A Multi-sensor School Violence Detecting Method Based on Improved Relief-F and D-S Algorithms

Liang Ye[1,2,*] [0000-0001-6076-0261], Jifu Shi[1], Hany Ferdinando[2,3], Tapio Seppänen[4], Esko Alasaarela[2]

1. Department of Information and Communication Engineering, Harbin Institute of Technology, No.2 Yikuang Street, Harbin 150080, China

2. Health and Wellness Measurement research group, OPEM unit, University of Oulu, Pentti Kaiteran katu 1, Oulu 90014, 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, Oulu 90014, Finland

yeliang@hit.edu.cn, +86-451-86413514-8224

Abstract: School bullying is a common social problem, and school violence is considered to be the most harmful form of school bullying. Fortunately, with the development of movement sensors and pattern recognition techniques, it is possible to detect school violence with artificial intelligence. This paper proposes a school violence detecting method based on improved Relief-F and Dempster-Shafe (D-S) algorithms. Two movement sensors are fixed on the object's waist and leg, respectively, to gather acceleration and gyro data. Altogether nine kinds of activities are gathered, including three kinds of school violence and six kinds of daily-life activities. After wavelet filtering, 39 time-domain features and 12 frequency-domain features are extracted. To reduce computational cost, this paper proposes an improved Relief-F algorithm which selects features according to classification contribution and correlation. By drawing boxplots of the selected features, the authors find that the frequency-domain energy of the y-axis of acceleration can distinguish jumping from other activities. Therefore, the authors build a two-layer classifier. The first layer is a decision tree which separates jumping from other activities, and the second layer is a Radial Basis Function (RBF) neutral network which classifies the remainder eight kinds of activities. Since the two movement sensors work independently, this paper proposes an improved D-S algorithm for decision layer fusion. The improved D-S algorithm designs a new probability distribution function on the evidence model and builds a new fusion rule, which solves the problem of fusion collision. According to the simulation results, the proposed method has increased the recognition accuracy compared with the authors' previous work. 89.6% of school violence and 95.1% of daily-life activities were correctly recognized. The accuracy reached 93.6% and the precision reached 87.8%, which were 29.9% and 2.7% higher than the authors' previous work, respectively.

Keywords: improved Relief-F, improved D-S, school violence, activity recognition, artificial intelligence

1. Introduction

Activity recognition has become a popular topic in the fields of machine learning and artificial intelligence as movement sensor techniques become more and more mature. Activity recognition has a wide application prospect, e.g. smart home and smart city [1-3]. 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* [4] and *TipOff* [5] are two anti-bullying smartphone applications [6]. 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 [7] 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 [8] 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 [9], 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 [10] in this group focused on HRV (Heart Rate Variability) and ECG (Electrocardiography) signals, and improved the average accuracy from 70% [11] to 88% [12]. Of course it is also possible to combine multiple models together for bullying detection, e.g. Ye [6] 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: Section 2 describes how the authors gathered data with multiple movement sensors; Section 3 describes the extracted features and proposes an improved Relief-F feature selecting method; Section 4 proposes a DT-RBF multi-layer classifier and an improved D-S decision layer fusion algorithm; Section 5 shows the simulation results; and Section 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 [13, 14]. For a single motion sensor, the best position is the waist [15], 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 [16] 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 [6], 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-axis and the z-axis of each motion sensor point to two orthogonal horizontal directions. Since the authors extract features from the combined horizontal direction of each sensor, they do not care to which exact horizontal directions the x-axes and the z-axes point. The combined horizontal vector is calculated as,

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

(2) Vertical vector of acceleration: The positive y-axis of each motion sensor points straight up, so the y-axis is the vertical vector.

(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
$Mean$	Mean
MAD	Median absolute deviation
Var	Variance
Max_{diff}	Maximum of differential
$Mean_{diff}$	Mean of differential
Max	Maximum of amplitude
Min	Minimum of amplitude

PCC	Pearson's Correlation Coefficient
$Kurtosis$	Kurtosis
$Skewness$	Skewness
$Quan^{25^{th}}$	1/4 Quantile
$Quan^{75^{th}}$	3/4 Quantile
ZCR	Zero Cross Ratio

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

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

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

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

Skewness is a measure of data symmetry, and calculated as

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

3.2 Frequency Domain features

Table 3 lists the extracted frequency domain features.

Table 3 Frequency domain features

Feature	Meaning
$Mean_{\mathrm{FFT}}$	Mean
MAD_{FFT}	Median absolute deviation
$Energy_{\mathrm{FFT}}$	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 [16] 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 [6]. 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 am improved Relief-F algorithm which eliminates the redundancy of similar features during the feature selection procedure. Fig. 1 shows the framework of the improved Relief-F algorithm.

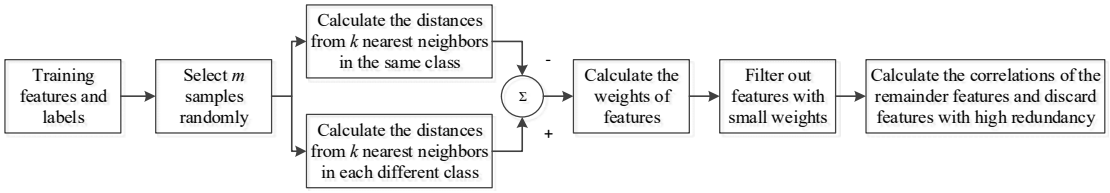

Fig. 1 Framework of the improved Relief-F algorithm

The detailed description of the algorithm is given as follows.

(1) Assume that the training set is D, the number of features is N, the weight of a feature is W, and the weight threshold is δ. Select m samples from D, and for each sample, perform the following steps.

(2) For a feature F of the given sample R, find its k nearest neighbors in the same class (marked as H_j, where j=1, 2, … , k), and its k nearest neighbors in each different class (marked as $M_{c,j}$, where j=1, 2, … , k, c=1, 2, …, C, C is the number of classes and $c \neq \text{class}(R)$). The difference between two samples R_1 and R_2 on feature F is calculated as

$$diff(F, R_1, R_2) = \frac{|R_1(F) - R_2(F)|}{\max(F) - \min(F)} \tag{6}$$

(3) Calculate the class distribution probability as

$$P = \frac{P(c)}{1 - P(class(R))} \tag{7}$$

where $P(c)$ is the probability of class c.

(4) For all the features of all the m samples, calculate their weights as

$$W(F) = \frac{\sum_{c \neq class(R)} \left[P \times \left(\sum_{j=1}^{k} diff(F, R, M_{c,j}) \right) \right]}{km} - \frac{\left(\sum_{j=1}^{k} diff(F, R, H_j) \right)}{km} \tag{8}$$

(5) Select the features of which $W(F) > \delta$. Assume that the number of selected features is n.

(6) Calculate the correlation coefficient matrix $\mathbf{A}$ of the n features as

$$\mathbf{A} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & & r_{2n} \\ \vdots & & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{9}$$

where r is the Pearson correlation coefficient and calculated as

$$r = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{X_i - \overline{X}}{\sigma_X} \right) \left(\frac{Y_i - \overline{Y}}{\sigma_Y} \right) \tag{10}$$

(7) Discard redundant features according to $\mathbf{A}$. If r_{ij} in $\mathbf{A}$ is larger than a given threshold, that means

either of the features *i* and *j* is unnecessary, and the one with the smaller weight is abandoned.

This paper uses two movement sensors, and the features of the data gathered by different sensors carry out the improved Relief-F algorithm independently.

4. Classifier Design

4.1 RBF Neural Network

In the authors' previous work [6], 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 [8] and KNN (K-Nearest Neighbor) based [9], the authors had made a comparison [6] 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.

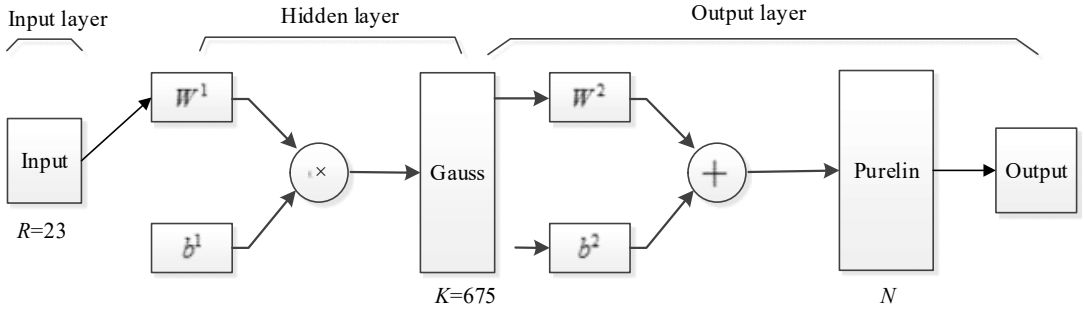

Fig. 2 The built RBF neural network.

The number of inputs *R*=23 which is the dimensionality of the feature vector. The number of neurons in the hidden layer *K*=675 (determined during the training procedure, which could achieve the best performance), 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 jumping from the other activities quite well (Fig. 3 shows the boxplot). Therefore, the authors design a two-layer classifier. The first layer is a DT classifier which separates jumping from other activities with this feature, and the second layer is an RBF which classifies the remainder eight kinds of activities.

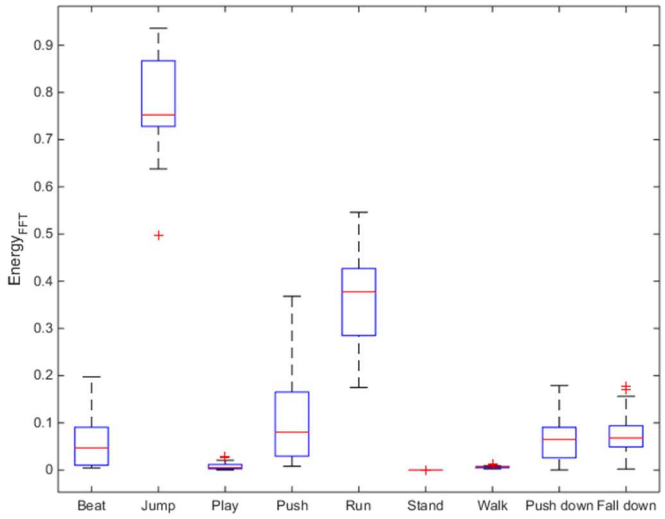

Fig. 3 The boxplot of $Energy_{FFT}$ of y-axis of acceleration on the waist.

4.2 Improved D-S Fusion Algorithm

This paper used two independent movement sensors for classification, so a fusion algorithm is necessary to produce the final decision. The D-S (Dempster-Shafe) is a kind of uncertain evidence reasoning theory [17-19]. It can make a fuzzy inference on things without knowing the prior probability, and has the ability to express the uncertainty directly. However, the classical D-S theory suffers from the problem of evidence collision, i.e., if the movement sensor on the waist and the one on the leg give two contradictory results, the system may give undesirable results.

This paper improves the classical D-S theory by proposing a new probability assignment function (PAF) according to the reliability of evidence sources. By adjusting the recognition probabilities of the neutral networks on different sensors, the improved D-S theory is able to reduce decision collisions. Furthermore, the authors improved the fusion rule in accordance with the new probability assignment function.

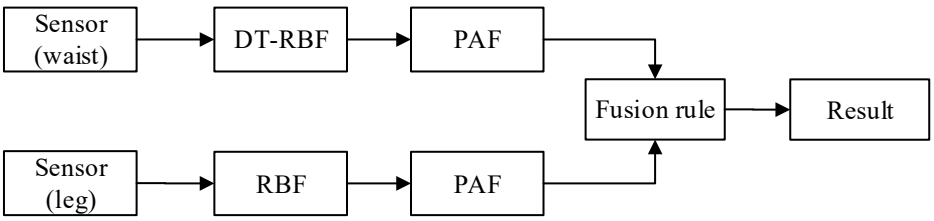

Fig. 4 Architecture of the decision-layer fusion algorithm.

Given an event A, if Θ is a set of all possible hypotheses of A, Θ is called the recognition frame of A, and $2^{\Theta}$is the set of all the subsets of Θ. Assume that m_1, m_2, …, m_n are n evidences belonging to Θ, Θ has N (N=9 in this paper) possible recognition results, i.e., $|\Theta|=N$, and $2^{\Theta}=\{A_i|i=1, 2, \ldots, 2^N\}$. Define the distance between m_1 and m_2 as,

$$d(m_1,m_2)=\sqrt{(M_1-M_2)^T D(M_1-M_2)/2} \tag{11}$$

where $M_i=\left[m_i(A_1),m_i(A_2),\cdots,m_i(A_{2^N})\right]^T$, i=1, 2; D is a $2^N\times 2^N$ matrix, and calculated by Jaccard

similarity coefficient as,

$$D_{ij} = \frac{\left|A_i \bigcap A_j\right|}{\left|A_i \bigcup A_j\right|}, \forall A_i, A_j \subseteq \Theta, i, j = 1, 2, \cdots, 2^N \tag{12}$$

then,

$$d\left(m_1, m_2\right) = \sqrt{\left(\left\langle M_1, M_1\right\rangle + \left\langle M_2, M_2\right\rangle - 2 * \left\langle M_1, M_2\right\rangle\right) D / 2} \tag{13}$$

where $\left\langle M_i, M_i\right\rangle = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} m_1(A_i) m_2(A_j) D_{ij}$.

The distance between two evidences is inversely proportional to the similarity. Define the similarity of the two evidences of the waist sensor and the leg sensor as,

$$sim(m_1, m_2) = 1 - d(m_1, m_2) \tag{14}$$

The degree of mutual support between two evidences is positively correlated with their similarity. That is to say, if the similarity between two evidences is high, the two evidences support each other well, and vice versa. Define the support of a certain evidence by other evidences as,

$$Sup(m_i) = \sum_{j=1, j \neq i}^{n} sim(m_i, m_j), \ \ i = 1, 2, \cdots, n \tag{15}$$

If an evidence has a high *Sup*, it has little conflict with other evidences and is believed to be reliable. Define the absolute credibility of an evidence as,

$$Crd_i^{\alpha} = \frac{Sup(m_i)}{\max\limits_{1 \leq j \leq n}[Sup(m_j)]}, \ \ j = 1, 2, \cdots, n \tag{16}$$

Normalize the absolute credibility, and one can get the relative credibility,

$$Crd_i^{\gamma} = \frac{Crd_i^{\alpha}}{\sum\limits_{j=1}^{n} Crd_i^{\alpha}}, \ i = 1, 2, \cdots, n \tag{17}$$

Then adjust the original evidence model with the absolute credibility,

$$m_i^{'}(A) = Crd_i^{\alpha} * m_i(A) \tag{18}$$

At present, on the basis of classical D-S theory, a unified reliability function combination model is proposed to deal with evidence conflict factors,

$$m(A) = \sum_{B \bigcap C = A} m_1(B) m_2(C) + K * \delta(A, m) \tag{19}$$

where $A, B, C \in 2^{\Theta}$, *K* is the total conflicts, $\delta(A, m)$ is the weight of *A* and $\sum_{A \subseteq \Theta} \delta(A, m) = 1$. In this paper, for consistent evidences, the fusion rule is “and”, whereas for conflicting evidences, probability assignment is performed on the conflicting parts. The improved fusion rule is given as,

$$m\left(A\right) = \sum_{\bigcap A_i = A} \prod_{1 \leq j \leq n} m_j^{'}\left(A_i\right) + K * \delta(A, m), \ A \subseteq \Theta, A \neq \varnothing \tag{20}$$

where

$$\delta\left(A,m\right)=\sum_{i=1}^{n}Crd_i^{\gamma}m_i(A) \tag{21}$$

For two events A and B, the probability assigned to A is given as,

$$P(A)=\frac{Crd_1^{\alpha}m_1^{'}(A)}{Crd_1^{\alpha}m_1^{'}(A)+Crd_2^{\alpha}m_2^{'}(B)}m_1^{'}(A)m_2^{'}(B) \tag{22}$$

Note that the proposed probability assignment function should meet the condition that $\sum_{A\subseteq\Theta}m(A)=1$.

Proof:

For $\forall A\subseteq\Theta, A\neq\varnothing$,

$$\begin{aligned}\sum_{A\subseteq\Theta}\delta(A,m)&=\sum_{A\subseteq\Theta}\sum_{i=1}^{n}Crd_i^{\gamma}*m_i(A)=\sum_{i=1}^{n}\sum_{A\subseteq\Theta}Crd_i^{\gamma}*m_i(A)\\&=\sum_{i=1}^{n}Crd_i^{\gamma}\sum_{A\subseteq\Theta}m_i(A)=\sum_{i=1}^{n}Crd_i^{\gamma}=1\end{aligned} \tag{23}$$

therefore,

$$\begin{aligned}\sum_{A\subseteq\Theta}m(A)&=\sum_{A\subseteq\Theta}\sum_{\cap A_i=A}\prod_{1\le j\le n}m_j^{'}(A_i)+\sum_{A\subseteq\Theta}K*\delta(A,m)\\&=\sum_{\cap A_i\neq\varnothing}\prod_{1\le j\le n}m_j^{'}(A_i)+K=1-K+K=1\end{aligned} \tag{24}$$

Proof completed.

5. Experiments

Altogether 1,160 sections of activities were recorded by means of role playing, including six kinds of daily-life activities and three kinds of school violence activities. Five-fold cross validation was used in the experiments.

Firstly, the authors tested the recognition accuracy of a single movement sensor. Table 4 gives the confusion matrix of activity recognition on the waist.

Table 4 Confusion matrix of activity recognition on the waist (%).

	Hit	Jump	Play	Push	Run	Stand	Walk	Push down	Fall down
Hit	**78.5**	0.0	7.4	3.7	0.7	1.5	2.2	3.7	2.2
Jump	0.0	**100.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Play	2.2	0.0	**92.2**	1.1	0.0	0.0	3.3	0.0	1.1
Push	15.7	0.0	11.4	**65.7**	0.0	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	0.7	0.0	5.3	0.0	0.0	**94.0**	0.7	0.0	0.0
Walk	0.0	0.0	2.1	2.1	0.0	0.0	**95.7**	0.0	0.0
Push down	6.7	0.0	4.4	2.2	13.3	8.9	0.0	**46.7**	17.8
Fall down	2.0	0.0	10.0	0.0	2.0	4.0	0.0	22.0	**60.0**

Since the purpose of this paper is school violence detection, the authors then classify the nine kinds

of activities into two classes, i.e. daily-life activities and school violence. “Hit”, “Push”, and “Push down” are school violence activities, and the remainder 6 kinds of activities belong to daily-life. Table 5 gives the 2-class confusion matrix.

Table 5 Two-class confusion matrix of activity recognition on the waist (%).

	School violence	Daily-life activity
School violence	79.2	20.8
Daily-life activity	3.0	97.0

According to Table 5, *accuracy*=92.0%, *precision*=91.2%, *recall*=79.2%, and F_1=84.8%. Similarly, Table 6 gives the 2-class confusion matrix of activity recognition on the leg.

Table 6 Two-class confusion matrix of activity recognition on the leg (%).

	School violence	Daily-life activity
School violence	79.2	20.8
Daily-life activity	13.4	86.6

As for Table 6, *accuracy*=84.5%, *precision*=70.0%, *recall*=79.2%, and F_1=74.3%. It is obvious that the sensor on the waist outperforms the one on the leg, so the former should take more weight than the latter in the fusion algorithm. Table 7 gives the 2-class confusion matrix after decision layer fusion with the improved D-S algorithm.

Table 7 Two-class confusion matrix with the improved D-S fusion (%).

	School violence	Daily-life activity
School violence	90.4	9.6
Daily-life activity	5.3	94.7

With the improved D-S fusion algorithm, *accuracy*=93.5%, *precision*=86.9%, *recall*=90.4%, and F_1=88.6%, which outperforms the results by a single sensor. Moreover, in terms of activity recognition, compared with the authors’ previous work [6] in which *accuracy*=63.7%, *precision*=85.1%, *recall*=71.7%, and F_1=76.6%, the improvement is significant.

In practical use, one can consider applying a dimensionality reduction algorithm to reduce the runtime overhead. This paper applies the LDA (Linear Discriminant Analysis) algorithm after feature selection, and Table 8 gives the classification result.

Table 8 Two-class confusion matrix with improved D-S and LDA (%).

	School violence	Daily-life activity
School violence	89.6	10.4
Daily-life activity	4.9	95.1

Accuracy=93.6%, *precision*=87.8%, *recall*=89.6%, and F_1=88.7%. The recognition performances of Table 7 and Table 8 are similar. However, the running time can be significantly reduced. Without LDA, the simulation ran 635s at a time, whereas with LDA, the simulation ran 310s at a time, i.e., LDA saved 51% running time.

6. Conclusion

School bullying is a common social problem, and school violence is considered to be the most harmful form of school bullying. This paper proposed a multi-sensor school violence detecting method based on improved Relief-F and D-S algorithms. Two movement sensors worked independently on the object’s waist and leg. Altogether 39 time-domain features and 12 frequency-domain features were extracted from acceleration and gyro for each sensor. The authors proposed an improved Relief-F algorithm to

select helpful and low redundancy features according to the classification contribution and correlation. LDA was used to further decrease the feature dimensionality. The authors found that one extracted feature was able to distinguish "jump" from the other activities, so they built a two-layer classifier. The first layer was a decision tree which separated "jump" from the other activities, and the second layer was a RBF neutral network which classified the remainder eight kinds of activities. In order to combine the two recognition results of the two sensors together, and solve the problem of evidence collision, the author proposed a new probability assignment function and improved the D-S fusion rule. According to simulation results, 89.6% of school violence and 95.1% of daily-life activities were correctly recognized, which showed an improvement compared with the authors' previous work.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61602127 and 4181101180), with significant contributions by the Directorate General of Higher Education, Indonesia (2142/E4.4/K/2013) and by the North Ostrobothnia Regional Fund of the Finnish Cultural Foundation.

The authors would like to thank Tuija Huuki and Vappu Sunnari (University of Oulu, Finland) for educational and psychological guidance in the school bullying experiments, teachers Taina Aalto and Pekka Kurttila and principal Maija Laukka (Oulunlahti School, Finland) for arranging the experiments and pupils in the 2nd and 6th grades of Oulunlahti School for acting in the experiments, Seppo Laukka and Antti Siipo (University of Oulu, Finland) for hardware and technical support during the experiments, and Tian Han, Zhu Zhang (Harbin University of Science and Technology, China) for assisting the experiments.

References

[1] Lu W, Gong Y, Liu X, *et al*. (2018) Collaborative Energy and Information Transfer in Green Wireless Sensor Networks for Smart Cities, IEEE Transaction on Industrial Informatics, 14(4):1585-1593.

[2] Han S, Zhang Y, Meng W, *et al*. (2019) Full-Duplex Relay-Assisted Macrocell with Millimeter Wave Backhauls: Framework and Prospects, IEEE Network, 33(5): 190-197.

[3] Han S, Huang Y, Meng W, *et al*. (2019) Optimal Power Allocation for SCMA Downlink Systems Based on Maximum Capacity, IEEE Transactions on Communications, 67(2): 1480-1489.

[4] https://apps.apple.com/cn/app/id469623408

[5] http://www.intouchmobile.ca/uploads/TipOff_Brochure.pdf

[6] Ye L, Wang P, Wang L, *et al*. (2018) A Combined Motion-Audio School Bullying Detection Algorithm. International Journal of Pattern Recognition and Artificial Intelligence, 32(2): 1-20.

[7] Ye L, Ferdinando H, and Alasaarela E (2014) Techniques in Pattern Recognition for School Bullying Prevention: Review and Outlook. Journal of Pattern Recognition Research, 9: 50-63.

[8] Ye L, Ferdinando H, Seppänen T, *et al.* (2014) Physical violence detection for preventing school bullying. Advances in Artificial Intelligence, 2014: 1-9.

[9] Ye L, Ferdinando H, Seppänen T, *et al*. (2015) An instance-based physical violence detection algorithm for school bullying prevention. 2015 International Wireless Communications and Mobile Computing Conference (IWCMC): 1384-1388.

[10] Ferdinando H, Ye L, Seppänen T, *et al*. (2014) Emotion recognition by heart rate variability, Australian Journal of Basic and Applied Sciences, Special 8(14): 50-55.

[11] Ferdinando H, Seppänen T and Alasaarela E (2017) Enhancing emotion recognition from ECG

signals using supervised dimensionality reduction, Proceeding of 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM): 112-118.

[12] Ferdinando H, Ye L, Han T, *et al*. (2017) Violence detection from ECG signals: a preliminary study. Journal of Pattern Recognition Research, 12(1): 7-18.

[13] Coskun D, Incel O D, Ozgovde A (2015) Phone position/placement detection using accelerometer: Impact on activity recognition, IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing. IEEE:1-6.

[14] Mannini A, Intille SS, Rosenberger M, *et al*. (2013) Activity recognition using a single accelerometer placed at the wrist or ankle, Medicine and Science in Sports and Exercise, 45: 2193.

[15] Yang CC and Hsu YL (2010) A review of accelerometry-based wearable motion detectors for physical activity monitoring, Sensors, 10: 7772-7788.

[16] Chowdhury A, Tjondronegoro D, Chandran V, *et al*. (2017) Physical Activity Recognition using Posterior-adapted Class-based Fusion of Multi-Accelerometers data. IEEE Journal of Biomedical & Health Informatics, (99):1-1.

[17] Li W, Bao J, Fu X, *et al*. (2012) Human Postures Recognition Based on D-S Evidence Theory and Multi-sensor Data Fusion. IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. IEEE, 2012:912-917.

[18] Liu X, Jia M, Na Z. (2018) Multi-Modal Cooperative Spectrum Sensing Based on Dempster-Shafer Fusion in 5G-Based Cognitive Radio. IEEE Access, 6: 199-208.

[19] Liu X, Zhang X, Jia M. (2018) 5G-based green broadband communication system design with simultaneous wireless information and power transfer. Physical Communication, 25: 539-545.