P02-2 ID86

THE USE OF MOTION SENSORS AND SUPPORT VECTOR MACHINE FOR CLASSIFYING SIMULATED ANKLE SPRAIN AND NORMAL MOTIONS

Kate Sin-Ki Lai[1], Daniel Tik-Pui Fong[2]

[1]Department of Physics, Faculty of Science, The Chinese University of Hong Kong, Hong Kong, China
[2]Department of Orthopaedics and Traumatology, Prince of Wales Hospital, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

Ankle sprain is one of the most common sports injuries. Our research team has developed an intelligent system to prevent the injury, and the system relies on a method to identify an ankle sprain motion. The purpose of this study is to increase the accuracy of Support Vector Machine (SVM) in classifying ankle sprain from normal motions and investigate the feasibility to employ SVM in the intelligent system. Fourteen subjects performed trials of (a) walking, (b) vertical jump, (c) stepping down a stair, and (d) jumping off a stair. Data from a motion sensor at the posterior calcaneus were used and trimmed to 230 (0.4s) and 60 (0.12s) window size, and were transformed from time to frequency domain by discrete Fourier Transform. Motion data from eleven subjects (11 out of 14) were used for training the SVM. A Radial Basis Function kernel function was employed in the SVM. Accuracy was tested on the data from another three subjects, which reached 96.1% and 93.1% for window size 230 and 60 respectively.

KEY WORDS: sports medicine, injury biomechanics, motion sensing, biomedical engineering

INTRODUCTION: Our research team has developed an intelligent system to prevent a supination type ankle sprain injury (Fong, 2012), which accounts for 14% of all sports injuries and is considered the most common sport trauma (Fong, Hong, Chan, Yung, & Chan, 2007). The system consists of a motion sensor placed inside the footwear. The motion sensor includes a tri-axial gyroscope and accelerometer to measure and monitor the kinematic motion of the ankle joint in three directions. When the ankle inversion velocity exceeds a pre-set threshold (Chu et al., 2010), a small electrical current is sent to the peroneal muscle in lower leg to stimulate muscle contraction to stop further inversion of the ankle, therefore to prevent ankle sprain (Fong, Chu, & Chan, 2012). For the device to function ideally, it should be able to discriminate ankle sprain from other types of common sport motions, such that not only would it be able to recognize ankle sprain accurately, but also not to falsely recognize common sport motions as injurious spraining motion. The accuracy of the current device is 91.3% and window size employed is 500 (1s) (Chan et al., 2010), which means the error was about 8.7% and window size is too large to allow real-time application , there is thus a room for improvement.
The objective of this study is to improve the accuracy of the device in discriminating ankle sprain and common sports motion, which is done by using support vector machine, a binary classification method. This study would also seek possibility to apply the method in real-time by using only the beginning part of the motions and reducing the window size to 60 (0.12s).

METHODS: Fourteen adult subjects participated in this study. Subjects wore four motion sensors, each consists of a tri-axial gyroscope and accelerometer, on the left or right foot. The purpose of using multiple sensors is to investigate the variation in strength of signal for sprain classification from different parts of the foot. The sensors are placed at including the dorsal area of the third metatarsal, the medial aspect of the first metatarsal, the lateral aspect of the fifth metatarsal, and the posterior calcaneus. The sampling frequency of the sensors was 500Hz. Subjects first performed simulated sprain using the supinated sprain machine (Chan, Fong, Yung, Fung, & Chan, 2008). The machine can simulate ankle sprain with different

Figure 1: Setup for data collection of simulated ankle sprain

degrees of inversion and plantarflexion. This study adopted four combinations, namely: 0° (pure inversion), 23° supination, 45° supination and 67° supination. 10 trials were performed for each combination. Afterwards, subjects performed 10 trials of each of the following common sports motions in random order: (a) walking, (b) vertical jump, (c) stepping down from a stair, and (d) jumping off a stair. Long range common motions like running and cutting are not performed due to the restriction of the wire connecting the sensor and the computer.

The recorded data underwent several processes before being used for training and classification in the support vector machine. The first procedure was to trim out data corresponds to the beginning part of motion. Since the purpose of the sprain-free device is to identify and then correct a sprain before it has completed, training and classification should only involve the beginning part of motions. The window size employed is 60, which corresponds to 0.12s. However, window size 230 (~0.4s), which normally include the majority part of motion, is also employed in order to investigate the effect of a bigger window size on the classification accuracy. Since the signal strength of the sensor at the posterior calcaneus is the strongest, the data from only this sensor was used. The second procedure was to carry out a discrete Fourier Transform (FFT) on each data parameter to transform the data from time domain to frequency domain. This allows the waveform characteristics to be visualized and characterized. In this study, only the first ten FFT components in each data parameter are trimmed out.

Support vector machine (SVM) is a supervised learning algorithm. The use of SVM involves two stages, which are the training of model, and then classification using the trained model. Training of SVM is to draw an optimal hyperplane that can separate data effectively into two classes, which are sprain and common motions in this study. Before training, a kernel function has to be specified. SVM performs only linear classification, i.e. classify data using linear function, however, most situations in real world are complicated and data cannot simply be classified with high accuracy using a linear function. A kernel function must be used to map the data to a high dimensional feature space where data can then be linearly classified. The kernel used in this study is the Radial Basis Function (RBF). Studies conducted by Keerthi and Lin (2010) showed that the RBF kernel is the most reliable kernel in many situations. Before training, data is scaled to [-1,1]. The purpose of scaling is to facilitate the dot product in the kernel function to prevent the appearance of extremely large or small numbers. SVM using RBF kernel involves two major parameters: Cost (C) and Gamma (g). Cost is related to the way the hyperplane is drawn. Gamma, appears in the RBF kernel, controls how data are mapped to the feature space. In this study, the two parameters are optimized by cross validation and grid search. Parameter selection is performed before each training. Training and classification are done in Matlab (Version R2012a, MathWorks, Inc., Natick, Massachusetts, USA).

In this study, data from 11 subjects (out of 14) was used for training stage of the SVM. They contributed 726 trials on simulated ankle sprain motions and common sporting motions. Another three subjects contributed 200-220 trials on simulated ankle sprain motions and common sporting motions for testing the accuracy of the trained SVM model.

Three investigations were carried out in the study. First is the investigation on the number of training subjects with accuracy for window size 60 and 230. The test data set was kept the same for all trials. The number of training subjects starts from 4, and one more subject would be added to the training data set in the next trial. One more trial would be done and then ends if the accuracy exceeds 90%, which is regarded as a satisfactory accuracy. Separate training and testing were done for each trial and window size. Second is the investigation on the effect of window size on accuracy. Two window sizes were studied, which are 230 (0.4s) and 60 (0.12s). Window size 60 includes only the beginning part of the motion while 230 includes half or most part of the motion. Training data set and test data set were the same for both window sizes; they differ only by the duration of time included in each data parameter. Separate training and testing were done on each window size. Third is the investigation of the effect of parameter selection on accuracy. Training and test data set were kept the same in each trial. In using cross validation (CV) and grid search to search for best parameter, the searching step size for log_2C and log_2g are set to 2, then a combination of trial best parameter would be output from CV and grid search (Table 4). Step size 2 is used since further reducing the step size may increase the duration of CV and grid search. Then C and g are searched for values close to the trial best parameter to determine if better accuracy can be obtained.

RESULTS: The investigation in the influence of number of training subjects showed that for window size 230, accuracy increased with number of training subjects, and 4 training subjects can already achieve accuracy >90% (Table1). For window size 60, accuracy showed no consistent increase in accuracy. This demonstrated that individual variation in performing same motion affected accuracy significantly when window size was reduced. Up to 11 training subjects, accuracy was still lower than 90% (Table 2). It was suspected that a large number of training subjects was required to overcome the individual variation. The investigation in window size showed that accuracy was lowered when widow size was reduced from 230 to 60 (Table 3), where the source of training data and test data were the same in the two cases. For small window size, better parameter selection was required to achieve high accuracy.

Table 1: No. of training subjects with accuracy for window size 230

Window Size	No. of Training Subjects	No. of Training Data	No. of Test Subjects	Accuracy of Classification
230	4	243	3	93.98% (203/216)
230	5	296	3	95.83% (207/216)
230	6	363	3	95.83% (207/216)

Table 2: No. of training subjects with accuracy for window size 60

Window Size	No. of Training Subjects	No. of Training Data	No. of Test Subjects	Accuracy of Classification
60	8	510	3	83.25% (169/203)
60	9	589	3	75.86% (154/203)
60	10	658	3	88.67% (180/203)
60	11	726	3	88.67% (180/203)

Table 3: Effect of window size and accuracy

Window Size	No. of Training Subjects	No. of Training Data	No. of Test Subjects	Accuracy of Classification
230	6	379	3	95.58% (195/204)
60	6	379	3	86.27% (176/204)

The investigation in parameter selection demonstrates that Cost (C) is less sensitive to Gamma (g). The search of log_2C may keep in step size 2 but the search of log_2g was reduced to step size 0.1 in order to get optimal parameters (Table 4). The final classification result for window size 230 was 95.83% accuracy with C=32 and g=0.00781. For window size 60, the accuracy is 93.10% with C=32, and g=1.3.

Table 4: Parameter selection and accuracy for window size 60

Parameters	Accuracy of Classification	Parameters	Accuracy of Classification
*C=32, g=0.5	88.67%(180/203)	C=16,g=1.30	93.10% (189/203)
C=16, g=1.0	92.12% (187/203)	C=16, g=1.40	91.63% (186/203)
C=16, g=2.0	87.68% (178/203)	C=32,g=1.40	91.63% (186/203)
C=4, g=1.0	92.12% (187/203)	C=16, g=1.35	92.12%(187/203)
C=16, g=1.5	91.13%(185/203)	C=32, g=1.35	92.12%(187/203)
C=16, g=1.1	92.61%(188/203)	C=64, g=1.30	93.10% (189/203)
C=32, g=1.3	93.10% (189/203)	C=128,g=1.30	93.10% (189/203)

*Best parameters with step size 2 for log_2C and log_2g

DISCUSSION: The team had previously conducted studies to classify ankle sprain and normal motions by SVM. (Chan et al., 2010) The window size adopted in that study was 500 (~1s) and the best classification accuracy was 91.3%. This study was a following-up of the previous study. In this study, classification accuracy was increased and window size was reduced to 60 to seek real-time application. It was expected that accuracy can be further increased by employing more training subjects in the case of window size 60. Errors may have been induced for trimming the wrong part of data and also from problems in data recording where it was sometimes found that motion sensors malfunctioned in the middle of data collection. The most time consuming part in using SVM was the training stage. Once a model was trained up, classification was quick and easy. To increase possibility of real-time application of SVM in sprain-free shoes, further studies could be done to reduce the window size to smaller than 50 (~0.1s). Furthermore, a reliable wireless system has to be attained to transfer data from sensor to computer in real-time, such that real-time classification can be carried out.

CONCLUSION: The final classification result for window size 230 (~0.4s) and 60 (~0.12s) were 95.83% and 93.1% respectively, which was satisfactory. Classification accuracy would decrease with smaller window size, and parameter C was found to be less sensitive than g. Step size for g has to be reduced to 0.1 to get optimal g-value.

REFERENCES:

Chan, Y.Y., Fong, D.T.P., Chung, M.M.L., Li, W.J., Liao, W.H., Yung, P.S.H.,… Chan, K.M. (2010). Identification of ankle sprain motion from common sporting activities by dorsal foot kinematics data. *Journal of Biomechanics*, 43(10), 1965-1969.

Chan, Y.Y., Fong, D.T.P., Yung, P.S.H., Fung, K.Y., & Chan, K.M. (2008). A mechanical supination sprain simulator for studying ankle supination sprain kinematics. *Journal of Biomechanics*, 41(11), 2571-2574.

Chu, V.W.S., Fong, D.T.P., Chan, Y.Y., Yung, P.S.H., Fung, K.Y., & Chan, K.M. (2010). Differentiation of ankle sprain motion and common sporting motion by ankle inversion velocity *Journal of Biomechanics*, 43, 2035-2038.

Fong, D.T.P. (2012). An intelligent sport shoe to prevent ankle inversion sprain injury. *Journal of Foot and Ankle Research*, 5 (Supp 1), K6.

Fong, D.T.P., Chu, V.W.S., & Chan, K.M. (2012). Myoelectric stimulation on peroneal muscles resists simulated ankle sprain motion. *Journal of Biomechanics*, 45(11), 2055-2057.

Fong, D.T.P., Hong, Y., Chan, L.K., Yung, P.S.H., & Chan, K.M. (2007). A systematic review on ankle injury and ankle sprain in sports. *Sports Medicine*, 37(1), 73-94.

Keerthi, S.S., & Lin, C.J. (2010). Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel*. Neural Computation, 15* (2003), 1667-1689.