International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 4, No. 1, 15-27

On Enhancing the Accuracy of Nearest Neighbour Time Series Classifier Using Improved Shape Exchange Algorithm

Imen Boulnemour, Bachir Boucheham, and Abdelmadjid Lahreche

Université 20 août 1955 de Skikda
BP 26, Skikda, DZ21000, Algérie
boulnemourimen@hotmail.fr
bachir.boucheham@hotmail.com
majid.lahreche@gmail.com

Abstract. Several methods have been proposed for time series alignment and classification. In particular our previously published method I-SEA (Improved Shape Exchange Algorithm) has been proposed as a rival method to the SEA (Shape Exchange Algorithm) method for time series alignment. The aim of this work is to improve the accuracy of the SEA method for time series classification by proposing a 1NN-ISEA (1 Nearest Neighbor-Improved Shape Exchange Algorithm) classifier. Results of the proposed method show to be better as compared to those of the 1NN-SEA and the 1NN-ED classifiers (Euclidian Distance). All results have been obtained using the UCR (University of California at Riverside) time series Dataset, universally admitted as the first Benchmark in time series classification and clustering.

Keywords: Time Series · Classification · KNN · SEA · I-SEA · ED.

1 Introduction

Time series are sequences of numbered observations of the same phenomenon, ordered over time. They can be biomedical (ECG, EEG, etc.), natural (atmospheric, seismic, etc.), technological (traffic density in a network, etc.), to cite some examples of time series natures and domains.

Time series applications include, among others, handwritten number recognition [1], abnormality detection in physiological signals [2, 3], time series clustering and compression [4], time series classification [8, 9], time series alignment [5, 6], speech recognition [10], alignment of gene expression [11], classification of cardiac arrhythmias [12].

Now, as a process, time series classification consists in assigning a time series to one of the classes present in a predefined set [13]. For this specific task, several methods are used in the literature to classify time series. The categorization of these approaches established by Amr [14] includes feature-based classification, distance-based classification and model-based classification. Feature-based classification discretizes data into symbolic representations such as Boss approach(Bag-of-SFA-Symbols) [15] or directly extract descriptive features such as mean or standard deviation to input them into a classifier such as Support Vector Machine (SVM) [16], Convolutional Neural Networks (CNN) [17] or decision trees [18]. These classifiers require the point of view of experts. In this family, there is also the Ensemble of classifiers such as HIVE-COTE (the collective of transformation-based ensembles) [19]. They combine the results of several classifiers and are known to have very good classification results, but they often have very long execution times [20].

Distance based classification uses numerical time series with similarity measures such as Euclidean distance, DTW (Dynamic Time Warping) similarity measure [21, 22] and generally a One Nearest Neighbor classifier (1NN). The third family concerns model-based approaches such as Hidden Markov Models (HMM) [23]. These approaches build a model for the time series within a cluster (class in our case) and classify the new series according to the model that best characterizes them [24]. The difficulty with these latter approaches is learning initial states and transition probabilities [25].

Several studies show that a simple 1NN classifier used with an appropriate similarity measure is more effective and surpasses many sophisticated classification approaches [26]. In particular, the Euclidean distance (ED) is reported as being one of the most used distances for the KNN classifier [27]. The ED distance is defined by the following equation (1).

$$D_{Euc}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

In another context, the SEA (Shape Exchange Algorithm) [28, 7] and the ISEA (Improved Shape Exchange Algorithm) [5, 6] have been previously proposed as competitive similarity measures for time series alignment.

The aim of this work is to improve the accuracy of the SEA method for time series classification by proposing a 1NN-ISEA (1 Nearest Neighbor Improved Shape Exchange Algorithm) classifier. We will conduct experimental comparisons of the 1NNED, 1NN-SEA and 1NN-ISEA classifiers on the universal benchmark UCR (University of California at Riverside), universally used to test and evaluate time series classification and clustering algorithms [29].

The reminder of this paper is organized as follow. In the second section, we will give an overview of main related works. In the third section, the proposed method 1NN-ISEA for the classification of time series is presented and described. The 4th section is devoted to experimental tests. Finally, Section 5 concludes this paper and gives perspectives of our research.

2 Related works

2.1 K-Nearest Neighbour Classifieur

Shape Exchange Algorithm The K nearest neighbour (KNN) is considered one of the top 10 methods in datamining [30]. It usually yields efficient performance and, in certain cases, its accuracy is greater than state-of the-art classifiers [31].

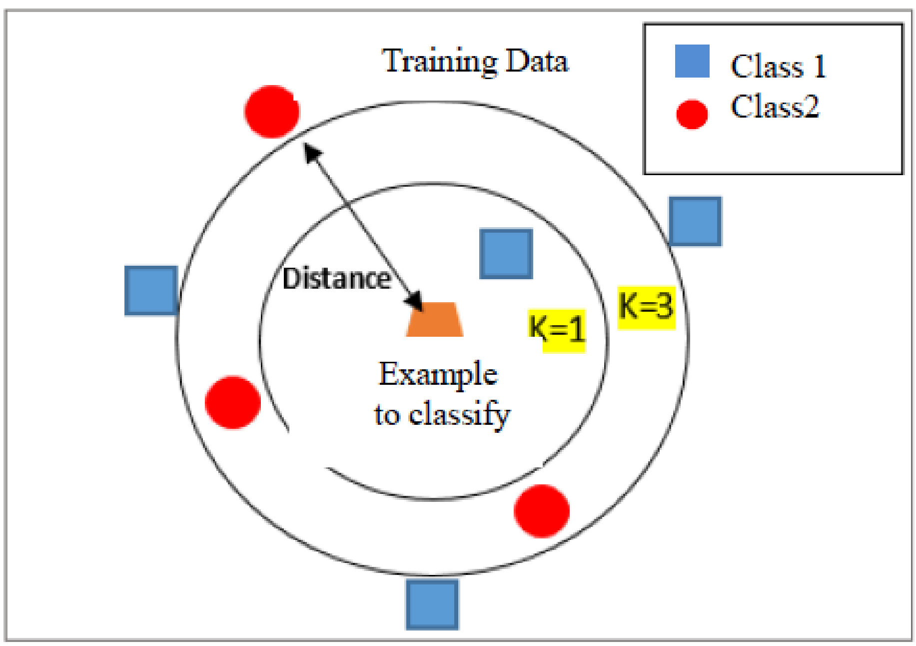

Fig. 1. Illustration of the KNN method: The method classifies the trapezoid in class 1 (square) because it is closer to it. For k = 3, the new example would be placed in class 2(circle) because it has the most nearest neighbors in the K value set.

The idea behind the KNN approach, introduced by Fix and Hodges [32], consists in assigning a new element whose label is unknown to the class of which belongs the majority of the elements which are the closer in the sense of a given metric. This method, by its simplicity and efficiency, is still very widespread.

In the example illustrated by Fig. 1, we have two classes of elements (blue squares and red circles), and we try to place a new element (represented by the brown trapezoid) in one of these classes. We look at the "nearest" neighbors and vote on what the more recent element is. If we take k = 1, the new example would be placed in class 1, for k = 3, the new example would be placed in class 2.

2.2 Shape Exchange Algorithm

SEA method is a robust similarity measure. It was proposed by Boucheham [28] and it was shown to be more appropriate than the famous DTW similarity measure for the matching of quasi-periodic time series [33]. It is be worth noting that quasi-periodic time series are concatenations of quasi-similar forms called periods (Cycles) [28]. The key of this method is the exchange of signatures (sorted magnitude axes) between the time series to be matched. The rationale behind this is that the sorting of time series on magnitude confers them a kind of a robust characteristic signature. More formally, let X and Y be two times series. The principle of the SEA method is as follows:

1. Sort X and Y according to their magnitude indexes. This operation is performed on both magnitude and temporal indexes in the sense that each magnitude entry is associated with a temporal entry;

2. Exchange sorted magnitude indexes between X and Y without changing their temporal indexes which are in disorder. The linear interpolation is used here to shrink or to expand magnitude indexes to match with the target temporal indexes in the case where X and Y have unequal lengths;

3. Sort X and Y on their temporal indexes to form the reconstructed versions of X and Y, namely: XREC and YREC;

4. Compare X with XREC and Y with YREC. For example, using a distance measure like Euclidean Distance (Eq.1) which is the case in our approach.

3 Proposed method

Our proposed method for the classification of time series is a 1NN-ISEA classifiers. It combines the I-SEA method with the 1-NN classifier.

3.1 Improved Shape Exchange Algorithm

Improved Shape Exchange Algorithm (I-SEA) is a similarity measure. It combines the SEA [28] and the DTW [21] methods. It was especially proposed by Boulnemour et al [5] and [6] for the alignment of quasi-periodic time series and especially for heartbeats time series provided from the ECG (Electrocardiogram) records. I-SEA method is described in Fig.2.

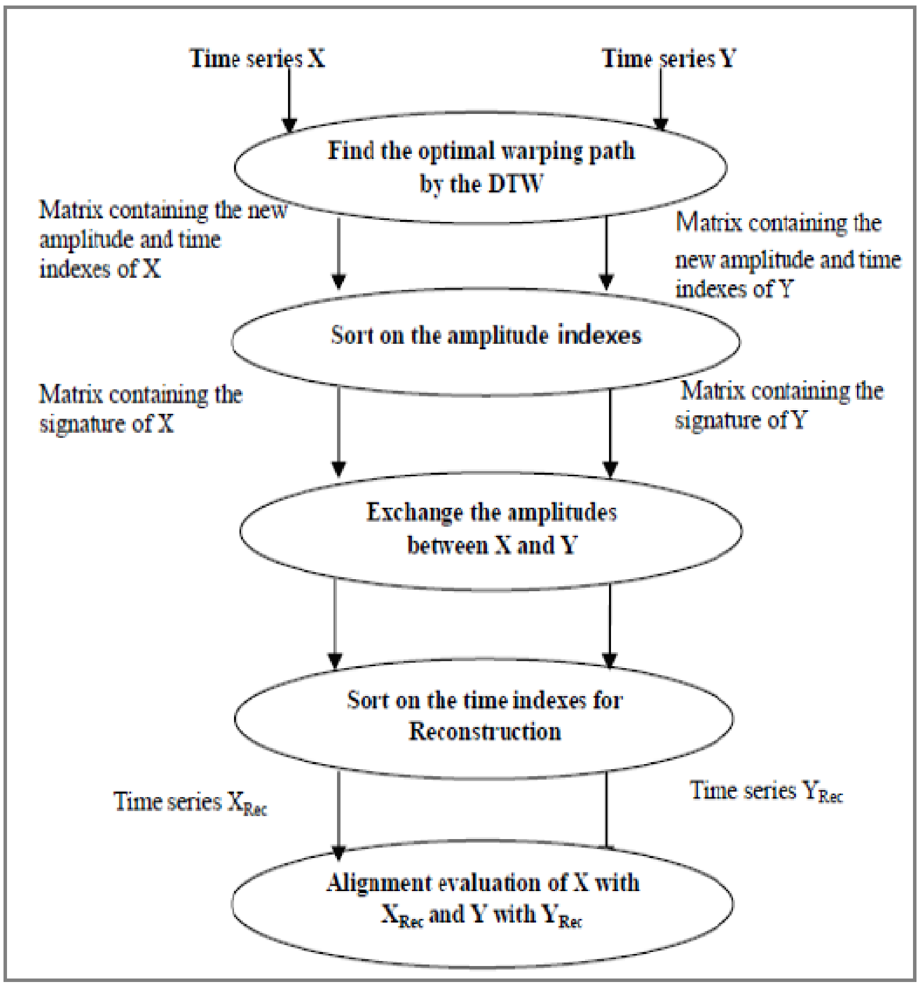

Fig. 2. Illustrative diagram of the I-SEA method.

Step 1: Finding the optimal warping path by the DTW:
The algorithm DTW is executed to equalize the length of the two time series

and make the first mapping between them by the replication of certain values which are close in both time series.

Step 2: Sorting on the amplitude:
The sorting of the stretched time series, on the coordinates of their amplitude indexes is established to give them a stable signature. The result of this step is a matrix for each time series containing the sorted amplitudes in ascendant or descendent order with their equivalents temporal coordinates (not sorted). In this way, if the time series are similar but phase shifted, they will always have almost the same trace which rep-resents their signatures. An example of signatures is illustrated in Fig.3.

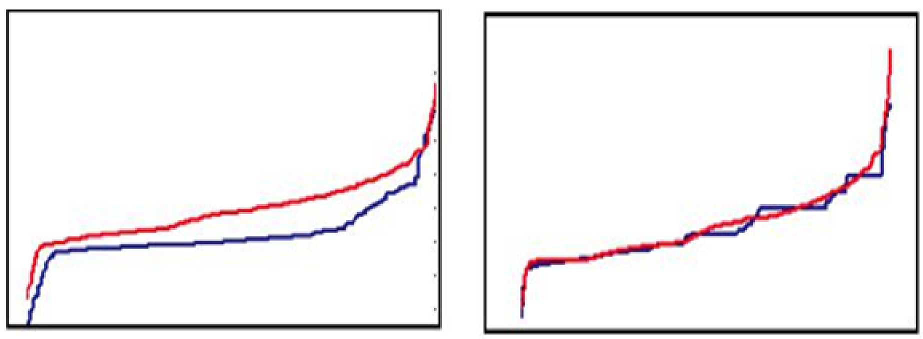

Fig. 3. Signatures of the time series X and Y established by SEA (a) and I-SEA (b) methods.

Step 3: Signatures exchange:
This step is divided into two parts. The first one consists in making the exchange of signatures between both time series. The first time series will receive the coordinates of amplitudes (signature) of the second time series and vice versa for the second series. The second part consists in restoring the normal size of the time series after the exchange of signature which requires that the time series are of equal length. Thus, the temporal indexes replied in every new series must be deleted with their equivalent amplitudes. This part does not affect the results of alignment; it just allows having a clearer visual inspection of the traces.

Step 4: Sorting on the time indexes and alignment:
Signatures represent well the characteristics of the time series but they do not consider the shifts and the differences of timescales. The sorting of the time series on the respective temporal indexes allows reconstructing them. Every time series will then be aligned with its reconstructed time series by using the correlation and PRD (percent root difference) factors as objective criteria and the visual inspection as subjective criteria.

3.2 One-Nearest Neighbor Classifiers Using I-SEA

The proposed 1NN-ISEA classifier combines the I-SEA method with the 1-NN classifier.the 1-NN classifier was chosen because it does not require any parameters and its relevance depends on the similarity measure used for the comparison. 1NN-ISEA classifieris is described in the following algorithm (Fig. 4).

```
Inputs: P = {P1, P2, ..., PN}: Training Set, Q: TS
Outputs: Class_Label of Q
Begin
    bsf←∞ //initialize the best_so_far distance
    Class_Label←NaN
    For i←1 to N do
        Distance(Q,Pi)←ISEA(Q,Pi)
        If (Distance < bsf)
            bsf←Distance
            Class_Label←Class of Pi
        Endif
    EndFor
    Return (Class_Label)
End.
```

Fig. 4. Time series classification algorithm 1NN-ISEA

4 Experimental Evaluation

The aim of this section is to evaluate the accuracy of our proposed classifier 1NN-ISEA with comparison to 1NN-SEA and 1NN-ED classifiers. The used time series dataset for the experiments are publically available from the UCR Time Series Data Archive [29]. Note that the UCR data archive includes real life, synthetic and generic time series come from different application domains and provides a time series datasets as a benchmark for testing and evaluating time series classification algorithms. The UCR contains 85 time series dataset, each having one standard partition for training and one for testing.

All applications in this study have been performed on an experimental environment with the following characteristics: PC Intel (R) Core ^{TM} i3 CPU, 2.40

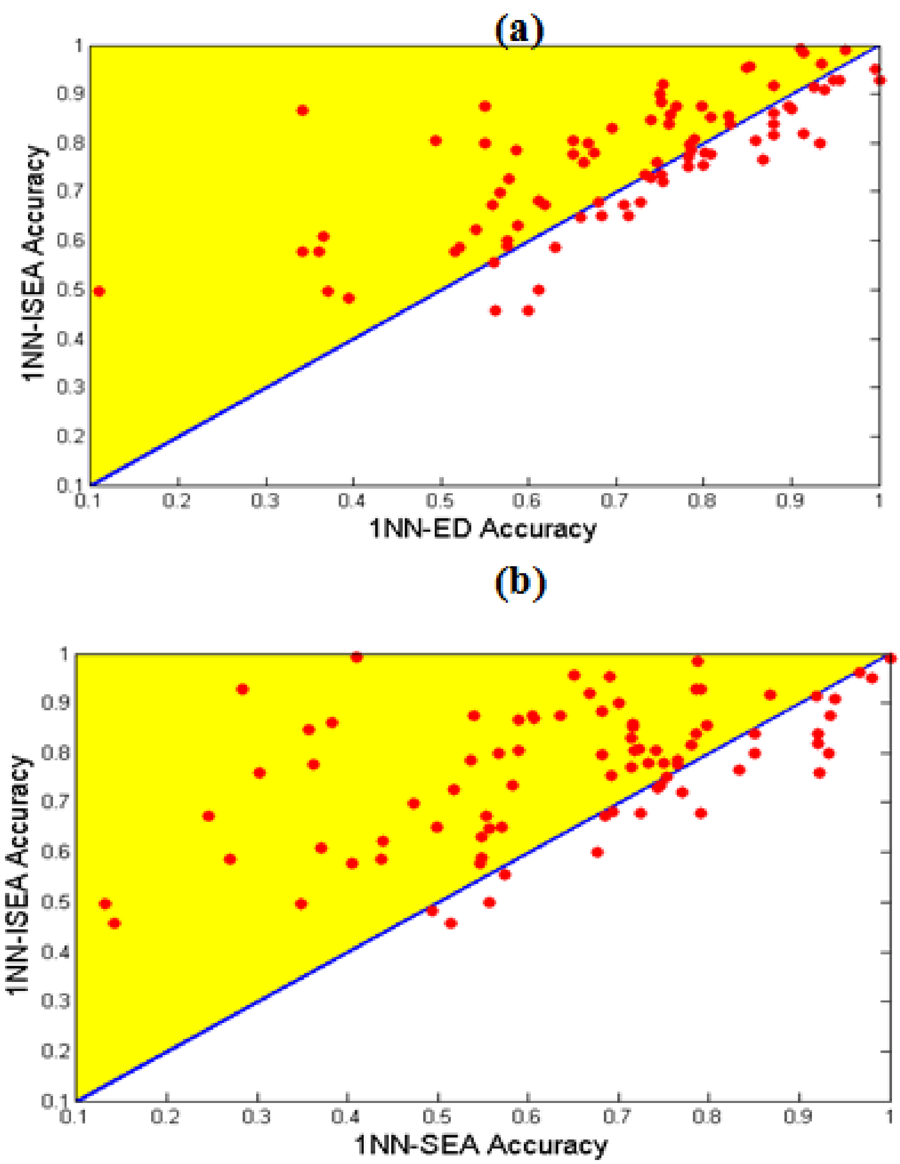

Fig. 5. Accuracy comparison of the 1NN-ED, 1NN-SEA and 1NN-ISEA classifiers.

GHz, 4Go of main memory, OS Windows 7 (64 bits) and MATLAB 8 development environment. Figure 5 provides scatter plots of the relative accuracies, of each of the 1NN-ED, 1NN-SEA and 1NN-ISEA classifiers. Each point represents a different UCR dataset. Fig.5.a shows that 1NN-ISEA is more accurate than 1NN-ED for a large number of datasets.

Precisely, 1NN-ISEA wins in 48 datasets compared to ED which wins in 37 data-sets with 1 tie. Fig.5.b shows that 1NN-ISEA is largely superior to 1NN-SEA. Precisely, 1NN-ISEA wins in 58 datasets compared to 1NN-SEA which wins in 27 datasets with 1 tie.

Table 1 illustrates the average error rates of the 1NN-ED, 1NN-SEA and 1NN-ISEA classifiers. As we can see from table 1, the 1NN-ISEA classifier has the lowest error rate followed by the1NN-ED and the 1NN-SEA classifiers. On the other hands, we see that 1NN-ISEA is significantly more accurate than 1NN-ED and 1NN-SEA. 1NN-ISEA wins on 42 datasets followed by 1NN-ED which wins on 27 datasets and 1NN-SEA which wins on 16 datasets. We deduce that, 1NN-ISEA is a good upgrade of the SEA method for time series classification.

$$Errorrate = \frac{Size\,of\,testing\,dataset - Number\,of\,TS\,correctly\,classified}{Size\,of\,testing\,dataset}. \tag{2}$$

Table 1: Comparison of 1NN-ED, 1NN-SEA and 1NN-ISEA classifiers on 85 time series dataset.

Datasets	1NN-ED	1NN-SEA	1NN-I-SEA
50words	0,369	0,729	0,413
Adiac	0,389	0,306	0,319
ArrowHead	0,2	0,308	0,245
Beef	0,333	0,433	0,2
BeetleFly	0,25	0,3	0,1
BirdChicken	0,45	0,15	0,2
Car	0,267	0,416	0,266
CBF	0,148	0,348	0,044
ChlorineConcentration	0,35	0,409	0,196
CinC ECG torso	0,103	0,365	0,125
Coffee	0	0,214	0,071
Computers	0,424	0,328	0,4
Cricket X	0,423	0,482	0,273
Cricket Y	0,433	0,525	0,302
Cricket Z	0,413	0,451	0,369
DiatomSizeReduction	0,065	0,032	0,039
DistalPhalanxOutlineAgeGroup	0,218	0,247	0,247
DistalPhalanxOutlineCorrect	0,248	0,253	0,266
DistalPhalanxTW	0,273	0,275	0,32
Earthquakes	0,326	0,267	0,219
ECG200	0,12	0,15	0,16

ECG5000	0,075	0,081	0,086
ECGFiveDays	0,203	0,066	0,27
ElectricDevices	0,45	0,461	0,125
FaceAll	0,286	0,501	0,35
FaceFour	0,216	0,318	0,204
FacesUCR	0,231	0,395	0,125
FISH	0,217	0,285	0,228
FordA	0,341	0,443	0,351
FordB	0,442	0,446	0,327
Gun_Point	0,087	0,08	0,18
Ham	0,4	0,485	0,543
HandOutlines	0,199	0,251	0,22
Haptics	0,63	0,652	0,503
Herring	0,484	0,453	0,422
InlineSkate	0,658	0,596	0,423
InsectWingbeatSound	0,438	0,858	0,542
ItalyPowerDemand	0,045	0,21	0,072
LargeKitchenAppliances	0,507	0,282	0,195
Lighting2	0,246	0,229	0,278
Lighting7	0,425	0,452	0,41
MALLAT	0,086	0,212	0,016
Meat	0,067	0,066	0,2
MedicalImages	0,316	0,43	0,348
MiddlePhalanxOutlineAgeGroup	0,26	0,257	0,27
MiddlePhalanxOutlineCorrect	0,247	0,333	0,273
MiddlePhalanxTW	0,439	0,426	0,443
MoteStrain	0,121	0,218	0,183
NonInvasiveFatalECGThorax1	0,171	0,202	0,143
NonInvasiveFatalECGThorax2	0,12	0,132	0,082
OliveOil	0,133	0,166	0,233
OSULeaf	0,479	0,561	0,413
PhalangesOutlinesCorrect	0,239	0,284	0,142
Phoneme	0,891	0,869	0,503
Plane	0,038	0	0,009
ProximalPhalanxOutlineAgeGroup	0,215	0,234	0,214
ProximalPhalanxOutlineCorrect	0,192	0,233	0,223
ProximalPhalanxTW	0,292	0,315	0,327
RefrigerationDevices	0,605	0,506	0,518
ScreenType	0,64	0,595	0,422
ShapeletSim	0,461	0,561	0,377
ShapesAll	0,248	0,318	0,117
SmallKitchenAppliances	0,659	0,411	0,134
SonyAIBORobotSurface	0,305	0,286	0,169

SonyAIBORobotSurfaceII	0,141	0,259	0,196
StarLightCurves	0,151	0,31	0,045
Strawberry	0,062	0,06	0,092
SwedishLeaf	0,211	0,276	0,192
Symbols	0,1	0,393	0,129
synthetic_control	0,12	0,616	0,14
ToeSegmentation1	0,32	0,21	0,32
ToeSegmentation2	0,192	0,284	0,146
Trace	0,24	0,08	0,16
Two_Patterns	0,09	0,59	0,007
TwoLeadECG	0,253	0,077	0,239
uWaveGestureLibrary_X	0,261	0,642	0,152
uWaveGestureLibrary_Y	0,338	0,697	0,241
uWaveGestureLibrary_Z	0,35	0,638	0,224
UWaveGestureLibraryAll	0,052	0,717	0,072
Wafer	0,005	0,019	0,048
Wine	0,389	0,444	0,5
WordsSynonyms	0,382	0,753	0,326
Worms	0,635	0,629	0,392
WormsTwoClass	0,414	0,464	0,215
Yoga	0,17	0,214	0,16
Average	0,288	0,355	0,239
Wins	27	16	42

5 Conclusion

In this paper, we proposed a 1NN classification method based on our ISEA time series alignment technique. We compared the so-obtained 1NN-ISEA classifier with the widely used Euclidian distance based 1NN and the 1NN-SEA classifiers. Results of experiments performed on the UCR universal time series classification benchmark that includes recorded time series from different fields of application show that the proposed 1NN-ISEA technique is a promising method for time series classification. As a perspective we will compare our method with more recent works of the state of the art.

References

1. Serfass, D.: Dynamic Biometric Recognition of Handwritten Digits Using Symbolic Aggregate Approximation, In Proceedings of the ACM SE'17, South East Conference, pp. 1-4 (2017).
2. Boucheham, B.: Abnormality Detection in Electrocardiograms by Time Series Alignment, Communications in Information Science and Management Engineering (CISME), Vol.1 No.3, pp. 7-11 (2011).

3. Boulnemour, I., Boucheham, B., Benloucif, S.: Improved dynamic time warping for abnormality detection in ECG time series, In F. Ortuo and I. Rojas, Eds., Bioinformatics and Biomedical Engineering: 4th International Conference, IWBBIO 2016, Granada, Spain, April 20-22. 2016, Proceedings, Cham: Springer International Publishing AG, pp. 242-253 (2016).
4. Aghabozorgi, S., Wah, T.Y.Effective Clustering of Time-Series Data Using FCM, In International Journal of Machine Learning and Computing, vol.4, n2, pp. 170-176 (2014).
5. Boulnemour, I., Boucheham, B.: I-SEA: Improved shape exchange algorithm for quasi-periodic time series alignment, In IEEE International Conference on Computer Vision and Image Analysis Applications (ICCVIA), Sousse, Tunisia, pp. 1-6 (2015).
6. Boulnemour, I., Boucheham, B., QP-DTW: upgrading dynamic time warping to handle quasi-periodic time series alignment, Journal of Information Processing Systems, vol. 14, n4, pp. 851-876 (2018).
7. Lahreche, A., & Boucheham, B.: FastSEA: A very fast and very effective matching technique for very complex time series, In IEEE International Conference on Mathematics and Information Technology (ICMIT), Adrar, Algeria, pp. 286-293 (2017).
8. Lahreche, A., & Boucheham, B.: LMDS-SEA: Upgrading the Shape Exchange Algorithm (SEA) to handle general time series classification by local matching and distance selection, In IEEE 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS) , Tbessa, Algeria, pp. 1-6 (2018).
9. Lahreche, A., & Boucheham, B. A fast and accurate similarity measure for long time series classification based on local extrema and dynamic time warping, Expert Systems with Applications, vol. 168, 114374, (2021). https://doi.org/10.1016/j.eswa.2020.114374.
10. Graves, A., Mohamed, A., Hinton, G.: Speech recognition with deep re-current neural networks, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (2013).
11. Smith, A.A.: Classification and Alignment of Gene-Expression Time-Series Data, Thse de doctorat, Universit de WisconsinMadison (2009).
12. Singha, S., Pandey, S. K., Pawar, U., Janghel, R. R.: Classification of ECG Arrhythmia using Recurrent Neural Networks, In Proceedings Computer Science, Elsevier, vol. 132, pp. 1290-1297 (2018).
13. Ziat, A.Y. : Apprentissage de reprsentation pour la prdiction et la classifi-cation de sries temporelles, Thse de Doctorat en Rseau de neurones, Universit Pierre et Marie Curie - Paris VI (2017).
14. Amr, T.: Survey on time-series data classification, In TSDM journal (2012).
15. Schfer, P The boss is concerned with time series classification in the presence of noise, Data Mining and Knowledge Discovery, vol. 29, n 6, pp. 15051530 (2015).
16. Huerta, R., Vembu, S., Muezzinoglu, M.K., Vergara, A.: Dynamical SVM for Time Series Classification, Pattern Recognition. DAGM/OAGM. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, vol. 7476, pp. 216-225 (2012).
17. Liu, C. L., Hsaio, W. H., Tu, Y. C.: Time Series Classification With Multi-variate Convolutional Neural Network, In Proceedings of the IEEE Transactions on Industrial Electronics, vol. 66, n 6 (2019).
18. Nguyen, T.L., Gsponer, S, Ifrim, G.: Time series classification by sequence learning in all-subsequence space, In Proceedings of the 33th International Con-ference on Data Engineering, pp. 947958. IEEE Press, San Diego (2017).

19. Lines, J., Taylor, S., Bagnall, A: HIVE-COTE: The Hierarchical Vote Collective of Transformation-based Ensembles for Time Series Classification, In proceedings of the IEEE 16th International Conference on Data Mining (ICDM), 2016.
20. Bagnall, A., Lines, J., Hills, J., Bostrom, A.:Time-series classification-with cote: the collective of transformation-based ensembles. IEEE Transactions on Knowledge and Data Engineering, vol. 27, No.9, pp. 25222535 (2015).
21. Sakoe, H, Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, In proceedings of the IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, n1, pp. 43-49 (1978).
22. Berndt, D. J., Clifford, J.: Using dynamic time warping to find patterns in time series, In AAI Workshop on Knowledge Discovery in Databases (KDD94), pp. 359-370, Seattle, Washington (1994).
23. Owsley, L., Atlas, L., Bernard, G.: Self-organizing feature maps and hidden markov models for machine-tool monitoring, In IEEE Transactions on Signal Processing vol. 45, n11, pp. 27872798 (1997).
24. Liao,T. W.:Clustering of time series data - a survey, Pattern Recognition, vol. 38, n11, pp. 1857-1874 (2005).
25. Ravinder, K.:Comparison of hmm and dtw for isolated word recognition system of punjabi language, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 244252 (2010).
26. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E Querying and mining of time series data: experimental comparison of representations and distance measure, In Proceedings of the VLDB Endowment, vol.1, n2, pp. 15421552 (2008).
27. Xing, Z., Pei, J., Keogh, E.; A brief survey on sequence classification. ACM SIGKDD Explorations Newsletter, vol. 12, n1, pp. 40-48(2010).
28. Boucheham, B.: Matching of quasi-periodic time series patterns by exchange of block-sorting signatures. Pattern Recognition Letters (Elsevier). Vol. 29, pp. 501514 (2008).
29. Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G.: The UCR time series classification archive, (2015).
30. Wu,C.L., Chau, K.W,: "Data-driven models for monthly streamflow time series prediction, Engineering Applications of Artificial Intelligence, Vol.23, Issue 8, pp.1350-1367 (2010).
31. Alkasassbeh, M., Altarawneh,G., Hassanat, A.B.: On enhancing the performance of nearest neighbour classifers using Hassanat distance metric, Canadian Journal of Pure and Applied Sciences, 9 (1), pp 3291-3298 (2015).
32. Fix, E, Hodges,J.L,: Discriminatory analysis, non-parametric discrimination: Consistency properties, In Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas (1951).
33. Boucheham, B.: Efficient matching of very complex time series. In International Journal of Machine Learning and Cybernetics, Springer, Vol. 4, Issue 5, pp. 537-550 (2013).