Scoring Performance on the Y-Balance Test

Vivek Mahato, William Johnston, and Pádraig Cunningham

School of Computer Science
University College Dublin
Dublin 4, Ireland
vivek.mahato@ucdconnect.ie, padraig.cunningham@ucd.ie

Abstract. The Y-Balance Test (YBT) is a dynamic balance assessment commonly used in sports medicine. In this research we explore how data from a wearable sensor can provide further insights from YBT performance. We do this in a Case-Based Reasoning (CBR) framework where the assessment of similarity on the wearable sensor data is the key challenge. The assessment of similarity on time-series data is not a new topic in CBR research; however the focus here is on working as close to the raw time-series as possible so that no information is lost. We report results on two aspects, the assessment of YBT performance and the insights that can be drawn from comparisons between pre- and post- injury performance.

Keywords: Wearable Sensors, Y Balance Test, Time Series Data.

1 Introduction

This research addresses the challenge of assessing human physical activity as measured using wearable sensors. We focus on the Y Balance Test (YBT) a test for assessing dynamic balance used in clinical and research settings [9]. We address two tasks, the task of scoring performance based on the sensor data (a regression task) and a classification task that identifies abnormal performance. We also seek to provide insight into *how* performance on a test is abnormal.

The YBT produces a normalised reach score which quantifies performance (section 3). Our first objective is to see if we can estimate this directly from the sensor data. We report on what data streams from the sensor are most effective for this. This regression task is performed using k-Nearest Neighbour (k-NN) and we analyse a number of similarity mechanisms for identifying neighbours. The motivation for our first objective is to eliminate the manual task of measuring the reach distance.

In the second part of our evaluation, we examine data from six athletes recovering from a concussion. We explore the hypothesis that an in-depth analysis of the YBT sensor data provides insight into the extent to which the individual has recovered from the concussion. While the results we report are preliminary, this seems a promising strategy.

This paper is structured as follows. The next section provides an overview of relevant research on similarity measures for time-series. The Y Balance Test is

described in section 3. Our evaluation is presented in section 4 and conclusions and directions for future work are presented in section 5.

2 Similarity Measures for Time-Series

In reviewing relevant research on similarity measures it is worth separating research in a CBR context from the wider research in this area. In the next subsections we review the dominant methods for measuring similarity on time-series before focusing on CBR research on time series at the end of this section.

2.1 Dynamic Time Warping

To find the distance between two time series, the Euclidean distance formula is an obvious choice. But when dealing with time series data where the series may be displaced in time, the Euclidean distance may be large when the two series are similar, just off slightly on the time line (see Figure 1(a)). To tackle this situation Dynamic Time Warping (DTW) offers us the flexibility of mapping the two data series in a non-linear fashion by warping the time axis [15]. It creates a cost matrix where the cells contain the distance value of the corresponding datapoints and then finds the shortest path through the grid, which minimizes the total distance between them. Sakoe-Chiba [22] global constraint is introduced to the model to increase its performance and reduce time complexity.

The following are the steps DTW executes to find the optimum mapping path with forward Dynamic Programming (DP), which provides us with the minimum distance:

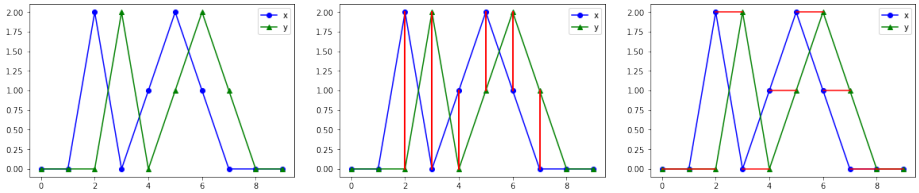
- Let t and r be two time-series vectors; then define $D(i, j)$ as the DTW distance between $t(1 : i)$ and $r(1 : j)$, with the mapping path starting from $(1, 1)$ to (i, j).
- With initial condition as $D(m, n) = |t(m) - r(n)|$, recursively calculate:

$$D(i, j) = |t(i) - r(j)))| + min \left\{ \begin{matrix} D(i-1, j) \\ D(i-1, j-1) \\ D(i, j-1) \end{matrix} \right\} \quad (1)$$

- The minimum distance then is $D(1, 1)$.

Simply said, we will construct a matrix D of dimensions $m \times n$ (where m and n are the sizes of time-series vectors t and r), and then insert the value of $D(1, 1)$ by using the initial condition. Using the recursive formula the whole matrix gets filled one element at a time, either following a column-by-column or row-by-row order. When completed, the minimum cost or distance between t and r will be available at $D(m, n)$. Thus, the computational complexity of DTW is $O(mn)$ when the time-series are unidimensional.

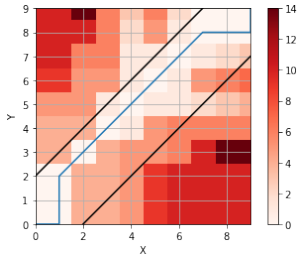
In many cases, DTW may not provide the best mapping as required as it strives to find the minimum distance which can result in forming an unwanted



(a) Two similar time-series displaced in time.

(b) Mapping of data-points without warping.

(c) Mapping of data-points with warping.



(d) DTW cost-matrix with global constraint.

Fig. 1: An example of DTW non-linearly mapping two time-series displaced in time, with Sakoe-Chiba global constraint.

path, which does not assist in discriminating two time-series belonging to different classes. Fixing this issue requires limiting the possible warping paths utilizing a global constraint. Sakoe-Chiba band [Figure 1(d)], is one of the simplest and most popular global constraints applied to DTW. The warping path then is limited to the zone that falls under the band indices. Initially, when we restrict our algorithm with no warping allowed, the data points are linearly mapped between the two data series based on the common time axis value. As seen in Figure 1(b), the algorithm does a poor job of matching the time series. But when we grant the DTW algorithm the flexibility of considering a warping window, the algorithm performs remarkably well when mapping the data-points following the trend of the time series data, which can be visualized in Figure 1(c).

2.2 Symbolic Aggregate approXimation

Several symbolic representations of a time series data have been developed in recent decades with the objective of bringing the power of text processing algorithms to bear on time series problems. Keogh et *al.* provide an overview of these methods in their 2003 paper [16].

Symbolic Aggregate Approximation (SAX) is one such algorithm that achieves dimensionality and numerosity reduction and provides a distance measure that

is a lower bound on the distance measures on the orignal series [16]. In this case numerosity reduction refers to a more compact representation of the data.

Piecewise Aggregate Approximation SAX uses Piecewise Aggregate Approximation (PAA) in its algorithm for dimensionality reduction. The fundamental idea behind the algorithm is to reduce the dimensionality of a time series by slicing it into equal-sized fragments which are then represented by the average of the values in the fragment.

PAA approximates a time series X of length n into vector $\bar{X} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_m)$ of any arbitrary length $m \leqslant n$, where each of x_i is computed as follows:

$$\bar{x}_i = \frac{m}{n} \sum_{j=\frac{n}{m}(i-1)+1}^{\frac{n}{m}i} x_j \tag{2}$$

This simply means that in order to reduce the size from n to m, the original time series is first divided into m fragments of equal size and then the mean values for each of these fragments are computed. The series constructed from these mean values is the PAA approximation of the original time series. There are two cases worth noting when using PAA. When $m = n$ the transformed representation is alike to the original input, and when $m = 1$ the transformed representation is just the mean of the original series [14]. Before the transformation of original data into the PAA representation, SAX also normalizes each of the time series to have a mean of zero and a standard deviation of one, given the difficulty of comparing time series of different scales [16,13].

(a) Raw time series data. (b) Time series after PAA dimensionality reduction. (c) Mapping of discretized bins to symbols in SAX.

Fig. 2: Symbolic Aggregate Approximation; The raw time-series in (a) will be represented by the sequence 'abfedbc' in (c) [17].

After the PAA transformation of the time series data, the output goes through another discretization procedure to obtain a discrete representation of the series. The objective is to discretize these levels into a bins of roughly equal size. These levels will typically follow a Gaussian distribution so these bins will get larger away from the mean. The breakpoints separating these discretized bins form a

sorted list $B = \beta_1, ..., \beta_{a-1}$, such that the area under a $N(0, 1)$ Gaussian curve from β_i to $\beta_{i+1} = \frac{1}{a}$. β_0 and β_a are defined as $-\infty$ and ∞ respectively [16].

When all the breakpoints are computed, the original time series is discretized as follows. First, the PAA transformation of the time series is performed. Then each of the PAA coefficients less than the smallest breakpoint β_1 is mapped to the symbol s_1, and all coefficients between breakpoints β_1 and β_2 (second smallest breakpoint) are mapped to the symbol s_2, and so on, until the last PAA coefficient gets mapped. Here, s_1 and s_2 belongs to a set of symbols $S = (s_1, s_2, ..., s_m)$ to which the time series is mapped by SAX, where m is the size of symbol pool.

SAX also has a sliding window implemented in its algorithm, the size of which can be adjusted. It extracts the symbols present in that window frame and creates a word, which is just the concatenated sequence of symbols in that frame. This sliding window is then shifted to the right and another word is extracted corresponding to the new frame. This goes on until the window hits the end of the time series, yielding a"bag-of-words" representing the series.

Once the data is converted to this symbolic representation, one can use this bag-of-words representation for calculating the distance between two time series using a string distance metric such as Levenshtein distance [26].

2.3 Symbolic Fourier Approximation

SFA was introduced by Schäfer et *al.* in 2012 as an alternative method to SAX built upon the idea of dimensionality reduction by symbolic representation. Unlike SAX which works on the time domain, SFA works on the frequency domain. In the frequency domain, each dimension contains approximate information about the whole time series. By increasing the dimensionality one can add detail, thus improving the overall quality of the approximation. In the time domain, we have to decide on a length of the approximation in advance and a prefix of this length only represents a subset of the time series [23].

Discrete Fourier Transform In contrast to SAX which uses PAA as its dimensionality reduction technique, SFA, focusing on the frequency domain, uses the Discrete Fourier Transform (DFT). DFT is the equivalent of the continuous Fourier Transform for signals known only at N instants by sample times T, which is a finite series of data.

Let $X(t)$ be the continuous signal which is the source of the data. Let N samples be denoted $x[0], x[1], ..., x[N-1]$. The Fourier Transform of the original signal, $X(t)$, would be:

$$F(\omega_k) \triangleq \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n}, k = 0, 1, 2, ..., N-1 \tag{3}$$

Simply stated, DFT analyzes a time domain signal $x(n)$ to determine the signal's frequency content $X[k]$. This is achieved by comparing $x[n]$ against signals known as sinusoidal basis functions, using correlation. The first few basis

functions correspond to gradually changing regions and describe the coarse distribution, while later basis functions describe rapid changes like gaps or noise. Thus employing only the first few basis functions yields a good approximation of the time series [23].

The DFT Approximation is a part of the preprocessing step of SFA algorithm, where all time series data are approximated by computing DFT coefficients. When all these DFT coefficients are calculated, multiple discretisations are determined from all these DFT approximations using Multiple Coefficient Binning (MCB) as shown in Figure 3. MCB helps in mapping the coefficients to their symbols, and concatenates it to form an SFA word. Thus, this converts the time series into its symbolic representation.

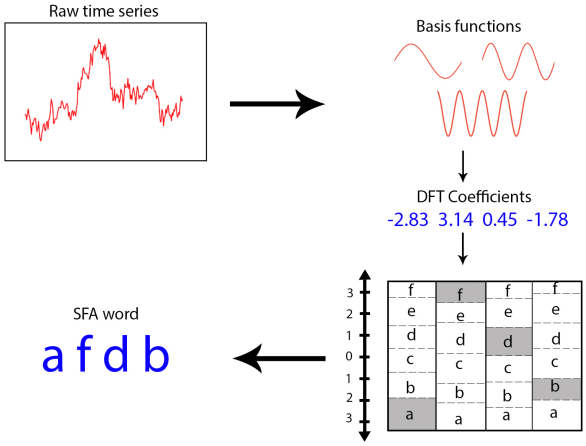


Fig. 3: Symbolic Fourier Approximation; The raw time-series will be represented by the sequence 'afdb' [17].

As in SAX, there is a sliding window present here which serves the same purpose of extracting a word representing the data in that frame. Thus, the output of SFA for a given source time series is a bag-of-words symbolically representing the entire series in lower dimension.

2.4 Time-Series Similarity in CBR

Research on temporal analysis within the CBR community has been strongly influenced by the Temporal Abstractions (TA) methodology. The idea with TA is to map low-level temporal data into higher level concepts that are meaningful for the domain in question [24]. This idea has its roots in the *The Knowledge Level*[20] view of Artificial Intelligence which fits well with the CBR paradigm. There has been significant research on TA in CBR with a particular focus on applications in medical decision support [18,19].

The objective with TA is to produce a high-level symbolic representation of the time-series that will reduce the dimension of the data by providing a high-level abstraction that fits with a *feature-value* case representation [18]. The attractiveness of TA for early CBR research was two fold, it delivered a feature value representation and it avoided the computational problem of dealing with the raw data. It is worth saying that while SAX and SFA also produce symbolic representations the motivations are different as the objective is not to produce *knowledge level* representations.

On the question of computational tractability, early work by Penta *et al.* [21] recognised the benefit of using DTW to quantify similarity on time-series for CBR but dismissed it as an option because of the computational cost.

More recently Elsayed *et al.* [5] did use DTW in CBR to classify *pseudo*-times-series in medical image analysis. This shows that the computational cost of DTW is no longer an issue. Bregón *et al.* [2] also report good results using DTW with CBR on a fault classification problem.

3 Y Balance Test

The Y Balance Test (YBT) is the most common dynamic balance assessment used within the sports medicine clinical context [6]. It requires an individual to transition from a position of bilateral to unilateral stance, perform a maximal reach excursion with the non-stance limb in three standardised directions (anterior; posteromedial; posterolateral), while maintaining controlled balance [6] (see Figure 4). The individual is then required to return to the starting position in a controlled manner. A trial is deemed a fail if they remove their hands from their hips, make contact with the ground, weight bear through the slider, raise the stance leg heel or kick the slider forward for extra distance. Participants typically complete four practice trials prior to completion of three recorded trials in each direction (randomized order), bilaterally [6].

The traditional balance score is obtained by manually measuring the distance the individual reaches outside of their base of support and normalising it to their leg length, allowing for appropriate comparison between individuals. Previous research has demonstrated the ability of this protocol to identify differences in dynamic performance between control and pathological groups, in conditions such as acute lateral ankle sprain [4] and anterior cruciate ligament injuries [7].

It has also been suggested that the YBT may have a role in evaluating concussed athletes. It can provide a means to challenge the sensorimotor subsystems of injured athletes, highlighting deficits that may increase their risk of sustaining further injury [10].

Johnston *et al.* have shown that a very good assessment of YBT performance can be obtained from a single wearable sensor. [9] Normal and abnormal balance performance can be assessed with a moderate level of accuracy.

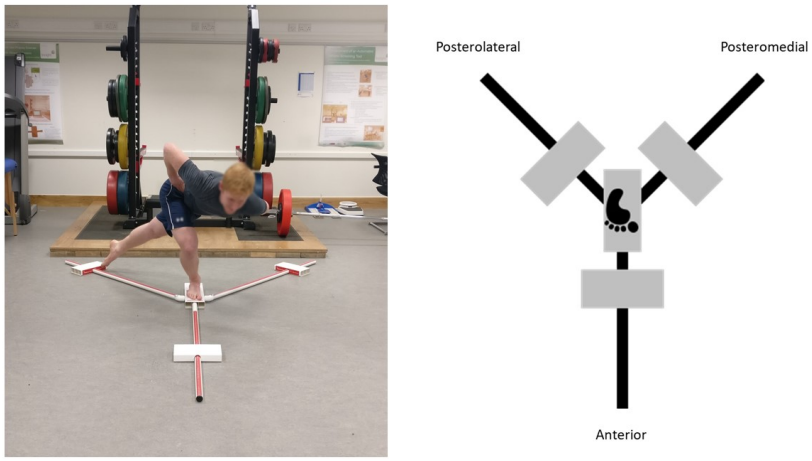


Fig. 4: A demonstration of the YBT test in operation.

3.1 The Datasets

The data set consists of data collected from two cohorts:

1. 29 young healthy adults (aged 23.3 ± 2.1 years; height 174.7 ± 9.2 cm; weight 71.6 ± 13.3 kg; left leg length 95.4 ± 4.8 cm; right leg length 95.5 ± 5.1 cm) were tested on one occasion in a university biomechanics laboratory.[1] In our evaluation we use 21 subjects for training and model/parameter selection and 8 for testing. For each subject there are 18 samples (3 trials, 3 directions, and 2 stances). So we have 378 training samples and 143 test samples (one sample is missing). The length of the time-series ranges from 114 to 645 data-points, having an average of ≈298 ticks.
2. Six elite rugby union players (aged 21 ± 1.5 years; height 182 ± 6.3 cm; weight 91 ± 15.4 kg; right leg length 95 ± 4.2 cm; left leg length 95 ± 4.22 cm) were baseline tested as part of a wider study protocol, as described in Johnston *et al.* [12]. These six athletes later went on to sustain a concussive injury, and were follow-up tested using the inertial sensor quantified YBT 48-hours post-injury and at the point of medical clearance to return to full contact training (RtP). The length of the time-series here ranges from 158 to 648 data-points, having an average of ≈345 ticks. Reliability control data was also obtained from two healthy young adults who were repeat tested on two occasions, separate by 7-10 days, as described in [11].

Ethical approval was sought and obtained from the university research ethics board, and all participants provided informed consent prior to completion of

[1] This dataset is available at `http://mlg.ucd.ie/ybt`

the testing protocol. Additional consent was provided from the young healthy participants (dataset 1) to allow open-access publication of the dataset.

3.2 Sensor Methods

The sensor used was a Shimmer3 sensor[2] that returns 10 data streams; accelerometer, gyroscope and magnetometer in three dimensions and an altimeter to provide the 10^{th} data stream. The altimeter data was not used in our study; however pitch, roll and yaw were derived from the other data streams to provide 12 streams in all. The sensor was mounted at the level of the 4^{th} lumbar vertebra, in line with the top of the iliac crests using a custom-made elastic belt. The sensor was configured to collect tri-axial accelerometer data ($\pm 2g$) and tri-axial gyroscope data ($\pm 500^{\circ}/s$) at a sampling frequency of 51.2Hz during each YBT reach excursion. The data collection procedure was consistent with previously describe methods [11,12].

4 Evaluation

In this evaluation we consider two questions:

1. Can we score YBT performance without actually measuring the reach?
2. Does a visual inspection of the sensor plots offer insights into performance?

For the first question we need to determine which data streams from the sensor are predictive of performance (section 4.1) and identify which similarity measures are best for this task (section 4.2). We take the tasks in this order, first we identify the best data streams then we tackle the similarity measures.

4.1 Feature Selection

The first task was to identify which subset of the 12 data streams would be effective for the regression task. A meta-analysis by Mitsa (2010) states that when it comes to time-series classification, 1-NN-DTW is challenging to beat. Therefore, we employ k-NN-DTW to evaluate the features (i.e. data streams) individually, based on its prediction capability of the reach distance.

The results in terms of Mean Absolute Percentage Error (MAPE) are shown for each of the 12 time-series in Figure 5. The results show that the Z-axis of the accelerometer proves to be most informative, with Y and Z-axes of the magnetometer being the next best features. A magnetometer is sensitive to external disturbances such as the earths magnetic field, the location of the experiment, or electrical systems present in proximity, therefore 'Accel Z' is selected as the single best feature because it is robust to such interference.

Next we use a Forward Sequential Search strategy [1] to see if adding other features (i.e. time-series) will improve accuracy. It is clear from the results shown in Figure 6 that the impact is minimal. Adding two features (Mag Y and Mag X) reduce the MAPE from 6.24cm to 6.04cm.

[2] `http://www.shimmersensing.com`

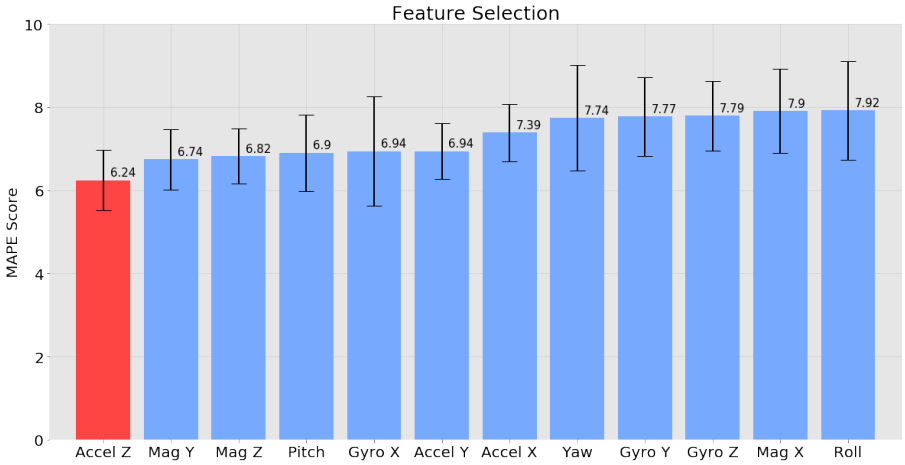


Fig. 5: Feature selection over the performance of k-NN-DTW in predicting reach distance using only one dimension.

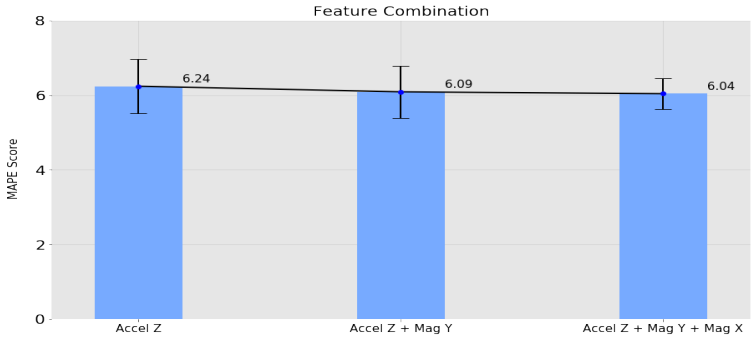


Fig. 6: Performance of kNN-DTW model when a combination of features is given.

4.2 Comparing Similarity Measures

We move on now to consider the performance of the other similarity measures (SAX and SFA) compared with DTW[3]. Given the results of the feature selection analysis we consider the Accel Z time series only.

As explained in section 2, both SAX and SFA turn time-series matching into a sequence matching problem. So we have some choices on how we measure sequence similarity. Here we consider two options, standard Edit Distance (Levenshtein Distance) [3] and the Wagner-Fischer algorithm [25][4].

[3] Similarity computation with DTW between two time-series of unequal length is handled by padding the shorter time-series with zeroes.

[4] Edit Distance and Wagner-Fischer measures requires no size matching as it handles the unequal length of the sequences.

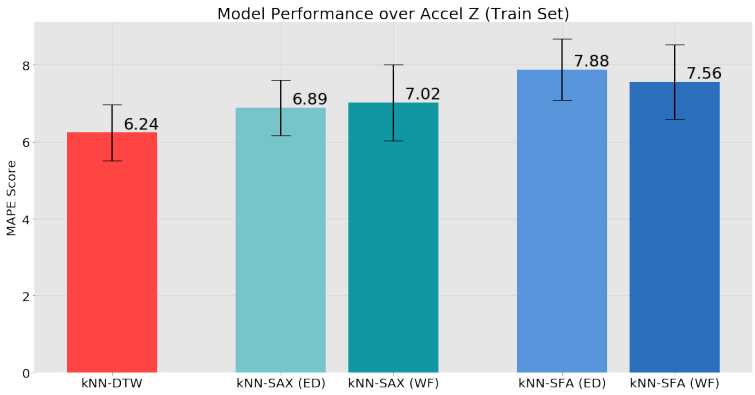


Fig. 7: Evaluation of models with 10-fold cross-validation over 'Accel Z' in the train set, and MAPE score as its evaluation metric.

The Wagner-Fischer algorithm is more nuanced than standard Edit Distance as it allows for different penalties for insertion, deletion and substitution and for distances within the alphabet to be included in the penalty score. For example, the underlying implementation of Edit distance measures the distance between "boat" and "coat" as 1, and the distance between "coat" and "goat" is also computed as 1, because these strings are only one edit away. Whereas, the Wagner-Fischer algorithm measures the distance between "coat" and "goat" as four because 'g' is 4-steps away from 'c' in the alphabet series.

We compare the three methods (DTW, SAX and SFA) used in time series regression in a k-NN model. k-NN-SAX and k-NN-SFA were evaluated on both vanilla Edit distance and Wagner-Fischer version with custom penalties. We report two sets of results, results on the training data (21 subjects) which we are using for model and parameter selection and results on the test data (8 subjects) which gives us an estimate of generalisation accuracy.

Figure 8 illustrates the performance of the kNN-models on the reach estimation task. Our conclusions are as follows:

1. k-NN-DTW beats SAX and SFA on this reach estimation task. This is consistent with earlier work that shows that DTW will beat SFA and SAX when similarity depends on the overall signal rather than local features [17].
2. Edit Distance performs better than Wagner-Fischer when used with SAX and SFA. This may be because of overfitting in the parameter setting process.

4.3 Insights

Our next objective is to see if the sensor data offers any insight into recovery from concussion. The second dataset (section 3.1) contains sensor readings from six athletes who suffered concussions. There are readings, Pre-, Post-injury and on Return-to-Play (RtP) with three readings (i.e. trials) for each category. The

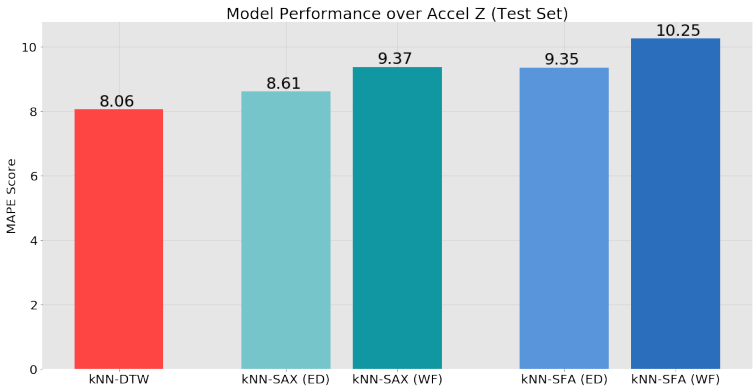


Fig. 8: Performance of the models over 'Accel Z' in the test set, with MAPE score as its evaluation metric.

Post-injury measurements were taken immediately after injury when athletes were excluded from playing due to concussion. Thus the RtP signals should be similar to the Pre- data and not the Post-.

As a baseline we have data on 'healthy' subjects which shows us what normal variations between test sessions should look like. The first two plots in Figure 9 show data on two such subjects. The data shows two sets of three repetitions measured one week apart. It is clear from the plots that the strategies are reasonably consistent, Subject 6 has a steady acceleration while subject 11 increases acceleration through the movement and then slows sharply.

The next two plots in the Figure present the picture for two of the concussed athletes. The plots for the four other concussed athletes are shown in the Appendix. The signals for the concussed athletes were selected as follows:

- We take each of the three RtP examples and calculate the similarity to the three Pre and three Post examples.
- We present the RtP signal and its closest match (Pre- or Post-) and all the non matching signals for comparison.

We would like to see RtP signals that are similar to the Pre-injury signals. We have this for Athlete 517 and not for Athlete 400. Athlete 517 has an RtP signal similar to his Pre-injury performance and different to his Post- signal (shown in red). By contrast the signal for Athlete 400 looks like his Post-injury signal. While it would be expected that both athletes dynamic balance performance should have returned to baseline levels at the point of 'clinical recovery' (return to play), concussion presentation is multi-factorial and variable in nature, where no two injured athletes present the same. Furthermore, there is increasing evidence that neuromuscular control deficits may persist beyond resolution of symptoms, increasing their risk of future injury [8]. This may help explain why athlete 400s RtP signal looked most similar to their post-injury signal, while athlete 517 appeared to have returned to pre-injury levels of performance.

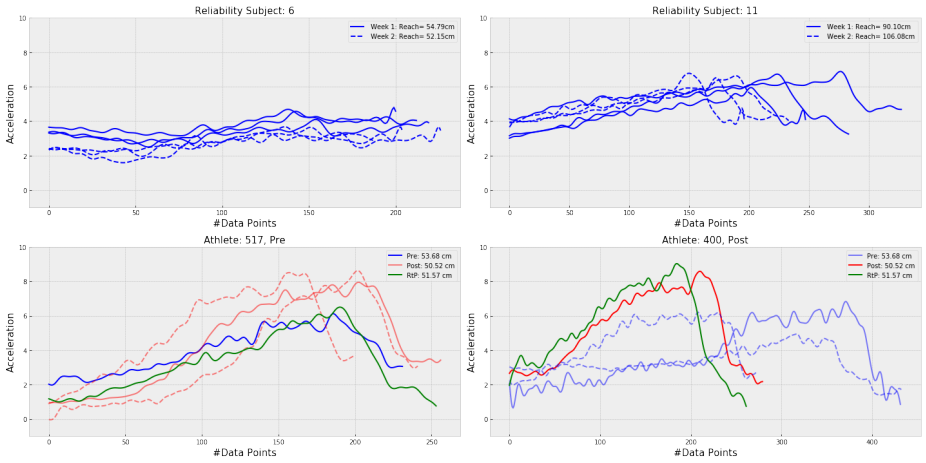


Fig. 9: A comparison between 'healthy' subjects and two subjects who suffered concussion (standing on non-dominant leg, anterior reach, Accel Z data).

5 Conclusion

Our analysis shows that using k-NN with DTW on the Accelerometer Z-axis data is effective for predicting performance on the YBT.

While the analysis relating to concussion is preliminary, we feel that the sensor data can have a role is assessing recovery from concussion. The strategy would be to gather 'healthy' baseline data during the pre-season training period, and use this to help determine when an athlete may have fully recovered post-injury. This approach may help health care professionals identify players who have not fully recovered post-concussion, facilitating the implementation of additional rehabilitation strategies to aid in the reduction of future re-injury. Due to the large degree of inter-subject variability within the data, our analysis suggests that this pre/post injury comparison needs to be player specific, and cannot be generalised between subjects.

Acknowledgments

This publication has resulted from research supported in part by a grant from Science Foundation Ireland (SFI) under Grant Number 16/RC/3872 and is co-funded under the European Regional Development Fund.

References

1. Aha, D.W., Bankert, R.L.: A comparative evaluation of sequential feature selection algorithms. In: Learning from data, pp. 199–206. Springer (1996)

2. Bregón, A., Simón, M.A., Rodríguez, J.J., Alonso, C., Pulido, B., Moro, I.: Early fault classification in dynamic systems using case-based reasoning. In: Conference of the Spanish Association for Artificial Intelligence. pp. 211–220. Springer (2005)
3. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data. Proceedings of the VLDB Endowment 1(2), 15421552 (Jan 2008)
4. Doherty, C., Bleakley, C.M., Hertel, J., Caulfield, B., Ryan, J., Delahunt, E.: Laboratory measures of postural control during the star excursion balance test after acute first-time lateral ankle sprain. Journal of athletic training 50(6), 651–664 (2015)
5. Elsayed, A., Hijazi, M.H.A., Coenen, F., García-Fiñana, M., Sluming, V., Zheng, Y.: Time series case based reasoning for image categorisation. In: International Conference on Case-Based Reasoning. pp. 423–436. Springer (2011)
6. Gribble, P.A., Hertel, J., Plisky, P.: Using the star excursion balance test to assess dynamic postural-control deficits and outcomes in lower extremity injury: a literature and systematic review. Journal of athletic training 47(3), 339–357 (2012)
7. Herrington, L., Hatcher, J., Hatcher, A., McNicholas, M.: A comparison of star excursion balance test reach distances between acl deficient patients and asymptomatic controls. The Knee 16(2), 149–152 (2009)
8. Howell, D.R., Lynall, R.C., Buckley, T.A., Herman, D.C.: Neuromuscular control deficits and the risk of subsequent injury after a concussion: a scoping review. Sports medicine 48(5), 1097–1115 (2018)
9. Johnston, W., O'Reilly, M., Dolan, K., Reid, N., Coughlan, G., Caulfield, B.: Objective classification of dynamic balance using a single wearable sensor. In: 4th International Congress on Sport Sciences Research and Technology Support 2016, Porto, Portugal, 7-9 November 2016. pp. 15–24. SCITEPRESS–Science and Technology Publications (2016)
10. Johnston, W., OReilly, M., Argent, R., Caulfield, B.: Reliability, validity and utility of inertial sensor systems for postural control assessment in sport science and medicine applications: A systematic review. Sports Medicine pp. 1–36 (2019)
11. Johnston, W., OReilly, M., Coughlan, G.F., Caulfield, B.: Inter-session test-retest reliability of the quantified y balance test. In: 6th International Congress on Sports Sciences Research and Technology Support. pp. 63–70 (2018)
12. Johnston, W., OReilly, M., Duignan, C., Liston, M., McLoughlin, R., Coughlan, G.F., Caulfield, B.: Association of dynamic balance with sports-related concussion: A prospective cohort study. The American journal of sports medicine 47(1), 197–205 (2019)
13. Keogh, E., Kasetty, S.: On the need for time series data mining benchmarks. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 02 (2002)
14. Keogh, E.J., Pazzani, M.J.: Scaling up dynamic time warping for datamining applications. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 00 (2000)
15. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: Proceedings of the 2001 SIAM International Conference on Data Mining. pp. 1–11. SIAM (2001)
16. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery - DMKD 03 (2003)

17. Mahato, V., O'Reilly, M., Cunningham, P.: A comparison of k-NN methods for time series classification and regression. In: Brennan, R., Beel, J., Byrne, R., Debattista, J., Junior, A.C. (eds.) Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, 2018. CEUR Workshop Proceedings, vol. 2259, pp. 102–113. CEUR-WS.org (2018), `http://ceur-ws.org/Vol-2259/aics_11.pdf`
18. Montani, S., Bottrighi, A., Leonardi, G., Portinale, L.: A CBR-based, closed-loop architecture for temporal abstractions configuration. Computational Intelligence 25(3), 235–249 (2009)
19. Montani, S., Leonardi, G., Bottrighi, A., Portinale, L., Terenziani, P.: Supporting flexible, efficient, and user-interpretable retrieval of similar time series. IEEE Transactions on Knowledge and Data Engineering 25(3), 677–689 (2013)
20. Newell, A., et al.: The knowledge level. Artificial intelligence 18(1), 87–127 (1982)
21. Penta, K.K., Khemani, D.: Satellite health monitoring using CBR framework. In: European Conference on Case-Based Reasoning. pp. 732–747. Springer (2004)
22. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 26(1), 43–49 (February 1978)
23. Schäfer, P., Högqvist, M.: SFA: a symbolic fourier approximation and index for similarity search in high dimensional datasets. In: Proceedings of the 15th International Conference on Extending Database Technology. pp. 516–527. ACM (2012)
24. Shahar, Y.: A framework for knowledge-based temporal abstraction. Artificial intelligence 90(1-2), 79–133 (1997)
25. Wagner, R.A., Fischer, M.J.: The string-to-string correction problem. Journal of the ACM 21(1), 168173 (Jan 1974)
26. Yujian, L., Bo, L.: A normalized levenshtein distance metric. IEEE transactions on pattern analysis and machine intelligence 29(6), 1091–1095 (2007)

Appendix

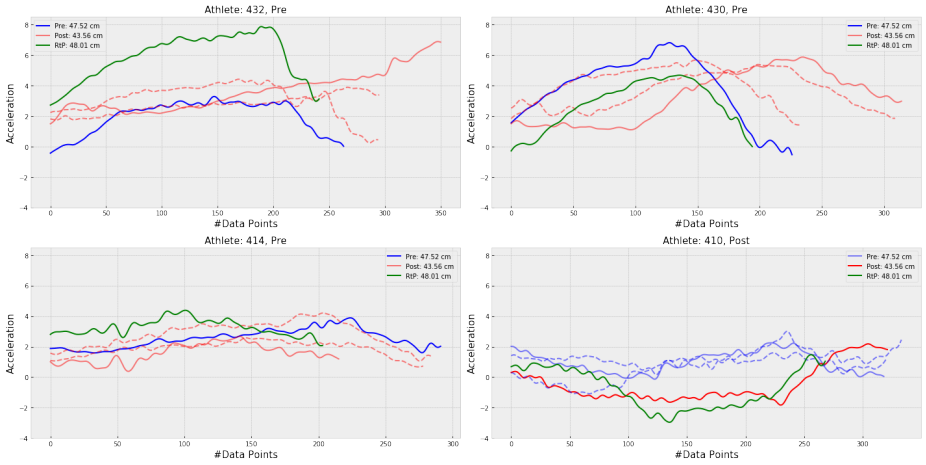


Fig. 10: A comparison between the rest four athletes in the dataset who suffered concussion. (Standing on non-dominant leg, anterior reach, Accel Z data.)