Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/67/20240607

Investigating the capability of recurrent neural networks for identifying text-based cyberbullying incidents on social platforms

Yukun Qi

Chengdu Foreign Language School, Chengdu, China

15388103770@163.com

Abstract. As cyberbullying escalates worldwide, the emotional toll on those targeted is profound, demanding attention. This question allows the framing of the detection of cyberbullying as a categorization puzzle, leveraging a spectrum of detection strategies from classical machine learning to advanced deep learning techniques via Natural Language Processing (NLP). It zeroes down to that capability which characterizes RNNs in parsing sequential data and interpreting the contextual nuances. This hence underscores their effectiveness and accuracy in the flagged detection of cyberbullying content. Directly compared to classical algorithms, it shows the best performance of RNN regarding accuracy, speed, and universality over different languages. The overall result of this research does, therefore, affirm the very strong promise and effectiveness of RNN frameworks toward the discrimination of cyberbullying across the various linguistic online environments, setting a firm ground for further development in cybersecurity.

Keywords: Recurrent Neural Networks, Text-Based Cyberbullying Detection, Social Platforms, Natural Language Processing, Machine Learning.

1. Introduction

1.1. Problem presentation

The reality, in fact, is that cyberbullying exists in the digital world, where the battleground is laid with harassment, intimidation, or even naked injury, in our very much networked world. The risk this assumes, especially in its textual nature—the form of sharp comments or messages full of venom—it leaves on the addressee, to underscore the severity of the influence. Textual cyberbullying is a concept to be known because of its insidious nature and pervasiveness across demographic contexts. As shown in the graph of Figure 1, worldwide growth of online violence, from slight to severe, is on the rise [1].

DOI: 10.54254/2755-2721/67/20240607

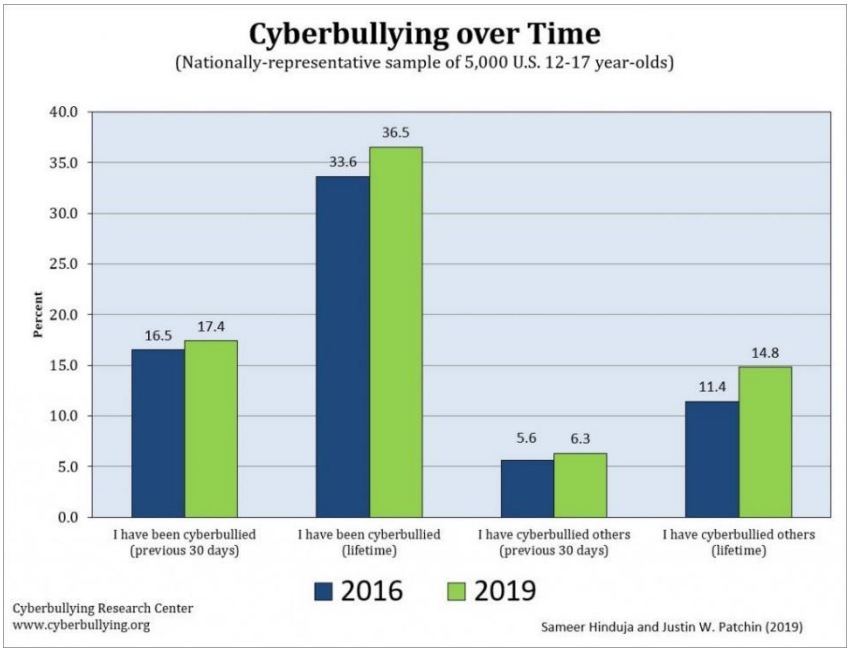

Figure 1. Evolution and Escalation of Digital Harassment

Text-based online violence takes an exceptional rate of occurrence among the various forms of cybercrimes and carries deep psychological impacts. Detection of the same is the area of focus through an advanced algorithm developed specifically for this violence approached through text classification within natural language processing (NLP). The method presupposes to encode the to-be-checked sentences by a model, which is to categorize these as "cyberbullying" or "not cyberbullying." A description of the procedure is given in the accompanying Figure 2.

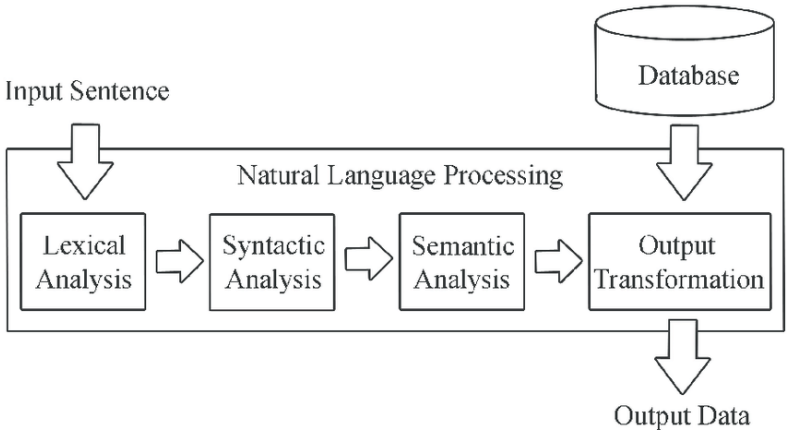

Figure 2. Mechanics of Machine Understanding of Human Language

1.2. Introduction to textual cyberbullying detection

Textual cyberbullying detection include different approaches: traditional and the deep-learning-based, primarily differing either in their encoding approaches or models. The following sections discuss NLP and the various methodologies.

NLP resides at the intersection of computer science and linguistics, with a focus on making it possible for a computer to process and understand human language in a way similar to that of human beings—hence "natural language." The potential of these includes language translation, sentiment analysis, and even chatbots and speech recognition systems. Thus, the challenge of automatically detecting textual

DOI: 10.54254/2755-2721/67/20240607

cyberbullying is approached as an NLP problem, and many other attractive opportunities are under current investigation.

Generally, the majority of techniques used are categorized as either machine learning or deep learning. Some of the conventional methods employed in classical approaches include statistical and machine learning techniques such as linear regression models, random forests, SVMs, and others in order to detect malicious content over digital communications. These, of course, all have been effective, but the very nature of cyberbullying has changed to the point that much more sophisticated means of dealing with it need to be found. One of the disadvantages of such deep learning models is that a deep neural network normally works in an end-to-end manner, from an input sentence to an output classification of the sentence. Therefore, the semantics of the sentence are not taken into account. Recurrent Neural Networks (RNNs) represent a subset within the category of deep neural networks, processing the data by a sequential mechanism and at the same time storing language context dependencies. Unlike classical networks, RNNs are accountable not only for the present situation but also for the history of inputs. They help learn and understand sequential and contextual data. They can be useful in identifying any sequential changes happening in context to online conversation, another area that could help identify the texts in cyberbullying. Speech analysis in the context of cyberbullying involves reviewing the content of speech to be able to determine its context and where it is coming from. In general, RNNs, therefore, become a very effective and useful approach in NLP used to safeguard digital spaces. Hopefully, figure 3 reveals the benefit as a deep learning method (namely, RNNs in this task) over classic methods.

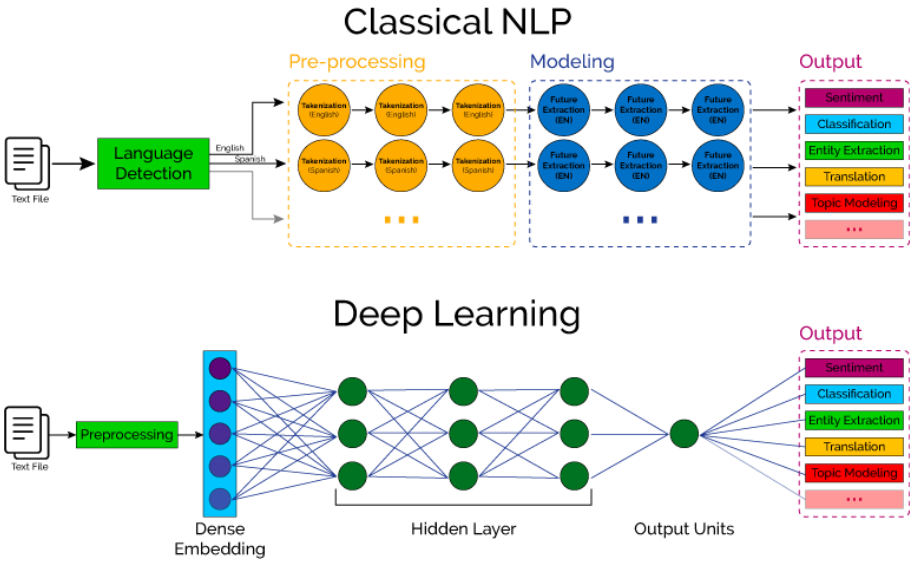

Figure 3. Superiority of Advanced Neural Networks in Data Analysis

1.3. Motivation to do this research

Motivation to do this research Our work really takes an important step forward in including RNNs—a class of deep learning neural networks well appreciated in their aptitude to sequential data analysis—in our implementation. The inherent ability of the RNN model to deal with both temporal dependencies and contextual shades of meaning fits well with the complexities of language as mirrored in the situations of cyberbullying.

Our framework integrates the Recurrent Neural Network (RNN) model to meticulously analyze sequential patterns in textual data, effectively identifying the intricate nuances and evolving tactics prevalent in cyberbullying incidents. This structural flexibility of the RNN model, therefore, in the diverse linguistic and cultural contexts, points at its effectiveness as a robust tool in the search for abusive behavior online. At the core of our research approach lies an anchorage to the thorough scrutiny and refining of the RNN model, designed explicitly for cyberbullying detection. It exposes the model to

DOI: 10.54254/2755-2721/67/20240607

a large training corpus, largely made up of English comments, so as to fine-tune its responses to the different linguistic idiosyncrasies. Then, a more comprehensive model includes a customized dataset from Zhihu, one of the leading Chinese social media, such that the effectiveness of the model to detect across diversities of language environments the different features of cyberbullying is improved.

That is precisely what this study intends to assess in the RNN model against conventional algorithms such as linear regression, Random Forest, and SVM. It is in this comparative perspective that the present research will be conducted for the deducing of the exact advantages that RNN holds with respect to its precision, efficiency, and adaptability in the most complex dimension of application: cyberbullying identification.

1.4. Structure of the paper

We start our paper with an overview of cyberbullying and its need to devise stronger tools to stop detecting this illegal and deplorable behavior. This study involves exploration of literature discussion on cyberbullying with detection approaches, portraying the traditional models against potential use of RNNs. The section of methodology includes a dataset obtained by the researchers, architectural design of models, and the modes of its training. We enable the return of performances in both methods, and we enable the performance of both methods. One is able to come up with a decision from an informed point on how effective the RNNs are for cyberbullying detection. At the end of the conclusion, main contributions to the research are summarized for the reader.

2. Literature Review

2.1. Traditional methods

They are using linear regression, one of the traditional methods in the detection of cyberbullying, which uses statistical methods to point out the correlation of possible signs of online abuse. The linear regression models of Yokoyama and Sanada focus on language markers of salient features of cyberbullying: profanity, threats, hate [2]. These models take into account the contribution of each single linguistic element to the probability of cyberbullying in terms of a linear relationship with metric values assigned to each form of expression. The basic point, however, of the drawback of linear regression relates to the intrinsic presumption that linearly features with the outcome variable. This may be taking off some of the more nuanced layers of complexity that are embodied in the bulwark data of cyberbullying and instead leaving one to wonder if, in a small way, the data is being missed.

Random Forest models, famous for their accuracy in detecting cyberbuljson, encapsulate ensembles' strength. These, therefore, would underscore the prowess of Novalita et al., in deecting episodes of cyberbullying and henceforth proved the competency of the models in handling high-dimensional data with lots of features [3]. For this very reason, it is very good at learning the nested complex patterns generally found in textual data, a very important feature for correctly detecting cyberbullying. However, this model depends mostly on the goodness and variability of the features used in the model.

This, therefore, would mean that the use of the SVM algorithm in cyberbullying detection turns out to be very efficient, since it turns out to be highly appropriate for binary classification and easily copes with the high dimension of the attributes of data. With regard to this, it is deemed that SVM-based frameworks work effectively in the differentiation of both offensive and non-offensive content, where they use hyperplanes to help them in stark discrimination [4]. It is often found that their best strength emanates from their sophisticated modeling of complex relationships that exist within textual data, especially if the considered features go beyond the confines of simple one-dimensional metrics. On the other hand, the crux of the success of SVMs lies in the choice and fine-tuning of their kernel functions. Both need to be finely optimized as they are of great importance.

2.2. RNN algorithms

Deep learning has come to dominate the field of cyberbullying detection. It can be really good at multifaceted patterns of sequential data, which is a very critical aspect in scrutiny of the vast and nuanced

DOI: 10.54254/2755-2721/67/20240607

array of exchanges taking place online. This can be a very powerful way for understanding subtler intra- and interconnections, often hierarchically and non-linearly contained in text data, with respect to complex cyberbullying phenomena. This holds focus on the kind of models which decode these fine details correctly, hence supports their centrality in the effective identification and moderation of these cyberbullying phenomena.

The advent of the Recurrent Neural Networks (RNNs) to the deep-learning space is based on their ability to process sequences of information while keeping a contextual remembrance at any given time. This is also the reason for the finding by Murshed et al. [5] that RNNs could capture dynamic patterns and more ambiguous linguistic signs that hint at online harassments, hence prove effective in their identification. Other systems to be in place include recurrent connections equipped with memory-like function, intrinsic to RNNs, that ensure the system in place recognizes temporal dependencies that are key to ensuring the correct detection of cyberbullying instances.

The introduction of Receptance Weighted Key-Value (RWKV) models with Recurrent Neural Networks (RNNs) will, by and large, contribute much to bettering cyberbullying detection, especially in interpretability and speed improvement. The models are an interesting blend of the lexical framework with contextual comprehension, thus offering full analysis of text with an array of insights in several layers. As Peng et al. highlight, the prowess of RWKV models in interpreting the sentence structures and nailing down semantic hierarchical links shines through to provide valuable augmentation with respect to the sequential analytical prowess of RNNs [6]. These interactions with the RWKV models would fine-tune the combined capabilities of the two and hence augment their effectiveness in being able to identify when a cyberbullying incident is taking place over the social media platform.

3. Methodology

The aim of the study is to develop and apply an RNN model, particularly the RWKV version, for the aim of unearthing cyberbullying in online commentaries, more specifically based on English and Chinese content. First, we pre-train with broad multilabel English Twitter data and then further fine-tune with Chinese context using Zhihu-sourced data. The dual-phase training regime unfolds within the framework methodological with meticulous concern articulating about how this very contemporary model is fine-tuned to handle the nuances of cyberbullying that proficiently transcends beyond the realms of linguistic boundaries.

3.1. The RWKV Model

Recognized as the RWKV algorithm, this wave-based technology opens a new frontier for the Recurrent Neural Network (RNN) family that otherwise accounts for more dynamic, time-sensitive interactions opposed to the static norm observed in conventional RNNs. In this approach, both the memory bottleneck problems and the expensive quadratic scaling that are traditionally related to transformers are directly handled and mitigated. In this approach, linear, efficient scaling is achieved without losing key benefits offered by transformers, such as parallelizable training, features of scaling. Details of these advances are described in the 2023 paper "Reinventing Recurrent Neural Networks for the Transformer Era." RWKV is remarkable since it uses different mechanisms that provide the model with the analytic depth for data points, which may enable use-cases to be supported much more than current architectures afford [6]. The ability of the model in bringing out the best performance across all NLP tasks was considered above what was previously considered in the State-of-The-Art (SOTA) with fewer resources, hence very much contributed to our decision to implement this model in our research.

RWKV accumulates by having an accumulating array for the residual block. Each of the blocks is made up of two discrete sub-blocks—one for the time-mixing part and the other for channel mixing. These variables are very important to maintain the recurrent structure of the model, allowing it to merge smoothly between the insights derived from past and the upcoming data. Figure 4 shows the blocks of time-mixing and channel-mixing, representing their critical roles to the operational framework of the model.

DOI: 10.54254/2755-2721/67/20240607

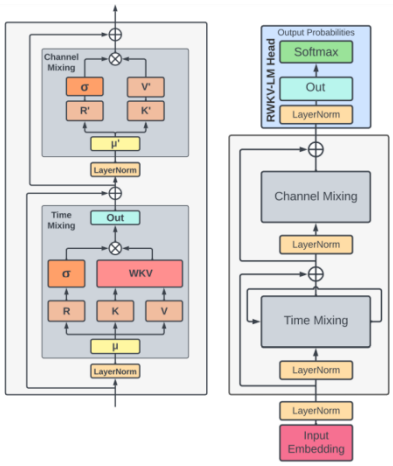

Figure 4. Structural Design of the Rotate-Wide-Kick-Vanish (RWKV) Architecture

The novelty of the mixing of time block stands out regarding the ability of its attention mechanisms to be enriched by a new way of computation for the weighted sum. This innovation will be of much importance in refining the ability of the model in the dynamic assimilation and processing of the information, marking a great leap in refining efficiency and effectiveness of neural network architectures. It then goes ahead to decode these hidden states by trying to establish patterns and relationships within them, going on further to transfer and transform the information represented in the hidden states. Time-mixing block vectors for time-mixing are formulated as linear projections of linear combinations of both current and previous inputs, in Equation (1) deduced by Peng et al [6].

$$\begin{aligned} r_t &= W_r \cdot (\mu_r \otimes x_t + (1-\mu_r) \otimes x_{t-1}) \\ k_t &= W_k \cdot (\mu_k \otimes x_t + (1-\mu_l) \otimes x_{t-1}) \\ v_t &= W_r \cdot (\mu_v \otimes x_t + (1-\mu_v) \otimes x_{t-1}) \end{aligned} \tag{1}$$

The block of channel-mixing aims to minimize the coefficient "aller" for the system to have a better capacity in knowledge representation. Primarily, this one sets a platform for encoding and processing information. Specifically, vectors for channel-mixing are also linear projections of the linear combinations of the block's current and past input [6], as expressed in Equation (2).

$$\begin{aligned} r'_t &= W'_r \cdot (\mu'_r \otimes x_t + (1-\mu'_r) \otimes x_{t-1}) \\ k'_t &= W'_k \cdot (\mu'_k \otimes x_t + (1-\mu'_k) \otimes x_{t-1}) \end{aligned} \tag{2}$$

To be brief, the RWKV model has competed parallelization with the time parallel method because these costs are given out with the tensor result,and a linear approximation for analogous matrix is included.

3.2. Implementation Details

For the tokenizer implementation, we plan to use a third-party package that converts the input textual sequence into the output sequence of integer numbers from 0 up to 50276, where each one presents a unique word, symbol, or token in our vocabulary, encoded as encode(context).ids. Then, these sequences of tokens are folded with the state of the RWKV model. In short, RWKV is an operation applied to a function that takes a token and gives the probability distribution for the next token and, at the same time, updates its current state. We do this using a pretrained RWKV model and only the state is initialized to be zeroed. The model is then iteratively operated with all the tokens.

While "state" holds the encoder's state for the text data, "probs" details the prediction probabilities by the model to guide in taking the next token. Such elements are captured for the purpose of getting the comprehensive summary of the text given as input. The first layer of this network architecture is a

DOI: 10.54254/2755-2721/67/20240607

3-layer MLP used to act as a trainable classifier for binary outcomes. In this first layer of the network architecture, along with the mentioned classifier and its operational logic, it is represented through pseudo-code and corresponding diagram in Figure 5.

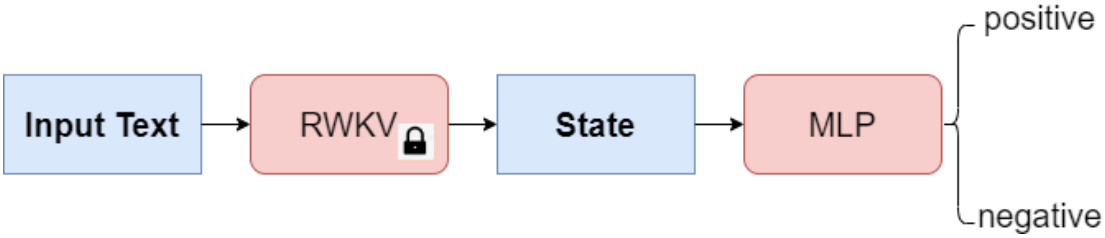

Figure 5. Schematic Representation of RWKV-Enhanced Detection System

3.3. Training Model

The next key phase for the best model effect is the choice of a useful training measure. Improving the training procedure and crucially selecting the proper loss function are involved, which is key to instructing the model towards precise predictions.

3.3.1. Loss function

Wang Q, Ma Y, and Zhao K in 2020 stressed that in the field of machine learning algorithms, the question of the loss function becomes quite an important area for research, as it has an impact on the development of algorithms and their improvement in performance. Many scientists have considered and studied this area at length [7]. The problem at hand, a multi-label classification problem of cyberbullying, requires that we define the loss function to be categorical cross-entropy and train in the supervised setting. Here, the model picks up the right "answers" to make this objective function as low as possible. The multi-label categorical cross-entropy loss is given by:

$$Loss = -\sum_{i=0}^{C-1} y_i \, lg(p_i) = -lg(p_c) \tag{3}$$

where $\mathbf{p} = [\mathbf{p_0}, \ldots, \mathbf{p_{c-1}}]$ is a distribution over the elements, which represents the probabilities of a sample belonging to category i, and $\mathbf{y} = [\mathbf{y_0}, \ldots, \mathbf{y_{c-1}}]$ is the one-hot representation of labels in the sense that $y_i = 1$ if the sample belongs to category i.

According to Peng.B's article at 2023 [6], the loss function at position T can be written as:

$$L_T = l(f(wkv_T), y_T) \tag{4}$$

Because wkv_T relates to $(W_k)_{i,j}$ and $(W_v)_{i,j}$ only through the i-th channel $(wkv_T)_i$, we have

$$\frac{\partial L}{\partial (W_v)_{i,j}} = \frac{\partial L}{\partial (wkv_T)_i} \frac{\partial (wkv_T)_i}{\partial (W_v)_{i,j}} \tag{5}$$

Where the first part of the above equation, the others operate at layers of output or others at layers of time mixing, therefore containing trivial ones, which is proved inductively. The second part of the above equation can be bounded as follows:

$$\left| \frac{\partial (wkv_T)_i}{\partial (W_v)_{i,j}} \right| = \left| \frac{\partial E_i[(v_t)_i]}{\partial (W_v)_{i,j}} \right| = |E_i[(x_t)_i]| \leq \max_t |(x_t)_i| \tag{6}$$

which is irrelevant to T Similarly,

$$\begin{aligned} \frac{\partial (wkv_T)_i}{\partial (W_k)_{i,j}} &= \partial \frac{S_i[(v_t)_i]}{S_i(1)} / \partial (W_k)_{i,j} \\ &= \frac{S_i[(x_t)_j (v_t)_i]}{S_i(1)} - \frac{S_i[(x_t)_j] S_i[(v_t)_i]}{S_i(1)^2} \\ &= E_i[(x_t)_j (v_t)_i] - E_i[(x_t)_j] E_i[(v_t)_i] \end{aligned}$$

DOI: 10.54254/2755-2721/67/20240607

$$= cov_i((x_t)_j, (v_t)_i) \tag{7}$$

can also be bounded. Note that in the softmax of *wkv*, each channel's minimum contains at least two nonzero terms (*u* and *w*), so the "cov" of above will not degenerate into 0.

3.3.2. Two steps of training

Since the current Chinese dataset was of a very small and relatively closed size, we, on the first note, had to train the process built on the relatively large sizes and the high accessibility of English comments. The model will then be adjusted and fine-tuned using the collected Chinese comment dataset.

The present work uses an elaborate English multi-label Twitter dataset by Salawu et al., which is purpose-designed for detecting online abuse behaviors. Preprocessing downstream includes tokenization, cleaning, and vectorizing of text comments. These are central procedures that refine data into the right format for the RNN architecture. Therefore, that gives an ability for the model to have better accuracy concerning subtleties and anomalies in online communication. Such preprocessing can help in successfully distinguishing between appropriate and inappropriate content. Illustrative examples of this data are set to be displayed in Figure 6.

1955	Non-offens	If you a hoe ima treat you like one just like if you're a good girl ima treat you like one. I don't discriminate
1956	Non-offens	If you are descended from anything less than European nobility and/or royalty, then you are non-titled common gutter trash.
1957	Non-offens	If you didn't have tattoos as a diversion for your bird chest and malnourished body sir... your instagram would be trash
1958	Non-offens	If you enjoy hunting for sport, then you're trash. If you hunt for food, you deserve to starve.
1959	Non-offens	If you must call me a cracker, please call me a Cheez-It, not a saltine.
1960	Non-offens	If you think that the term #redskins being offensive is a political issue, you are ignorant.
1961	Non-offens	If you're not directly descended from European royalty and upper nobility, then you're not fully human. You're just common gutter trash.
1962	Non-offens	If you've ever been to drunk to fish, you might be a redneck

Figure 6. Exemplary Data Entries with Categorized Annotations

The contents of the open-source datasets, COLDATASET, were classified in regard to the three themes: gender antagonism, regional discrimination, and racism, from which they were sourced from Weibo and Zhihu. The optimization of data matching becomes not easy because, from source to source, the length and quality of the data fields and their context tend to vary. However, such fully automatic annotation is quite questionable, because a noise level becomes inevitable. In the following section, in order to make the dataset more reliable, we have proceeded through a merging process by removing duplicates and further refining the training data to be confident about its precision, through manual re-labeling. If the re-labeling exercise showed discrepancies between the new and original labels, then the one that fits best was chosen so as to have consistency.

It fine-tunes the parameters of a sentiment analysis model with a Chinese dataset, updating them in a new linguistic context. Our model is made adaptable, so it could easily be able to capture the pattern of cyberbullying in comments by Chinese students and exhibit how flexible is its powerful capacity to transform it to differences in languages and peculiarities of nationalities.

3.3.3. Evaluation

The evaluation metrics give essential benchmarks to the performance of the model in an online anti-bullying battle, aiming at balanced results through accuracy, precision, recall, and f-1. Detailed implementations for these metrics and more can be found in Figure 7 below.

DOI: 10.54254/2755-2721/67/20240607

```
def pipeline(learner_list, X_train, y_train, X_test, y_test):
    '''
    inputs:
       - learner: the learning algorithm to be trained and predicted on
       - X_train: features training set
       - y_train: income training set
       - X_test: features testing set
       - y_test: income testing set
    '''
    size = len(y_train)
    results = {}
    final_results = []
    for learner in learner_list:
        results['Algorithm'] = learner.__class__.__name__
        start = time()
        print("Training {}".format(learner.__class__.__name__))
        learner = learner.fit(X_train, y_train)
        end = time()
        results['Training Time'] = end - start
        start = time()
        predictions_test = learner.predict(X_test)
        predictions_train = learner.predict(X_train)
        end = time()
        results['Prediction Time'] = end - start
        results['Accuracy: Test'] = accuracy_score(y_test, predictions_test)
        results['Accuracy: Train'] = accuracy_score(y_train, predictions_train)
        results['F1 Score: Test'] = f1_score(y_test, predictions_test)
        results['F1 Score: Train'] = f1_score(y_train, predictions_train)
        results['Precision: Test'] = precision_score(y_test, predictions_test)
        results['Precision: Train'] = precision_score(y_train, predictions_train)
        results['Recall: Test'] = recall_score(y_test, predictions_test)
        results['Recall: Train'] = recall_score(y_train, predictions_train)
        print("Training {} finished in {:.2f} sec".format(learner.__class__.__name__, results['Training Time']))
        print('-------------------------------------------------')

        final_results.append(results.copy())

    return final results
```

Figure 7. Script for Assessing and Validating Model Performance

The accuracy literally measures the proportion of well-classified instances against all categorization; therefore, it contrasts valid detection and false alert with benchmark data. An accuracy rate above 90% deems it sufficient for adding functionality to the detection system by classifying cases. This is of importance: a model should separate accurately between a real incident of cyberbullying from a non-incident, but more importantly, this should reduce the rate of false positives; a case where a benign action is flagged as cyberbullying. High precision would thus mean few instances of the cases wrongly classified and, therefore, better detection of real cases of cyberbullying.

The second measure—recall, also called sensitivity—accounts for the model's performance to retrieve every instance of cyberbullying among all available real cases. It represents the proportion of actual cyberbullying incidents that were correctly identified in the total actual cyberbullying events contained in the data set. The fact that the high percentage of recall reflects the ability to provide comprehensive detection of the phenomenon of cyberbullying and significantly decreases the possibility of its omission, which is expressed in a small percentage of false negatives.

The F1 score provides a measure that tries to balance the tradeoff between precision and recall into one metric value, hoping for a balance where the score close to 1 would indicate an optimum performance at certain precision or recall thresholds. The harmonic mean of both precision and recall metrics equally weights both and hence gives a good summary score of accuracy in the model. High F1 score: the model is very good in discriminating cases of cyberbullying from false positive and negative cases.

The methodology will comprise implementation with RWKV that will start with training on the English datasets and fine-tuning on the Chinese comments. The approach will tap the potential ability of RNN in capturing sequential dependencies in the development of a universal model for the recognition of cyberbullying beyond linguistic and cultural diversification. The process finishes with comprehensive validation, ensuring the model is robust and adaptable to diverse scenarios.

DOI: 10.54254/2755-2721/67/20240607

4. Results

From collecting data and training models, we construct some representative traditional and RNN-based models and estimate the performance of the model according to different standards of evaluation, as depicted in Table 1.

Table 1. Results base on varied performance metrics.

	SGD	Logistic Regression	Decision Tree	Linear SVC	RWKV
accuracy	0.9572	0.9555	0.9540	0.9561	0.9652
Precision	0.9741	0.9731	0.9721	0.9735	0.9811
recall	0.9817	0.9764	0.9829	0.9775	0.9871
F1-score	0.9666	0.9699	0.9614	0.9695	0.9726
Time/ms	2300	1900	1930	1780	930

5. Conclusion

In conclusion, our research focuses on utilizing RNNs frameworks, especially the RWKV variant, to pinpoint cyberbullying incidents across English and Chinese digital landscapes. Our side-by-side evaluation with conventional detection methodologies underscores that the RNN-centric strategy not only elevates accuracy levels but also amplifies processing efficiency. This superior operational performance stems from the RNN's intricate representational ability, masterfully capturing complex linguistic patterns. Furthermore, the adaptability of RNN-based models in various linguistic settings is distinctly evident, gaining significantly from the incorporation of additional training datasets. Therefore, we propose that these RNN-based methods in that line bear enormous potential and are very effective techniques for cyberbullying detection, supplying at the same time a very powerful but flexible tool against this prevalent problem in the online, linguistic environment.

References

[1] Aoyama I and Talbert TL 2010 Cyberbullying Internationally Increasing: New Ihallenges in the Technology Generation Adolescent Online Social Communication and Behavior: Relationship Formation on the Internet ed R Zheng J Burrow-Sanchez et al (Pennsylvania: IGI Global) chapter 12 pp 183-201

[2] Yokoyama S and Sanada H 2009 Issues in Quantitative Linguistics Logistic Regression Model for Predicting Language Change ed R Köhler pp 176-192

[3] Novalita N Herdiani A et al 2019 J. Phys.: Conf. Ser. 1192 012029

[4] Purnamasari NM Fauzi MA et al 2020 Cyberbullying identification in twitter using support vector machine and information gain based feature selection Indon. J. Electrical Engineering and Computer Science vol 18 pp 1494-500.

[5] Murshed BA Abawajy J et al 2022 DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform IEEE vol 10 pp 25857-71.

[6] Peng B Alcaide E Anthony Q et al 2023 Rwkv: Reinventing rnns for the transformer era arXiv preprint arXiv:2305.13048

[7] Wang Q Ma Y et al 2020 A comprehensive survey of loss functions in machine learning Annals of Data Science pp 1-26.