



School Violence Detection Based on Multi-sensor Fusion and Improved Relief-F Algorithms

Liang Ye^{1,2(✉)}, Jifu Shi¹, Hany Ferdinando^{2,3}, Tapio Seppänen⁴,
and Esko Alasaarela²

¹ Department of Information and Communication Engineering,
Harbin Institute of Technology, No. 2 Yikuang Street, Harbin 150080, China
yeliang@hit.edu.cn

² Health and Wellness Measurement Research Group, OPEM Unit,
University of Oulu, Pentti Kaiteran katu 1, 90014 Oulu, Finland

³ Department of Electrical Engineering, Petra Christian University,
Siwalankerto 121-131, Surabaya 60236, Indonesia

⁴ Physiological Signal Analysis Team, University of Oulu,
Pentti Kaiteran katu 1, 90014 Oulu, Finland

Abstract. School bullying is a common social problem around the world, and school violence is considered to be the most harmful form of school bullying. This paper proposes a school violence detecting method based on multi-sensor fusion and improved Relief-F algorithms. Data are gathered with two movement sensors by role playing of school violence and daily-life activities. Altogether 9 kinds of activities are recorded. Time domain features and frequency domain features are extracted and filtered by an improved Relief-F algorithm. Then the authors build a two-level classifier. The first level is a Decision Tree classifier which separates the activity of jump from the others, and the second level is a Radial Basis Function neural network which classifies the remainder 8 kinds of activities. Finally a decision layer fusion algorithm combines the recognition results of the two sensors together. The average recognition accuracy of school violence reaches 84.4%, and that of daily-life reaches 97.3%.

Keywords: Improved Relief-F · Activity recognition · School violence · Artificial intelligence · Pattern recognition

1 Introduction

Activity recognition has become a popular topic in the fields of machine learning and artificial intelligence (AI) as movement sensor techniques become more and more mature. Activity recognition has a wide application prospect, e.g. smart home, smart city [1]. On the other hand, as scientific techniques develop, the human society also develops. However, an undesired problem has also grown, namely school bullying. School bullying is a kind of offensive behavior which hurt another person physically and/or mentally. School bullying can happen in various forms, e.g. physical violence, verbal bullying, destroying personal properties, among which physical violence is

considered to be the most harmful to teenagers. Traditional anti-bullying methods are passive and man-driven. For example, *Stop Bullies* and *TipOff* are two anti-bullying smartphone applications. The user needs to operate the smartphone to send an alarm when he/she is being bullied, which is obviously very difficult for him/her.

Fortunately, as mentioned above, since activity recognition is already available, an active and information-driven bullying detecting method comes to people's mind. The authors' research group has started this research work ever since 2013. Alasaarela [2] firstly argued that a smartphone embedded with a 3D accelerometer and a 3D gyroscope was able to detect school bullying events automatically. Then Ye in 2014 [3] developed an experimental classifier FMT (Fuzzy Multi-Threshold) and recognized some typical violent activities and daily-life ones with an average accuracy of 92%. However, as the number of activity types increased and players of different ages were involved, FMT failed to work because it was difficult to find unified thresholds. Later in 2015 [4], Ye developed an instance-based classifier named PKNN (Proportional K-Nearest Neighbor) which could deal with players of different ages and more activity types. However, the average accuracy dropped to 80%. On the other hand, besides movement data, physiological characteristics can also be used for bullying detection. Ferdinando [5] in this group focused on HRV (Heart Rate Variability) and ECG (Electrocardiography) signals, and improved the average accuracy from 70% [6] to 88% [7]. Of course it is also possible to combine multiple models together for bullying detection, e.g. Ye in 2018 [8] combined motion features and acoustic features together, and precision = 92.2% whereas recall = 85.8%.

In the authors' previous work, they only used one single movement sensor fixed on the player's waist. In this paper, the authors are going to apply multiple movement sensors fixed on the player's waist and leg. Simulation results will show that the proposed method outperforms single sensor methods, even better than the multi-model one.

The remainder of this paper is organized as follows: Sect. 2 describes how the authors gathered data with multiple movement sensors; Sect. 3 describes the extracted features and proposes an improved Relief-F feature selecting method; Sect. 4 proposes a DT-RBF two-layer classifier; Sect. 5 shows the simulation results; and Sect. 6 finally draws a conclusion.

2 Data Acquisition

Activity recognition with a single motion sensor of the authors' research work has hit a bottleneck when they tried to improve the recognition accuracy, so this paper chooses to use multiple motion sensors for activity recognition. The numbers and positions of the motion sensors affect recognition accuracy [9, 10]. For a single motion sensor, the best position is the waist [11], whereas for multiple sensors, there are several possible positions for the sensor, e.g. the waist, the chest, the wrist, the arm, and the leg. However, in daily-life activities, the motions of wrists and arms are very random, and are difficult for recognition. Therefore, the authors consider the waist, the chest, and the leg. Chowdhury [12] has given a comparison of the recognition performance of the three positions as shown in Table 1.

Table 1. Recognition performance of different positions.

Position(s)	Average accuracy
Leg	83.6%
Waist	84.1%
Chest	79.6%
Leg+chest	81.7%
Leg+waist	91.6%
Chest+waist	90.5%
Waist+leg+chest	88.6%

It can be seen from Table 1 that the combination of the leg and the waist outperformed the others. Therefore, the authors chose to use two motion sensors to collect motion data, i.e. one on the leg and the other on the waist.

Motion data were gathered by role playing of school violence and daily life. Eight people of different ages participated in the role playing, and they acted three kinds of school violence activities, namely beat, push, and push down, and six kinds of daily life activities, namely stand, walk, run, jump, play, and fall down. These people took turns to act different roles, e.g. the bullied and the bullies. During the role playing, the authors used a camera to record every action and made fine synchronization with the motion sensors. After the role playing, the authors extracted the activity samples according to the video recording, and the total number of activity samples was 1,160.

3 Feature Extraction and Feature Selection

As the authors did in their previous work [8], both time domain features and frequency domain features are considered, but more features have been extracted in this experiment. All the features are extracted from the following three items:

- (1) Combined horizontal vector of acceleration: As mentioned above, two motion sensors are used in this experiment, one of which is fixed on the waist whereas the other on the leg. The x-axes and z-axes of the motion sensors point to two orthogonal horizontal directions, and this paper combines them together, i.e.

$$ACC_{Hori}(i) = \sqrt{(ACC_{x-axis}(i)^2 + ACC_{z-axis}(i)^2)} \quad (1)$$

- (2) Vertical vector of acceleration: The y-axes of the motion sensors are the vertical vectors.
- (3) Combined vector of gyro: This paper combines all the three axes of gyro together, i.e.

$$Gyro = \sqrt{(Gyro_{x-axis}(i)^2 + Gyro_{y-axis}(i)^2 + Gyro_{z-axis}(i)^2)} \quad (2)$$

For example, when the authors extract the *Mean* feature, they extract $Mean_{AccHori}$ (the mean of the combined horizontal vector of acceleration), $Mean_{AccVert}$ (the mean of the vertical vector of acceleration), and $Mean_{Gyro}$ (the mean of the combined vector of gyro).

3.1 Time Domain Features

Table 2 lists the extracted time domain features.

Table 2. Time domain features.

Feature	Meaning
<i>Mean</i>	Mean
<i>MAD</i>	Median absolute deviation
<i>Var</i>	Variance
<i>Max_{diff}</i>	Maximum of differential
<i>Mean_{diff}</i>	Mean of differential
<i>Max</i>	Maximum of amplitude
<i>Min</i>	Minimum of amplitude
<i>PCC</i>	Pearson’s Correlation Coefficient
<i>Kurtosis</i>	Kurtosis
<i>Skewness</i>	Skewness
<i>Quan^{25th}</i>	1/4 Quantile
<i>Quan^{75th}</i>	3/4 Quantile
<i>ZCR</i>	Zero Cross Ratio

PCC (Pearson’s Correlation Coefficient) here is used to describe the correlation between two axes, and calculated as

$$PCC_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \tag{3}$$

Kurtosis is a measure of the flatness of the data distribution, and calculated as

$$Kurtosis = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \tag{4}$$

Skewness is a measure of data symmetry, and calculated as

$$Skewness = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{\frac{3}{2}}} \tag{5}$$

3.2 Frequency Domain Features

Table 3 lists the extracted frequency domain features.

Table 3. Frequency domain features.

Feature	Meaning
$Mean_{FFT}$	Mean
MAD_{FFT}	Median absolute deviation
$Energy_{FFT}$	Energy
Max_{FFT}	Maximum

All the frequency domain features are extracted with FFT (Fast Fourier Transformation) from the 3 items (i.e. the combined horizontal vector of acceleration, the vertical vector of acceleration, and the combined vector of gyro). Features like main lob center frequency [8] have proven to be useless, so this paper does not take them into consideration.

3.3 Improved Relief-F Feature Selecting Algorithm

The authors used a Wrapper method for feature selection in their previous work [8]. However, Wrapper is very time-consuming, especially when the number of features is large. Therefore, this paper chooses to use a Filter-based feature selecting algorithm.

The Relief algorithm selects features according to the inter-class distances and intra-class distances. Relief itself is designed for 2-class classification, and Relief-F is an improvement on Relief for multi-class classification. Relief-F selects feature sets in the same way as Relief, i.e. it only considers the discrimination of the samples, but ignores the redundancy of similar features. Therefore, this paper proposes an improved

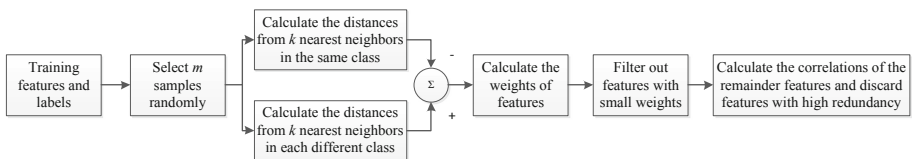


Fig. 1. Framework of the improved Relief-F algorithm.

Relief-F algorithm which eliminates the redundancy of similar features during the feature selection procedure. Figure 1 shows the framework of the improved Relief-F algorithm.

After feature selection, the dimensionality of the feature vector is reduced from 51 (39 time domain features and 12 frequency domain features) to 23.

4 Classifier Design

In the authors' previous work [8], BPNN (Back Propagation Neural Network) was used for classification. However, the learning speed of BPNN is very slow because of global approximation, so BPNN is not suitable for practical use when re-training is needed. Deep learning based classifiers such as CNN (Convolutional Neural Network) need quite a lot of training samples and thus are also unsuitable for practical use for the same reason. As for other classifiers such as DT (Decision Tree) based [3] and KNN (K-Nearest Neighbor) based [4], the authors had made a comparison [8] and found that they were not even as good as BPNN in terms of accuracy.

Therefore, this paper chooses the RBF (Radial Basis Function) Neural Network as the basic classifier. The RBF Neural Network has several advantages compared with the authors' previously used classifier BPNN. Firstly, the generalization ability of RBF is superior to that of BPNN in many aspects. Secondly, the approximation accuracy of RBF is higher than that of BPNN. It can almost achieve complete approximation, and it is extremely convenient to design. The network can automatically increase neurons until it meets the accuracy requirements. RBF is a kind of feedforward neural network with excellent performance. RBF can approximate any nonlinear function with arbitrary accuracy and has the ability of global approximation. It solves the local optimum problem of BPNN. Moreover, its topological structure is compact, and its convergence speed is fast.

In an RBF network, the number of neurons in the hidden layer can be smaller than that of the training samples, which will reduce the time cost. Add training samples into the hidden until the error meets the demand. The built RBF neural network is given in Fig. 2.

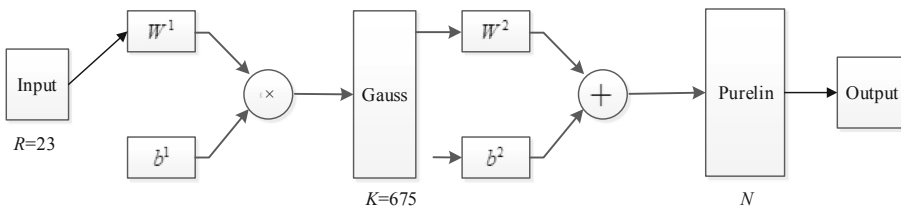


Fig. 2. The built RBF neural network.

The number of inputs $R = 23$ which is the dimensionality of the feature vector. The number of neurons in the hidden layer $K = 675$, and the radial basis function of the

hidden layer is Gauss. The transfer function of the output layer is Purelin, and the number of outputs N equals to the kinds of activities, i.e. if the activities are to be classified into 9 classes, $N = 9$.

Among all the remainder features after feature selection, the authors find that the energy of the y-axis of acceleration in frequency domain can distinguish “jump” from other activities quite well. Therefore, the authors design a two-level classifier. The first level is a DT classifier which separates “jump” from other activities, and the second level is an RBF classifier which classifies the remainder 8 kinds of activities.

The authors use two separate movement sensors for activity recognition, so a decision-layer fusion algorithm is needed to combine the recognition results together. The Dempster-Shafe theory is used in this paper as the decision-layer fusion algorithm.

5 Experiments

The authors gathered 1,160 sections of activities which are classified into 9 kinds. Feature filtering by the improved Relief-F algorithm were performed after feature extraction. Then the proposed two-layer DT-RBF classifier tried to classify the samples with 5-fold cross-validation. The recognition results are given in Table 4.

Table 4. Confusion matrix of school violence detection (unit: %, classified into 9 kinds)

	Beat	Jump	Play	Push	Run	Stand	Walk	Push down	Fall down
Beat	85.2	0.0	3.0	0.7	3.7	1.5	2.2	1.5	2.2
Jump	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Play	1.1	0.0	88.9	2.2	0.0	2.2	4.4	0.0	1.1
Push	12.9	0.0	5.7	71.4	2.9	0.0	7.1	0.0	0.0
Run	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Stand	1.3	0.0	2.7	0.0	0.0	95.3	0.7	0.0	0.0
Walk	0.7	0.0	2.1	0.0	0.0	0.0	97.1	0.0	0.0
Push down	35.6	0.0	2.2	0.0	4.4	4.4	0.0	40.0	13.3
Fall down	10.0	0.0	4.0	0.0	12.0	0.0	0.0	12.0	62.0

Since the aim of this paper is to recognize school violence, then the authors classify the activities into 2 kinds, i.e., school violence and non-violence. The confusion matrix is given in Table 5.

Table 5. Confusion matrix of school violence detection (unit: %, classified into 2 kinds)

	Violence	Non-violence
Violence	84.4	15.6
Non-violence	2.7	97.3

The average accuracy of violence recognition is 84.4%, and that of non-violence is 97.3%. Moreover, *Accuracy* = 93.7%, *Precision* = 92.6%, *Recall* = 84.4%, and *F₁_score* = 88.3%, which shows an improvement compared with the authors' previous work in the aspect of activity recognition [8].

6 Conclusion

This paper proposed a school violence detecting method based on multi-sensor fusion and improved Relief-F algorithms. The authors gathered 9 kinds of activities including school violence and daily-life activities and extracted time domain features and frequency domain features. Then the authors proposed a feature selecting algorithm – improved Relief-F, which takes the redundancy of similar features into consideration. Different features fit different classifiers, so the authors proposed a two-level classifier DT-RBF. The first level DT separated “jump” from the other activities, and the second level RBF classified the remainder. The Dempster-Shafe theory was used as the decision-layer fusion algorithm. The final recognition accuracy of violence recognition is 84.4%, and that of non-violence is 97.3%, which showed an improvement compared with the authors' previous work.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61602127), the Directorate General of Higher Education, Indonesia (2142/E4.4/K/2013) and the North Ostrobothnia Regional Fund of the Finnish Cultural Foundation.

References

1. Weidang, L., Yi, G., Xin, L., et al.: Collaborative energy and information transfer in green wireless sensor networks for smart cities. *IEEE Trans. Industr. Inf.* **14**(4), 1585–1593 (2018)
2. Liang, Y., Hany, F., Esko, A.: Techniques in pattern recognition for school bullying prevention: review and outlook. *J. Pattern Recogn. Res.* **9**, 50–63 (2014)
3. Liang, Y., Hany, F., Tapio, S., et al.: Physical violence detection for preventing school bullying. *Adv. Artif. Intell.* **2014**, 1–9 (2014)
4. Liang, Y., Hany, F., Tapio, S., et al.: An instance-based physical violence detection algorithm for school bullying prevention. In: 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1384–1388 (2015)
5. Hany, F., Liang, Y., Seppänen, T., et al.: Emotion recognition by heart rate variability. *Aust. J. Basic Appl. Sci.* **8**(14), 50–55 (2014)
6. Hany, F., Seppänen, T., Esko, A.: Enhancing emotion recognition from ECG signals using supervised dimensionality reduction. In: Proceeding of 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp. 112–118 (2017)
7. Hany, F., Liang, Y., Tian, H., et al.: Violence detection from ECG signals: a preliminary study. *J. Pattern Recogn. Res.* **12**(1), 7–18 (2017)
8. Liang, Y., Peng, W., Le, W., et al.: A combined motion-audio school bullying detection algorithm. *Int. J. Pattern Recogn. Artif. Intell.* **32**, 1850046 (2018)
9. Coskun, D., Incel, O.D., Oztgovde, A.: Phone position/placement detection using accelerometer: impact on activity recognition. In: IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 1–6. IEEE (2015)

10. Mannini, A., Intille, S.S., Rosenberger, M., et al.: Activity recognition using a single accelerometer placed at the wrist or ankle. *Med. Sci. Sports Exerc.* **45**, 2193 (2013)
11. Yang, C.C., Hsu, Y.L.: A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors* **10**, 7772–7788 (2010)
12. Chowdhury, A., Tjondronegoro, D., Chandran, V., et al.: Physical activity recognition using posterior-adapted class-based fusion of multi-accelerometers data. *IEEE J. Biomed. Health Inform.* **99**, 1 (2017)