Karbala International Journal of Modern Science

Manuscript 3321

CyberVandalism Detection in Wikipedia Using Light Architecture of 1D-CNN

Azha Talal Mohammed Ali

Huda Hallawi

Noor D. Al-Shakarchy

CyberVandalism Detection in Wikipedia Using Light Architecture of 1D-CNN

Abstract

The rapid expansion of human-software-agent interaction has come with new issues. Accordingly, different engage-ments are necessary to adapt to changing human needs in dynamic socio-technical systems. Generally, cybervandalism is the act of leaving any negative impact on any piece of writing in an attempt to modify it. In Wikipedia, vandalism is any attempt to modify an article in a way that negatively affects the article's quality. Recently, several automatic detec-tion techniques and related features have been developed to address this issue. This work introduces a deep learning model with a new and light architecture to detect vandalism in Wikipedia articles. The proposed model employs a one-dimensional convolutional neural network architecture (1D CNN) that can determine the type of modification in Wikipedia articles based on two main stages: the feature extraction stage and the vandalism detection stage, preceded by the data-resampling step, which is used to address class imbalance issues in the dataset. Features are extracted from edits and their associated metadata, as well as new features (reviewers' trust), and then only the salient features are adopted to make a decision about the article; regular or vandalism can contribute to improving the accuracy of predic-tion. The experiments were conducted on a benchmark dataset, the PAN-WVC-2010 corpus, taken from a vandalism detection competition hosted at the CLEF conference. The proposed system, with the new features added, has achieved an accuracy of 100%.

Keywords

Cyber vandalism; Wikipedia; Vandalism; Light CNN; Deep Learning.

Creative Commons License

RESEARCH PAPER

CyberVandalism Detection in Wikipedia Using Light Architecture of 1D-CNN

Azha T.M. Ali*, Huda Hallawi, Noor D. Al-Shakarchy

Department of Computer Science, College of Computer Science & Information Technology, University of Kerbala, Kerbala, Iraq

Abstract

The rapid expansion of human-software-agent interaction has come with new issues. Accordingly, different engagements are necessary to adapt to changing human needs in dynamic socio-technical systems. Generally, cybervandalism is the act of leaving any negative impact on any piece of writing in an attempt to modify it. In Wikipedia, vandalism is any attempt to modify an article in a way that negatively affects the article's quality. Recently, several automatic detection techniques and related features have been developed to address this issue. This work introduces a deep learning model with a new and light architecture to detect vandalism in Wikipedia articles. The proposed model employs a one-dimensional convolutional neural network architecture (1D CNN) that can determine the type of modification in Wikipedia articles based on two main stages: the feature extraction stage and the vandalism detection stage, preceded by the data-resampling step, which is used to address class imbalance issues in the dataset. Features are extracted from edits and their associated metadata, as well as new features (reviewers' trust), and then only the salient features are adopted to make a decision about the article; regular or vandalism can contribute to improving the accuracy of prediction. The experiments were conducted on a benchmark dataset, the PAN-WVC-2010 corpus, taken from a vandalism detection competition hosted at the CLEF conference. The proposed system, with the new features added, has achieved an accuracy of 100%.

Keywords: Cybervandalism, Wikipedia, Vandalism, Light CNN, Deep learning

1. Introduction

Since the establishment of Wikipedia in 2001, more than 56 million articles have been generated by millions of contributors. People use the most popular online encyclopedia for a wide range of purposes, including research, background knowledge, learning about areas of interest for school or work, and fact-checking [1,2]. Wikipedia has a distinctive feature in which people can freely edit and publish articles. In spite of the benefits of this feature, which contribute to the vast expansion of the encyclopedia, it is the main cause of vandalism [3]. According to the Merriam-Webster dictionary, “vandalism” is defined as the intentional or malicious damage or defacement of private or public property. Within Wikipedia, vandalism occurs in the form of a deliberate edit with the intent of providing incorrect information or hiding information by deleting content, abusive language, advertisements, and/or irrelevant text. Identifying and correcting the defaced article can distract the editor from developing or extending new articles or other vital activities, so the reader will receive no or incorrect information due to the deletion [4,5]. The user community is responsible for the first attempts of identifying vandalism, which resulted in the creation of several bots. These bots analyze newly created revisions, implement hand-crafted rule sets, and identify instances of vandalism. The use of a wide variety of statistical and machine-learning techniques has contributed to the extremely complicated nature of these approaches as time has passed [6]. The community that acts as a knowledge

Received 12 March 2023; revised 21 August 2023; accepted 24 August 2023.
Available online 12 October 2023

* Corresponding author.
E-mail address: azha.t@s.uokerbala.edu.iq (A.T.M. Ali).

https://doi.org/10.33640/2405-609X.3321

base would greatly benefit from an automated solution to assist in reviewing such things as vandalism [7,8]. Deep learning is regarded as one of the most recent advances in machine learning. It assisted researchers in developing solutions that could only have been imagined a decade ago due to a lack of data management and processing capacity [9,10]. Convolutional Neural Networks (CNNs) have grown as the de facto paradigm for a wide range of Computer Vision and machine learning tasks [11,12]. When the NN's depth is deep enough, the CNN is considered universal. This implies that it may be used to approximate any continuous function to arbitrary accuracy [13]. The key advantage of CNNs seems to be that they combine feature extraction and classification operations into a single machine-learning body that can be adjusted concurrently to enhance classification performance. This eliminates the need for hand-crafted elements or other post-processing [14,15]. Recent research has shown that, with a suitable systematic approach, compact 1D CNNs can outperform all old and conventional techniques [11]. The primary benefits of the 1D CNN classifier are its low-complexity architecture and practical, cost-effective real-time hardware implementation [16,17]. The Deep Learning approach has been successful in many topics [18] close to or even similar to the topic of detecting vandalism in Wikipedia, such as Wikipedia vandal [19], Fake News detection [20], Anomaly Detection [21], one-class classification problems [22], and Cyberbullying [23]. In this paper, Wikipedia's article editing framework that deals with the most common specific crowdsourcing problems is presented. A vandalism detector model is proposed based on the extraction of multi-features using a 1D-CNN applied in a new light architecture to distinguish Wikipedia's articles' vandalism edits from regular edits. The proposed system contributed to overcoming the challenges associated with the detection of patterns and text by making a decision based on extracting multiple features. As well as from several subjects (text, metadata, and reviewer trust) so that if the features extracted from one subject fail to predict the state of modification in the article, the features extracted from others can detect the right class. Thus, the contributions of this study are summarized below:

- Vandalism Detection system for Wikipedia's article editing with major modifications, taking into account the use of Deep neural networks
- Present a light model by applying 1D CNN with a new architecture that can make the model useable with reasonable hardware capabilities.
- Adopting features extracted from a new subject, which is reviewers' information, and employing them to make a decision about the reviewers' trust.
- Balancing reviewers' trust and vandalism by giving more weight to the state of distrust in reviewers, so that trust is approved with 80% of the trusted features.

The paper is structured as follows: A literature review is presented in Section 2. Section 3 presents and describes the mechanisms used in the proposed framework. Section 4 discusses the results and the conclusions given in Section 5. Finally, some recommendations for future work are proposed.

2. Literature review

Authors in Ref. [24] suggested a structure combining techniques for ensemble and incremental learning that utilizes interaction data to understand and update the meaning of a norm violation. Ensemble learning, which is used to deal with the imbalanced dataset and incremental learning, handles the process of updating the ensemble models. Moreover, by using Wikipedia article-edits cases, the evaluation of the proposed approach is achieved. In addition, both techniques achieved acceptable results, while mini-batch learning beat online learning in detecting vandalism modifications. In the second trial, the mini-batch strategy demonstrated greater learning stability and superior performance in classifying vandalism activities, whereas online learning showed a significant decline in performance for vandalism classification due to a bias towards the majority class. Authors in Ref. [3] states that three levels of new characteristics are derived from various approaches employed to investigate the usability of leading technology such as deep learning for vandalism detection. While the first set of features was gained by developing a vocabulary of vandals using current semantic-similarity ties in word embedding and DNN, the next set of characteristics, specifically stacked denoising autoencoders (SDA), was gained by using DL techniques to minimize the number of parameters of a BOW model derived from a set of Wikipedia edits. The third set employs graph-based ranking techniques to build a list of vandalism phrases from a Wikipedia vandalism corpus. The above sets of novel features were tested independently and in combination to determine their complementarity; thus, enhance the state-of-the-art outcomes. The Authors in Ref. [25] designed unique vandalism detectors based on machine learning to save manual review time. To achieve this goal, a large-scale vandalism dataset was developed. Besides, a high-predictive, low-bias vandalism detector against specific editors was evaluated on a variety of parameters. Then, the authors tried to compare them to the state-of-art indicated by the Wikimedia Foundation's Objective Modification Evaluation Service and Wikidata Abuse Filter. This machine-learning approach instantly assigns a vandalism mark to each edit made, prompting action against vandalism in multiple procedures of operation. Edits with top marks are automatically changed back, medium-mark edits are manually reviewed based on their marks, and low-mark edits are not reviewed at all. Moreover, 47 computing features are suggested to discover vandalism, taking both content and context data into consideration. For detecting vandalism, bagging and random forests, as well as multiple-instance learning, are used. The authors in Ref. [26] investigated vandalism on Wikipedia sites and methods for preventing it. Because of the flexibility of Wikipedia, the early analysis of this study suggests that unregistered individuals are the cause of 90% of vandalism or incorrect modifications. The study also attempts to create a better environment for Wikipedians by controlling individuals' behavior and the way the community reacts to vandalism. In this study, the authors discovered that not all unregistered users are vandals, and it is possible to depend on their modifications. However, this discovery steadily changed over time, demonstrating that anonymous users with fewer edits are sometimes more reliable than registered ones. This research is considered an attempt to tackle vandalism problems. The authors in Ref. [27] looked at two Wikipedia language editions: simple English and Albanian. He found no differences in the results of these classifiers aside from the fact that they had to under-sample the non-vandalism observations to fit the dataset's number of vandalism incidents. The findings show that vandals' viewing and editing behaviors are comparable across languages. As a result of these findings, vandalism models are trained in a language and then applied to others.

3. Proposed vandalism detection model

The proposed framework used in this paper is based on an 1D CNN model for vandalism detection in Wikipedia's articles, focusing on selected features. Initially, a new CNN architecture was built in the experiments through training and testing the pan-WVC-10 dataset. Fig. 1 shows a general block diagram of the vandalism detection model.

3.1. Dataset preparing

The Wikipedia PAN Vandalism Corpus (PAN-WVC-10) was created by using annotations received from Amazon's Mechanical Turk. It contains (32.452) edits distributed on (28.468) pages, resulting in (2.391) modifications caused by vandalism. There are 753 human annotators who voted on the edits, totaling (193.022) votes, guaranteeing that each update was reviewed by at least three annotators. The results gained from edits analysis help to determine whether an alteration was “regular” or “vandalism” [28]. To improve the performance and generalization capabilities of the proposed model, the data-resampling step must be done in the dataset in order to ensure that each class has a similar number of samples. The data-resampling step is used to address class imbalance issues in the dataset, where some classes have significantly fewer samples than others. It involves manipulating the dataset's distribution by under-sampling the majority class (decreasing the number of samples). Some missing values were removed during data preparation due to the removal of some fields in which the annotators were not sure of their answers.

3.2. Feature extraction stage

As shown in Fig. 1, the proposed system relies on carefully selected features that are different from the conventional features applied in previous research. The given features are divided into two types: Textual and Meta-Data. While the textual features are taken from the text files (revisions of articles) in the dataset, the metadata features are extracted from the CSV files, which are also in the dataset. Furthermore, a new feature, which is based on annotators' information, was added to meta-data, relying on annotators' decisions about whether or not the respective edit is vandalism. Giving trust to the annotator who rates the most articles. The annotator also decides whether more than one article is vandalised, and his assessments are correct, with points of trust given for the annotator's age and gender. The details for both types of applied features are depicted in Table 1.

3.3. Vandalism detection stage

This stage is implemented using the proposed CNN model and consists of seven layers: two one-dimensional convolutional layers have a depth size of 64 and a kernel size of three. Each one merged with the activation (non-linear) layer using non-saturating rectified linear units (ReLU). In order to

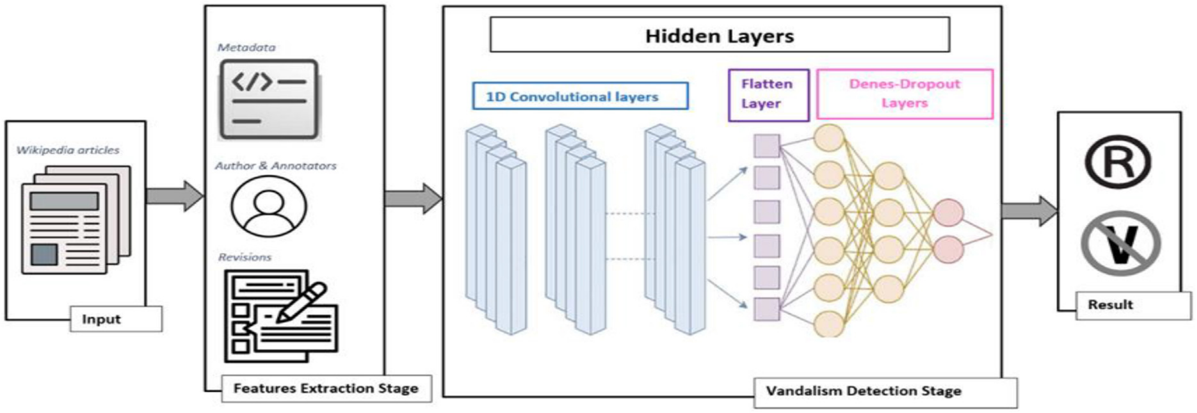


Fig. 1. A general block diagram of the Vandalism Detection Model (VDM) stages.

handle sample noise, ReLU was combined with the CNN layer to collect important features and reject weak ones. ReLU performed this by only keeping the parts of the features that have positive values and throwing away the rest as noise. The third layer is Batch normalization, which can improve training stability, convergence speed, and overall performance. Batch normalization can normalize the activations of each layer in the CNN by adjusting and scaling them, which makes the model more resilient to variations in the input data distribution and thus helps reduce the impact of noise. The 1D max-pooling (4th) layer, with pool size (2), creates an effective and efficient feature hierarchy, which is a crucial sub-step in feature extraction and generalization, leading to effective classification tasks. This layer implements downsampling and reduces the computational complexity of subsequent layers. In order to reduce overfitting and improve noise tolerance in the proposed model, a 0.5% dropout

Table 1. Features for vandalism.

Vandalism Features		
Features Type	Feature	Description
Textual	Size increment	Article, absolute size increment, i.e., |new|-|old|
	Size ratio	Calculate size ratio Calculate digit ratio Calculate upper to lower case ratio Calculate upper to all ratio Calculate non alphanumeric ratio
	Longest character sequence	Vandalism often consists of long strings of the same character, such as (aaggggghhhhhh!!!!!, sssoooo huge).
	Character distribution	Character distribution of the inserted text with respect to the expectations useful for detecting nonsense.
	Character diversity	Expression giving the length of the inserted text as a percentage of the total length of characters.
	Word categories ratio	Vulgarism, Biased and Pronoun ratio are defined and listed.
	Average term frequency	Average frequency of inserted terms relative to the old revision text
	Longest word	Longest inserted word (links are not considered)
Meta-Data	Is anonymous	The editor is anonymous or not.
	Comment character distribution	Character distribution of the comment.
	Comment character diversity	Measure of different characters compared to the length of comment.
	Comment longest character sequence	Long strings of the same character.
	Comment longest word	Longest inserted word in the comment.
	Comment word categories	Vulgarism, Biased and Pronoun in the comment.
	Comment length	Append a length of the comment.
	Annotator Trust	Depend on annotator information to build trust feature.

layer was added. The final two layers are dense layers that have 100 and 1 neurons and implement ReLU and sigmoid activation functions, respectively. High-level representations are learned during these two layers based on the extracted features, and final predictions are made by leveraging the fully connected nature to capture global information and relationships, leading to effective classification. Table 2 presents a summary representation of the proposed model architecture.

4. Results

Accuracy and loss functions are two key metrics used to evaluate performance in both training and evaluation modes. For each epoch, the outcomes of evaluating the training and validation datasets are displayed in Fig. 3, presenting a visual representation of the given model's learning behavior on a certain dataset. Several experiments were conducted on the proposed system, during which different features were used in order to achieve the best accuracy. The two most prominent experiences that have passed through the system are reviewed to compare the use of traditional features with the new proposed features. In addition, Table 3 presents a comparison between the results of previous research and the proposed model.

4.1. The first experiment

In this experiment, the traditional features presented in Table 1 were used. These features are among the most important features related to vandalism that were used in previous research.

These features showed results of an estimated accuracy of 92% when used with the architecture of the proposed system. Fig. 2 shows the accuracy and loss function of the proposed model for the first experiments. While (a) displays the model accuracy, (b) displays the model loss for the proposed system.

Table 2. The proposed model architecture.

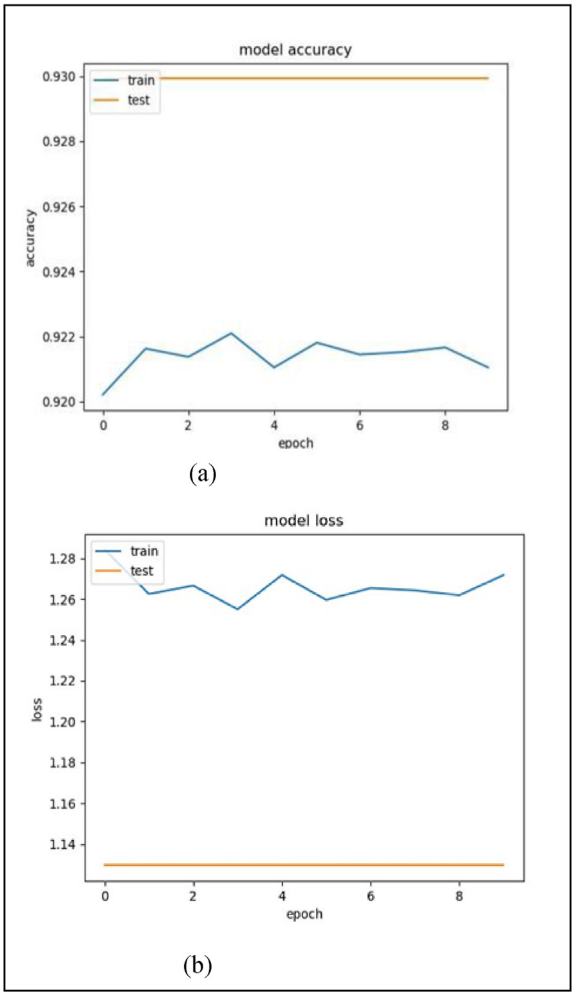
Layer (type)	Output Shape	Parameters
reshape_1 (Reshape)	(None, 19, 1)	0
conv1d_1 (Conv1D)	(None, 17, 64)	256
conv1d_2 (Conv1D)	(None, 15, 64)	12352
batch_normalization_1	(None, 15, 64)	256
max_pooling1d_1 (MaxPooling1)	(None, 7, 64)	0
dropout_1 (Dropout)	(None, 7, 64)	0
flatten_1 (Flatten)	(None, 448)	0
dense_1 (Dense)	(None, 100)	44900
dense_2 (Dense)	(None, 1)	101
Total no. of parameters: 57,865		
Trainable no. of parameters: 57,737		
Non-trainable no. of parameters: 128		

Table 3. Vandalism Detection System (VDS) Accuracies for different scenarios of the PAN-WVC-10 Dataset.

No.	Methods	Classifier	Recall	F1	Precision	AUC
1	Ref. [29]	Logit Boost	47%	58%	73%	93%
2	Ref. [30]	Random Forest	60%	-	60%	-
3	Ref. [3]	Random Forest	96%	79%	85%	-
4	Ref. [31]	Logistic Model Trees + K-Means	63.8%	-	78.1%	-
5	VDS	1D-CNN	100%	100%	100%	100%

4.2. Model behavior with additional new features

In this experiment, the new proposed features were used In addition to the traditional features. With the use of the new features called (trust features) and the traditional (standard features, making the architecture of the proposed system is directed towards a correct classification of the results with a high accuracy of up to 100. Fig. 3 below shows the



(a)

(b)

Fig. 2. The proposed model's accuracy and loss function - first experiment (a) accuracy and (b) loss.

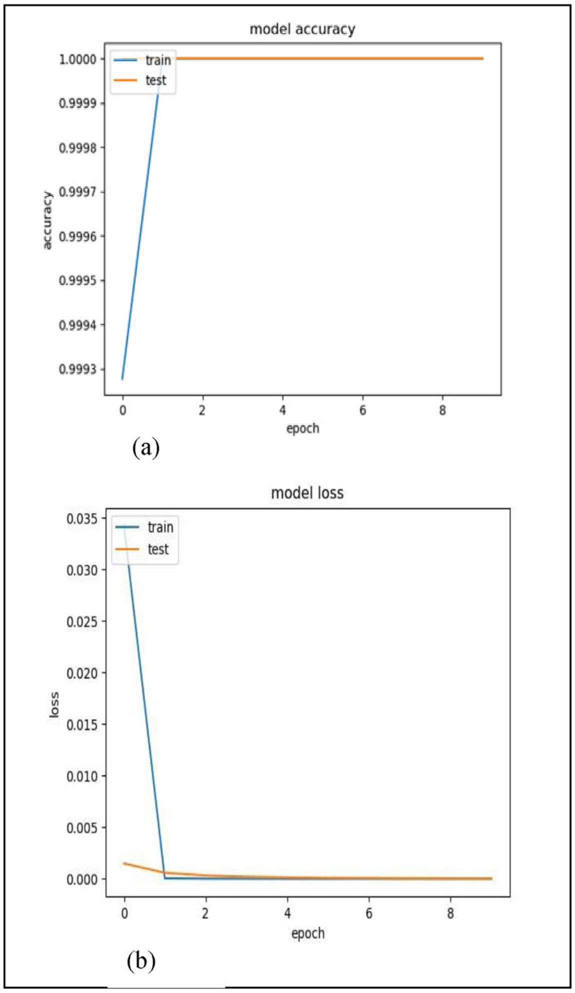


Fig. 3. The proposed model's accuracy and loss function - second experiment (a) accuracy and (b) loss.

accuracy and loss function of the proposed model for the second experiment. While (a) displays the model accuracy, (b) displays the model loss for the proposed system. Adding new features (new variables) helps clarify invisible relationships between data. In addition, these features are selected after processing the missing and uncertain data. All these factors help improve the model's accuracy.

4.3. Evaluation model

Because of their stochastic nature, deep learning models produce slightly different predictions. This results in slightly different overall abilities each time the same model is applied to the same data. Since the same trained model can be applied to different data, thus producing various evaluation results, the k-fold cross-validation approach is used to estimate the model skill while controlling the model variance. The second method involves the estimation of the skill of a stochastic model (model stability control) by repeatedly evaluating a non-stochastic model and then averaging the resulting estimates of the model's performance. This accounts for the fact that different models can produce widely different results when applied to the same data. Table 4 displays the model's precision, recall, and F1-measure results.

Table 4. The proposed model's evaluation metrics.

Vandalism	Precision	Recall	F1-measure	Support
Accuracy	1.00	1.00	1.00	64

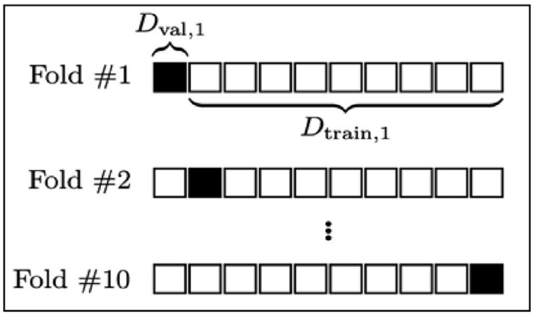


Fig. 4. Using a 10-Fold Cross-Validation procedure [32].

4.4. K-folds cross-validation

Cross-validation is a technique used to determine a machine-learning model's performance on new data by comparing the performance of multiple machine-learning models on a subset of that data using a single parameter called k. In other words, the small training set is used to make an educated guess to show how well the model performs when predicting new, unseen data. Fig. 4 shows the 10-fold cross-validation procedure for training and testing. Given a value of k = 10, the proposed models will divide the dataset into three parts: Training (80%), Validation (10%), and Testing (10%), checking testing performance based on the metric of choice (adopted accuracy). Finally, an average of all the results is computed to demonstrate how things

Table 5. The proposed model's cross-validation results.

No. Of Fold	Accuracy of each fold	Model accuracy
#1	100.00%	**100.000% (+/-0.000)**
#2	100.00%	
#3	100.00%	
#4	100.00%	
#5	100.00%	
#6	100.00%	
#7	100.00%	
#8	100.00%	
#9	100.00%	
#10	100.00%	

turned out. The results of the test can be summed up by looking at Table 5 below and the following 10-fold implementation.

5. Conclusions and future works

Some social disorder problems, such as disturbing the peace and trespassing, are often associated with vandalism problems. At the same time, reducing and limiting the damage resulting from the occurrence of vandalism can be achieved through early detection and rapid resolution of solutions that can end vandalism. In terms of accuracy and loss functions, the detection system that employs deep neural networks is superior to other traditional approaches. The main conclusion of this research is that we achieve successful detection of the Wikipedia article status to indicate whether it is vandalism edits or regular edits with our proposed model. In particular, the light 1D CNN model provides low complexity, which leads to the possibility of applying the model with limited computational capability and requirements as well as a reasonable run time.

Additionally, the implementation of the ReLU activation function in the nonlinear layer integrated with the convolution layer provides the ability to process the noise associated with inputs. This integrated layer extracts salient features and neglects the weak ones, including noise, by eliminating all weak and noisy entries from the series and leaving just those with a positive value. The suggested vandalism detection approach allows us to expand Wikipedia's editing and viewing features. Finally, one contribution of this paper is extracting new (reviewers' trust) features that are used together with the conventional vandalism features to increase the accuracy and reach of an efficient model. In the future, the proposed system can be improved to determine vandalism types such as add, delete, and change in addition to its detection. It can also be improved to determine the location of vandalism in articles. Other datasets that deal with the problem of vandalism can also be applied to extract more features and evaluate the model. This process can also be achieved by combining more than one dataset.

Conflicts of interest

The authors declares that they have no competing interests.

Acknowledgements

This work is supported by the University of Kerbala.

References

[1] M. Steinkasserer, T. Ruprechter, D. Helic, Investigating western bias in wikipedia articles about terrorist incidents, in: 17th Int. Symp. Open Collab., 2021, pp. 1–5.

[2] G.A. Al-Sultany, H.J. Aleqabie, Enriching tweets for topic modeling via linking to the wikipedia, Int. J. Eng. Technol. 7 (2018) 144–150.

[3] J.R. Martinez-Rico, J. Martinez-Romo, L. Araujo, Can deep learning techniques improve classification performance of vandalism detection in wikipedia? Eng. Appl. Artif. Intell. 78 (2019) 248–259.

[4] M. Shulhan, D.H. Widyantoro, Detecting vandalism on English wikipedia using LNSMOTE resampling and cascaded random forest classifier, Int. Conf. Adv. Info. Concept (2016) 1–6.

[5] M. Lehto, P. Neittaanmaki, Cyber Security: Analytics, Technology and Automation, first ed., Springer Cham, Switzerland. 2015.

[6] B.T. Adler, U.P. De Valencia, P. Rosso, A.G. West, Wikipedia vandalism detection : combining natural language , metadata , and reputation features, in: 12th International Conference on Intelligent Text Processing and Computational Linguistics 6609, 2011, pp. 277–288.

[7] S. Heindorf, M. Potthast, B. Stein, G. Engels, Towards vandalism detection in knowledge bases: corpus construction and analysis, in: 38th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (SIGIR'15), 2015, pp. 831–834.

[8] K.-N.D. Tran, Detecting Vandalism on Wikipedia across Multiple Languages, Australian National University. 2015.

[9] E.U.H. Qazi, A. Almorjan, T. Zia, A one-dimensional convolutional neural network (1D-CNN) based deep learning system for network intrusion detection, Appl. Sci. 12 (2022) 1–14.

[10] S. Dong, P. Wang, K. Abbas, A survey on deep learning and its applications, Comput. Sci. Rev. 40 (2021) 1–22.

[11] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: a survey, Mech. Syst. Signal Proc. 151 (2021) 1–21.

[12] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, H. Ghayvat, Cnn variants for computer vision: history, architecture, application, challenges and future scope, Electron 10 (2021) 1–28.

[13] D.X. Zhou, Universality of deep convolutional neural networks, Appl. Comput. Harmon. Anal. 48 (2020) 787–794.

[14] S. Kiranyaz, M. Zabihi, A.B. Rad, T. Ince, R. Hamila, M. Gabbouj, Real-time phonocardiogram anomaly detection by adaptive 1D convolutional neural networks, Neurocomputing 411 (2020) 291–301.

[15] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8 (2021) 1–74.

[16] T. Ince, Real-time broken rotor bar fault detection and classification by shallow 1D convolutional neural networks, Electr. Eng. 101 (2019) 599–608.

[17] S.M. Shahid, S. Ko, S. Kwon, Performance comparison of 1D and 2D convolutional neural networks for real-time classification of time series sensor data, Int. Conf. Inf. Netw. Jeju-si, Korea (2022) 507–511.

[18] N.C. Thompson, K. Greenewald, K. Lee, G.F. Manso, The computational limits of deep learning, MIT Initiat. Digit. Econ. Res. Br. 4 (2020) 1–33.

[19] S. Yuan, P. Zheng, X. Wu, Y. Xiang, Wikipedia vandal early detection: from user behavior to user embedding, in: M. Ceci, J. Hollmén, L. Todorovski, C. Vens, S. Džeroski, eds., Machine Learning and Knowledge Discovery in Database, Springer Cham. 2017, pp. 832–846.

[20] W.H. Bangyal, R. Qasim, N.U. Rehman, Z. Ahmad, H. Dar, L. Rukhsar, Z. Aman, J. Ahmad, Detection of fake news text classification on COVID-19 using deep learning approaches, Hindawi Comput. Math. Meth. Med. 2021 (2021) 1–14.

[21] G. Pang, C. Shen, L. Cao, A. Van Den Hengel, Deep learning for anomaly detection: a review, ACM Comput. Surv. 54 (2020) 1−36.

[22] P. Zheng, S. Yuan, X. Wu, J. Li, A. Lu, One-class adversarial nets for fraud detection, thirty-third AAAI conference on artificial intelligence and thirty-first innovative applications of artificial intelligence conference and ninth AAAI symposium on educational advances in artificial intelligence (AAAI'19/IAAI'19/EAAI'19), 2019, pp. 1286−1293.

[23] S. Neelakandan, M. Sridevi, S. Chandrasekaran, K. Murugeswari, A. Kumar, S. Pundir, R. Sridevi, T.B. Lingaiah, Deep learning approaches for cyberbullying detection and classification on social media, Hindawi Comput. Intell. Neurosci. 2022 (2022) 1−13.

[24] T.F. dos Santos, N. Osman, M. Schorlemmer, Ensemble and incremental learning for norm violation detection, 21st Int. Conf. Auton. Agents Multiagent. Syst. (AAMAS '22) (2022) 427−435.

[25] S. Heindorf, Y. Scholten, G. Engels, M. Potthast, Debiasing Vandalism Detection Models at Wikidata, World Wide Web Conf. (WWW '19), 2019, pp. 670−680.

[26] A. Alkharashi, J. Jose, Vandalism on collaborative web communities: an exploration of editorial behaviour in wikipedia, in: 5th Spanish Conf. Inf. Retr. (CERI'18), 2018, pp. 1−4.

[27] A. Susuri, M. Hamiti, A. Dika, Detection of vandalism in wikipedia using metadata features − implementation in simple English and Albanian sections, Adv. Sci. Technol. Eng. Syst. 2 (2017) 1−7.

[28] M. Potthast, Crowdsourcing a wikipedia vandalism corpus, in: 33rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (SIGIR'10), 2010, pp. 789−790.

[29] M. Harpalani, M. Hart, S. Singh, R. Johnson, Y. Choi, Language of vandalism: improving wikipedia vandalism detection via stylometric analysis, 49th Annual Meeting ofthe Association for Computational Linguistics:shortpapers 2, 2011, pp. 83−88.

[30] P.C. Götze, Advanced Vandalism Detection on Wikipedia, Bauhaus-Universität Weimar. 2014.

[31] T. Freitas dos Santos, N. Osman, M. Schorlemmer, Learning for detecting norm violation in online communities, in: A. Theodorou, J.C. Nieves, M. De Vos, eds., Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIV, Springer, Cham. 2022, pp. 127−142.

[32] D. Berrar, Cross-validation, in: Encycl. Bioinforma Comput. Biol. ABC Bioinfo. 1, 2018, pp. 542−545.