A Physical and Verbal Bullying Detecting Algorithm Based on K-NN for School Bullying Prevention

Shangbin Gao[1,2] and Liang Ye[2(✉)]

[1] Harbin Institute of Technology, Weihai, China
1270565172@qq.com
[2] Harbin Institute of Technology, Harbin, China
yeliang@hit.edu.cn

Abstract. School bullying is a common social problem around the world which affects teenagers, and physical violence is considered to be the most harmful while verbal bullying is the most frequent. This paper proposes an automatic physical and verbal bullying detecting method in the field of artificial intelligence. Dozens of features were extracted from acceleration and gyro data to train the physical bullying recognition while the mean value of each frame of samples was calculated for verbal bullying detection. The authors used the *k*-NN algorithm as the classifier. The final test accuracies of physical and verbal bullying detecting were 70.4% and 78.0%, respectively, indicating that activity recognition and speech emotion recognition can be used for detecting bullying behaviors as an artificial intelligence technique, and speech emotion recognition appeared to be better than activity recognition.

Keywords: Physical bullying · Verbal bullying · *k*-NN · Artificial Intelligence · School bullying

1 Introduction

School bullying is a common social problem among teenagers. It affects the victims both mentally and physically and is considered as one of the main reasons for depression, dropping out of school and adolescent suicide [1, 2]. In view of this, anti-bullying is a significant as well as timeless topic. New approaches for preventing school bullying become available as classification methods develop. School bullying can take various forms, such as cursing, physical violence, and so on. Physical violence is considered to be the most harmful to teenagers while the verbal bullying is the most frequent. Consequently, this paper will focus on detecting both physical and verbal bullying.

Following the popularity of smartphones, several anti-bullying applications can be found on Internet. Nevertheless, the victim or a witness needs to take out a smartphone, run the application and press a button to send an alarm message. To photograph the event, they must hold the camera toward the bullies. However, this is not convenient for the victim, especially when bullied physically. Therefore, Alasaarela [3] and his

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 150–157, 2019.
https://doi.org/10.1007/978-3-030-22971-9_13

team proposed some algorithms to detect physical bullying by using a 3D accelerometer and a 3D gyroscope. Once activated, it could run in the background and detect physical bullying automatically.

There are also some experiments focusing on emotion recognition and daily-life activity recognition. For instance, Ferdinando had found a method that was able to recognize emotions indicative of bullying incidents with ECG and Heart Rate Variability (HRV), reaching an average accuracy of 47.69% for arousal and 42.55% for valence [4]. Zalluhoglu and Ikizler-Cinbis had reached the accuracy of 72.4% [5]. Moreover, Seyed Ali and Rokni got an accuracy of 82.2% in activity recognition [6].

This paper applied an algorithm for detecting physical and verbal bullying—the *k*-Nearest Neighbor (*k*-NN) algorithm. It is a non-parametric method used for classification and regression in pattern recognition. In both cases, the input consists of *k* closest training examples in the feature space. And the output depends on whether *k*-NN is used for classification or regression.

The remainder of this paper is constructed as follows: Sect. 2 shows the process of activity recognition; Sect. 3 shows the process of speech emotion recognition; Sect. 4 gives out the classification results by simulation; and Sect. 5 draws a conclusion.

2 Activity Recognition

Nowadays activity recognition with various sensors is a very popular topic. For the convenience of life, many researchers try to use this technique to assist people with their daily life, e.g. elder care [1, 2], smart home [7, 8], and other artificial intelligence (AI) environments. As a matter of fact, there are quite a number of classifiers as well as feature selection methods for activity recognition. However, it is nearly impossible to draw a conclusion which activity recognition algorithm is absolutely the best, considering the differences of databases used by different researches, as well as the difference sources of activities provided by different actors/actresses. This paper focuses on the activity recognition by accelerometers and gyroscopes.

The following is the concrete procedures in activity recognition.

The data were collected by Ye, *et al.* in Finland. A movement sensor (integrated accelerometer and gyroscope) was fixed on the subjects' waist in order to collect 3D accelerations and 3D gyros at 50 Hz. All experiments were video-recorded for synchronization. Each activity was repeated several times by different actors or actresses.

The authors had finally gathered more than 250 fragments of movement in total and then selected 221 of them as samples. The data were the acceleration in x-axis, y-axis and z-axis, and the gyro in x-axis, y-axis and z-axis. The y-axis was a vertical vector, while the x-axis and the z-axis were horizontal vectors. Classified by movement, it can be divided into 9 kinds, i.e. falling down, jumping, playing, pushing, pushing down, running, shoulder-hit, standing and walking.

The authors firstly loaded these data from Excel to documents ending with ".mat". These data were equally and randomly divided into 2 groups, namely the training group and the testing group (if one certain kind of movement cannot be divided equally, let the training set contain one more sample than the testing group) (Table 1).

Table 1. The exact number of each movement in the training group and the testing group.

	Training group	Testing group
Falling down	8	7
Jumping	6	6
Playing	16	15
Pushing	30	29
Pushing down	22	21
Running	5	4
Shoulder-hit	15	15
Standing	6	5
Walking	6	5
Sum	114	107

Then the authors calculated the maximum, minimum, median absolute deviation, mean value, variance, sum and energy of each movement in both two groups. Each feature value concluded 6 sections, e.g. the maximum concluded the maximum acceleration in x-axis, y-axis, and z-axis, and the maximum gyro in x-axis, y-axis, and z-axis. Table 2 shows the extracted features.

Table 2. All the 42 features that should be calculated.

Maximum from accelerations in *x*-axis	Maximum from accelerations in y-axis	Maximum from accelerations in z-axis	Maximum from gyros in x-axis	Maximum from gyros in y-axis	Maximum from gyros in z-axis
Minimum from accelerations in x-axis	Minimum from accelerations in y-axis	Minimum from accelerations in z-axis	Minimum from gyros in x-axis	Minimum from gyros in y-axis	Minimum from gyros in z-axis
Median absolute deviation from accelerations in x-axis	Median absolute deviation from accelerations in y-axis	Median absolute deviation from accelerations in z-axis	Median absolute deviation from gyros in x-axis	Median absolute deviation from gyros in y-axis	Median absolute deviation from gyros in z-axis
Mean value from accelerations in x-axis	Mean value from accelerations in y-axis	Mean value from accelerations in z-axis	Mean value from gyros in x-axis	Mean value from gyros in y-axis	Mean value from gyros in z-axis
Variance from accelerations in x-axis	Variance from accelerations in y-axis	Variance from accelerations in z-axis	Variance from gyros in x-axis	Variance from gyros in y-axis	Variance from gyros in z-axis

(*continued*)

Table 2. (*continued*)

Sum from accelerations in x-axis	Sum from accelerations in y-axis	Sum from accelerations in z-axis	Sum from gyros in x-axis	Sum from gyros in y-axis	Sum from gyros in z-axis
Energy from accelerations in x-axis	Energy from accelerations in y-axis	Energy from accelerations in z-axis	Energy from gyros in x-axis	Energy from gyros in y-axis	Energy from gyros in z-axis

All values from the maximum to the energy were extracted and finally formed a 1 row by 42 column matrix (Table 3).

Table 3. The features extracted in each document

Maximum (*1* row by 6 column)	Minimum (*1* row by 6 column)	Median absolute deviation (*1* row by 6 column)	Mean value (*1* row by 6 column)	Variance (*1* row by 6 column)	Sum (*1* row by 6 column)	Energy (*1* row by 6 column)

Next, the authors used the same procedure to calculate the values mentioned above in others documents.

Add up all 1 row by 42 column matrices in each kind of movement in each group. For example, in the training group, the number of documents of falling down is 8, so add up that 81 row by 42 column matrices to form an 8 row by 42 column matrix. Table 4 shows the exact number of row and column of all movements.

Table 4. The number of row and column of all movements

Movement	Training group		Testing group	
	Number of rows	Number of columns	Number of rows	Number of columns
Falling down	8	42	7	42
Jumping	6	42	6	42
Playing	16	42	15	42
Pushing	30	42	29	42
Pushing down	22	42	21	42
Running	5	42	4	42
Shoulder-hit	15	42	15	42
Standing	6	42	5	42
Walking	6	42	5	42

Add all the 9 matrices in the training group into one sum-up matrix, as well as all the 9 matrices in testing group. Then, a 114 row by 42 column matrix $\mathbf{A}_{144\times42}$ and a 107 row by 42 column matrix $\mathbf{A}_{107\times42}$ were built.

Two transition matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ were produced by using PCA (Principal Component Analysis).

The matrix after dimensionality reduction was created by multiplying the sum-up matrix and the transition matrix, written as $\mathbf{M}_1$ and $\mathbf{M}_2$.

$$\mathbf{M}_1 = \mathbf{A}_{144\times 42} \times \mathbf{T}_1 \tag{1}$$

$$\mathbf{M}_2 = \mathbf{A}_{107\times 42} \times \mathbf{T}_2 \tag{2}$$

The authors then constructed two n row by 2 column matrices which referred to the two sum-up matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively. Bullying behaviors (pushing, pushing down and shoulder-hit) matched matrix [1 0] while non-bullying behaviors (falling down, jumping, playing, running, shaking, standing and walking) matched matrix [0 1]. Two 0-1 matrices were got (written as $\mathbf{Z}_1$ and $\mathbf{Z}_2$, respectively).

A k-NN model NET was created with input dimension NIN, output dimension NOUT and k neighbours where k is a user-defined constant.

Took a matrix $\mathbf{X}$ of input vectors (one vector per row) and uses the k-NN rule on the training data contained in NET to produce a matrix $\mathbf{Y}$ of outputs and a matrix $\mathbf{L}$ of classification labels.

Two-fold cross validation was used to calculate the accuracy, i.e., in the first round, the training group was used to train the classifier and the testing group was used to test the classifier, but in the second round, the testing group was used to train the classifier and the training group was used to test the classifier. The accuracies of the two rounds were recorded as Accuracy_1 and Accuracy_2, respectively. Then the final accuracy was calculated as

$$\text{Accuracy} = \frac{1}{2}\left(\text{Accuracy}_1 + \text{Accuracy}_2\right) \tag{3}$$

Change the value of k, and repeat the steps above. The simulation results are given in Sect. 4.

Compared with Seyed Ali's team [6] and other teams [9, 10], the accuracies achieved were not better, so the authors decided to add 2 kinds of features into the features aggregation. The 2 features were the sum vectors of the absolute value of acceleration and the angle of gyroscope. The simulation results are also given in Sect. 4 as a comparison.

3 Speech Emotion Recognition

Nowadays, speech emotion recognition is a very popular topic. In the field of artificial intelligence, a system that is able to recognize the voice of a consumer and to control a robot's movements with verbal instructions has been developed [11]. However, most of the speech emotion recognition has not been used in detecting bullying behaviors. Considering that it is significant to prevent bullying as soon as possible, the authors also used k-NN to detect verbal bullying.

The following is the concrete procedure in speech emotion recognition.

(1) The data were collected by Ye in Finland, too. Several pupils provided the sound recording in 4 kinds, namely, bullying, cry, happy and normal. The pupils were asked to pretend that they were under these circumstances so that the data were divided by class in advance.
(2) The authors finally gathered hundreds of seconds of sound recordings, and cut them equally into hundreds of parts. These data were divided into 4 classes—bullying, cry, happy and normal. Bullying and cry were classified into bullying while happy and normal were classified into non-bullying for convenience. These data were in form of mp4.
(3) These two groups of data were equally and randomly divided into 2 groups (the training group and the testing group), and in each of them there were 80 s bullying recordings and 81 s non-bullying recordings.
(4) The authors then extracted the feature parameters from the selected voice sequences by using the MFCC (Mel Frequency Cepstral Coefficients).

$$\mathrm{Mel}(f) = 2595 \times \lg\left(1 + \frac{f}{700}\right) \tag{4}$$

12 MFCC parameters, 12 first-order differential MFCCs and 12 second-order differential MFCCs were extracted and classified by k-NN. The classification results are also given in Sect. 4.

4 Classification Results

Table 5 shows the first activity recognition accuracy (k = 1). Table 6 shows the average accuracy of physical bullying detection (k = 1–6), indicating that the accuracy was the highest when k equaled to 1. Nevertheless, the highest average accuracy ass 52.8%, which is not high enough.

Table 5. Accuracy of physical bullying detection (k = 1)

	Bullying	Non-bullying
Bullying (real)	55.5%	45.5%
Non-bullying (real)	50.0%	50.0%

Table 6. Average accuracy of physical bullying detection (k = 1–6)

The value of k	1	2	3	4	5	6
Accuracy	52.8%	32.9%	20.0%	16.2%	8.7%	4.8%

Table 7 shows the activity recognition accuracy after adding 2 kinds of sum vectors (k = 1). And Table 8 shows the average activity recognition accuracy after adding 2 kinds of sum vectors (k = 1–6). Similarly, when k equaled to 1, the accuracy was the highest, which is 70.4%, higher than the first one.

Table 7. Accuracy of physical bullying detection after adding 2 kinds of sum vectors (k = 1)

	Bullying	Non-bullying
Bullying (real)	72.4%	31.6%
Non-bullying (real)	27.6%	68.4%

Table 8. Average accuracy of physical bullying detection after adding 2 sum vectors (k = 1–6)

The value of k	1	2	3	4	5	6
Accuracy	70.4%	50.5%	35.5%	19.0%	12.1%	7.8%

Table 9 shows the accuracy of speech emotion recognition (k = 1). Table 10 shows the average accuracy of speech emotion recognition (k = 1–6). The highest average accuracy is 78.0%, a bit higher than that of the activity detection.

Table 9. The accuracy of speech emotion recognition (k = 1)

	Bullying	Non-bullying
Bullying (real)	74.5%	18.5%
Non-bullying (real)	35.5%	81.5%

Table 10. The average accuracy of speech emotion recognition (k = 1–6)

The value of k	1	2	3	4	5	6
Accuracy	78.0%	62.5%	47.6%	34.7%	16.8%	8.5%

5 Discussion and Conclusion

Considering that physical bullying and verbal bullying have the worst impact on teenagers among all school bullying types but most of the existing anti-bullying methods are unrealistic and inconvenient, this paper applied an algorithm to detect physical bullying and verbal bullying events automatically. Time domain features and frequency domain features of activities and MFCC features of speeches were extracted to describe the characteristics of bullying events. k-NN was used as the classifier. A preliminary result of 52.8% of activity recognition was achieved, which was not satisfying. After adding the sum of the absolute value of acceleration and the angle of gyroscope, the accuracy was increased to 70.4%, indicating that the sum of the absolute value have a great impact on recognition. Additionally, the authors thought that speech emotion recognition could play a better role in bullying detection since the accuracy of speech emotion recognition was higher than that of the activity recognition. Gathering more samples to train the classifier might be helpful for improving recognition accuracy in future work.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61602127). The authors would like to thank the pupils from the second and the sixth grades who acted in the school bullying experiments.

References

1. Kim, Y.S., Leventhal, B.: Bullying and suicide. A review. Int. J. Adolesc. Med. Health **20**(2), 133–154 (2008)
2. Hinduja, S., Patchin, J.W.: Bullying, cyberbullying, and suicide. Arch. Suicide Res. **14**(3), 206–221 (2010)
3. Ye, L., Ferdinando, H., Seppänen, T., Alasaarela, E.: Physical violence detection for preventing school bullying. Adv. Artif. Intell. **2014**, f1–f9 (2014)
4. Ferdinando, H., Ye, L., Seppänen, T., et al.: Emotion recognition by heart rate variability. Aus. J. Basic Appl. Sci. Special **8**(14), 50–55 (2014)
5. Zalluhoglu, C., Ikizler-Cinbis, N.: Region based multi-stream convolutional neural networks for collective activity recognition. J. Vis. Commun. Image Represent. **60**, 170–179 (2019)
6. Rokni, S.A., Ghasemzadeh, H.: Autonomous training of activity recognition algorithms in mobile sensors: a transfer learning approach in context-invariant views. Hassan Source IEEE Trans. Mob. Comput. **17**(8), 1764–1777 (2018)
7. Chen, L.M., Nugent, C.D., Wang, H.: A knowledge-driven approach to activity recognition in smart homes. IEEE Trans. Knowl. Data Eng. **24**(6), 961–974 (2012)
8. Wongpatikaseree, K., Ikeda, M., Buranarach, M., Supnithi, T., Lim, A.O., Tan, Y.: Activity recognition using context-aware infrastructure ontology in smart home domain. KICSS **2012**, 50–57 (2012)
9. Liangying, P., Ling, C., Menghan, W., et al.: Complex activity recognition using acceleration, vital sign, and location data. Gencai Source IEEE Trans. Mob. Comput. **18**(7), 1488–1498 (2019)
10. Uddin, M.Z., Kim, J.: Human activity recognition using spatiotemporal 3-D body joint features with hidden markov models. KSII Trans. Internet Inf. Syst. **10**(6), 2767–2780 (2016)
11. Shuaibu, A.N., Abdullahi, Z.H., Buhari, M.D.: Mobile robot voice recognition in control movements. Int. J. Comput. Sci. Electron. (IJCSEE) **3**(1) (2015). ISSN: 2320–4028