks; Long-Short-Term Memory; Natural Language Processing.

1. Introduction

The rapid development in different fields of study is the result of the rapidly growing number of researches being conducted in these fields. The ease of accessing information has encouraged many researchers to investigate and propose new techniques, in order to propose improvements to the field of study these studies are conducted in. The highly available internet access and digital libraries are two of the main factors behind the ease of accessing these researches, where researchers have become able to search for and get the latest researches over the internet. Moreover, researchers have also been able to publish their article online, by submitting a digital version of the articles to the journals of their choice for publishing [1, 2].

Publishing a scientific article in a journal, especially the highly rated journal, requires peer review, which is the evaluation of reliable professionals in the same field of study that the article investigates. Those reviewers determine the suitability of the article, by means of scientific contribution and information presentation in the article. If the article is approved for publishing, the journal publishes the article and index it in its digital library, so that, it becomes accessible to other researchers. According to scientific research morality and ethics, the use of information acquired from any article requires citing that article. The citation is an acknowledgment from the author of an article that indicates the use of some information from the other article, which is included in the references of the article. Thus, scientific articles that have a significant contribution to the field of their study and good information presentation, i.e. better quality, receive more citations than others [3, 4].

In addition to the quality assurance for the articles published in a certain journal, these journals are rated using the Journal Impact Factors (JIFs), which is the average number of citations per article published in that journals. Thus, reviewers attempt to choose the articles

expected to get more citations, in order to improve the JIFs of the journal publishing those articles. Journals with higher JIFs are normally more selective than others, i.e. authors need to ensure the quality of their articles in order to get them approved by these journals [5].

In addition to the information presentation and contributions of an article, the field that the article studies also have a significant impact on the number of citations that the article gets. Certain fields of scientific research get more attention, which increases the chances of getting more citation. Thus, the number of citations per year of an article can be used as a measure of the quality of the article, hence, predicting this number can be used to predict the quality of the paper and the possibility of publishing it in a highly rated journal. However, evaluating the entire article is difficult, according to the huge number of words that an article may have. On the other hand, certain parts of the article can be used to produce an overview of the entire article, in order to estimate its quality [6, 7].

The title of the article is one of the important parts of an article, where the main objective of the article can be known, as well as the backbone methodology behind the used technique. The keywords also have an important role in recognizing the fields of study that the article investigates, the techniques and subfields included in the article. Moreover, the abstract can provide a better overview of the information presentation and the structure of the sentences, as well as the possibility of finding some results or conclusions from the study. In most of the journals, the size of these parts is limited, and published articles cannot exceed them. Thus, it is possible to define the maximum length of each of these parts a model built to measure the quality of the article.

2. Literature Review

The implementation of an Artificial Neural Network (ANN) is inspired by the distribution and connections among an enormous number of biological cells in the brains of humans, known as neurons. Depending on the biological connections among these neurons and the conductivity of these connections, different decisions are made from different humans, for that same set of information collected using different senses. The effect of one neurons decision, which is represented by an electrical impulse, on the next neurons that it is connected to, is controlled by electrochemical solutions, one per each connection. As the conductivity of the solution is increased, the effect of the decision made by the neuron increases on the next one, and vice versa [8].

Artificial neural networks also have neurons, which as the basic component of these networks. These neurons are also connected to each other, where different types of connections produce different types of decisions. Moreover, the effect of one neuron on another is controlled by numerical values, corresponding the connection, known as weights. Each neuron collects the outputs from neurons connected to it, multiplied by the weight of that connection. Then, the neuron passes the results of the summation into a function, known as activation function, which provides the neuron with non-linearity to produce better decisions. The values of the weights are updated during the training phase of the neural network, using the backpropagation technique. To simplify the representation, neurons in an artificial neural network are grouped in layers, where outputs from neurons in one layer are delivered as inputs to the neurons in the neurons of the next layer. However, neurons in the input layer receive their inputs from the external world that they are interacting with, while the output of the neurons in the output layer are connected to

the system the network is controlling. All remaining layers are denoted as hidden layers, as they are internal layers that are not visible to the external world [9].

Depending on the distribution of the neurons in a layer, and the connections among the neurons from different layers, different types of artificial layers can be implemented. Depending on the application that the artificial neural network is implemented for, specific types of these networks can have better performance in that application, compared to the other types. For example, a Convolutional Neural Network (CNN) has convolutional layers, where the weights are distributed in two-dimensional filters. This type of ANNs is inspired from the human's visual system, where visual features are detected in two dimensions and related to each other's position, rather than their relative position to the border of the two-dimensional input. Thus, this type of ANNs has shown better performance when used with images, according to the flexibility that local two-dimensional features detection provides [10].

Another type of ANNs is the Recurrent Neural Network (RNN), where the output of a neuron in a certain time instance affects its output in the next time instance. This behavior is achieved by feeding the output at the last time instance back to the same neuron. In traditional RNNs, this output is only multiplied by a weight value, which is also updated using the backpropagation, during the training of the neural network. Thus, the effect of the last output is also adjusted, depending on the behavior of that neuron and how dependent its current output on the last output. Figure 1 shows the neuron cell in an RNN and illustrates its operation relevance over time by unfolding its operations for three consequent time instances [11].

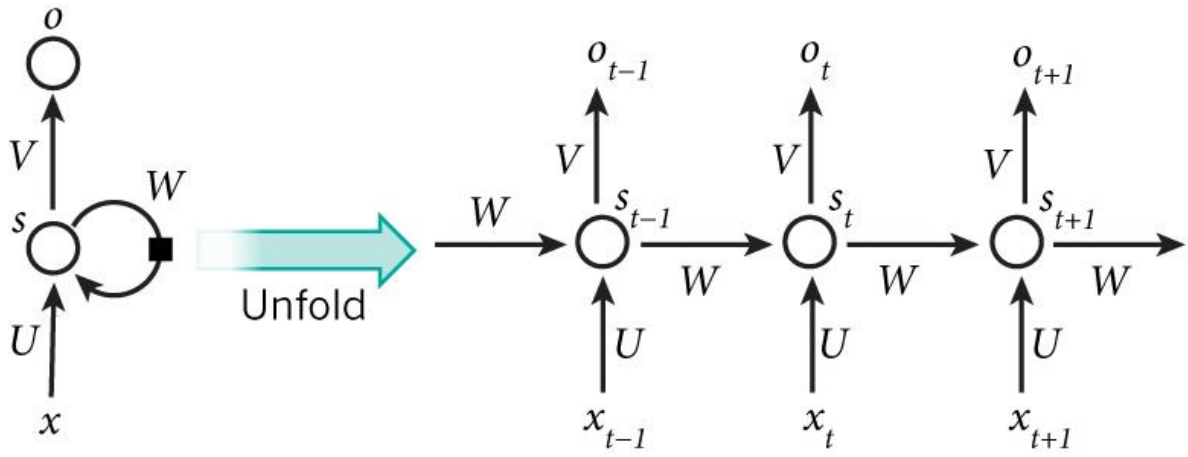

Figure 1: Structure of a neuron's cell in a traditional RNN neural network.

As the output of a neuron in an RNN is weighted and fed back to its inputs, the effect of the output at a certain time instance starts to fade out as more instances are processed. In some applications, such behavior may lower the performance of the neural network, as earlier inputs may have more impact on the output in the current time instance. A popular example of such applications is the weather forecasting, where the most effective values on the prediction made at a time instance are the most recent values and the values for the same date one year earlier. In traditional RNNs, even if such relevance can be detected, it is impossible to adjust the weights in the network to achieve it. Thus, a more complicated type of RNNs is employed for such applications, which is the Long- Short-Term Memory (LSTM). The block diagram shown in Figure 2 illustrates the basic components of a neuron in an LSTM neural network [12, 13].

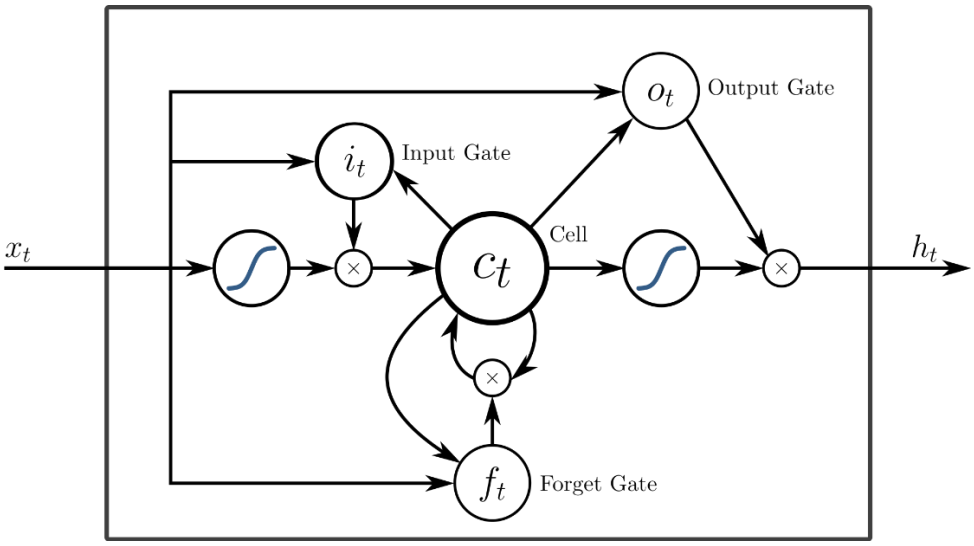

Figure 2: Illustration of an LSTM block.

The existence of the control gates in an LSTM block provides the neural network with the ability to control the time instances that are included in the output at the current time instance. The input gate controls the effect of the input at this time instance on the output, while the forget gate controls the effect of the previous outputs on the current output. These gates depend on the value of the inputs of the neuron and last calculated value, in order to make the required decision. Moreover, the output gate controls the direct effect on the input on the output of the neuron, after being passed through the activation function. Using such topology, the effect of the previous values can be adjusted based on their importance, rather than their position [14].

Different Natural Language Processing (NLP) methods are proposed based on ANNs, where the text is fed to the neural network in order to extract the required knowledge. Yin et al. [15] evaluate the performance of RNNs and CNNs in NLP. The results of this study show that the CNNs and LSTM neural networks have the highest performance, compared to other types of ANNs. Moreover, Dos and Gatti [16] illustrate the ability of CNNs in sentimental analysis of text, where the output indicates whether the input sentence has positive or negative sentiment.

3. Proposed Method

In order to assess the quality of scientific articles, a dataset is collected for existing articles. Per each article, the title, abstract and keywords are collected, in order to analyze these data. However, as a training dataset is required, the number of citations per year is also collected for each article. This average citation is used as the actual quality measure of the article, instead of relying on manual evaluation. This quality measure includes all the factors required from the article, in order to have a significant impact on the literature.

Three types of neural networks are used to predict the quality of the article, which are the CNN, RNN and LSTM. Per each type of neural networks, three models are implemented, one for the title, one for the abstract and one for the keywords. Each model predicts a quality measure, which is collected and fused in a single measure that describes the overall quality of the article. Two types of fusion techniques are going to be evaluated, which are statistical- and machine-learning-based. Statistical techniques use traditional functions, such as the average and median of the calculated measures, in order to provide the overall quality measure, while feed-forward artificial neural network is implemented to fuse these values, in the machine-learning-based technique.

For the machine learning based fusion approach, a single hybrid neural network is implemented, where the text corresponding each network is fed to that network, which is eventually connected to a feed-forward neural network. This approach enables the backpropagation method to update the weights of the entire network, based on the error measured between the required values and those calculated by the neural network. For statistical-based model, three separate neural networks are implemented, where each network is required to predict the average citations per year separately. This approach also assists in recognizing the

input that has the most accurate predictions, i.e. the part of the article that is most suitable for quality measurement. Figure 3 illustrates both statistical and machine learning based topologies.

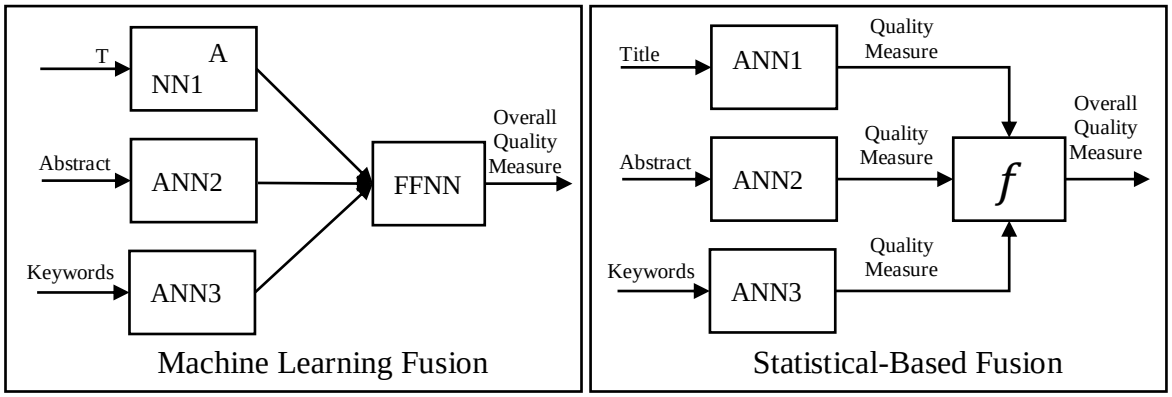

Figure 3: The proposed statistical and machine learning-based approaches.

4. Experimental Results

To evaluate the performance of the proposed method, two UCI [17] datasets, collected from the Association for the Advancement of Artificial Intelligence (AAAI) journal. The datasets contain the title, keywords and abstract of the articles accepted by that journal in 2013 and 2014, a dataset per each year. However, these datasets do not contain the number of citations each article has earned, which are used as the quality measured for these articles to train and evaluate the performance of the proposed methods. The number of citations per each article in the dataset is collected, one by one, from Google Scholar indexes search engine. The collected numbers of citations are then divided by the number of years the article has been published for and used as a measure of the quality of that article. These datasets are combined into a single dataset for the training and evaluation of the proposed method. Two of the articles in the resulting dataset have citation rates higher than 100/year. These numbers are reduced to 100, in order to avoid

overfitting the neural networks. Each of the components used for the evaluation is the tokenized, separately, and padded, so that, all the texts per each component have the same length of numerical values. All experiments are conducted using a Windows computer with Intel Core i7-7700HQ processer and 8GB of memory, with a Graphics Processing Unit with 8GB of memory. The GPU is used to accelerate the mathematical operations required by the artificial neural network, according to the ability of GPU to parallelize matrices operations. All evaluations are performed using 5-fold cross-validation, to avoid any biased results toward any of the implemented neural networks, where the methods are implemented using Python programming language.

4.1. Experiment A

In this experiment, the statistical model is trained and evaluate, where the title, keywords and abstract per each article are used to calculate the quality of that article. The neural networks in this approach are trained to predict the overall quality of the article based on each of these components, solely. Then, the resulting three values are summarized into one value, using the average and median functions. To select the appropriate type of neural networks for title processing, the Convolutional, simple Recurrent and Long- Short-Term Memory (LSTM) networks are evaluated. The Mean-Squared-Errors are calculated per each of the evaluated networks, as shown in Table 1.

Table 1: Mean-Squared-Error of the different neural networks used for title processing.

Neural Network	**MSE**
Convolutional	12.06
Simple-RNN	24.73
LSTM	18.15

As the order of the keywords do not have significant effect on the meaning of these keywords, where moving a keyword before or after another does not change the meaning of these keywords, a feed-forward neural network is used to process these keywords. Keywords that have more than one word in it are joint together and tokenized as a single word. The MSE of the predictions provided by this neural network is 28.86. Moreover, as the order of the words have significant effect on how the sentences, in the abstract, are interpreted, the text in the abstract is processed using the same types of neural networks used for the title, which are the CNN, simple-RNN and LSTM. The MSE per each type of these neural networks is shown in Table 2.

Table 2: Mean-Squared-Error of the different neural networks used for abstract processing.

Neural Network	**MSE**
Convolutional	7.37
Simple-RNN	16.25
LSTM	14.31

As the one-dimensional convolutional neural networks have achieved the highest performance, i.e. least MSE values, in both title and abstract processing, these networks are used alongside with the feed-forward fully connected network, used for the keywords. The resulting model has an overall MSE of 16.10 when the average function is used and 8.47 when the median function is used.

4.2. Experiment B

In this experiment, a single neural network is implemented to handle all the inputs and predict a single quality measure. Similar to the previous experiment, a fully connected feed-forward artificial neural network is used to process the keywords part of the inputs. However, the same three types of neural networks are evaluated, as shown in the left side of Figure 3, which are the convolutional, simple-RNN and LSTM. Each of these neural networks outputs eight values, i.e. have eight neurons in their last layer, which are connected to a single layer in the feed-forward neural network used to combine them. The output layer of the hybrid neural network consists of a single neuron, which outputs the overall quality measure of the article. The performances of the evaluated types of neural networks are shown in Table 3.

Table 3: Mean-Squared-Error of the different hybrid neural networks.

Neural Network	**MSE**
Convolutional	4.52
Simple-RNN	12.91
LSTM	7.66

The one-dimensional convolutional neural network has been able to outperform the LSTM networks, which normally have better performance in text processing. However, the results also show that the performance of the LSTM is the closest to the convolutional-based, with significantly better performance than the Simple-RNN networks. These performances are results of the ability of convolutional and LSTM layers to detect and combine words from different positions in the sentences, which is not the case in the simple-RNN network, where closer words have higher effect on the current value than older ones.

5. Conclusion

With the existence of digital libraries, an enormous number of scientific articles are being submitted to journals for publishing. These journals require per-reviewing these articles to ensure that they match the qualifications of the journal before being submitted, so that, journals can maintain their reputations. Accordingly, the time that each article is queued for peer-reviewing has increased dramatically as each article must be reviewed regardless of its quality. However, as a huge percentage of these articles are being rejected, assessing the quality of these articles prior to making the decision can significantly improve the efficiency of this process. Such assessment can assist authors improving the quality of the articles before submission, reviewers to have an estimation of the quality of the paper and journals to have certain threshold for papers to be peer-reviewed.

In this study, a new quality assessment for scientific articles is proposed, based on the significance of the field of the study the article investigates and the quality of the writing. The proposed method makes use of artificial neural networks, according to their ability to process texts written in natural language. This assessment is conducted based on the title, keywords and abstract of the articles. Two approaches are evaluated for the proposed method, the first attempt to estimate the quality based on each component, separately, and calculate the overall quality using statistical methods and the three measures of these components. The second approach uses a single hybrid neural network that processes all the inputs simultaneously. Three types of neural networks are evaluated for the title and abstract, which are the CNN, Simple-RNN and LSTM, as the position of the words in these components affect the semantic meaning of the phrases. The hybrid method has shown higher performance, where the CNN has achieved the best performance among the evaluated neural networks. These results also confirm the results of the

comparisons conducted in [18-20], which encourage the use of both the CNN and LSTM layers in the same networks. Moreover, the proposed method has been able to achieve 4.52 MSE, which is achieved by the hybrid network with the CNN layers. Thus, the proposed method can significantly improve the efficiency of peer-reviewing process by reducing the time required to process low-quality articles that are most probably are rejected.

In future work, both CNN and LSTM layers are going to be employed in the hybrid network in order to improve the performance and reduce the MSE of the proposed method. The performance of such network is going to be evaluated in means of improvement in accuracy and reduction in execution speed, as the use of these layers in the same network significantly increases the complexity of the computations. Another important scheme to improve the accuracy of the predictions is to include more information in the quality assessment procedure, such as the experimental result and conclusion of the article. Increasing the amount of data can increase the accuracy of the extracted knowledge, hence, the accuracy of the provided predictions. However, this can also increase the complexity of the computations. Thus, a balance between the complexity of the computations, i.e. the prediction speed, and the accuracy of the model is investigated.

References

[1] V. Larivière, S. Haustein, and P. Mongeon, "The oligopoly of academic publishers in the digital era," *PloS one*, vol. 10, p. e0127502, 2015.

[2] P. Lyman, "What is a digital library? Technology, intellectual property, and the public interest," in *Books, Bricks and Bytes*, ed: Routledge, 2017, pp. 1-34.

[3] C. J. Haug, "Peer-review fraud—hacking the scientific publication process," *New England Journal of Medicine*, vol. 373, pp. 2393-2395, 2015.

[4] V. M. Nguyen, N. R. Haddaway, L. F. Gutowsky, A. D. Wilson, A. J. Gallagher, M. R. Donaldson, *et al.*, "How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals," *PloS one*, vol. 10, p. e0132557, 2015.

[5] L. Bornmann and W. Marx, "The journal Impact Factor and alternative metrics: A variety of bibliometric measures has been developed to supplant the Impact Factor to better assess the impact of individual research papers," *EMBO reports*, vol. 17, pp. 1094-1097, 2016.

[6] D. Hicks, P. Wouters, L. Waltman, S. d. Rijcke, and I. Rafols, "Bibliometrics: the Leiden Manifesto for research metrics," 2015.

[7] Y. Suzuki, "Quality assessment of Wikipedia articles using h-index," *Journal of Information Processing*, vol. 23, pp. 22-30, 2015.

[8] K. Gurney, *An introduction to neural networks*: CRC press, 2014.

[9] I. N. Da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, "Artificial neural networks," *Cham: Springer International Publishing*, 2017.

[10] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[11] K.-L. Du and M. Swamy, "Recurrent neural networks," in *Neural Networks and Statistical Learning*, ed: Springer, 2014, pp. 337-353.

[12] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, pp. 2222-2232, 2017.

[13] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[14] H. Zen, Y. Agiomyrgiannakis, N. Egberts, F. Henderson, and P. Szczepaniak, "Fast, compact, and high quality LSTM-RNN based statistical parametric speech synthesizers for mobile devices," *arXiv preprint arXiv:1606.06061*, 2016.

[15] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017.

[16] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69-78.

[17] C. Blake, "UCI repository of machine learning databases," *http://www.ics.uci.edu/~ mlearn/MLRepository.html*, 1998.

[18] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Advances in neural information processing systems*, 2015, pp. 649-657.

[19] P. He, W. Huang, Y. Qiao, C. C. Loy, and X. Tang, "Reading scene text in deep convolutional sequences," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[20] X. Wang, W. Jiang, and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2428-2437.

