

# 데이터 마이닝을 이용한 피칭과 팔 동작의 분류\*

정인섭<sup>1)\*\*</sup> · 이기광<sup>1)</sup> · 최민호<sup>2)</sup>

1) 국민대학교 2) (주)직토

## ABSTRACT

In-sub Jeong, Ki-kwang Lee, Min-ho Choi. Classification of Pitching and Arm Motion Using Data Mining. *KINESIOLOGY*, 2017, 19(2): 75-81. **[PURPOSE]** The purpose of this research is to classify pitching motions using the data mining method, which aims to help injury prevention overuse. **[METHODS]** One healthy person participated in the experiment. Subject performed six actions like pitching including pitching by wearing a smart band with IMU sensor built in the wrist. We converted the IMU data of each of the six motion into 5 Datasets. We performed data mining using the WEKA program to find the Dataset with the highest classification probability among the five Datasets and the appropriate classification model. **[RESULTS]** Among the 5 Datasets, Peak value Dataset when changing to Frequency domain through FFT showed the highest classification probability of each classification model, and NaiveBayes of each classification model had appropriate advantages for classification of pitching motion. Therefore NaiveBayes has decided on an appropriate classification model to classify pitching motion. **[CONCLUSIONS]** The data of the acceleration sensor and the gyroscope of the six actions are best classified for conversion using FFT and the NaiveBayes classification model is an appropriate classification model for classifying each motion.

**Key words :** overuse, IMU sensor, data mining, pitching, classification

**주요어 :** 과사용, 관성센서, 데이터 마이닝, 피칭, 분류

## 서론

경쟁이 요구되는 스포츠 선수에게 있어 부상은 피할 수 없으며, 이러한 부상은 선수의 경기력 저하와 선수의 생명을 위협할 수 있는 위험요인 중 하나이다. Kim(1997)은 배구 종목에서 훈련 시 75.8%, 경기 시 24.2%로 훈련 시 더 큰 비중으로 부상을 당한다고 보고하였다. 또한 Junge, et al.(2008)은 2008 베이징 올림픽 동안 선수의 72.5%가 부상을 입었으며, 그 중 22%는 과사용으로 인한 부상이라고 보고하였다. 과사용으로 인한 부상은 과사용 증후군이라 불리우며, 최대한의 부하가 지속적으로 정상적인 조직에 발생하는 것이 원인이다(Stanitski, 1989). 과사용으로 인한 부상은 연부 조직의 접합부, 힘이 집중되는 곳 등에서 발생한다. Nam, et al.(2002)은 한국 프로야구투수들의 부상 발생 원인 및 분류에 관한 연구를 진행했으며, 그 결과 과

다피칭과 오버트레이닝으로 인한 투수 부상이 50%이상 차지한다고 보고하였다.

과사용으로 인한 부상은 야구에서도 중요 부상 원인 중 하나이다. 프로야구 선수들은 과도한 반복훈련으로 인한 과사용성 상해가 매우 빈번하게 발생하며, 그 중 75%가 회전근개 건염 부상이다. 특히 투수의 경우 주로 장시간의 반복적인 피칭으로 인해 신체 곳곳에 발생하는 미세손상이 누적되어 부상이 발생한다(Wang et al., 2004). Kim, et al.(2007)과 Matsuo, et al.(2001)은 반복적인 투구는 견관절과 연부조직에 지속적인 스트레스를 주어 손상을 야기할 수 있다고 보고하였다. 이처럼 야구를 포함한 여러 스포츠 종목에서 과사용으로 인한 부상이 빈번하게 발생하고 있으며, 이러한 부상을 방지하기 위한 노력이 필요하다.

특정 동작에 대한 동작수를 측정할 수 있다면 과사용으로 인한 부상을 예방하는데 많은 도움이 될 것이다. 인간의 동작은 각각 특성이 있으며, 이는 가속도센

\* 이 논문은 2015년 문화체육관광부의 지원을 받아 연구되었음[s0720105082015].

\*\* 교신저자: junginsub03@naver.com

서로 측정이 가능하다. Park, et al.(2015)은 3축 가속도계의 착용 위치에 따른 동작별 특징 분석 연구를 진행했으며, 더 나아가 Kim, et al.(2011)과 Lee(2009)는 3축 가속도센서 하나만을 가지고 걸음 수 측정, 앉기&서기, 걷기&뛰기 등을 Fuzzy 알고리즘을 통해 동작을 인식하는 연구를 진행하였다. 각 동작의 출력값을 SVM(Signal Vector Magnitude)을 사용하여 분류했으며, 위와 같은 연구들은 센서를 사용한 동작 인식과 더 나아가 동작의 분류까지 가능하다는 것을 보여준다.

기존에 있던 대량의 데이터를 이용한 분류기준을 통해 새로 입력된 데이터를 분류시키는 것은 가장 많이 쓰이는 분류 방법이며 이는 데이터 마이닝의 분야 중 하나이다. Yang(2009)은 스마트 폰에 내장된 가속도센서의 데이터를 이용해 사용자의 앉기, 서기, 걷기, 뛰기, 자전거타기, 운전 등의 동작을 분류하였다. 이 연구에서 동작의 특징점을 찾기 위해 수직/수평 특징, 수직 특징, 수평 특징 등 여러개의 Dataset을 만들었으며, 데이터마이닝 도구를 통해 가장 이상적인 Dataset과 분류 특성에 맞는 분류모델인 Decision tree를 찾을 수 있었다. 또한 Cherbunroong, et al.(2011), Kao, et al.(2009), Siirtola, et al.(2009) 등은 한쪽 손목에 착용한 가속도센서를 통한 동작분류에 대한 연구를 진행하였으며, 그 중 Cherbunroong, et al.(2011)은 WEKA 프로그램을 통해 분류에 알맞은 Dataset과 분류모델을 찾아냈다.

선행 연구들에선 센서를 사용한 동작의 인식과 데이터 마이닝을 이용한 동작 분류에 대한 실험이 각각 진행되었다. 센서를 이용해 동작에 대한 데이터를 수집하고, 데이터마이닝을 이용해 특정 동작을 분류할 수 있다면 특정 동작에 대한 동작수를 측정할 수 있을 것이다. 따라서 본연구의 목적은 관성센서의 동작데이터를 데이터마이닝 방법을 이용해 피칭 동작을 분류하는 적절한 분류모델을 찾는 것이며, 이는 과사용에 대한 부상 예방에 도움이 될 것이다.

## 연구방법

### 연구대상

본 연구의 대상자는 최근 6개월 동안 근신경학적 및 정형외과적 병력이 없고, 주어진 실험과제를 수행하는데 문제가 없는 일반인을 대상으로 실시하였다. 대상자는 실험과정에 대한 충분한 설명을 들은 후 동의서에 서명하였다. 대상자의 신체적 특성은 다음과 같다<Table 1>.

Table 1. Physical characteristic of subjects (M±SD)

age (yr)	height (cm)	weight (kg)
31	177.3	75

### 실험설계 및 절차

과제 수행 시 피칭과 유사한 팔 동작을 측정하기 위해 양쪽 손목에는 IMU센서(관성센서)가 포함된 스마트 밴드(ZiktoBand, (주)직토, Korea)를 착용하였다. IMU센서의 사양은 가속도센서 : 16g, 자이로센서 : 2000deg/s, Sampling rate : 20Hz 이다. 스마트 밴드를 부착한 후 동작 중 떨어짐이나 이동을 방지하기 위해 키네지오 테이프를 이용해 손목에 고정시켰다. <Figure 1>과 같이 피험자는 Normal : 팔을 머리 위에서 아래로 움직이는 동작(피칭시 팔동작), Under : 팔을 아래에서 위로 움직이는 동작, Up & Down & Left & Right : 곧게 서있는 상태에서 팔을 위, 아래, 좌, 우로 빠르게 움직이는 동작 등을 각각 10회씩 총 60회 진행했다.

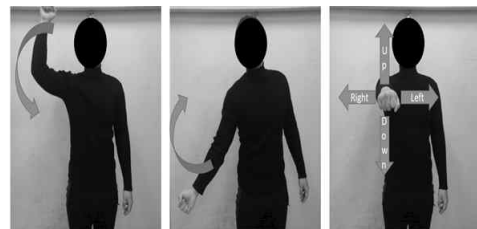


Figure 1. Experimental motion (Normal, Under, Up & Down & Right & Left)

### 자료 처리

Table 2. 5 kinds of Dataset

Frequency domain	①	maxf_ax	maxf_ay	maxf_az	maxf_gx	maxf_gy	maxf_gz
	②	minf_ax	minf_ay	minf_az	minf_gx	minf_gy	minf_gz
Time domain	③	int_ax	int_ay	int_az	int_gx	int_gy	int_gz
	④	max_ax	max_ay	max_az	max_gx	max_gy	max_gz
	⑤	min_ax	min_ay	min_az	min_gx	min_gy	min_gz

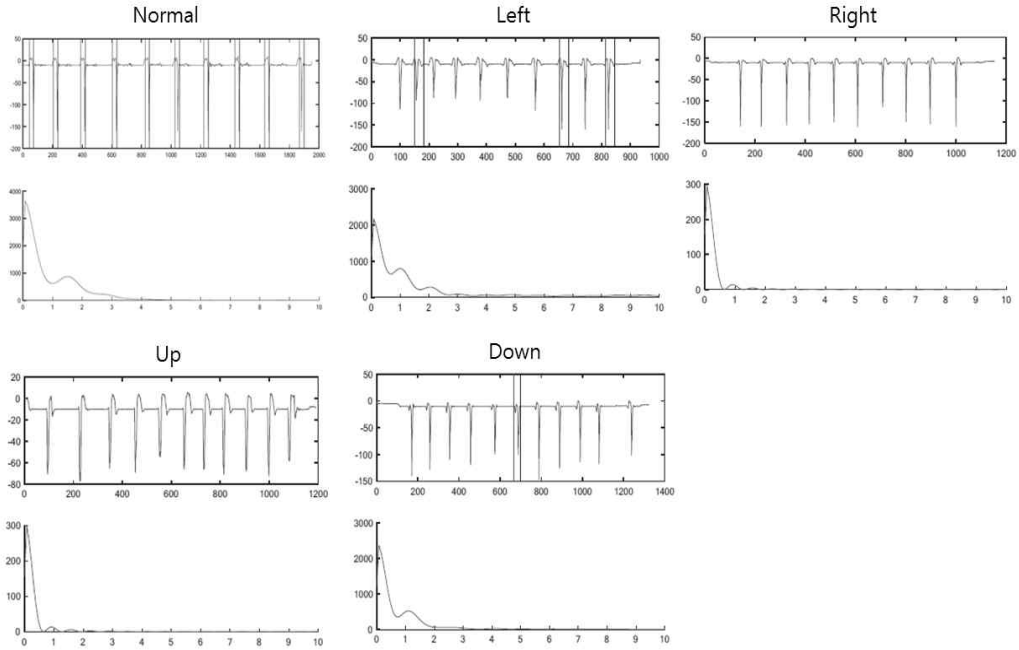


Figure 2. Graph of x-axis acceleration data of each motion (above : time domain , blow : frequency domain)

**분석변인**

본 실험에서 사용된 변인은 <Table 2>와 같이 ① & ② : FFT변환 된 가속도센서와 자이로센서의 3축(x,y,z)에 대한 데이터의 최대 및 최소값, ③ & ④ : 가속도센서와 자이로센서의 3축(x,y,z)에 대한 최대 및 최소값, ⑤ : 가속도센서와 자이로센서의 3축(x,y,z) 최댓값에 대한 기울기 등 동작 분류의 가능성이 있는 다섯 가지 Dataset을 변인으로 정했다.

**분석방법**

5가지의 Dataset 중 피칭을 포함한 피칭과 유사한 6 가지 동작을 분류하는데 적절한 Dataset과 분류모델을 정하기 위해, 데이터 마이닝 도구인 WEKA(Waikato

Environment for Knowledge Analysis)를 사용했다.

**자료처리 방법**

스마트 밴드에 내장되어 있는 IMU센서의 데이터는 Bluetooth방식을 이용해 Database에 저장되었으며, 저장된 모든 데이터는 Matlab R2014a(The MathWorks inc, USA)를 사용하여 <Figure 2>와 같이 처리하였다.

**결과**

본 실험의 결과 <Table 3>와 같이 각 축의 FFT를 통한 Frequency domain으로 변경했을 때의 Peak값인

① Dataset을 이용한 것이 실험에 사용된 4가지의 모델 중에서 가장 높은 분류확률을 보였다. 각각의 분류 모델은 J48(Decision Tree), IBk(k-NN), NaiveBayes, MulticlassClassifier 순으로 분류확률이 높았다. 그 중 본 실험에선 3번째로 높았던 NaiveBayes은 분류 확률이 가장 높지는 않았지만, 분류 모델의 특징 상 피칭 동작을 분류하는데 장점을 가지고 있기 때문에 분류 모델로 선택했다. NaiveBayes 분류 모델에 대한 분류 정보는 <Figure 3>에서 자세히 설명되어 있다.

Table 3. Classification probability of each model.

Classification model	Classification probability
NaiveBayes	89.83
IBk(k-NN)	93.22
MulticlassClassifier	88.13
J48(Decision Tree)	94.91

총 60개의 동작에 대한 분류는 Confusion Matrix를 통해 자세히 나타났으며, 6개의 동작만이 이상 분류로 나타났다. 또한 분류 모델의 성능을 평가하는 지표인 ROC Area는 높은 값을 가지고 있어 분류 모델의 성능이 높다는 것을 알 수 있다.

## 논의

본 연구는 한 사람을 대상으로 진행하였다. 피칭을 포함한 피칭과 유사한 6가지 동작을 분류하여 적절한 Dataset과 분류 모델을 찾기 위해 WEKA프로그램을 사용하였다.

동작의 특징을 세분화하는데 있어 측정된 데이터를 어떻게 설정할 것인가가 중요했다. 데이터는 가속도센서 데이터와 자이로센서 데이터 두 종류이며, 각 데이

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      53          89.8305 %
Incorrectly Classified Instances     6          10.1695 %
Kappa statistic                     0.8778
Mean absolute error                  0.0351
Root mean squared error              0.1848
Relative absolute error              12.6371 %
Root relative squared error          49.5825 %
Total Number of Instances           59

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.556   0.000   1.000     0.556   0.714     0.717   0.918    0.784    normal
1.000   0.122   0.625     1.000   0.769     0.741   0.959    0.751    under
1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    left
0.800   0.000   1.000     0.800   0.889     0.877   0.994    0.977    right
1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    up
1.000   0.000   1.000     1.000   1.000     1.000   1.000    1.000    down
Weighted Avg.   0.898   0.021   0.936     0.898   0.898     0.892   0.980    0.921

=== Confusion Matrix ===

 a  b  c  d  e  f  <-- classified as
5  4  0  0  0  0 | a = normal
0 10  0  0  0  0 | b = under
0  0 10  0  0  0 | c = left
0  2  0  8  0  0 | d = right
0  0  0  0 10  0 | e = up
0  0  0  0  0 10 | f = down
    
```

Figure 3. Result of WEKA (NaiveBayes)

터는  $x, y, z$  3축으로 이루어져 있다. 분류를 위해 SVM(Sum Vector Magnitude)방법을 이용했으며, SVM 방법은  $x, y, z$  축에 대한 데이터를 하나로 만들어서 처리하는 방법으로 가속도 혹은 자이로센서 데이터에 대한 처리에 많이 쓰이는 방법 중 하나이다. Chung(2016)은 3축 가속도계의 데이터를 SVM 방법을 통해 데이터 처리를 했으며, 이를 통해 뇌졸중 환자의 자세 흔들림에 대한 연구를 진행했다. 하지만 이러한 방법은  $x, y, z$  3축에 대한 데이터를 하나로 만들어서 처리하기 때문에, 몸을 풀거나, 팔을 움직이는 동작 등 피칭에서 팔을 휘두르는 동작과 유사한 동작을 구분하기 어렵다. 따라서 보다 정확한 분류를 위해서 가속도센서와 자이로센서의 각 축 데이터 값들을 이용하는 것이 효율적이라고 판단했다.

데이터 분류를 위해서 분류의 특징이 잘 들어나는 Dataset과 Dataset에 맞는 분류모델이 필요하다. Chernbumroong, et al.(2011)와 Ravi, et al.(2005)는 WEKA프로그램과 가속도센서를 이용한 신체활동분류 실험을 진행하였다. 분석을 위해 여러 종류의 Dataset을 만들었으며, WEKA 프로그램을 이용해 각 분류 모델에서 적용시켜 분류확률을 계산했으며, 그 결과 각 분류 모델에서 높은 분류 확률을 가지는 Dataset을 선택할 수 있었다.

시간 함수일 때의 Peak기울기, Max(Peak), Min과 주파수 함수일 때의 Max(Peak), Min 등 분류의 가능성이 있는 5종류의 Dataset을 임의로 정한 뒤, WEKA 프로그램을 사용해 분류 확률이 가장 높은 Dataset을 선택했다. WEKA의 분석 결과, 가속도데이터와 자이로데이터를 Frequency domain으로 FFT변환시킨 데이터의 Max(Peak)값이 J48(Decision Tree), NaiveBayes, IBk(k-NN), MultiClassClassifier에서 높은 분류 확률을 보였다. 따라서 동작 분류를 위한 Dataset은 가속도데이터와 자이로데이터를 FFT변환시킨 데이터의 Max값을 사용했다.

또한 NaiveBayes = 89.83%, IBk(k-NN) = 93.22%, MultiClassClassifier = 88.13%, J48(Decision Tree) = 94.91%의 분류 확률을 보였으며, 가장 분류 확률이 높은 J48(Decision Tree)가 가장 좋은 분류 모델일 것이다. 하지만 각 분류 모델은 각각의 장단점을 가지고 있으며, Dataset의 특징과 분류 환경 등을 고려해서 정해야 한다

(Kim, 2005). IBk(k-NN)은 가장 간단한 기법 중의 하나이며 비교적 높은 정확도를 보이지만 실행속도가 매우 느리다는 단점을 가지고 있다(Yang, 1994). NaiveBayes는 저장된 데이터를 이용해 새로운 정보에 대해 분류를 하는 방법으로, 데이터 분류에 효과적인 분류 모델이다(Dumais, 1998). J48(Decision Tree)는 복잡한 계산절차가 필요하지 않아 실용성이 높다(Ragan, 2005). 하지만 분류 모델이 불안하고, 숫자의 데이터셋으로 만들어진 트리들은 복잡해질 수 있다는 단점을 가지고 있다. Galathiya, et al.(2012)는 DecisionTree 알고리즘의 개선을 위한 연구를 진행하였으며, NavieBayes, J48(Decision Tree), IBk(k-NN), SVM(Support Vector Machine), NeuralNetworks 등의 알고리즘에 대해 분석하였다. 본 연구에서 사용된 NavieBayes, J48(Decision tree), IBk(k-NN)의 알고리즘에 대해 IBk(1k-NN)은 느린 처리속도, J48(DecisionTree)는 빠른 처리속도, Nave ayes는 매우 빠른 처리속도로 분류하였다. 따라서 IBk(k-NN)과 J48(Decision Tree)는 높은 분류확률에도 불구하고 느린처리속도와 숫자로 이루어진 Dataset에 대해 처리가 복잡해질 수 있다는 단점으로 인해 평균 1.5초인 피칭동작과 가속도 데이터 값이 기반인 본 연구와는 맞지 않는 분류 모델이다. 따라서 NaiveBayes가 큰 Dataset에 적합한 장점과 빠른 처리속도로 인해 이번 연구에 알맞은 분류모델로 보여진다. NaiveBayes의 분류 결과, Correctly Classified Instances : 89.83%이며, Confusion matrix에서 알 수 있듯이 6개의 동작이 10번씩 총 60번 측정되어 있는 데이터 중 7개의 잘못된 분류를 제외하고 53개의 성공적인 분류를 보였다.

## 결론

동작에 대한 분류를 위해서, 측정된 데이터를 동작의 특징이 나타나게 변환시켜야 하며, 변환시킨 데이터에 알맞은 분류 모델을 찾아야 한다. 데이터 변환 방법인 SVM방법을 사용했지만, 피칭을 포함한 피칭과 비슷한 6가지 동작에 대한 분류가 불가능했다. IMU센서의 가속도데이터와 자이로데이터의  $x, y, z$  등 3축 데이터를 다섯 가지 Dataset으로 변환시켰으며, 그 중 주

과수 변환한 데이터의 Max값인 Dataset이 NaiveBayes 모델에서 상당히 높은 분류확률을 보였다. NaiveBayes가 가장 높은 분류 확률은 아니었지만 분류모델의 특징으로 인해 피칭 동작의 분류에 적절한 분류모델이라고 생각되며 타당성 검증에 위한 다양한 후속 연구가 필요할 것으로 보여진다.

## 참고문헌

- Chernbumroong, S., Atkins, A. S., & Yu, H. (2011, September). Activity classification using a single wrist-worn accelerometer. In *Software, Knowledge Information, Industrial Management and Applications (SKIMA)*, 2011 5th International Conference on (pp. 1-6). IEEE.
- Chung, J. Y. (2016). Effect of Visual and Somatosensory Information Inputs on Postural Sway in Patients With Stroke Using Tri-Axial Accelerometer Measurement. *Physical Therapy Korea*, 23(1): 87-93.
- Dumais, S., Platt, J., Heckerman, D., & Sahami, M. (1998, November). Inductive learning algorithms and representations for text categorization. In *Proceedings of the seventh international conference on Information and knowledge management. ACM* pp. 148-155.
- Galathiya, A. S., Ganatra, A. P., & Bhensdadia, C. K. (2012). Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning. *International Journal of Computer Science and Information Technologies*, 3(2): 3427-3431.
- Junge, A., Engebretsen, L., Alonso, J. M., Renström, P., Mountjoy, M., Aubry, M., & Dvorak, J. (2008). Injury surveillance in multi-sport events: the International Olympic Committee approach. *British journal of sports medicine*, 42(6): 413-421.
- Kao, T. P., Lin, C. W., & Wang, J. S. (2009, July). Development of a portable activity detector for daily activity recognition. In *Industrial Electronics*, 2009. ISIE 2009. IEEE International Symposium on. pp. 115-120.
- Kim, C. S. (1997). Sport injury status in Korean volleyball players during the 4 olympic games. *Korean Journal of sports Science*, 6(1): 177-201.
- Kim, K. S (2005). A Study of the Integration of Individual Classification Model in Data Mining for the Credit Evaluation. *Information processing society journal D* 12(2): 211-218.
- Kim, Y. G, Han, K. J, & Park, J. Y. (2007). Clinical Article : The Study on Changes of Shoulder, Elbow and Forearm ROM with increasing Age in Youth Baseball Players. *The Korean Journal of Sports Medicine* 25(1): 45-52.
- Kim, Y. K, Lho, H. S, & Cho, W. D. (2011). Real-Time Step Count Detection Algorithm using a Tri-Axial Accelerometer. *The Institute of Electronics and Information Engineers*. 48: 127-137.
- LEE, M.H. (2009). Physical activity recognition using a single tri-axis accelerometer. In *Proceedings of the world congress on engineering and computer science*
- Matsuo, T., Escamilla, R. F., Fleisig, G. S., Barrentine, S. W., & Andrews, J. R. (2001). Comparison of kinematic and temporal parameters between different pitch velocity groups. *Journal of Applied Biomechanics*, 17(1): 1-13.
- Nam, J. C, Park, W. H, Cho, J. H, Park, Y. E & Kim, H. S. (2002). Case of Rehabilitation Program after Repairing of Shoulder

- Labrum for Professional Baseball Pitchers. *The Korean Journal of Sports Medicine* 20(2): 235-240.
- Park, H. S, Yoon, I. C, Lee, Y. J, & Jung, I. Y. (2015). Study on the Features of 3-axis Accelerometer Signal by Change of Sensor Positioning. 'KSPE 2015 Spring Conference., *Korean Society for Precision Engineering* 5: 918 ~919.
- Ragan, B. G. (2005). Construction of a Classification/Decision Tree. *Korean Society of Measurement and Evaluation for Physical Education and Sports Science*, 7(1), 61-75.
- Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005, July). Activity recognition from accelerometer data. *In AAAI* 5: 1541-1546.
- Siirtola, P., Laurinen, P., Haapalainen, E., Roning, J., & Kinnunen, H. (2009, March). Clustering-based activity classification with a wrist-worn accelerometer using basic features. *In Computational Intelligence and Data Mining* 2009. CIDM'09. IEEE Symposium on. pp. 95-100.
- Stanitski, C. L. (1989). Common injuries in preadolescent and adolescent athletes recommendations for prevention. *Sports Medicine*, 7(1): 32-41.
- Wang, H. K., Juang, L. G., Lin, J. J., Wang, T. G., & Jan, M. H. (2004). Isokinetic performance and shoulder mobility in Taiwanese elite junior volleyball players. *Isokinetics and Exercise science*, 12(2): 135-141.
- WEKA HP, <http://www.cs.waikato.ac.nz/ml/weka>
- Yang, J. (2009, October). Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics. pp. 1-10.