Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/21/20231114

MLOffense: Multilingual offensive language detection and target identification on social media using graph attention transformer

Grant Wang

University at Buffalo, the State University of New York, Buffalo, NY 14260

grantwan@buffalo.edu

Abstract. With the increasing use of social media in our daily lives, it is crucial to maintain safe and inclusive platforms for users of diverse backgrounds. Offensive content can inflict emotional distress, perpetuate discrimination towards targeted individuals and groups, and foster a toxic online environment. While natural language processing (NLP) has been employed for automatic offensive language detection, most studies focus on English only, leaving languages other than English understudied due to limited training data. This project fills this gap by developing a novel multilingual model for offensive language detection in 100 languages, leveraging existing English resources. The model employs graph attention mechanisms in transformers, improving its capacity to extend from English to other languages. Moreover, this work breaks new ground as the first study ever to identify the specific individuals or groups targeted by offensive posts. Statistical analysis using F1 scores shows high accuracy in offensive language classification and target recognition across multiple languages. This innovative model is expected to enable multilingual offensive language detection and prevention in social media settings. It represents a significant step forward in the field of offensive language detection, paving the way for a safer and more inclusive social media experience for users worldwide.

Keywords: offensive language detection, multilingual, graph attention, target identification.

1. Introduction

Social media has become an important part of everyday life. However, offensive language on social media has become a serious issue. First, offensive language can have serious consequences for the individuals who are the target. It can lead to feelings of fear, anxiety, and isolation and can even contribute to mental health problems such as depression and anxiety [1][2]. In addition, offensive language can also be used to harass, bully, or discriminate against certain groups of people, which can have serious consequences for those individuals and for society as a whole. Finally, offensive language on social media can also have negative impacts on the overall online environment. It can create a toxic atmosphere that can discourage participation and engagement, and can even contribute to the spread of discrimination and biases.

With the increasing use of social media in our daily lives, it is important to ensure that these platforms are safe and welcoming for all users regardless of their background. Being able to automatically identify offensive language and take appropriate action (e.g., remove and/or warn the author) on social media,

DOI: 10.54254/2755-2721/21/20231114

we can help to protect individuals from harm, provide a safer online environment, and promote a more positive and inclusive society.

Various methodologies have been employed in the research on multilingual offensive language identification. Early works relied on manually crafted rules and regular expressions for offensive language identification [3], such as extracting lexical and syntactic features. Traditional machine learning algorithms, such as support vector machines (SVM) [4], Naïve Bayes [5], and Decision Trees [6], were used with features like n-grams, bag-of-words, and sentiment scores. Recently, the rise of deep learning has led to the adoption of Convolutional Neural Networks (CNN) [7], Recurrent Neural Networks (RNN) [8][9], and Long Short-Term Memory (LSTM) [10] for offensive language classifications. More recently, pretrained language models, such as BERT [11], GPT [12], and XLM-RoBERTa [13], have been fine-tuned for offensive language identification tasks, leading to significant improvements in performance [14][15]. To train these models, a variety of datasets have been developed for multilingual offensive language identification. A comprehensive summary of such datasets can be found in [16]. These datasets include hate speech, interpersonal harassment, toxic language, trolling, aggression, bullying, and profanity. Some of the most prominent datasets include OLID [17][18], HatEval [19], HASOC [20], and Large-scale Hate Speech [21].

Among these existing datasets, English has abundant resources with large-size training datasets and various types of labels. As a result, offensive language detection has been extensively studied in English, but rarely studied in other languages due to the lack of training data. Recently, with zero-shot cross-lingual transfer learning, more studies have investigated offensive detection in languages other than English. For example, DeepOffense fine-tuned XLM-R, a state-of-the-art large model for offensive language detection in Bengali, Hindi, and Spanish [15]. Cross-lingual transferability of mBERT and XLM-R was also studied between English and Turkish [21]. Despite the research effort, multilingual offensive language identification is still a challenging topic. Models need to better understand the context in which offensive language is used, as it can significantly impact the interpretation of the content.

This study aims to improve offensive language detection in other languages by developing and training a new multilingual model. The new model introduces *graph attention* on top of XLM-R, a pre-trained model that represents words or phrases in a numerical form using the context in which they appear. Graph attention is not only able to capture long-range dependencies for languages like Arabic, but it is also independent of the language word order because it uses syntactic distance instead of linear distance to model the relationships between words [22][23]. For example, the regular order of a verbal sentence in Arabic is the verb, the subject, the objects, and then the adverbs and the prepositional phrases, which is typologically different from English. The word-order-agnostic features of graph attention can improve cross-lingual transferability over XLM-R when offensive language detection is fine-tuned for English only but then applied directly to other languages such as Arabic. In addition, graph attention enables the exchange of information between different heads within the multi-head attention structure. This information propagation is based on syntactic distances and facilitates the learning of correlations between various mention types and target labels. This is particularly important for offensive language identification because allows our new model to classify offensiveness based more on its long-range contextual information rather than just offensive words.

The new model is first trained on multilingual datasets of general purpose so that the trained model can "understand" the grammar, vocabulary, and syntax of each language. Taking advantage of existing English resources, the model is then fine-tuned using English datasets labeled with offensive/non-offensive to gain an understanding of what "offensive" means. During this fine-tuning process, zero-shot transfer learning is being used to extend the capability of offensive language detection to other languages and project predictions in other languages without the need for labeled offensive language data in these low-resource languages. Besides offensive detection, the model is also trained to identify if the offensive post is targeted or not and if the target is an individual, group, or others. Most importantly, the new model is able to identify the names of the targets. To the best of our knowledge, this is the first time name recognition is studied in offensive language identification. Such a capability is significant

DOI: 10.54254/2755-2721/21/20231114

because offensive posts targeting specific individuals or groups are the most damaging, so immediate steps can be taken to ensure their safety and well-being. In addition, this information may help understand the extent and nature of discrimination against the targets and is essential for implementing strategies to combat and prevent such harmful posts in the future.

The main contributions of this paper can be summarized as follows:

- A novel graph attention transformer for offensive language detection is proposed such that more contextual information can be extracted and cross-lingual transferability is improved.
- The model is able to perform offensive language identification in 100 languages, independent of the word order in different languages, and capable of capturing long-range dependencies for languages with long sentences.
- This is the first study on name recognition of the individual or group that is being targeted in the offensive language.

2. Materials and methods

2.1. Model

The proposed new model is composed of the following components.

First, the input sentence is converted into a language-universal parse tree using UDPipe6 [24]. Specifically, the process breaks the sentence into its constituent words and punctuation marks, annotates each word in the sentence with its part of speech, such as noun, verb, adjective, etc., and extracts in a tree structure the grammatical representation of a sentence that defines the relationships between the words (see Fig. 1 for an example of a bilingual sentence). To make the parse tree language-universal, the process needs to achieve consistent annotation of grammar across different human languages. The open-source, customizable UDPipe6 is used in this work for multilingual tree parsing.

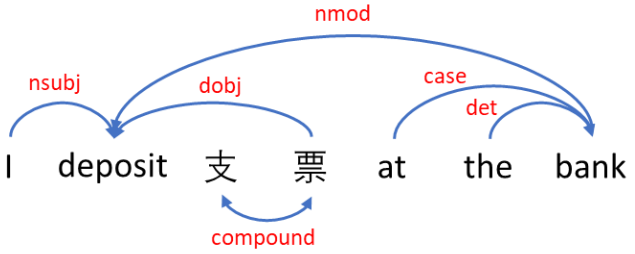

Figure 1. Tree structure of grammatical representation of a sentence.[1]

Second, the words in the sentence are embedded into a shared semantic space across different languages using the proposed architecture, named graph attention XLM-R (GA-XLM-R). The new architecture is built upon Cross-Lingual Language Representation Model – RoBERTa (XLM-R), a state-of-the art multilingual language representation model that was pre-trained on 2.5 TB of text data in about 104 languages extracted from CommonCrawl [13]. The XLM-R contains approximately 125M parameters with 12 layers, 768 hidden states, 3072 feed-forward hidden states, and 8 heads, and it is based on a transformer with self-attention.

A transformer is a type of deep neural network with several layers. Its role is to learn how to represent each word in a sentence using a high-dimensional numerical vector such that words that are more related contextually are closer in the vector space. For example, the word “bank” is close to the word “支票” (Chinese word for check) in the context of the sentence in Fig. 1, but becomes closer to the word “river” for the sentence “I arrived at the bank of the river” in the numerical vector space, as illustrated in Fig. 2.

[1] This paper is a study on multilingualism, and other languages must appear. Chinese words are used here to provide an example of multilingual sentences. “支票” is Chinese for “check”.

DOI: 10.54254/2755-2721/21/20231114

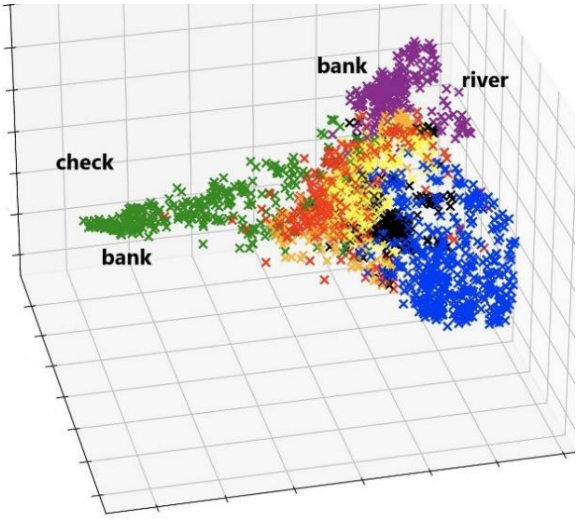

Figure 2. Illustration of vector representations of words.

In a transformer, after tokenization and vectorization of the words in the input sentence, we obtain an input matrix **X** of size $n \times d_{\text{model}}$, where n is the number of tokens and d_{model} is the number of dimensions of the vectors. The attention mechanism, similar to our human cognitive attention, learns the contextual correlation by using Query (**Q**), Key (**K**), and Value (**V**) matrices. These Query, Key, and Value matrices are created by multiplying the input matrix **X**, by weight matrices W_Q ($d_{\text{model}} \times d_k$), W_K, W_v. The weight matrices are randomly initialized and then learned during training. The self-attention A is calculated as [25]

$$A = \text{softmax}(\frac{\mathbf{QK}^T}{\sqrt{d_k}})\mathbf{V} \tag{1}$$

where **Q**, **K**, **V**, d_k are matrices of queries, keys, values, and the number of dimensions of each key. The softmax function is defined as

$$\text{softmax}(P)_{ij} = \frac{e^{P_{ij}}}{\sum_i e^{P_{ij}}} \tag{2}$$

In the case of multiple transformer layers, each layer generates different levels of latent representations recursively. Typically, the latent representations generated by the last layer are used as the contextual representations of the input words (i.e., replacing **X** with the latent representation from the last layer). Multiple self-attention heads are employed in each transformer layer.

The key innovation of the new GA-XLM-R over the existing XLM-R is the use of graph attention. The main concept involves manipulating the mask matrix to establish the desired graph structure and adjusting the attention weights using a so-called syntactic distances. Specifically, in the new architecture, we extend self-attention to graph attention A_g [22]:

$$A_g = F\left(\text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}} + \mathbf{M}\right)\right)\mathbf{V} \tag{3}$$

Here, **M** is a mask matrix to incorporate syntactic structure and distance information:

$$M_{ij} = \begin{cases} 0, & D_{ij} \le \rho \\ -\infty, & \text{otherwise} \end{cases} \tag{4}$$

with D_{ij} being the syntactic distance between two tokens and ρ being a threshold. The function F is used to modify the attention weights such that more attention is paid to words that have a shorter distance in the parse tree:

DOI: 10.54254/2755-2721/21/20231114

$$F(H)_{ij} = \frac{H_{ij}}{Z_i D_{ij}} \tag{5}$$

where $Z_i = \sum_j \frac{H_{ij}}{D_{ij}}$ is used for normalization. The universal dependency parse tree in Fig. 1 is used to compute the syntactic distances between every pair of words in a sentence. An example is illustrated in Fig. 3. The benefits of using graph attention over self-attention include being able to capture long-range dependencies and being independent of word order since it uses syntactic distance instead of linear distance to model the relationships between words.

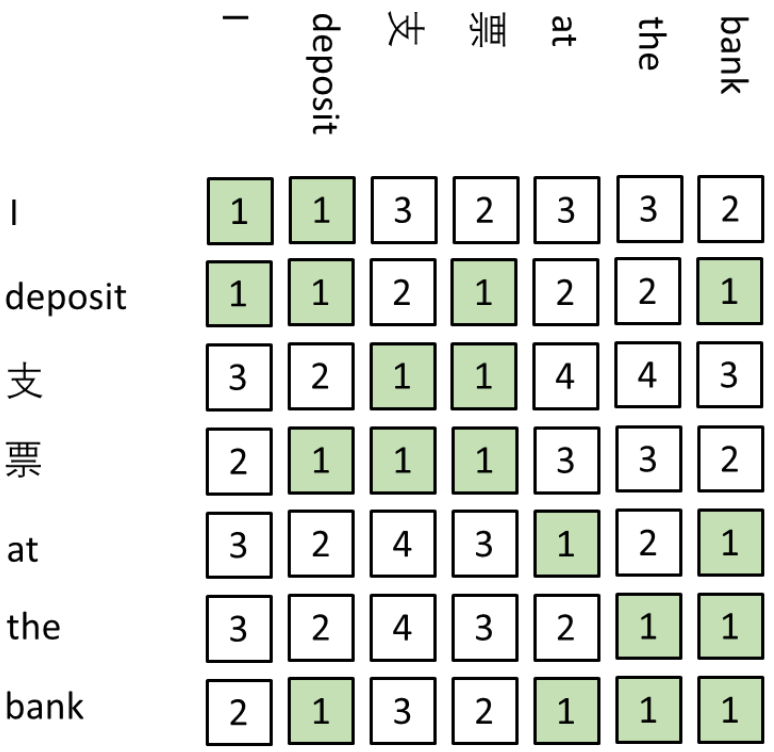

Figure 3. Syntactic distance matrix showing the shortest path distances between all pairs.

Finally, a classifier is used to perform the following tasks:

A. Classify if the sentence is offensive or not;
B. If offensive, classify if the offensive language is targeted or not;
C. If targeted, classify if the target is an individual, group, or something else;
D. In the case of individual or group, identify the name of the target (NER).

For each of the classification tasks, a simple softmax classifier is added to the top of the network to predict the probability of label *c* (e.g., in task A, *c* = 1 means offensive, and *c* = 0 means non-offensive):

$$p(c \mid \mathbf{h}) = \text{softmax}(\mathbf{Wh}) \tag{6}$$

where **h** is the final latent state of the network, **W** is the task-specific (tasks A, B, C, and D) parameter matrix. Task D performs named entity recognition (NER), which involves identifying and labeling named entities, such as person names, organizations, and geographical locations, in a sentence.

The complete representation of the above-mentioned three steps, preprocessing, graph attention transformer, and classifier is illustrated in Fig. 4. The preprocessing designates each word's language, position, part of the speech, and grammatical role in the sentence. The graph attention transformer completes cross-lingual embedding, and the softmax unit calculates the classification probabilities and performs the classification tasks.

DOI: 10.54254/2755-2721/21/20231114

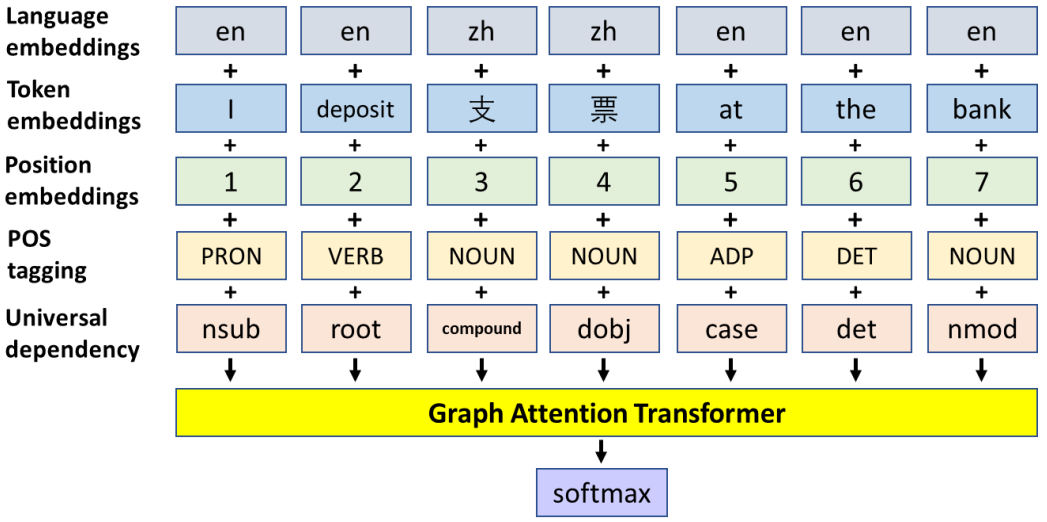

Figure 4. Diagram of the main structure of the proposed model.

Fig. 5 illustrates the details of the tasks to be accomplished by the trained model. The classification model takes a hierarchical structure. When a multilingual sentence is inputted into the new model GA-XLM-R, the network is able to perform the aforementioned four tasks, and the trained model cannot only detect offensive, targeted language, but also identify who the target of the offensive language is.

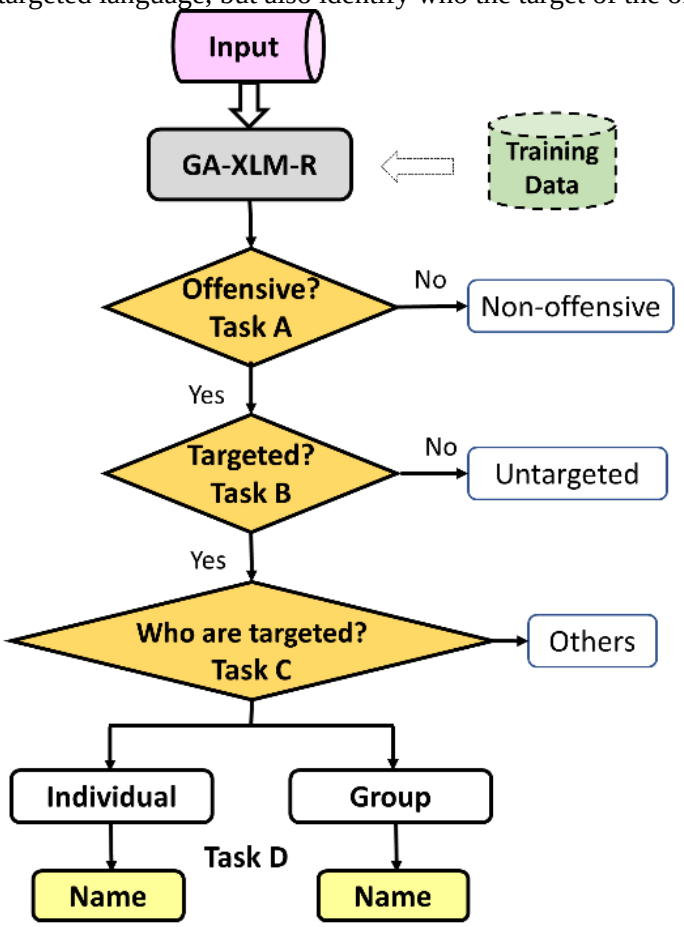

Figure 5. Flow chart of offensive language detection and target identification.

DOI: 10.54254/2755-2721/21/20231114

2.2. Training

Training (to determine the weights in GA-XLM-R) was performed to implement each of the four tasks in a multilingual setting. Supervised learning (i.e., the label for each of the four tasks is given in the training data) was used here.

The model was first pre-trained to understand or "speak" different languages. This pre-training was performed using the multilingual pre-processed CommonCrawl dataset that contains 100 languages (CC-100) [26]. The datasets include 2.5 TB of training data with general content only.

Then in the fine tuning step, the Offensive Language Identification Dataset (OLID) [17][18], one of the most popular labeled offensive language datasets with 14,100 tweets, was used as the training dataset for tasks A, B, and C of offensive language identification. Hate Speech and Offensive Content (HASOC) dataset [20] was also used to train task A only because labels for other tasks are not available. Because the model has learned to understand the contextual information in the first pre-training step, this fine-tuning step trains the network to detect offensive content in various contexts even without offensive words. As a major distinction from existing studies, the TweetNER7 [27] dataset with 7,110 tweets was used as the training data for named entity recognition (NER) in task D. The datasets used in the fine-tuning were all in English only. There are no other languages that have such abundant training data available.

We then applied the so-called zero-shot cross-lingual transfer learning so that the above fine-tuned model has the offensive language identification and NER capabilities in 100 languages. This process of training on one language and applying the knowledge to other languages is illustrated in Fig. 6. The zero-shot cross-lingual transfer learning allows for offensive language detection in low-resource languages without the need for a labeled offensive language dataset in those low-resource languages, which rarely even exist. The multilingual capability is achieved through the aforementioned pre-training step with general content datasets in 100 languages. After the above training, the new multilingual model is called MLOffense.

The dataset was divided into a training set and a validation set using a 0.8:0.2 split on the dataset. The learning rate and the number of epochs were tuned manually to obtain the best results for the validation set.

The model was trained on Google Colab. Specifically, the training is performed on an NVidia T4 GPU with 16GB memory, 8.1 TFLOPS, and dual CPU cores (12GB RAM).

3. Results

The trained model was tested in all 100 languages. When public datasets are not available in a language, Python googletrans library was used to translate an English dataset to that language. Among available datasets, four languages, English, Arabic, Chinese, and Marathi [16] were chosen to statistically evaluate the performance of MLOffense and identify any errors or biases. Arabic and Chinese were selected because they are very different from English (the language the model is trained on) in terms of writing systems and even reading directions for Arabic. Marathi was chosen because it is a low-resource language with very limited training data available. Public datasets [16] in Spanish, Italian, German, Hindi, and code-mixed languages (Hindi-English and German-English mixing) were also used to evaluate task A.

For each dataset/language, the weighted F1 score, defined as follows, was used for evaluation.

$$\text{F1 score} = \frac{\text{TP}}{\text{TP}+\frac{1}{2}(\text{FP}+\text{FN})} \tag{7}$$

where FN is the number of false negatives, FP false positives, TN true negatives, TP true positives, and the weighted scores are the average of the scores for all labels with each label weighted by its support. For comparison, the same testing was performed for a state-of-the-art method DeepOffense [15], which has the same settings as MLOffense except that the conventional self attention was used. Tables 1-4 summarize the macro and weighted F1 scores for all four tasks.

DOI: 10.54254/2755-2721/21/20231114

Table 1. Statistical results of Task A.

Task A: Offensive or Not					
		Macro F1	**Weighted F1**	**Offensive F1**	**Non-Offensive F1**
English	MLOffense	0.8461	0.8531	0.8252	0.8670
	DeepOffense	0.8126	0.8228	0.7818	0.8433
Arabic	MLOffense	0.7848	0.7862	0.7793	0.7904
	DeepOffense	0.7148	0.7198	0.6941	0.7356
Chinese	MLOffense	0.8361	0.8388	0.8192	0.8530
	DeepOffense	0.7743	0.7794	0.7426	0.8060
Marathi	MLOffense	0.6924	0.6936	0.6614	0.7233
	DeepOffense	0.6087	0.6088	0.607	0.6104
Spanish	MLOffense	0.8348	0.8365	0.8228	0.8468
	DeepOffense	0.7788	0.7801	0.7692	0.7883
Italian	MLOffense	0.8198	0.8211	0.8123	0.8272
	DeepOffense	0.7616	0.7619	0.7601	0.7632
German	MLOffense	0.8333	0.8344	0.8291	0.8375
	DeepOffense	0.7650	0.7656	0.7629	0.7672
Hindi	MLOffense	0.7937	0.7991	0.769	0.8183
	DeepOffense	0.7051	0.7084	0.6902	0.7200
Code-mixing	MLOffense	0.8081	0.8081	0.8082	0.8080
	DeepOffense	0.7123	0.7167	0.6755	0.7491

Table 2. Weighted F1 scores for Task B.

Task B: Targeted or Not				
	English	**Arabic**	**Chinese**	**Marathi**
MLOffense	0.7963	0.7127	0.7828	0.6394
DeepOffense	0.7609	0.6583	0.7359	0.5780

Table 3. Weighted F1 scores for Task C.

Task C: Individual, Group, or Other				
	English	**Arabic**	**Chinese**	**Marathi**
MLOffense	0.7468	0.7091	0.7262	0.6117
DeepOffense	0.7289	0.6522	0.6950	0.5568

Table 4. Weighted F1 scores for Task D.

Task D: Name Recognition				
		English	**Arabic**	**Chinese**
ML-Offense	Person	0.8118	0.7426	0.7696
	Group	0.7312	0.6621	0.6807

DOI: 10.54254/2755-2721/21/20231114

The tables show that the model trained in English is able to provide accurate predictions in other languages, achieving high scores in quantitative metrics.

4. Discussion and conclusion

This work is the first study to apply a graph attention transformer in multilingual offensive language identification. We have shown that the new model is able to achieve accurate results statistically in classifying a sentence as offensive or not, targeted or not, and whether the target is an individual, a group, or something else in different languages. More importantly, for the first time, we demonstrated the capability to identify the name of the individual or group that is targeted in the offensive sentence. This is important to address the challenge that people from varied backgrounds perceive offensiveness differently. Although formal statistical testing is performed on only 9 languages due to the limited testing datasets, the model works for all languages in the CC-100 dataset that were used in pre-training. The testing results also show the MLOffense model is superior to the state-of-the-art model for offensive detection.

The model would be useful to detect and eliminate offensive language in social media and to provide a safer multilingual online environment worldwide. Most social media platforms offer APIs that allow third-party services to interact with their systems. The model can be integrated with these APIs to scan and analyze text from posts, comments, messages, etc. so that appropriate actions are taken such as hiding or deleting the post, sending a warning to the user, or, in extreme cases, suspending the offending user's account based on the terms of use. Web browser plugins can be an alternative application of the model for users who are sensitive to mildly offensive languages. Given the diverse and mixed languages used in social media, the multilingual model will significantly reduce the workload of human moderators and prevent damage to users in different languages. In addition, the model can be used to extract data for behavioral and social science research, such as analyzing the prevalence and causes of racial hate and identification and support for potential victims. Because offensive language can have a wide spectrum ranging from mildly inappropriate, hate speech and harassment, future work will investigate how to classify a spectrum of offensive language by leveraging the available datasets.

References

[1] Bonanno, R.A. and Hymel, S. (2013). Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying. *Journal of Youth and Adolescence*, 42(5):685–697.

[2] Bannink, R., Broeren, S., van de Looij-Jansen, P.M., de Waart, F.G., and Raat, H. (2014). Cyber and traditional bullying victimization as a risk factor for mental health problems and suicidal ideation in adolescents. *PLOS ONE*, 9(4), e94026.

[3] Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. *Proceedings of the ASE/IEEE International Conference on Social Computing*, p. 71-80

[4] Nandhini, B. S. and Sheeba, J. I. (2015). Cyberbullying detection and classification using information retrieval algorithm. Proceedings of the 2015 international conference on advanced research in computer science engineering & technology, p. 1-5.

[5] Dadvar, M., Jong, F. D., Ordelman, R., and Trieschnigg, D. (2012) Improved cyberbullying detection using gender information. Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop. University of Ghent.

[6] Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P.C., Carvalho, J.P., Oliveira, S., Coheur, L., Paulino, P., Veiga Simão, A.M., and Trancoso, I. (2019). Automatic cyberbullying detection: A systematic review. Computers in Human Behavior, 93:333–345.

[7] Malmasi, S. and Zampieri, M. 2017. Detecting hate speech in social media offensive language on twitter: Analysis and experiments. Proceedings of Recent Advances in Natural Language Processing, p. 467–472.

[8] Agrawal, S. and Awekar, A. (2018). Deep learning for detecting cyberbullying across multiple social media platforms. Advances in Information Retrieval. Lecture Notes in Computer Science, 10772.

[9] Cheng, Z. Q., Wu, X., Huang, S., Li, J. X., Hauptmann, A. G., & Peng, Q. (2018). Learning to transfer: Generalizable attribute learning with multitask neural model search. Proceedings of the 26th ACM international conference on Multimedia, p. 90-98.

[10] Kumar, A., Tyagi, V., and Das, S. (2021). Detection of Offensive Language in Social Networks Using LSTM and BERT Model. IEEE 6th International Conference on Computing, Communication and Automation. P. 546-548.

[11] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of naacL-HLT, p. 2.

[12] Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI Blog.

[13] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzman, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, p. 8440-8451.

[14] Chiu, K. L., Collins, A., and Alexander, R. (2021). Detecting hate speech with GPT-3. arXiv:2103.12407.

[15] Ranasinghe, T. and Zampieri, M. (2020). Multilingual offensive language identification with cross-lingual embeddings. In Proceedings of Conference on Empirical Methods in Natural Language Processing, p. 5838–5844.

[16] Vidgen, B. and Derczynski, L. (2020). Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLOS ONE, 15(12), e0243300.

[17] Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., and Kumar, R. (2019a). Predicting the type and target of offensive posts in social media. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, p.1415–1420.

[18] Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., and Kumar, R. (2019b.) SemEval-2019 Task 6: Identifying and categorizing offensive language in social media (OffensEval). In Proceedings of the 13th International Workshop on Semantic Evaluation, p. 75–86.

[19] Basile, V., Bosco, C., Fersini, E., Nozza, D., Patti, V., Pardo, F. M. R., Rosso, P., and Sanguinetti, M. (2019). Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. Proceedings of the 13th international workshop on semantic evaluation. p. 54-63.

[20] Mandl, T., Modha, S., Majumder, P., Patel, D., Dave, M., Mandlia, C., and Patel, A. (2019). Overview of the HASOC track at FIRE 2019: Hate speech and offensive content identification in indo-european languages. Proceedings of the 11th Forum for Information Retrieval Evaluation, p. 14-17.

[21] Toraman, C., Şahinuç, F., and Yilmaz. E. (2022). Large-Scale Hate Speech Detection with Cross-Domain Transfer. Proceedings of the Thirteenth Language Resources and Evaluation Conference, p. 2215–2225.

[22] Ahmad, W., Peng N., and Chang, K.-W. (2021). GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction. In Proceedings of the AAAI Conference on Artificial Intelligence, 35(14), 12462-12470.

[23] Cheng, Z. Q., Dai, Q., Li, S., Mitamura, T., & Hauptmann, A. (2022). Gsrformer: Grounded situation recognition transformer with alternate semantic attention refinement. Proceedings of the 30th ACM International Conference on Multimedia, p. 3272-3281.

[24] Straka, M. and Strakova, J. (2017). Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, p. 88-99.

DOI: 10.54254/2755-2721/21/20231114

[25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, p. 5998–6008.
[26] CC-100: Monolingual Datasets from Web Crawl Data. https://data.statmt.org/cc-100/
[27] Ushio, A., Neves, L., Silva, V., Barbieri, F., and Camacho-Collados, J. (2022). Named entity recognition in Twitter: A dataset and analysis on short-term temporal shifts. In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, p. 309-319.