Campus bullying detection based on motion recognition and speech emotion recognition

Zhao Zihan¹, Zhao Zhanfeng²

¹ College of Information Science and Engineering, Harbin Institute of Technology at Weihai, Weihai, Shandong 264200, China
² College of Information Science and Engineering, Harbin Institute of Technology at Weihai, Weihai, Shandong 264200, China

Abstract. In many areas, school bullying incidents often occur. Such incidents not only damage the learning environment of the school but also cause great harm to the students who are bullied. The occurrence of the campus bullying incident is generally hidden and not easy to be discovered by teachers or parents in time. Aiming at the problem that the campus bullying incident needs to be discovered in time, a campus bullying detection method based on motion recognition and speech emotion recognition is proposed. The electronic device such as smart phone and smart watch worn by people are used to collect the data of human motion and voice in real time. Thus, it is possible to detect in time whether the wearer is being bullied. In this paper, six characteristics of human motion data and MFCC features of speech data are extracted. Then the PCA algorithm is used to reduce the dimension of the obtained feature matrix. Finally, the KNN algorithm is used to form the motion and speech recognition network. After cross-validation, the average recognition rate of the bullying movement and speech emotion of the system in this paper are 77.8% and 81.4%, respectively. The experimental results show that the campus bullying detection method based on KNN algorithm has obtained a good recognition rate.

1. Introduction
Campus bullying behaviors is usually divided into physical bullying, speech bullying, social bullying, and cyber bullying. Among them, physical bullying and speech bullying cause the most serious damage to the people who are bullied. The domestic scholars’ research on the phenomenon of campus bullying mostly evaluates the psychological and physiological damage caused by the bullying behaviors of the campus and explores the reasons for the formation of the campus bullying phenomenon. Domestic scholars mainly study the methods that can stop the occurrence of campus bullying events from the psychological and social aspects [1]. In general, the research of domestic scholars in the field of detecting campus bullying incidents started late and were not mature enough. In paper [2], the acceleration sensor is used to collect the acceleration signals of the human body in the X, Y and Z directions under different actions. The feature extraction is performed by blind selection method and then the PCA dimensionality reduction processing is performed. The improved SVM is used to identify and classify. Better recognition, but longer recognition time. In paper [3], a behavioral pattern algorithm based on acceleration time domain features is used, and the acceleration time domain feature is used as the unique feature quantity, thus reducing the recognition time and
improving the real-time recognition of the human body behavior. However, this system recognizes fewer types of actions and the recognition rate is not very high. Literature [4] proposed a multi-level SVM sentiment classification algorithm based on principal component analysis, which classifies the resulting speech into corresponding categories, and realizes the classification of seven speech emotions, but the average recognition rate is not high, and has a certain rate of confusion.

In order to detect the bullying events happening at campus in time and accurately, this paper proposes a campus bullying detection method based on motion recognition and speech emotion recognition, which can acquire and process the wearer's movements and voice data in real time through smart phones or watches with a high recognition speed. The six characteristics of human motion data and the MFCC features of speech data are extracted, and the feature matrix is reduced by using PCA (principal component analysis) algorithm. Then, KNN algorithm is used to form the recognition model of motion recognition and speech emotion recognition. The detection method proposed in this paper has good real-time and recognition rate by analyzing the wearer's real-time motion data and voice data.

2. Integrated detection system based on motion recognition and speech emotion recognition

2.1. PCA-based feature matrix dimensionality reduction

In this paper, the eight kinds of gyroscope triaxial angular acceleration data and the motion sensor triaxial acceleration data are extracted, and the maximum, minimum, variance, mean, median and sum of these six types of data are calculated as the feature vector of the action. In this paper, the covariance matrix of the data matrix is processed by the PCA algorithm, and then the eigenvalues and eigenvectors of the covariance matrix are obtained by calculation, and the eigenvectors corresponding to the N features with the largest eigenvalue (that is, the largest variance) are selected. Figure 1 shows the main steps of the PCA algorithm.

![Figure 1. The main steps of the PCA algorithm.](image-url)
Firstly, according to the eigenvalues and eigenvectors of the covariance of the data matrix, the largest \( N \) eigenvalues are obtained. The matrix composed of the corresponding feature vectors is used as a compression matrix, and the previous feature matrix is multiplied by the compression matrix, so that the original data feature vector is transformed into a new feature vector, and the data feature is reduced in dimension.

### 2.2. Motion recognition based on KNN algorithm

In this paper, the KNN algorithm is used to find the \( k \) records closest to the new data from the training data set, and then the category to which the data belongs is determined according to the main types of these records. The research team fixed the motion sensor at the waist of the tester and let the tester do the jump, fall, play, walk, and run, which are non-bullying actions. And then let another tester push the tester and hit the shoulders, which are bullying actions. The authors collected data from gyroscopes and motion sensors over multiple time periods. In this paper, the KNN classifier and the dimensionality-reduced feature matrix of the test motion data are used to generate the recognition model, and then the model and the dimension-reducing feature matrix of the test motion data are input into the KNN classifier, and the output is the classification result of the test sample. In this paper, when the KNN classifier is used, the action data set is divided into the bullying and the non-bullying. The KNN algorithm proposed in this paper takes the categories of the three closest points of the test samples to classify the test samples. Figure 2 shows the flow of the KNN algorithm.

![Figure 2. The flow chart of KNN.](image)

### 2.3. Speech emotion recognition based on MFCC feature and KNN algorithm

MFCC (Mel-frequency cepstral coefficients) is a kind of coefficient that simulates the auditory characteristics of human ear, and is one of the important characteristic parameters describing speech. It is widely used in speech recognition, and extracting MFCC features can make the system have good recognition effect. The research team collected four kinds of emotional voice data of three sixth-grade primary school students. After endpoint detection processing, the MFCC features of these data were extracted, and then the PCA was used to reduce the dimension, and the KNN algorithm was used to
classify the voice data used for the test. The results of the identification were analyzed after cross-validation. Figure 3 shows the process of extracting MFCC features.

![Diagram of MFCC feature extraction process]

**Figure 3. Steps to extract MFCC features.**

### 3. Simulation

#### 3.1. Simulation condition

In this paper, MATLAB is used as the simulation platform, and the PCA algorithm is used to reduce the dimension of the feature matrix. At the same time, the KNN classification algorithm is used to establish the recognition model. When collecting the raw data needed for motion recognition, the research team invited 12 primary school students as testers, and let several primary school students wear motion sensors at their waists and jump, play, run, walk, and fall. Then let other testers apply the bullying action to the tester wearing the motion sensor. The research team collected data of eight kinds of actions. In this paper, 180 action segments were used as training and test samples. In terms of speech emotion recognition, the research team collected four kinds of voice data of three primary school students. There are 160 voice clips of crying, happy, bullying and normal voice data. In this paper, the voices of crying and being bullied are divided into bullying voices, and the voices that are happy and normal are non-bullying voices. The system was cross-validated with these 160 samples, and then calculated the average value, which was taken as the recognition rate of the system. At the same time, the recognition rate of the recognition model for different speech was calculated.

#### 3.2. Motion recognition simulation results

After cross-validation, the average recognition rate of the motion recognition system is 77.8%. Table 1 shows the recognition rate of the algorithm for each cross-validation. And table 2 shows the average recognition rate of different actions of the system.
Table 1. Cross-validation results of motion recognition

|                | the first time | the second time | the third time | the fourth time | the fifth time |
|----------------|---------------|----------------|---------------|----------------|---------------|
| Recognition rate | 83.3%         | 61.1%          | 88.9%         | 83.3%          | 66.7%         |

|                | the sixth time | the seventh time | the eighth time | the ninth time | the tenth time |
|----------------|---------------|-----------------|---------------|---------------|---------------|
| Recognition rate | 77.8%         | 88.9%           | 83.3%         | 72.2%         | 72.2%         |

Table 2. Average recognition rate of eight actions

| Being pushed down | Run | Walk | Play | Being beaten | Being pushed | Jump | Fall down |
|-------------------|-----|------|------|--------------|--------------|------|----------|
| Recognition rate  | 90.0% | 80.0% | 50.0% | 30.0% | 90.0% | 91.7% | 60.0% | 40.0% |

4. Speech emotion recognition simulation results

In terms of speech emotion recognition, the average recognition rate of the speech recognition system based on KNN classifier is 81.4% after cross-validation. Table 3 gives the results of each cross-validation of the speech emotion recognition system. And table 4 shows the average recognition rate of the system for four different speech emotions.

Table 3. Cross-validation results of the speech emotion recognition

|                | the first time | the second time | the third time | the fourth time | the fifth time |
|----------------|---------------|----------------|---------------|----------------|---------------|
| Recognition rate | 81.3%         | 87.5%          | 87.5%         | 68.8%          | 87.5%         |

|                | the sixth time | the seventh time | the eighth time | the ninth time | the tenth time |
|----------------|---------------|-----------------|---------------|---------------|---------------|
| Recognition rate | 76.7%         | 93.8%           | 83.3%         | 68.8%         | 81.3%         |

Table 4. Average recognition rate of four speech emotions

| Normal situation | Happy | Being bullied | Cry |
|------------------|-------|---------------|-----|
| Recognition rate | 88.0% | 66.2%         | 85.0% | 86.0% |

5. Conclusion

In this paper, the PCA-based dimensionality reduction method is used in the motion recognition and speech emotion recognition to reduce the dimension of the motion and speech data. Then the KNN classification algorithm is used to process the training samples to form the recognition network. The experimental results show that the method has a good recognition effect on most bullying movements, but the average recognition rate of fall down, play and walk are low. And the recognition rate of the emotion of ‘happy’ in the speech emotion recognition model needs to be improved. The model in this paper has a low recognition rate for some actions and speech, which is related to the less samples of related types of speech emotion and motion. So the number of training samples can be increased by collecting more various types of motion data and voice emotion data, further improving the recognition rate of such method.

Acknowledgement:

Thanks to the support of Harbin Institute of Technology and the Fund of Weihai.

References:

[1] Wang Chao, Ma Yinghua 2005 Research progress on campus violence interventions Chinese school health vol 26 pp 979-980
[2] Wang Jian, Chen Yi, Deng Shuai 2016 Motion recognition method based on improved SVM classifier. Journal of Chongqing University vol 39 pp 12-17
[3] Liu Yu, Jiang Hongyi, Wang Shiliang, Wang Yibing, Chen Yanping 2015 Real-time human behavior pattern recognition based on acceleration time domain features *Journal of Shanghai Jiaotong University* vol 49 pp 169-172

[4] Ren Hao, Ye Liang, Li Yue, Sha Xuejun 2018 Speech emotion recognition algorithm based on multi-level SVM classification. *Application Research of Computers* vol 34 pp 1682-1684

[5] Zhang Xiong, Liu Rong, Liu Ming 2018 Research on Speech Emotion Recognition Based on Convolution Feature Extraction and Fusion. *Electronic measurement technology* vol 41 pp 138-142

[6] Zhang Xueying, Zhang Ting, Sun Ying, Zhang Wei 2018 Concatenated Classification Emotional Speech Recognition Based on PAD Model *Journal of Taiyuan University of Technology* vol 49 pp 731-735

[7] Lukas Köping, Kimiaki Shirahama, Marcin Grzegorzek 2018 A general framework for sensor-based human activity recognition. *Computers in Biology and Medicine* vol 95 pp 248-260

[8] Lei Gao, A.K.Bourke, John Nelson 2014 Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering and Physics* vol 36 pp 779-785

[9] Sun Xiaobing, Liu Haiming 2015 Rational Cognition Campus Bullying: From Campus Violence to Campus Bullying *Educational theory and practice* vol 35 pp 26-29

[10] Akram Bayat, Marc Pomplun, Duc A. Tran 2014 A Study on Human Activity Recognition Using Accelerometer Data from Smartphones *Procedia Computer Science* vol 34 pp 450-457

[11] John J. Guiry, Pepijn van de Ven, John Nelson, Lisanne Warmerdam, Heleen Riper 2014 Activity recognition with smartphone support *Medical Engineering and Physics* vol 36 pp 670-675

[12] F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou, Y. Amirat 2013 Joint segmentation of multivariate time series with hidden process regression for human activity recognition *Neurocomputing* vol 120 pp 633-644

[13] Ronald Poppe 2009 A survey on vision-based human action recognition *Image and Vision Computing* vol 28 pp 976-990

[14] Ma Fangyuan, Wang Xiaoying 2011 The Causes of Juvenile Campus Violence Crime and the Construction of Its Prevention Mechanism *Journal of Jilin Institute of Chemical Technology* vol 28 pp 113-115