ORIGINAL RESEARCH

Research on the psychological classification of violent crime based on a convolutional neural network

Hui Li[1,3] · Guandong Gao[2,3] · Ke Xiao[1,4] · Shengzun Song[5]

Abstract

This paper uses a text classification algorithm in the context of criminal psychology to scientifically evaluate the correctional effects among offenders in the hopes of preventing recidivism. The proposed algorithm uses an improved convolutional neural network (CNN) and feature fusion to carry out the psychological attribution of violent crime based on criminal fact texts, which are divided into three types: the premeditated type, the impulsive type, and the pathological type. First, a text set of criminal facts with labels was constructed by criminal psychology experts. Then, the numerical features are extracted by the word count and a keyword extraction method based on the criminal feature dictionaries, which are established by criminal psychology experts. The improved CNN model is used to extract the text features by replacing a fully connected layer with the global average pooling layer. Finally, after fusing the numerical features and the text features, softmax function is used to calculate the probability of each category and divide the attribution type. To verify the effect of the algorithm, the proposed model is evaluated and compared with the original CNN, the SVM based on TF-IDF and word count, random forest, and some integration methods. Experimental results show that the proposed algorithm is better in accuracy, recall, precision, and F1 in the micro and macro scopes, with the highest accuracy of 98%.

✉ Guandong Gao
  xiaoke@hebau.edu.cn

  Hui Li
  youthhui@163.com

  Ke Xiao
  xiaoketeaching@sina.com

  Shengzun Song
  songshengzun@163.com

1 College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, Hebei, China

2 Department of Information Management, The National Police University for Criminal Justice, Baoding 071001, Hebei, China

3 The Centre of Data Science and Intelligent Correction Technology, The National Police University for Criminal Justice, Baoding 071001, Hebei, China

4 Hebei Key Laboratory of Agricultural Big Data, Hebei Agricultural University, Baoding 071001, Hebei, China

5 Department of Penology, The National Police University for Criminal Justice, Baoding 071001, Hebei, China

1 Introduction

In criminal psychology, explaining "which types of people are more likely to commit crimes" has a certain macroscopic value. Judging a person's crime risk, which is the likelihood of a person committing a crime, can effectively prevent the occurrence of a crime (Ma et al. 2020). It is important to analyse the criminal psychological attribution types of offenders and provide a scientific basis for prison management departments to reasonably formulate a corrective education programme and assess the possibility of recidivism. Additionally, the scientific prediction and assessment of offenders are carried out to reduce crime to a certain extent, thus promoting the continuous stability of social security, improving the happiness and safety index of Chinese citizens' daily lives, and further promoting the modern development of the national governance system and governance capabilities.

Attribution analysis is used to deduce personal purpose based on behaviour and its results, which means that motivation and psychological status can be inferred to determine

the causal relationship between various factors based on a specific personality characteristic or behaviour (Jones and Harris 1967). In terms of the attribution classification of violent crime, violence is an important manifestation of criminal behaviour, which is mainly manifested as "aggressive", which can be divided into premeditated, impulsive, and pathological types (Barratt et al. 1991). The attribution classification of violent crime provides an important basis for crime risk assessment. In criminal psychology, traditional crime risk assessment tools are used for psychological attribution analysis, which mainly use historical data analysis, scale assessment, behaviour analysis and clinical analysis for attribution analysis, combined with static factors and dynamic factors. The main assessment tools include the LSRP (Levenson 2014), OGCS (Copas and Marshall 1998) and PCL: YV (Dickson 2014). The assessment tools are bloated, and the results are distorted because of the introduction of a large number of assessment factors. Subjective factors are easily introduced into operations or explanations by experts for the assessment process. With the development of artificial intelligence technology, the intelligent assessment tools proposed in recent years can avoid the influence of subjective factors when collecting data and implement the intelligence of the assessment tools (Xiao and Du 2020). However, the results may contain objective factors, because of further explanation by experts.

With the development of natural language processing, text classification is applied to criminal psychology, which is an important task for natural language processing. On the basis of the criminal facts related to offenders, text classification technology can automatically attribute and classify types of violent crime. The related technique of shallow learning includes support vector machines (Joachims 1998), decision trees (Johnson et al. 2002), naive Bayes (Schneider 2004) and other traditional machine learning methods, which are mainly used in predicting crime categories (Hassani et al. 2016). Before using machine learning algorithms for classification, shallow learning models need to extract features or design rules by humans (Liu et al. 2017; Das et al. 2020; Ni et al. 2020), and vector text representation is used for feature representation (Cichosz 2018). However, there is a certain impact on feature extraction due to the linguistic differences in different regions. The manually designed features may neglect the context and semantic information, and the generalisation is limited in that the performance is usually not good with other datasets.

The text features can be extracted by deep learning models, which can be used to avoid the influence of human factors. In addition, there is no prior knowledge. Deep learning has been widely applied in various fields (Chen et al. 2021; Xiao et al. 2021a) and has made good progress in different kinds of tasks in the field of crime (Pereira-Kohatsu et al. 2019; Liang et al. 2021). The mainstream extraction models for text features include RNNs (Liu et al. 2016), CNNs (Kim 2014), HANs (Yang et al. 2016) and transformers (Vaswani et al. 2017). RNN requires time-sequence data to be input, which results in poor parallel computing capability (Lei et al. 2018); Transformer requires the performance of pretraining tasks, which produces a huge amount of calculation for long text sequences (Devlin et al. 2019). CNN is better than RNN and HAN with respect to time and accuracy and requires fewer training samples (Maheshwari 2018). Unfortunately, single features have been extracted for classification in most literature (Luo et al. 2017; Chen et al. 2020). Additionally, due to the different segmentation effects in various fields, it is difficult to obtain important segmentation features for semantic representation, which may decrease the accuracy of classification (Yilmaz and Toklu 2020). There is a contradiction between word granularity and semantics after word segmentation. If the word granularity is large, the word segmentation effect is not good, and if the word granularity is small, the semantic information will be lost. Moreover, the convolution operation is the core of CNN, which can better combine the segmentation with semantic information and obtain more accurate word vectors. Furthermore, there are few studies on the application of criminal psychological attribution analysis.

Based on the above research, to solve the problem of inaccurate word segmentation and a short description of criminal facts that cannot extract features effectively, a classification method is proposed for the psychological attribution of violent crime to help relevant departments intuitively categorize crimes and formulate corresponding corrective measures for offenders based on the radar chart. The improved CNN model is used to extract the text features, which introduces the global average pooling layer to replace the max pooling layer and the fully connected layer to prevent overfitting. To better model the criminal facts, the numerical features are integrated with the text features to fully describe the corresponding relationship between criminal facts and attribution types of violent crime. The softmax classifier, which can directly obtain the probability of each attribution category and reduce the computation, is used for classification. Furthermore, the criminal dataset has been constructed from prison, and the China Judgements Online has been sorted out for the experiment. The predefined feature dictionaries have been constructed by experts to improve the performance of the word segmentation, which can filter out the meaningless stop words and reduce the feature dimension.

The rest of this paper is organised as follows. Related work is described in Sect. 2. The proposed model is presented and explained in detail in Sect. 3. The experimental results compared with other methods and a discussion are provided in Sect. 4. Finally, the conclusions and future work are provided in Sect. 5.

2 Related work

2.1 Analysis of criminal psychological attribution

Research on attribution analysis has made considerable progress and has been applied to different fields (Zhang et al. 2021). Consequently, research on the psychological attribution of violent crime can provide knowledge about the changes in criminals' psychological state and emotion, predict the timing of a crime and to further elucidate the causal relationship between psychological changes and criminal behaviour or emotion (Kunst and Koster 2017; Keijsers et al. 2021; Ghosh et al. 2020). To date, scales and questionnaires have been used to study the attribution type of violent crime. Zhang (2020) applied risk/need to conduct attribution analysis to prevent juveniles from reoffending. However, attribution analysis can be regarded as a classification problem that is applied in criminal psychology based on data mining technology. Lin et al. (2020) compared the COMPAS and LSI-R Assessment System to evaluate the crime risk of subjects and accuracy. Ma and Song (2021) studied the feasibility of constructing a dynamic crime risk assessment tool based on affective computing. The introduction of artificial intelligence makes the collection of numerical information and affective computing automatic. With the development of deep learning, the text features can be extracted for attribution classification by neural networks, which can keep the semantics of the text unchanged.

2.2 The CNN model and feature fusion for analysing crime data

In recent years, the development of deep learning has greatly improved the performance of text classification, and a number of classification models with better performance have been studied to realise the transition from shallow learning to deep learning (Li et al. 2020a). The CNN model applied to the task of text classification has been used to extract text features by different convolution kernels and pooling operations. Word embedding, which can be pretraining by unlabelled corpora, can further improve the performance of classification to a certain extent (Wang and Mao 2019; Aydoğan and Karci 2020). The network structure of the CNN model can be adjusted to advance the accuracy of classification. Ibrahim et al. (2021) proposed that a covariance pooling layer was used to replace the max pooling layer, which had a stronger ability to distinguish classification labels. Li et al. (2020b) proposed integrating the attention mechanism into a CNN, and words irrelevant to the classification task were filtered

and pruned in the pooling layer. Yang (2020) proposed an integrated model to predict the crime category of certain behaviour, which fused CNN with SVM, linearSVC and other classifiers. Wu (2021) conducted research on predicting crime trends to analyse criminal psychology using the 3D-CNN. CNNs combined with other models have been widely applied to different kinds of classification tasks (Xiao et al. 2021b), and the improved models have better performance for text classification. CNN cannot effectively model long text. Therefore, manually designed features based on specific rules and frameworks will be fused to fully describe the corresponding relationship between criminal facts and attribution types of violent crime.

Feature fusion can effectively solve the problem that a single extraction method cannot fully describe the criminal text (Xiao et al. 2021c). Accordingly, manually designed features have been extracted based on specific rules and frameworks in some literature. Wang et al. (2021) proposed standardising manually extracted features and joining them with features extracted by deep learning models for classification. Chen et al. (2019) proposed extracting features by TextCNN and TextRNN for the task of predicting articles. Unfortunately, TF-IDF and other methods have been used to fuse features for criminal facts, which will increase the feature dimension and cause dimension disaster. Different from feature extraction with neural networks, a specific rule should be designed to select features that can represent the relationship between text content and attribution types.

3 Proposed method

To classify the psychological attribution of violent crime based on criminal facts and obtain the tendencies associated with the types of violent criminal behaviour, an improved CNN model and feature fusion are used to construct a three-classification model. First, a text set of criminal facts with labels and predefined feature dictionaries, which are matched with a premeditated type, an impulsive type, and a pathological type, are constructed by criminal psychology experts. Then, the preprocessed criminal facts are segmented, and the numerical features and the text features are extracted. The numerical features are extracted by the word count and keyword extraction method based on the predefined dictionaries, where the term frequency is considered the weight value. The text is converted into a vector matrix by one-hot encoding, and the text features are extracted by an improved CNN, which is constructed with an embedding layer, a convolutional layer, and a global average pooling (GAP) layer that is introduced to replace the fully connected layer. Finally, the text features and the numerical features are fused, and the softmax classifier is used to calculate the probability of each category for psychological attribution

Fig. 1 Overview of the model
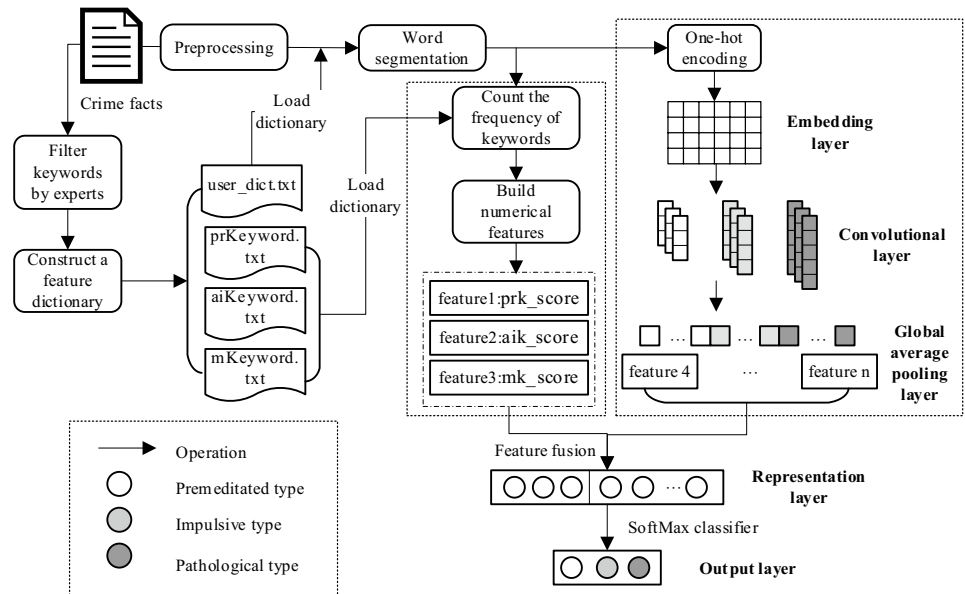
classification. The overall framework of the proposed model is shown in Fig. 1.

3.1 Construct the feature dictionaries

The keyword extraction method is used to extract the keywords of criminal facts, which belong to different types of violent crime. The result shows that the weights of representative words are distributed unevenly. Some representative keywords are segmented poorly due to the different performances of the tokenizer on different tasks. Therefore, criminal psychology experts screen the criminal facts about the offenders to select representative words and key phrases for different types of violent crime. The keyword dictionaries are built and named prKeyword.txt, aikeyword.txt and mkeyword.txt. The corresponding keyword dictionaries are premeditation, impulsiveness and pathotype, and these three types of dictionaries are summarised in user_dict.txt as a user dictionary. Figure 2 shows the word cloud display of the keywords in the user dictionary after the term frequency is counted in the dataset. The paper does not divide the prepositions into stop words different from other tasks. Some propositions are added to the dictionaries, which can represent the transition of action behaviour to a certain extent and express the emotional changes before and after the crime is committed.

3.2 Feature extraction

To obtain comprehensive information, two types of features are extracted: numerical features and text features. The numerical features are extracted by matching corresponding keywords in criminal facts based on different categories of

Fig. 2 Word cloud display

keyword dictionaries, and the word frequency is calculated to be the weight value of the features. The statistical term frequency of the different categories can indicate the attribution degree corresponding to the types of violent crime. The improved CNN is used for text feature extraction after the text data are vectorized. The local information of criminal facts is modelled for classification.

3.2.1 Algorithm for numerical feature extraction

In this paper, three numerical features are set up based on each feature dictionary to quantify the psychological attribution of violent crime. The corresponding keywords are extracted from the criminal facts based on the feature

Table 1 Description of numerical characteristic symbols

Symbols	Features	Explanation
Feature 1	pk_score	Related to premeditated type
Feature 2	aik_score	Related to impulsive type
Feature 3	mk_score	Related to pathological type

Table 2 Weight distribution chart of numerical characteristics of criminal facts

Features	Weights
pk_score	4
aik_score	5
mk_score	0

dictionaries, and the word frequency of the keywords from different categories is counted as the weight value of the corresponding numerical feature, which can indicate the attribution tendency to a certain kind of violent crime. The description of numerical characteristic symbols is shown in Table 1. For example, a criminal fact is shown in Fig. 3, which is taken from an award obtained from the China Judgements Online. After loading the predefined feature dictionaries to perform Chinese word segmentation and match keywords, the word frequency of the keywords is counted as the weight value of the features. The weight value of each feature is shown in Table 2.

3.2.2 Algorithm for text feature extraction

Attribution classification cannot be carried out accurately based on word frequency statistics alone, and it is necessary to analyse the semantic information and extract the text features from the criminal facts. The text data are converted into structured data. Then, the representative features are extracted for attribution classification to make the analysis results more accurate. The local feature information of text can be extracted by the CNN, which can be implemented simply, quickly, and in a robust manner. After the criminal facts are vectorized, feature reduction is performed in the embedding layer, and the text features are extracted by the improved CNN to represent the text information from the criminal facts. The feature vectors are combined for output classification in the fully connected layer after executing convolution and pooling operations in the traditional CNN model. It is prone to cause overfitting, and there are many parameter settings in the fully connected layer. Therefore,

this paper improves the use of the GAP layer instead of the fully connected layer, and a feature vector is extracted to represent a text feature in each feature map. The improved overall network structure is shown in Fig. 4.

(1) Convolutional layer

The CNN has the characteristics of translation invariance and can be used to extract local features of data. Each line of vectorized text data can represent a word segmentation feature different from image feature extraction, and words are the smallest granularity for text feature extraction. There is no mathematical meaning to truncate the feature dimension vertically, which will change the characteristic that words are the smallest granularity. Therefore, the convolution kernel only performs one-dimensional sliding, which means that the width of the convolution kernel is equal to the width of the input text vector.

Assume that the vectorized text data have a matrix $S = n * d$ in the input layer, the length of the text vectorisation is n, and the dimension is d. The size of the convolution kernel F is h*d, and h is the height of the convolution kernel. Then, after a convolution operation, the i-th dimension vector can be defined as Eq. (1),

$$c_i = (S * F)_i = \sum_{i+h-1} (S_{[i-h+1:i,:]} \otimes F)_{i+h-1} \qquad (1)$$

where $\otimes$ is a convolution operator and $[i - h + 1 : i, :]$ is a matrix of height h. A feature vector of dimension $(n-h+1)$ will be formed, as shown by Eq. (2).

$$C = [c_1, c_2, ..., c_{n-h+1}] \qquad (2)$$

Fig. 3 Example of criminal facts

xxxx / 因 / 女朋友 / xxxx / 在 / 年旧历 / 新年 / 期间 / 发生 / 感情 / 纠纷 / 并 / 提出 / 分手 / 而 / 怀恨在心 / 年月 / 日时 / 许 / xxxx / 回到 / 与 / 被告人 / xxxx / 共同 / 居住 / 的 / 位于 / 开平 市 / 街道 / 办事处 / 东路 / 号 / 房卧室 / 内 / 收拾 / 自己 / 的 / 行李 / 时 / 被告人 / xxxx / 遂 / 产生/ 杀害 / xxxx / 的 / 念头/ 随后 / 被告人 / xxxx / 使用 / 音响线 / 从 / xxxx / 背后 / 勒 / 住 / 其 / 脖子 / 致 / xxxx / 窒息 / 晕倒 / 昏迷 / 当 / xxxx / 苏醒 / 后 / 被告人 / xxxx / 又 / 使用 / 水果 / 刀割 / 伤 / xxxx / 颈部 / 处 / 损伤 / 创口 / 长度 / 累计 / 厘米 / 被告人 / xxxx / 见到 / xxxx / 颈部 / 流血 / 后 / 以为 / xxxx / 会 / 死亡 / 而 / 逃离 / 现场 / 并 / 上 / 到 / 涉案 / 地址 / 的 / 楼顶 / 准备 / 跳楼 / 自杀 / xxxx / 见 / 被告人 / xxxx / 离开 / 后遂 / 自行 / 逃 / 到 / 楼下 / 并 / 在 / 路人 / 的 / 帮助 / 下 / 报警 / 及 / 乘坐 / 出租车 / 到 / 医院 / 治疗 / 次日 / 凌晨 / 时 许 / 被告人 / xxxx / 在 / 其 / 朋友 / xxxx / 及 / 母亲 / 的 / 劝阻 / 下 / 回到 / 上述 / 涉案房 / 内 / 被 / 民警 / 抓获 / 经法医 / 鉴定 / 被害人 / xxxx / 的 / 损伤 / 程度 / 为 / 轻伤 / 一级

Fig. 4 Text feature extraction network structure diagram
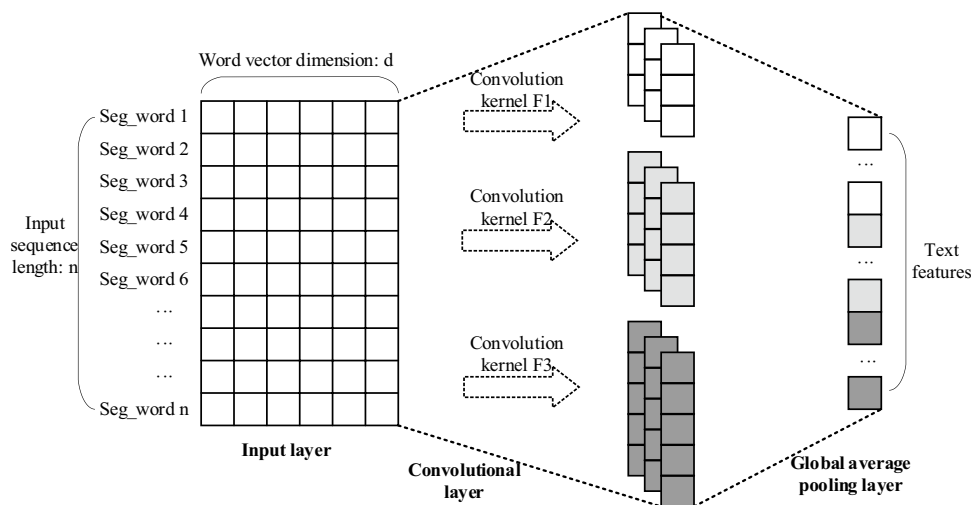
Since only one feature map can be extracted from a convolution kernel, different receptive fields should be obtained from multiple convolution kernels with different heights.

(2) Global average pooling layer

Multiple convolution kernels of different sizes are convolved on the vector matrix, and the filter is used to perform the pooling operation. Then, the feature vectors are combined in the fully connected layer and input into softmax for classification, as shown in Fig. 5(a). However, there are many training parameters in the fully connected layer, which makes the model more complicated and easily causes overfitting. In this paper, the fully connected layer is replaced with the GAP to calculate the average value of the feature map formed by each convolution operation, which can be regarded as the result vector of the corresponding feature. The formula can be defined as Eq. (3),

$$G^l_{avg-polling} = \frac{1}{x} \sum_{i=1}^{x} C^l_i \tag{3}$$

where $G^l_{avg-polling}$ is the value of the l-th feature map, l is the index of the feature map, C^l_i is the matrix of the feature map

and x is the total number of elements in the feature map. Finally, the dimension of the feature vectors is reduced from the $3*k(n-h+1)$ formed by the convolutional layer to $3*k$. The operation of the GAP layer is shown in Fig. 5(b). The corresponding relationship between the feature map and the category label can be represented better by the GAP layer, which can deliver the actual meaning of the category label. There is no need to set a large number of parameters for the GAP calculation, and the amount of calculation is greatly reduced, as shown in Fig. 5. The GAP layer can be regarded as a regularisation method to optimise the network structure, which can avoid the occurrence of overfitting.

3.2.3 Fusion feature classification based on softmax classifier

For criminal facts, the weight value of the numerical features indicates the attribution tendency of different crime types directly to a certain extent, and the text features can describe and model the local information about the criminal facts. As some descriptions of criminal facts lack standardisation due to human factors, the key words in the criminal facts cannot correspond to the feature dictionaries. In this

Fig. 5 Network structure diagram of the fully connected layer and the global average pooling layer

(a) The fully connected layer

(b) The global average pooling layer
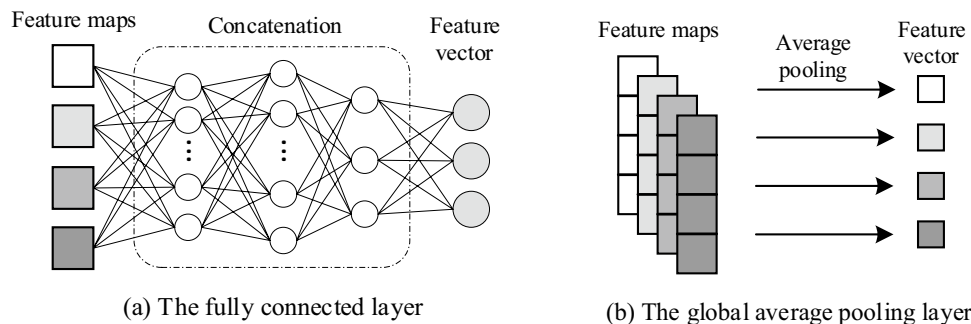
paper, the extracted text features and numerical features are fused to further analyse and accurately model the division of the psychological attribution types of violent crime. The $(3*k+3)$-dimensional feature vector is formed by connecting the $3*k$-dimensional text features with the numerical features directly after feature fusion.

The softmax classifier is used to classify and recognise the fusion features, which is a logistic regression classifier that performs multiple classifications. It maps the output of the category to (0,1), which can obtain the probability of each type, instead of calculating the score of each category. This paper designs a three-classifier. Assuming that the dataset contains m training samples $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$, where the input feature is $x^{(i)} \in \Re^{n+1}$, the three categories are a premeditated type, impulsive type, and a pathology type, which are labelled $y^{(i)} \in \{0, 1, 2\}$. Then, the category probability of each input sample is shown in formula Eq. (4).

$$p(y^{(i)} = j \,|x^{(i)};\theta) = e^{\theta_j^T x^{(i)}} / \sum\nolimits_{k=1}^{k} e^{\theta_j^T x^{(i)}} \tag{4}$$

where θ is the parameter vector of the model, and the category is marked as $j \in \{0, 1, 2\}$. Finally, the probability of each violent crime attribution category can be obtained after feature fusion, and the psychological attribution classification is realised for the three types of violent crime: premeditation, impulsiveness, and pathotype.

4 Experiment and result analysis

4.1 Construction of the experimental dataset

Violent crime mainly emphasise the violent nature of the methods when committing a crime. The Criminal Law of the People's Republic of China stipulates that crimes against the personal and democratic rights of citizens, crimes endangering national security, crimes endangering public security, crimes against property and crimes obstructing the administration of public order are considered violent crimes. The awards, which are related to the above types of criminal cases, are given from the China Judgements Online. Some deleted awards did not disclose the reason for the deletion. The content from "judgement determination" in the award is selected as the experimental data, which could describe the criminal facts about the offender in detail. Private information, such as names in the intercepted criminal facts, is processed. Then, the criminal fact labels are marked by criminal psychology experts corresponding to the violent criminal types: the premeditated type, the impulsive type, and the pathological type. Finally, there were 680 premeditated types, 287 impulsive types and 5 pathological types. The training set and test set are usually divided into 4:1 or 7:3 in
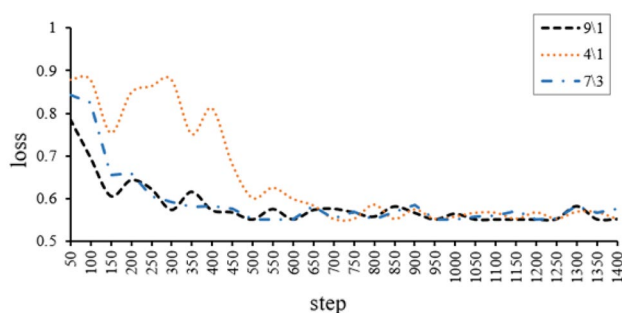

Fig. 6 Convergence of the dataset with different proportions

Table 3 Comparison of tokenizer

Tokenizer	Custom dictionary	Personalised word segmentation	Granularity	Accuracy	Time
Jieba	√	–	Large	0.937	0.654
LTP	√	√	Small	0.960	0.280

deep learning and traditional machine learning. The dataset is small, and the category distribution is uneven, so the experiment was carried out under different division ratios. As shown in Fig. 6, the convergence is faster when the ratio of the training set to the test set is 9:1. The dataset is sorted randomly and consequently divided into the training set and the test set at a ratio of 9:1.

4.2 Experimental environment and experimental settings

The experimental environment in the paper is as follows: the CPU was an Intel Core i7-7500, 1.9 GHz; the GPU is an NVIDIA GeForce MX150 with 4G memory, including a 16G memory configuration, and the operating system is Windows 10. The running environment is Python 3.6 and TensorFlow 1.10.

This paper compares the Jieba tokenizer and LTP using the criminal fact in Fig. 3 when choosing the tokenizer. LTP has a higher recognition accuracy for entity words, which can identify entity information effectively, such as names mentioned in the criminal facts. It takes less time and has a smaller granularity of word segmentation than the Jieba tokenizer. In particular, it is more accurate for LTP word segmentation in the judicial field. The comparison of the performance between Jieba and LTP word segmentation is shown in Table 3.

The criminal facts about offenders are converted into a vector matrix by one-hot encoding after word segmentation by LTP, and the vocabulary contains a total of 10,449 words. Different kinds of deep learning models have been compared

Table 4 Comparison of deep learning models

	CNN	LSTM	Bi-LSTM	Transformer
avg-acc	0.878	0.861	0.847	0.877
Time (s)	118.416	254.406	1,368.182	815.79

experimentally to select the appropriate model for feature extraction. As shown in Table 4, CNN is superior to other models in average accuracy and takes less time. The optimised parameters may reduce the training time of the classifier model and improve its accuracy. Since the length of each criminal fact is different, the maximum interception length is set to 150. The dimension of the word embedding is reduced to 128 by the embedding layer. According to the previous literature (Liu and Xu 2021; Hossain et al. 2021), three sizes of convolution kernels are set with heights measuring at 3, 4, and 5. The width is equal to the dimension of the word embedding, and the number of each convolution kernel is 128. The dropout value is set to 0.5. The training size of each batch is 64, totalling 50 epochs, and the best model parameters are saved every 50 training iterations when the network is trained.

4.3 Comparison and analysis of experiments on the classification model

4.3.1 Performance evaluation of the model

This paper selects accuracy and loss to evaluate the proposed model. To prevent the network from causing overfitting during the learning process, the cross-validation method is adopted. The training set data must be randomly sorted after each epoch of training. The test set is used to verify the training effect, and the best training model is retained after each epoch of training. To verify that the GAP layer replaces the fully connected layer and that the feature fusion could improve the performance of the model, the paper sets up the following experiments for comparison: (1) the numerical features are trained by a neural network for classification; (2) the text features are extracted by the CNN and the softmax classifier is used for classification; (3) the text features are extracted by the improved CNN for classification, which introduces the GAP to replace the fully connected layer. The accuracy and loss change graphs obtained by running the above different models for 50 epochs are given in Figs. 7 and 8. The test results are calculated every 50 steps, totalling 1400 steps. The accuracy of the comparative experiment is similar, maintained at approximately 86%, which is lower than the accuracy of the proposed model. The final loss of the experiment, which uses the traditional CNN for classification, is similar to the experiment, which uses the GAP layer to replace the fully connected layer in the CNN

model. They are higher than the loss of the proposed model. When GAP is introduced, the convergence speed becomes slower, and the convergence speed becomes faster with feature fusion. The accuracy of the proposed model is gradually improved and stabilised at 98%, and the loss of the test set tends to be flat.

The proposed model has the best accuracy performance compared with the above models. However, accuracy is not an effective parameter to evaluate performance if the dataset is unbalanced. Therefore, precision, recall and F1 in the micro and macro scopes of the different violent crime categories are selected to evaluate the model performance. Precision refers to the proportion of classified samples correctly among the samples in which the model is classified into one category; recall refers to the proportion of samples classified correctly by the model in a certain type of positive sample; F1 is the weighted harmonic average of precision and recall; the macro average refers to the arithmetic average of the performance scopes in each category; and the micro average is the arithmetic average of the performance scopes in each instance. The compared results are listed in Table 5, all with two decimal places. The proposed model is better in accuracy and recall, precision, and F1 in the micro average scope. Unfortunately, the model trained by numerical features is better in recall, precision, and F1 in the macro average scope, which is better than the proposed model.

4.3.2 Comparative experiment

To further verify that the proposed model is better in performance, a variety of shallow learning methods and deep learning models are tested for comparison. TF-IDF and
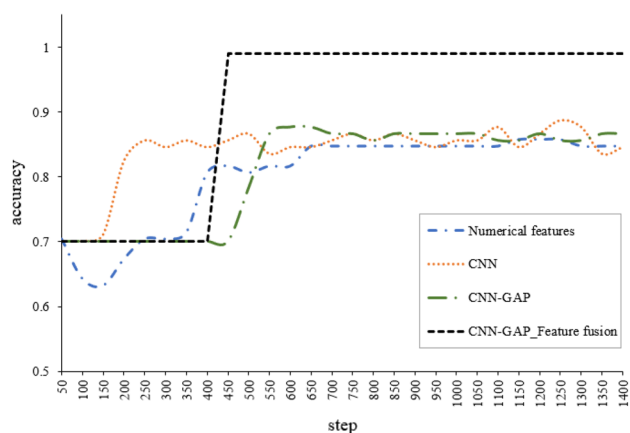

Fig. 7 The accuracy of different models with training steps. The curve of CNN is the experiment that the text features are extracted by CNN; the curve of CNN-GAP is the experiment that the text features are extracted by the improved CNN with the GAP layer for classification; the curve of CNN-GAP_Feature fusion is the proposed model in this paper; the curve of Numerical feature is that the numerical features are used for classification

Fig. 8 The loss of different models with training steps. The legend of the curve is the same as Fig. 7
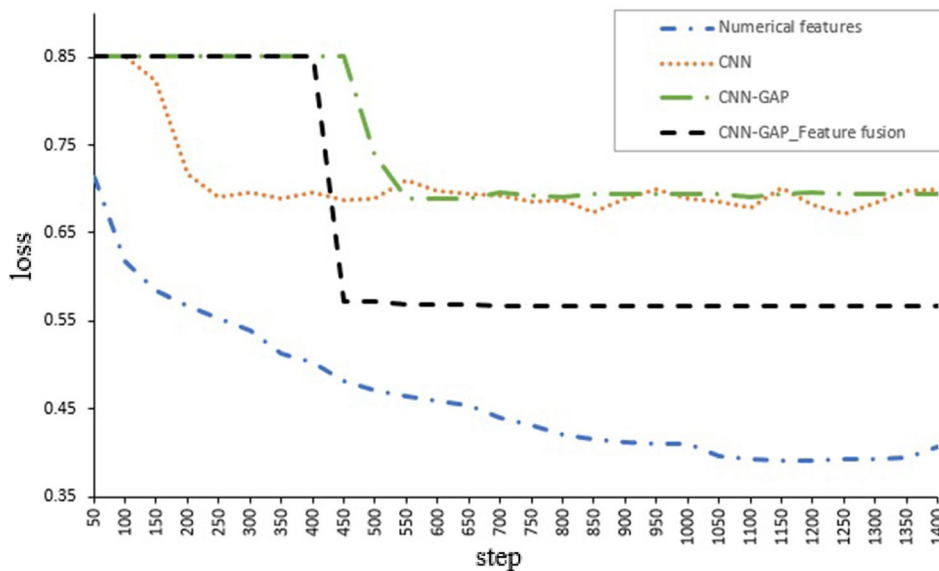

Table 5 Performance comparison of different models. avg-acc is the average accuracy of each model; avg-Micro_* is the micro average of each performance scope; avg-Macro_* is the Macro average of each performance scope. In addition, * is referred to precision, recall and F1

Model	avg-acc	avg-Micro_p	avg-Micro_ recall	avg-Micro_F1	avg-Macro_p	avg-Macro_ recall	avg-Macro_F1
Numerical features	0.87	0.87	0.88	0.88	**0.58**	**0.58**	**0.58**
CNN	0.8781	0.84	0.84	0.84	0.53	0.51	0.51
CNN-GAP	0.81	0.81	0.81	0.81	0.45	0.48	0.46
CNN-GAP_Feature fusion	**0.91**	**0.91**	**0.91**	**0.91**	0.54	0.57	0.55

Best results are given in bold

word count are selected to convert the text data into a vector matrix. TF-IDF is the product of term frequency (TF) and inverse document frequency (IDF), which considers the importance of each word segmentation to the text data as word embedding. Word Count counts the frequency of word segmentation in each piece of data as word embedding. Based on the above two methods, the converted vector matrix is tested by logistic regression, naive Bayes, decision tree, KNN, SVM, random forest and XGBoost methods. The results are shown in Table 6. The proposed model surpasses the machine learning model and deep learning models in accuracy and precision, recall, and F1 in macroscope. However, due to the data imbalance problem, the category with less data could not be classified correctly, which results in a low performance of precision, recall and F1 in macroscope. KNN has the best performance of precision, recall and F1 in the macroscope after converting the word embedding with TF-IDF.

Additionally, to verify the effectiveness of the proposed model, some deep learning models were tested for

comparison in terms of the number of parameters and run time on the testing set. The results are shown in Table 7. The proposed model is slower than CNN. Nevertheless, the proposed model is faster than the other models and is the best in terms of accuracy. Its parameters are lower than those of the CNN because of the introduction of the global average pooling layer.

4.3.3 Result analysis

The softmax classifier is used to obtain the probability of each type of violent crime, which could provide an analysis for the psychological attribution classification, and the distribution interval is [0,1]. The weight distribution can be measured by a hundred-mark system with an interval of [0,100], which is converted from the probability distribution of the three types of violent crime: premeditation, impulsiveness and pathotype. The result of the criminal fact in Fig. 3 is impulsive type, and the percentile weight distribution is converted from the probability distribution of each category

Table 6 Performance comparison of different machine learning and deep learning models

Model	acc	Micro_p	Micro_recall	Micro_F1	Macro_p	Macro_recall	Macro_F1
TF-IDF							
Logistic regression	0.90	0.89	0.90	0.89	0.59	0.57	0.58
Multinomial Naive Bayes	0.88	0.87	0.88	0.88	0.58	0.57	0.57
Bernoulli Naive Bayes	0.89	0.89	0.89	0.89	0.58	0.59	0.58
Gaussian Naive Bayes	0.83	0.83	0.83	0.82	0.56	0.50	0.52
Decision Tree	0.85	0.84	0.84	0.84	0.57	0.52	0.53
KNN	0.87	0.87	0.87	0.87	**0.90**	**0.89**	**0.89**
SVM	0.90	0.90	0.90	0.89	0.60	0.58	0.59
Random Forest	0.89	0.88	0.89	0.88	0.59	0.57	0.58
XGBoost	0.91	0.90	0.91	0.90	0.60	0.59	0.59
Word Count							
Logistic regression	0.90	0.89	0.89	0.89	0.59	0.58	0.59
Multinomial Naive Bayes	0.87	0.89	0.87	0.88	0.58	0.58	0.58
Bernoulli Naive Bayes	0.89	0.90	0.90	0.89	0.58	0.60	0.59
Gaussian Naive Bayes	0.83	0.83	0.83	0.82	0.56	0.50	0.52
Decision Tree	0.88	0.87	0.88	0.87	0.58	0.55	0.57
KNN	0.81	0.79	0.81	0.79	0.52	0.49	0.50
SVM	0.89	0.89	0.89	0.89	0.59	0.57	0.58
Random Forest	0.89	0.88	0.89	0.88	0.58	0.57	0.58
XGBoost	0.91	0.90	0.91	0.90	0.60	0.59	0.59
LSTM	0.89	0.89	0.89	0.89	0.57	0.59	0.58
Bi-LSTM	0.90	0.90	0.90	0.90	0.58	0.58	0.58
Transformer	0.92	0.92	0.92	0.92	0.60	0.61	0.61
CNN	0.91	0.91	0.91	0.91	0.59	0.60	0.59
CNN-GAP_Feature fusion	**0.98**	**0.98**	**0.98**	**0.98**	0.66	0.66	0.66

Best results are given in bold

Table 7 The number of parameters and run time results with different deep learning models

	LSTM	Bi-LSTM	Transformer	CNN	CNN-GAP_feature fusion
Time with GPU (s)	254.41	1368.18	815.75	118.42	199.84
Parameters	41,600	266,240	488,800	288,000	230,850

Table 8 Probability and weight distribution table

	Premeditated type	Impulsive type	Pathological type
Probability distribution	0.263201	0.736259	0.000539
Weight distribution	26.32	73.62	0.05

obtained by the softmax classifier shown in Table 8. The radar chart based on the probability distribution of the three categories is shown in Fig. 9. The offender's impulsive type score is 73.62, accounting for a high proportion, which indicates that a person shows impulsive behaviour when committing a crime and that their emotion cannot be controlled well. The score of the premeditated type is 26.32, indicating that the person is driven by a certain goal and that their

psychological state is free of guilt when committing a crime. The case in Fig. 3 shows that the criminal had a certain psychological impact due to romantic tussles for a long time and finally had impulsive violent behaviour under the influence of specific events. His psychology was further adjusted with the help of others, and he confessed to his criminal behaviour. Based on the interpretation of criminal psychology, we consider that the offender has been in a state of anxiety and
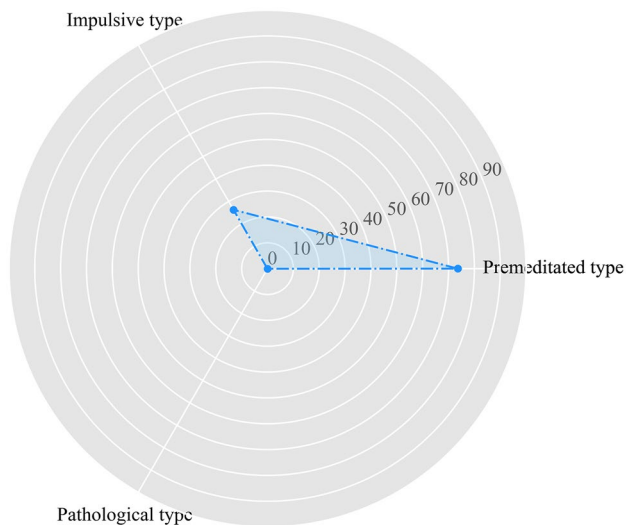
Fig. 9 Weight distribution diagram

fear for a long time, and his emotion changes substantially from fear to anger, which finally results in violence. As a consequence, relevant departments can formulate appropriate corrective education measures that focus on correcting the offender's impulsive emotion based on the weight distribution of the different types.

5 Conclusion

This paper introduces natural language processing technology to criminal psychology and proposes a model that uses the improved CNN and feature fusion for the psychological attribution classification of violent crime. Based on the analysis of the criminal facts, criminal psychology experts construct the dataset and criminal feature dictionaries, which are divided into a premeditated type, an impulsive type, and a pathology type. Then, the numerical features are extracted by the word count based on predefined dictionaries. The improved CNN, which replaces the fully connected layer with the GAP layer, is used for text feature extraction. Finally, the numerical features and the text features are fused for attribution classification. The proposed model proves to be better in accuracy and precision, recall, and F1 in the macroscope as well as the microscope compared with other traditional classification methods. It can also classify criminal facts about offenders for attribution analysis effectively.

However, the proposed model has poor performance in precision, recall and F1 of the microscope due to the data imbalance problem in the dataset. In the future, it is necessary to further improve the classification efficiency of the minority categories in the dataset. The automatic expansion of the predefined feature dictionaries should be further studied because the predefined dictionaries are obtained manually.

References

Aydoğan M, Karci A (2020) Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification. Physica A 541:123288. https://doi.org/10.1016/j.physa.2019.123288

Barrat ES, Kent TA, Bryant G, Felthous AR (1991) A controlled trail of phenytoin in impulsive aggression. J Clin Psychopharmacol 11(6):338–389. https://doi.org/10.1097/00004714-19911 2000-00017

Chen W-Z, Qin Y-B, Huang R-Z et al (2019) Legal text prediction method based on criminal behavior sequence. Comput Eng Appl 55(22):251-255+270

Chen H, Huang B, Zhu Y, Yu L, Yu Y (2020) Short text emotion classification method combining LDA and Self-Attention. Comput Eng Appl 56(18):171–176

Chen L, Cao J, Zhu G-X, Wang Y-Q, Liang W-C (2021) A multi-task learning approach for improving travel recommendation with keywords generation. Knowl-Based Syst 233:107521. https://doi.org/10.1016/j.knosys.2021.107521

Cichosz P (2018) A case study in text mining of discussion forum posts: classification with bag of words and global vectors. Int J Appl Math Comput Sci 28(4):787–801. https://doi.org/10.2478/amcs-2018-0060

Copas J, Marshall P (1998) The offender group reconviction scale: a statistical reconviction score for use by probation officers. J Roy Stat Soc Ser C (Appl Stat) 47(01):159–171. https://doi.org/10.1111/1467-9876.00104

Das P, Das AK, Nayak J (2020) Feature selection generating directed rough-spanning tree for crime pattern analysis. Neural Comput Appl 32:7623–7639. https://doi.org/10.1007/s00521-018-3880-8

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dickson SJ (2014) The psychopathy checklist youth version (PCL:YV): an investigation into its inter-rater reliability. Dissertation, University of Glasgow

Ghosh L, Saha S, Konar A (2020) Bi-directional Long Short-Term Memory model to analyze psychological effects on gamers. Appl Soft Comput 95:106573. https://doi.org/10.1016/j.asoc.2020.106573

Hassani H, Xu H, Silva ES, Ghodsi M (2016) A review of data mining applications in crime. Stat Anal Data Min. https://doi.org/10.1002/sam.11312

Hossain MR, Hoque MM, Dewan MAA, Siddique N, Islam MN, Sarker IH (2021) Authorship classification in a resource constraint language using convolutional neural networks. IEEE Access 9:100319–100338. https://doi.org/10.1109/ACCESS.2021.3095967

Ibrahim M, Shaawat A, Torki M (2021) Covariance pooling layer for text classification. Procedia Comput Sci 289:61–66. https://doi.org/10.1016/j.procs.2021.05.070

Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. Eur Conf Mach Learn 1398:137–142. https://doi.org/10.1007/BFb0026683

Johnson DE, Oles FJ, Zhang T, Goetz T (2002) A decision-tree-based symbolic rule induction system for text categorization. IBM Syst J 41(3):428–437. https://doi.org/10.1147/sj.413.0428

Jones EE, Harris VA (1967) The attribution of attitudes. J Exp Soc Psychol 3(1):1–24. https://doi.org/10.1016/0022-1031(67)90034-0

Keijsers M, Kazmi H, Eyssel F et al (2021) Teaching robots a lesson: determinants of robot punishment. Int J Soc Robot 13:41–54. https://doi.org/10.1007/s12369-019-00608-w

Kim Y (2014) Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751. https://doi.org/10.3115/v1/D14-1181

Kunst M, Koster NN (2017) Psychological distress following crime victimization: an exploratory study from an agency perspective. Stress Health 33(4):405–414. https://doi.org/10.1002/smi.2725

Lei T, Zhang Y, Wang S-I, Artzi Y (2018) Simple Recurrent Units for Highly Parallelizable Recurrence. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 4470–4481. https://doi.org/10.18653/v1/D18-1477

Levenson RW (2014) The autonomic nervous system and emotion. Emot Rev 6(02):100–112. https://doi.org/10.1177/1754073913512003

Li Q, Peng H, Li J, Xia C, Yang R, Sun L, Yu P-S (2020a) A survey on text classification: from shallow to deep learning. arXiv:2008.00364 (**preprint**)

Li Q, Li P-F, Mao K-Z, Lo EY-M (2020b) Improving convolutional neural network for text classification by recursive data pruning. Neurocomputing 414:143–152. https://doi.org/10.1016/j.neucom.2020.07.049

Liang W-C, Wang Y-Q, Tao H-C, Cao J (2021) Towards hour-level crime prediction: a neural attentive framework with spatial–temporal-categorical fusion. Neurocomputing. https://doi.org/10.1016/j.neucom.2021.11.052

Lin Z-Y, Jung J, Goel S, Skeem J (2020) The limits of human predictions of recidivism. Sci Adv 6(07):0652. https://doi.org/10.1126/sciadv.aaz0652

Liu C, Xu X-L (2021) AMFF: a new attention-based multi-feature fusion method for intention recognition. Knowl-Based Syst 233:107525. https://doi.org/10.1016/j.knosys.2021.107525

Liu P-F, Qiu X-P, Huang X-J (2016) Recurrent Neural Network for Text Classification with Multi-Task Learning. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 2873–2879. https://doi.org/10.5555/3060832.3061023

Liu Y, Wang N, Li B-H, Luo Q (2017) Application of fuzzy grammar method in crime text classification. Comput Eng Design 38(7):1965–1971

Luo B-F, Feng Y-S, Xu J-B, Zhang X, Zhao D-Y (2017) Learning to Predict Charges for Criminal Cases with Legal Basis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp 2727–2736. https://doi.org/10.18653/v1/D17-1289

Ma A, Song Y-Z (2021) How does emotional computing technology promote the development of crime risk assessment tools? J Psychol Sci 44(01):52–59

Ma A, Sun X, Song Y-Z (2020) Theory and practice of intelligent assessment for danger of crime. China Legal Publishing House, Beijing

Maheshwari A (2018) Report on text classification using CNN, RNN & HAN. Medium. https://medium.com/jatana/report-on-text-classification-using-cnn-rnn-han-f0e887214d5f. Accessed 15 Aug 2021

Ni YZ, Barzman D, Bachtel A, Griffey M, Osborn A, Sorter M (2020) Finding warning markers: leveraging natural language processing and machine learning technologies to detect risk of school violence. Int J Med Informatics 139:104137. https://doi.org/10.1016/j.ijmedinf.2020.104137

Pereira-Kohatsu JC, Quijano-Sánchez L, Liberatore F, Camacho-Collados M (2019) Detecting and monitoring hate speech in Twitter. Sensors 19(21):4654. https://doi.org/10.3390/s19214654

Schneider KM (2004) A new feature selection score for Multinomial Naïve Bayes text classification based on KL-Divergence. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions, pp 24–es. https://doi.org/10.3115/1219044.1219068

Vaswani A, Shazeer N, Parmar N et al (2017) Attention Is All You Need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, pp. 6000–6010. https://doi.org/10.48550/arXiv.1706.03762

Wang D-Z, Mao K-Z (2019) Task-generic semantic convolutional neural network for web text-aided image classification. Neurocomputing 329:103–115. https://doi.org/10.1016/j.neucom.2018.09.042

Wang P, Li M-Y, Li X-D, Zhou H-S, Hou J-R (2021) A hybrid approach to classifying Wikipedia article quality flaws with feature fusion framework. Expert Syst Appl 181:115089. https://doi.org/10.1016/j.eswa.2021.115089

Wu Y-L (2021) The impact of criminal psychology trend prediction based on deep learning algorithm and three-dimensional convolutional neural network. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-021-03455-8

Xiao X-Z, Du X-Y (2020) Design and implementation of community correction system based on mobile internet. Inform Technol 44(01):7-10+14

Xiao Z-W, Xu X, Xing H-L, Qu R, Song F, Zhao B (2021a) RNTS: Robust Neural Temporal Search for Time Series Classification. In: 2021a International Joint Conference on Neural Networks (IJCNN), pp. 1–8.

Xiao Z-W, Xu X, Xing H-L, Song F-H, Wang X-H, Zhao B-W (2021b) A federated learning system with enhanced feature extraction for human activity recognition. Knowl-Based Syst 229:107338. https://doi.org/10.1016/j.knosys.2021.107338

Xiao Z-W, Xu X, Xing H-L, Luo S-X, Dai P-L, Zhan D-W (2021c) RTFN: a robust temporal feature network for time series classification. Inf Sci 571:65–86. https://doi.org/10.1016/j.ins.2021.04.053

Yang X (2020) A study of automatically classifying text-described criminal behaviours based on a hybrid learning model. Dissertation, Guangxi Normal University

Yang Z-C, Yang D-Y, Dyer C, He X-D, Smola A, Hovy E (2016) Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480–1489. https://doi.org/10.18653/v1/N16-1174

Yilmaz S, Toklu S (2020) A deep learning analysis on question classification task using Word2vec representations. Neural Comput Appl 32:2909–2928. https://doi.org/10.1007/s00521-020-04725-w

Zhang H-W (2020) Application of risk /needs assessment in preventing juveniles' reoffending. Jinan J (Philosophy and Social Sciences) 40(02):68–75

Zhang Y-B, Chen Y-Q, Wang Y-S, Li H-Y (2021) Dual-attribution model of trust development between contracting parties in construction projects: psychological perspective. J Constr Eng Manag 147(11):04021161