Proceedings of AICS-2011, the 22nd Conference on Artificial Intelligence and Cognitive Science, Derry, August 2011.

Activity Recognition based on Accelerometer Data using Dynamic Time Warping with Ensembles

Conor O'Rourke, Michael Madden

College of Engineering and Informatics,
National University of Ireland, Galway, Ireland
c.orourke3@nuigalway.ie michael.madden@nuigalway.ie

Abstract. Presented here is an approach to predict what type of motion or activity a person is performing by using an Ensemble Dynamic Time Warping (DTW) Classifier based on data from accelerometers worn or used by a person. DTW is a method of comparing sequences of time series data, in which classifications are typically made with the *k*–Nearest Neighbour (*k*–NN) algorithm. We propose the use of an ensemble of DTW classifiers, to make greater use of the accelerometer data available. We construct separate classifiers for multiple parallel streams of data, each of which is a member of a committee that votes on classifications to make a final decision. We evaluate our approach using a dataset consisting of eight hand gestures from a single three axis accelerometer embedded into a remote device. Classification accuracy ranges from 67.8% to 77.3% when testing individual axes. When the three axes are used in an ensemble, accuracy increases to 86.1%.

Keywords: Ensemble, Dynamic Time Warping, Activity Recognition.

1 Introduction

Dynamic Time Warping (DTW) is an algorithm for comparing time-series, which has historically been used for speech recognition [1]. Nowadays it is often used in resource-constrained devices for activity recognition applications. In this work, we focus on ensembles of DTW classifiers, which have not been researched very much. Our motivation is to improve the classification accuracy of activity recognition applications. DTW computes similarities between two sequences of time series data and returns a distance value. The lower this value, the better the match, and a distance of zero means the sequences are identical. In order to classify an unknown sequence or query, it is compared with a number of different known template sequences. It is then classified as having the class as the template sequence to which it is most similar, using the *k*–Nearest Neighbour (*k*–NN) algorithm. The sequences do not need to be of equal length to use DTW, and the features of the sequences do not have to be exactly aligned at the same points in time. For this reason, DTW can be a useful algorithm for applications in activity recognition, as people will not perform the same activity precisely the same every time.

The use of an ensemble involves taking data streams from multiple sensors or multiple streams of data from one sensor (X, Y, Z axes) which all have corresponding time stamps. Each of these data streams will be used individually to perform a classification of the type of activity being performed, using DTW and *k*–NN. All the classifications made are combined in the ensemble, which takes the majority class to be the final decision.

Activity recognition based on motion sensor devices such as accelerometers is an area of research with application to a wide range of activity types. Recently a large number of applications have been developed for computer games and interactive devices such as smart phone apps and the Nintendo Wii, with Nintendo selling over 86 million Wii consoles since its release [2]. Another important application in this area is in effective monitoring of elderly and post-operative patients. The number of elderly people (defined as 60 years of age or older) is increasing rapidly. By 2050 there will be an elderly population of almost 2 billion which will outnumber the number of children [3]. Consequently, the average population age is also increasing [4], meaning there will be a rise in the portion of elderly compared with young or middle-aged people. This increase of older people will naturally increase the strain on the working population to care for the elderly around the world. Further strain will also be added to the care of the elderly as they will be mostly based in rural areas [5]. Many countries are already experiencing increasing costs in the healthcare system and predict it to worsen [6] [7]. A real time activity monitoring system would help reduce the need for medical staff to continually monitor patients and elderly people, thus reducing medical costs and improving people's quality of life. Our research into this area is primarily focused on developing the technology with the aim of it being applied to the medical field. In this paper the uWave dataset (See Section 2) has been used to evaluate the ensemble DTW method, which helps to illustrate that the technique is broadly applicable.

2 uWave: Gesture Recognition Dataset

The uWave application is designed to be embedded into a handheld device with a single 3-axis accelerometer. The primary purpose of this dataset is to distinguish between eight different gestures. The gestures in the dataset, as in Figure 1, were originally developed by a Nokia Research Group [8]. The dataset has been generated by the Rice University and Motorola Labs [9] with eight right-handed subjects using a Nintendo Wii remote as the device with the embedded accelerometer. It was collected over a number of weeks. The accelerometer itself is an ADXL330 model from Analogue Devices Inc. [10]. It can measure ±3g for each axis at up to 1600Hz. The particular version of the dataset presented by this paper was provided by Keogh [11]. It has been separated into X, Y and Z streams with 896 training objects and 3582 query objects in each. Each object has been processed to make each one have the same length of 315 points in time series. However prior to being processed the objects range from 124 to 315 points in time series. The i^{th} instance in the X axis is also the i^{th} instance for the Y and Z axes. The data has in addition been pre-processed using the standard z-score technique. This is a normalization method in which each time series in the dataset is normalized to have a mean of 0 and a standard deviation of 1. This dataset was created under controlled conditions; therefore the tilt of the device is not an issue for concern with this data. We are using the dataset for user-independent learning: the objects included are from all eight users.

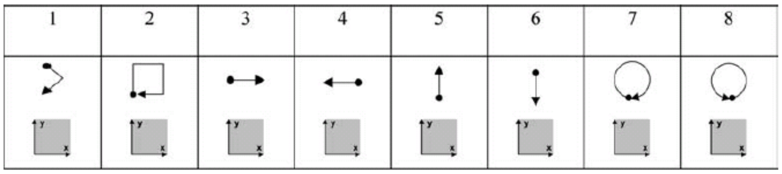


Figure 1. Gestures from the uWave dataset performed with a Nintendo Wii [8]

3 Related Research

DTW can be applied to a range of activity recognition applications and described here illustrates relevant experiments conducted on the dataset in question and other datasets. Past research of the ensemble DTW technique is also highlighted.

The Rice University and Motorola Labs [9], who created the uWave dataset, conducted experiments by combining the X, Y and Z values to find the magnitude at each point in time; therefore there was only one stream of data tested. Their preprocessing approach is different to that used in the majority of time series preprocessing techniques used, as they chose to quantize the data: discrete integer values are assigned based on the range the acceleration values fall under. This improved computation efficiency as it reduced the number of floating point computations. From that study it was found that it is difficult to classify activities correctly with user-independent tests. Even for user-dependent tests, they found that it is more difficult to correctly classify when tested on samples over multiple days as opposed to on the same day. Their accuracy with user-dependent tests conducted on samples from the one day reached 98.4%. However, when they performed tests with user-independent samples the accuracy reduced to 75.4%, due to variations in the way one person will perform a gesture compared with another person.

Keogh [11] has also tested the uWave dataset under the same conditions described in Section 2. In that study, the X, Y and Z axes were tested individually but the results were not combined. When using the Euclidean distance (this is the sum of the squares, it does not allow any warping i.e. points directly opposite each other are matched) they reported accuracies ranged from 65% to 73.9%, with an average accuracy of 68.3%. When testing with DTW and no warping window, the accuracies were reported to range from 65.8% to 72.8%, with an average accuracy of 67.3%. When optimized warping windows were used, they reported that the accuracy ranged from 67.8% to 77.3%, with an average accuracy of 71.7%.

There are limited examples of ensemble DTW classifiers researched. McGlynn & Madden [12] have researched recognition of Activities of Daily Living (ADL) from the Place Lab dataset [13]. This consists of a single user wearing accelerometers on the wrist, hip and thigh. Six different ADLs were identified, these are: Dressing, Washing, Preparing Food, Bathroom use, Phone use and Computer use. Each accelerometer had the X, Y and Z values combined to find the magnitude.

However because there are three accelerometers, there are still three streams of data available for testing. Muscillo et al. [14] have also used an ensemble DTW. The dataset contains five subjects, each with a 3 axis accelerometer on the wrist and a dual axis accelerometer on the arm near the shoulder, which means there are five streams available for testing with an ensemble. Also the use of a dual axis accelerometer as opposed to a three axis accelerometer on the shoulder means that there are five streams instead of six. This means for ensemble classification there will be less chance of conflict and gives a bias towards the decisions made by the wrist. In that application, there are five different activity types: Reaching and Grabbing, Drinking from a Jar, Writing with a Pencil, Stacking Pieces, and Locking a Door. Each of these activities was performed at two different speeds. The results from the study by McGlynn & Madden have accuracies ranging from 45.5% to 57.2% when testing individual sensors, however with the use of an ensemble accuracy increases to 84.3%. The study by Muscillo et al. reported an average accuracy of 68.7%, increasing to 76.6% with an ensemble when the tests were user-independent and included both slow and fast samples. When the tests were user-dependent and include both slow and fast samples, they reported that the results averaged 89.4% on individual sensors, and increased to 92.4% when using an ensemble.

4 Methodology

4.1 Dynamic Time Warping

As discussed in the Introduction, DTW is used to match similarities between two sequences (also referred to samples or objects) of time series data such as in Figure 2.

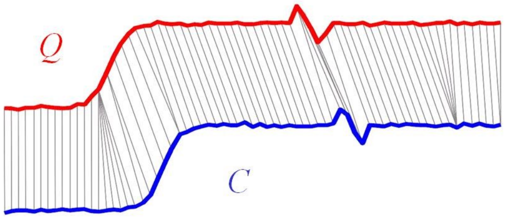


Figure 2. Dynamic Time Warping. Note how the similarities in shape are matched despite occurring at offset times. Source: [15]

It operates on two sequences, denoted Q and C, where $Q = \{q_1,q_2,...,q_N\}$ and $C = \{c_1,c_2,...,c_M\}$. A distance matrix d is calculated by finding the distance between each point of the two sequences where $d(n,m)=(q_n-c_m)^2$. The distance matrix is then used to calculate a cumulative distance matrix D. In the first row and first column, the cumulative distance at the previous step is added to the current value:

$D(n,1)=d(n,1)+D(n\text{-}1,1)$ and $D(1,m)=d(1,m)+D(1,m\text{-}1)$, respectively. The interior of the matrix is calculated by adding the current value to the smallest of the surrounding previous cumulative three distances: $D(n,m)=d(n,m)+min[D(n\text{-}1,m\text{-}1), D(n\text{-}1,1), D(1,m\text{-}1)]$. The value $D(N,M)$ is the minimum cumulative distance through the matrix.

When an unknown query is tested against every template, on each occasion the minimum distance is recorded and stored. After all the necessary comparisons have been made the *k*–NN algorithm is used, to choose the closest match. Here, *k* denotes the number of closest neighbours that will be used for classification. In this work we use *k*=1, which is the most often-used setting in conjunction with DTW, meaning that only the closest value is considered. We select 1–NN as the classification algorithm as it is a straightforward method but still a reliable classifier. Also many other DTW-based algorithms use this classifier, thus allowing for direct comparison with other research, such as [16]. Of the stored values, the smallest minimum distance is chosen to be the classification as it is the closest to zero (an identical match). This prediction will be used as one of the base classifiers for the ensemble.

Note that there are many techniques that can be employed to improve the accuracy of the DTW algorithm. One of the most commonly applied techniques is to use a warping window, of which there are a number of different types. However they typically all perform the same function which is exclude the outer sections of the distance matrix. This prevents extremely different sections of the two curves from being matched, and it also improves computational efficiency. In this paper, we use the Sakoe-Chiba band (SC band) [1]. The SC band works by only allowing the points to be matched a specific distance away from each other throughout the sequences, whereas other methods such as the Itakura Parallelogram allows large warping in the centre of the sequences but very little at the beginning and end. The SC band was chosen as it is the most commonly used constraining method which allows the proposed method to be tested under the same conditions as previous work, thus reducing variability between tests done and allowing results to be properly benchmarked.

4.2 Pre-Processing

Before DTW can be run, the time series sequences *Q* and *C* may be pre-processed. This is an essential part of data mining, in order to reduce noise, eliminate unnecessary data, keep attributes in a consistent format, and so on. The pre-processing techniques used on time series data of this type include:

- Down-sampling: This may be required as accelerometers can return approximately 60 samples per second. At this frequency, the time samples would most likely pick up meaningless motion data such as body tremors which has nothing to do with the overall motions of the activity or gesture. It would also be computationally inefficient to take so many data points into account.
- Interpolation: This is required so that all the time points are at regular intervals. While DTW is able to deal with time sequences of different lengths and it is not absolutely necessary to keep the time points at exact regular intervals, it is good

practice to keep the points consistent and will help to reduce the possibility of errors occurring in the algorithm.

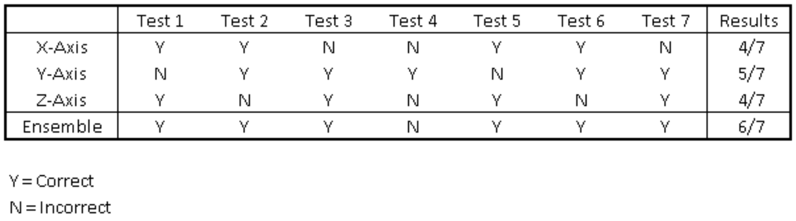
- Moving average: Also known as a rolling average, this is used to smooth the curve so that the general shape of the activity is used, while reducing noise and irrelevant spikes. A moving average can be applied in a number of different ways. The equation can be weighted to increase or decrease how much the curve will be smoothed. Too much smoothing will make different curves indistinguishable from each other, while too little will not eliminate noise.
- Normalization: This is a technique that takes all the values in a given set of data and puts them in a range of 0 to 1 proportionally. The highest value is assigned the value 1 and the lowest is assigned the value 0. The shape of the curve is not altered but the two sequences will then be compared on the same scale.

4.3 Ensemble

A machine learning ensemble is a form of statistical modeling. Each member (base classifier) of the ensemble makes its decisions independently of all others and without any weighting on the base classifiers. Although in reality not all classifiers may be equally accurate, simple schemes such as this give better results than would be intuitively expected in real world applications [17].

Condorcet's Jury Theorem describes mathematically how an ensemble works. Provided the votes (classifications made by base classifiers in this case) are independent of each other and better than random, then the greater the number of voters / base classifiers employed, the closer the accuracy will get to 100% [18]. Of course, an infinite number of classifiers that cannot be implemented as each accelerometer will return only three streams of data. Nonetheless, as long as the individual classifiers have high accuracy and are independent, we can assume that their errors will occur at different points in time [19]. Table 1 gives a hypothetical example of how an ensemble will improve the accuracy with the use of multiple classifiers.

Table 1. Hypothetical example of how an ensemble works to improve the accuracy with multiple models.



	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Results
X-Axis	Y	Y	N	N	Y	Y	N	4/7
Y-Axis	N	Y	Y	Y	N	Y	Y	5/7
Z-Axis	Y	N	Y	N	Y	N	Y	4/7
Ensemble	Y	Y	Y	N	Y	Y	Y	6/7

Y = Correct
N = Incorrect

5 Experiments and Results

In order to test the ensemble the techniques described in Section 4 must be applied appropriately. Our goal was to reproduce the results reported by Keogh [11] and follow on by using an ensemble to improve the accuracy.

5.1 Experimental Procedure

In order to correctly reproduce the results attained by Keogh [11], the same data was used. This meant the data had been preprocessed to make each object the same length and normalized with the standard z-score technique, these are described in Section 2. As noted earlier, the data is separated into a training set and a test set (which was also kept consistent with Keogh's data). Each test object is compared with all the training objects using DTW, which computes the minimum distance for every comparison. The lowest of all these minimum distance values is selected, and the corresponding training label is chosen as the classification prediction for the particular test. This is done for each test object. If a particular prediction is correct, the corresponding test label should be the same as the prediction. The accuracy can be tested be checking all the predictions.

The next step is to implement an ensemble where each prediction is used a base classifier and combined to make a final decision. As there are three streams in this case, if two or all three of the base classifiers have made the same prediction then this is used as the ensemble's prediction. However there are cases when all three base classifiers will make different predictions. In this situation the tie is broken by using the smallest absolute distance of the three base classifiers as the final decision.

Before the uWave dataset was tested with the described method, we used two other datasets to validate that our DTW algorithm implementation was correct (but without ensemble calculations, as these datasets have just one time series). These were the Gun-Point dataset [20] and the Trace dataset [16]. Our approach returned the same results as have been previously reported on these datasets.

5.2 Results and Analysis

We successfully replicated the individual results that were previously reported for the uWave X, Y and Z data [11]. Following from this, we combined their predictions in an ensemble to make overall predictions. Using the DTW Ensemble approach improved the accuracy for each variant of the algorithm that we considered:

- Euclidean Distance
- DTW with no warping window
- DTW with an optimized warping window

This can be seen in Table 2, which presents the results of classifying using each of the three streams (X, Y, Z) individually, the average accuracy of the three streams, and the performance of the DTW Ensemble.

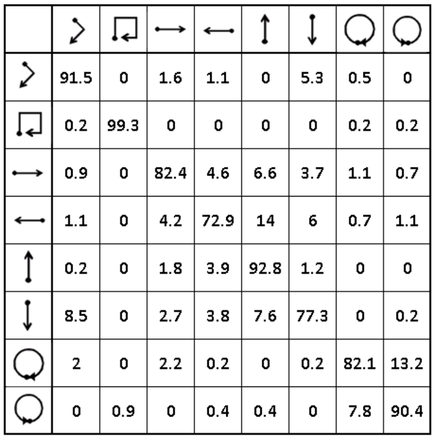
Table 2 Accuracy of tests performed on the uWave dataset

uWave (% accuracy)	Euclidean Distance	DTW No Window	DTW (% Window)
X	73.93	72.75	77.33 (4)
Y	66.16	63.40	69.91 (4)
Z	64.96	65.83	67.78 (6)
Average	68.35	67.33	71.67
Ensemble	84.45	82.75	86.07

From Table 2, we observe that the DTW Ensemble gives superior results than the single best stream in all cases, meaning that the weaker streams do not adversely affect the accuracy of the ensemble as a whole. This indicates that the ensemble performs as described in Section 2.3: it appears that the inaccuracies occurred at different points in time, and tend to be overruled by the majority vote, allowing for better final classifications.

Taking the most accurate classifier, which is the DTW with optimized warping windows, the DTW Ensemble's accuracy is 8.74% higher than that of the X axis and 14.4% higher than the average. Also, its accuracy is 10.67% higher than that reported in previously published research on the same data [9]. In 528 out of 3582 test cases (14.7%), the tie-breaker was needed.

The confusion matrix in Figure 3 shows the breakdown of percentage classifications for each gesture. This illustrates where misclassifications have occurred. In many cases there are reasonable explanations for these, as many of the classes have somewhat similar movements. In particular, Classes 3, 4, 5 and 6 are all single one-directional movements; classes 1 and 6 both involve downward movements and classes 7 and 8 are both circular movements. All these classes show some degree of confusion between each other. Class 2, on the other hand, is arguably the most complex and clearly defined gesture and as a result it has the highest accuracy rate of 99.3%.



	ᐳ	⮐	⟶	⟵	↑	↓	↻	↺
ᐳ	91.5	0	1.6	1.1	0	5.3	0.5	0
⮐	0.2	99.3	0	0	0	0	0.2	0.2
⟶	0.9	0	82.4	4.6	6.6	3.7	1.1	0.7
⟵	1.1	0	4.2	72.9	14	6	0.7	1.1
↑	0.2	0	1.8	3.9	92.8	1.2	0	0
↓	8.5	0	2.7	3.8	7.6	77.3	0	0.2
↻	2	0	2.2	0.2	0	0.2	82.1	13.2
↺	0	0.9	0	0.4	0.4	0	7.8	90.4

Figure 3: Confusion matrix for the DTW Ensemble that uses optimized warping windows. The rows illustrate the percentage classification of predictions made for each class. The average accuracy is 86.1%.

6 Conclusions and Future Work

This research demonstrates that combining multiple streams of data analyzed with DTW to form a DTW Ensemble increases the accuracy of activity recognition applications. In the application that has been studied here, when benchmarked against other similar tests on the same data, the accuracy of the DTW Ensemble is better on every occasion. The accuracy outperformed the single best-performing axis by 8.74%, outperformed the average of the three axes by 14.4% and outperformed by 10.67% with a classifier that used the magnitude of the three axes combined [9].

Based on the literature, it is clear that the task of recognizing movements correctly when they are performed by multiple subjects is a challenging one. We plan to conduct further research into the use of the DTW Ensemble by evaluating it on the Place Lab dataset that was described in Section 4. This will be a good dataset to test our method as it uses multiple sensors. Previously, McGlynn & Madden [12] used an ensemble with three streams using the magnitude for each sensor. However, we propose to use nine streams independently in a DTW Ensemble. The accuracy achieved previously on this dataset with ensembles was 84.3% with only one subject and six different classes. This illustrates the difficulty in classifying activities of daily living, as opposed to short, clearly-defined gestures. However we believe the accuracy can be improved on this dataset by optimizing the parameters such as the warping window, the sampling frequency, object length, ensemble method, etc.

Overall, we conclude that the DTW Ensemble method improves the accuracy of time series data classification and should be considered for gesture and/or activity recognition applications, particularly when the device in question will be used by multiple users.

References

1. Saoke, H. and Chiba,: Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 26, No. 1, pp. 43-49. (1987)
2. Nintendo: Investor Relations Information. Nintendo. March 2011. [Accessed: June 24, 2011.] http://www.nintendo.co.jp/ir/en/library/historical_data
3. United Nations: Population Ageing and Development 2009. Department of Economic and Social Affairs, Population Division. [Accessed: May 01, 2011.] http://www.un.org/esa/population/publications/ageing/ageing2009chart.pdf
4. Lutz, W., Sanderson, W. and Scherbov, S.: The coming acceleration of global population ageing. Nature 451: 716-719. (2008)
5. Central Statistics Office (CSO). Ageing in Ireland. Dublin : Central Statistics Office. (2007)
6. Orszag, P.R.: The Long-Term Budget Outlook and Options for Slowing the Growth of Health Care Costs. Congressional Budget Office. WASHINGTON, D.C. (2008)
7. Brockmann, H. and Gampe, J. The cost of population aging: forecasting future hospital expenses in Germany. Max Planck Institute for Demographic Research. Rostock. (2005)
8. Kela, J., Korpipaa, P., Mantyjarvi, J., Kallio, S., Savino, G., Jozzo, L., Di Marca, S.: Accelerometer-based gesture control for a design environment. Personal Ubiquitous Comput., Vol. 10, No. 5., pp. 285-299 (2006).
9. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer-based Personalized Gesture Recognition.Pervasive Computing and Communications, IEEE International Conference, Vol. 0, pp. 1-9. (2009)
10. Analog Devices: ADXL330. Analog. [Accessed: Apr 06, 2011.] http://www.analog.com/static/importedfiles/data_sheets/ADXL330.pdf
11. Keogh, E. Unpublished data in private communication. (Nov 2010)
12. McGlynn, D. and Madden, M.G.: An Ensemble Dynamic Time Warping Classifier with Application to Activity Recognition. Thirtieth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge. (2010)
13. MIT House_n. PlaceLab Datasets. House_n. [Accessed: Oct 20, 2010.] http://architecture.mit.edu/house_n/data/PlaceLab/PlaceLab.htm
14. Muscillo, R., Schmin, M. and Conforto, S.: The Median Point DTW Template to Classify Upper Limb Gestures at Different Speeds. IFMBE, pp. 63-66. (2008)
15. Ratanamahatana, C.A. and Keogh, E.: Making Time-series Classification More Accurate Using Learned Constraints. SIAM International Conference on Data Mining. (2004)
16. Keogh, E.: UCR Time Series Classification/Clustering Page. Eamonn Keogh. [Accessed: May 09, 2011.] http://www.cs.ucr.edu/~eamonn/time_series_data/
17. Witten, I.H. and Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers, San Francisco. (2005)
18. Rokach, L.: Ensemble-based classifiers. Artificial Intelligence Review, pp. 1-39. (2010)
19. Opitz, D. and Maclin, R.: Popular Ensemble Methods: An Empirical Study. Journal of Artificial Intelligence Research, Vol. 11. pp. 169-198. Montana/Minnesota (1999)
20. Ye, L. and Keogh, E.: Time Series Shapelets: A New Primitive for Data Mining. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 947-956. (2009)