A comparison of Multivariate Time Series clustering methods

Iago Vázquez[1], José R. Villar[2], Javier Sedano[1], and Svetlana Simić[3]

[1] Instituto Tecnológico de Castilla y León, Pol. Ind. Villalonquejar, 09001, Burgos, Spain
{iago.vazquez,javier.sedano}@itcl.es
[2] University of Oviedo, Computer Science Department,
{villarjose,delacal}@uniovi.es
[3] University of Novi Sad,
Department of Neurology
Clinical Centre of Vojvodina Novi Sad, Republic of Serbia
svetlana.simic@mf.uns.ac.rs

Abstract. Big Data and the IoT explosion has made clustering Multivariate Time Series (MTS) one of the most effervescent research fields. From Bio-informatics to Business and Management, MTS are becoming more and more interesting as they allow to match events the co-occur in time but that is hardly noticeable. In this paper, we compare four clustering methods retrieved from the literature analyzing their performance on five publicly available data sets. These methods make use of different TS representation and distance measurement functions. Results show that Dynamic Time Warping is still competitive; APCA+DTW and Compression-based dissimilarity obtained the best results on the different data sets.

Keywords: Time Series, Clustering, Multivariate

1 INTRODUCTION

Multivariate Time Series (MTS) have regained the focus of the research community with the effervescence of Big Data, Internet of Things and Cyber-Physical Systems. In many cases, there is no information that introduce relationships among the MTS instances. Until recently, the problem was focused on univariate TS clustering; for instance, [1] proposed use Dynamic Time Warping (DTW) and k-means to cluster the performance of a photovoltaic power plant, so to predict the meteorological conditions. Similarly, k-means was used to cluster TS and then predict the weather conditions [2]. Interested readers can refer to [3] for a good review on this topic. Nevertheless, when more than one Time Series (TS) is involved the clustering problem becomes much more challenging. Additionally, it is possible to choose between unsupervised and semi-supervised methods to perform the clustering.

Grouping MTS has been found interesting in order to perform complex event detection or to classify the current scenario. For instance, [4] proposed a Partitioning around Meriods and Fuzzy C-Meroids clustering for the problem of detecting high-value pollution records or alarms in the city of Rome. To group the instances, the similarity among the variables between two MTS instances is one of the most studied topics. As an example, the authors in [5] proposed Principal Component Analysis similarity factor combined with the average based Euclidean distance together with a fuzzy clustering scheme to group MTS instances. Discords have also been used in MTS instance clustering to identify anomalies [6]. Alternatively, hash functions have been proposed to index and to measure the similarities as well [7].

Interestingly, Machine Learning models have been also used in measuring the similarity between multivariate TS, i.e., Gaussian Mixture Models [8] or Recurrent Neural Networks [9,10]. A different approach is based on extracting features and then using these features to group the multivariate TS [11] or together with Self-Organized Maps [12], Hidden Markov Models [13] or Fuzzy Linear [14]. Still, this problem cannot be considered solved and a recent study found out that the combination of feature extraction and a classification stage performs better than the current approaches [15].

This paper shows a comparison among four MTS instance clustering methods. The MTS representation and the distance measurement are different from one method to the other. In all of them, hierarchical clustering is the algorithm responsible of the groupings according to the distance matrices; the obtained trees are cut to get the desired number of clusters k. In the experimentation, the 4 methods are compared using several published MTS data sets. Two different experiments are carried out: on the first hand, the best number of clusters is found using the elbow's rule; on the second hand, the number of groups are defined with the number of classes in each data set. These two experimentation set ups might provide some idea on the performance of the MTS clustering methods: the first one tackles the total ignorance of the problem (no knowledge in the number of classes) and how they behave with the elbow's rule; the second one represents the case of total knowledge, where the number of labels are known a-priori but not the grouping or the MTS patterns. The main goal of this study is to set the basis for a future research on merging the outcomes of different MTS data sets, giving some rules on how the different techniques perform and providing evidence on how to design the merging.

The structure of the paper is as follows. Next Section aims to give details of the 4 methods of this comparison, the data sets used in the comparison and the experimental set up. Sect. 3 discuss on the obtained results. Finally, the conclusions are drawn.

2 Material and Methods

This section describes the 4 methods in this comparison first, then the MTS data sets are introduced and, finally, the experimental set up is detailed.

2.1 MTS clustering methods

Let us call raw MTS the temporal sequence of values for each of the variables gathered from a certain source. Each instance in this raw MTS data set (ts^i) can be written as $< x_1^i, x_2^i, \cdots, x_M^i >$, where M is the number of variables and $x_m^i = x_{m1}^i, x_{m2}^i, \cdots, x_{mN}^i >$, N is the number of samples, m is the variable and i is the index on the MTS data set. We call $x^i[t] =< x_{1t}^i, x_{2t}^i, \cdots, x_{Mt}^i >$ the sample at time t. We assume a MTS data set as a collection of instances of raw MTS with arbitrary length. Note that we can store MTS for which the variables have different sampling rate provided there are some timestamps where all the sampling of all the variables coincide in time [16,17] using polynomial interpolation. Besides, long MTS are expected to be split in different instances; automatic segmentation of MTS can be employed in these cases to produce the set of suitable instances [18,17].

The four methods in this comparison are included in the following listing. In all of them, the distance between each pair of MTS instances in the data set are stored in a matrix; then, the hierarchical clustering (hclust) is employed to group the MTS instances.

- **Adaptive Piecewise Constant Approximation (APCA) plus MINDIST and hclust** [19], denoted as APCA-MINDIST. In this study, each variable j in a raw TS is represented by M segments ($APCA(ts^{ij})$=$\{< v_1^{ij}, p_1^{ij} >, \cdots, < v_M^{ij}, p_M^{ij} >\}$). The coefficients v_k^{ij} are the mean of the values of variable j in the interval $[p_(k-1)^{ij}, p_k^{ij}]$, with $p_0^{ij} = 0$. The limits of the intervals are computed with the Haar Discrete Wavelet Transform [20]. The MINDIST, defined by the authors, is used as the distance measurement.
- **APCA plus DTW and hclust** [19], denoted as APCA-DTW. The main variation is that DTW [21] is used as the MTS instances distance measurement.
- **Fast Fourier Transform (FFT) combined with hclust** [22] and denoted as FFT-hclust. The FFT is computed on the z-scored raw data, limiting the coefficients to the 10 components. The distance between two univariate TS is measured with the Energy of the differences between them.
- **Raw data and measuring similarities with the Compression-based dissimilarity measure (CMD) on the raw data** [23] and denoted as CMD-hclust. To overcome with the problem of TS of different lengths, the longer TS is windowed and the CDM is averaged. Let lng be the length of the shorter TS instance, then we propose to use a sliding window of size lng with a shift of lng samples; padding the window with the last TS sample whenever needed to avoid incomplete sliding windows. We consider two TS of similar length whenever the differences in length do not surpass the 1.5 ratio.

We have used the rule of the elbow to select the number of clusters [24]. To do so, the sum of squares distances of each point to its cluster center as the measure of quality Q_k of the current number of clusters k. Thus, if Cl_k is the set of every clusters found for every possible number of clusters k used to feed the clustering algorithm, then $Q_k = \sum_{C \in Cl_k} \sum_{p \in C} d(p, c_C)^2$, where c_C is the center of the cluster C and d corresponds to the Euclidean distance.

2.2 Experimental data sets

To illustrate the performance of the different clustering methods we have used several MTS data sets from the Time Series Classification site [25]. All the instances of the proposed data sets are labelled, which allows to evaluate the performance of the different solutions. The following MTS data sets are included in the experimentation stage:

- ArticularyWordRecognition (AWR) [26,27]: 25 train and 25 test instances of 12 variables, each with 143 samples.
- Cricket (Cr) [28,27]: records the movements of the hands of 4 cricket umpires using accelerometers. A total of 12 classes, with 6 variables and 1197 samples each per instance. The data set includes 108 train instances and 72 test instances.
- Epilepsy (EP) [29]: this data set includes triaxial accelerometer data recorded for several Activities of Daily Living and simulated Epileptic seizures. The data set includes 137 train intances and 128 test instances. Each instance includes 3 variables and 206 samples.
- Finger Movements (FM) [30]: this dataset has a correspondence to Benjamin Blankertz for the BCI II competition (Data set IV). The data set includes 316 train instances and 100 test instances. Each instance includes 28 variables, 50 samples each.
- HeartBeat (HB) [31,32]: this dataset is derived from the PhysioNet/CinC Challenge 2016. The data set includes 61 instance for training and 61 for testing. Each instance has 61 variables and 405 samples.

2.3 Assessment of the methods

We propose to use the following metrics to measure the performance of each method: Accuracy(ACC), Sensitivity (SEN), Specifity (SPE), and Kappa Factor (KPP). Therefore, we count the number of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) on the set of pairs of instances of a data set. In the context of clustering, we define these measures the following way, based on the work exposed in [33]:

- If the two instances are in the same cluster and belong to the same class, the pair counts as a True Positive.
- If the two instances are in different clusters and belong to different classes, the pair counts as a True Negative.
- If the two instances are in the same cluster and belong to different classes, the pair counts as a False Positive.
- If the two instances are in different clusters but they belong to the same class, the pair counts as a False Negative.

3 Results and Discussion

Results are included in Table 1 and Table 2. The former includes the results for the best number of clusters in each case; the latter shows the figures when the number of clusters is set to the number of labels in the data set.

	AWR					Cr				
Method	K	ACC	KPP	SEN	SPE	K	ACC	KPP	SEN	SPE
h-A-MIN	28	0.99	0.99	0.75	0.99	10	0.94	0.94	0.74	0.96
h-A-DTW	31	0.99	0.99	0.88	1.00	12	0.97	0.97	0.98	0.97
h-FFT	12	0.93	0.93	0.87	0.94	10	0.92	0.92	0.72	0.94
h-CMD	5	0.64	0.64	0.5	0.65	4	0.71	0.71	0.79	0.71
	EP					FM				
Method	K	ACC	KPP	SEN	SPE	K	ACC	KPP	SEN	SPE
h-A-MIN	5	0.70	0.68	0.26	0.84	5	0.70	0.68	0.26	0.84
h-A-DTW	6	0.71	0.70	0.37	0.83	3	0.50	0.37	0.41	0.509
h-FFT	5	0.64	0.61	0.32	0.74	5	0.64	0.61	0.32	0.74
h-CMD	5	0.80	0.79	0.64	0.86	5	0.80	0.79	0.64	0.86
	HB									
Method	K	ACC	KPP	SEN	SPE					
h-A-MIN	4	0.59	0.25	0.79	0.29					
h-A-DTW	3	0.61	0.23	0.87	0.23					
h-FFT	5	0.60	0.18	0.88	0.18					
h-CMD	5	0.47	0.37	0.28	0.75					

Table 1. Results for the **best number of clusters** found using the rule of the elbow.

As it can be seen, there is no clear winner among the different data sets. AWR shows a high Accuracy and Kappa Coefficient for APCA-MINDIST and APCA-DTW, with a significantly better sensitivity for the second method in the two run experiments. With the Cr data set, the best performance is observed for the APCA+DTW. Nevertheless, all the methods performed rather well with these two data sets. In the case of the EP data set, however, CMD-hclust is the best clustering method, followed by APCA+DTW in both experiments.

The results obtained with the FM and HB data sets are clearly poorer. In FM, for the first experiment, each method shows an accuracy of 0.5, while the sensitivity is higher for APCA+MINDIST and APCA+DTW and the specifity is higher for FFT-hclust and CMD-hclust. However, as Kappa coefficient is higher for these last two methods, their performance is based on their ability to find relevant clustering rules, while the APCA based methods seem to get clusters with more differences among their quantity of elements. In the second experiment, we have also a similar accuracy for each method, but the low specifity and high sensitivity for FFT-hclust, along the low value of the Kappa factor. APCA+DTW and CMD perform similarly, while APCA+MIN shows a less balanced result than the two previous methods.

Method	AWR				Cr			
Method	ACC	KPP	SEN	SPE	ACC	KPP	SEN	SPE
h-A-MIN	0.98	0.98	0.84	0.99	0.94	0.94	0.73	0.96
h-A-DTW	0.99	0.99	0.91	0.99	0.98	0.98	0.94	0.98
h-FFT	0.98	0.98	0.77	0.99	0.93	0.93	0.68	0.95
h-CMD	0.93	0.93	0.13	0.96	0.81	0.81	0.56	0.83
	EP				FM			
Method	ACC	KPP	SEN	SPE	ACC	KPP	SEN	SPE
h-A-MIN	0.64	0.62	0.31	0.75	0.50	0.27	0.62	0.37
h-A-DTW	0.69	0.67	0.37	0.79	0.50	0.33	0.501	0.49
h-FFT	0.62	0.59	0.33	0.72	0.50	0.15	0.82	0.18
h-CMD	0.79	0.78	0.71	0.82	0.50	0.33	0.501	0.48
	HB							
Method	ACC	KPP	SEN	SPE				
h-A-MIN	0.59	0.03	0.97	0.03				
h-A-DTW	0.61	0.23	0.87	0.23				
h-FFT	0.59	0.00	0.99	0.01				
h-CMD	0.51	0.29	0.52	0.49				

Table 2. Results obtained when the **number of clusters (K)** is set to the number of classes in the data set.

Finally, with the HB, the second experiment's results for FFT-hclust and APCA+MINDIST are the worst: the low Kappa Factor and specifity show that these two methods created two extremely imbalanced clusters, and their performance is similar to those obtained when clustering all the instances in the same cluster. APCA+DTW shows better performance, while CMD-hclust is the most balanced method considering the all the metrics. Overall, perhaps it can be concluded that the best two methods are APCA+DTW and CDM-hclust; however, what is really relevant is that the methods vary their performance according to the data set. More research is needed in obtaining MTS clustering methods that perform similarly among a wide variety of problems; perhaps an ensemble of techniques including some user feedback might help in driving the grouping process.

4 Conclusions

This study present a comparison of MTS clustering methods using publicly available MTS data sets. The aim of this research is to find which TS representation and distance measurements are more promising among APCA-DTW, APCA-MINDIST, FFT-hclust and CMD-hclust.

Results show that there is a strong variability in the results according to the data set, showing no clear winner method. Both APCA-DTW and CMD-hclust showed the best overall performance and were more balanced when considering all the metrics simultaneously. More research is needed in obtaining MTS clustering methods that perform similarly among a wide variety of problems;

perhaps an ensemble of techniques including some user feedback might help in driving the grouping process.

ACKNOWLEDGMENT

This research has been funded by the Spanish Ministry of Science and Innovation under project MINECO-TIN2017-84804-R and by the Grant FCGRUPIN-IDI/2018/000226 project from the Asturias Regional Government.

References

1. Liu, G., Zhu, L., Wu, X., Wang, J.: Time series clustering and physical implication for photovoltaic array systems with unknown working conditions. Solar Energy **180** (2019) 401 – 411
2. Lee, Y., Na, J., Lee, W.B.: Robust design of ambient-air vaporizer based on time-series clustering. Computers and Chemical Engineering **118** (2018) 236 – 247
3. Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y.: Time-series clustering – a decade review. Information Systems **53** (2015) 16 – 38
4. D'Urso, P., Giovanni, L.D., Massari, R.: Robust fuzzy clustering of multivariate time trajectories. International Journal of Approximate Reasoning **99** (2018) 12 – 38
5. Fontes, C.H., Budman, H.: A hybrid clustering approach for multivariate time series – a case study applied to failure analysis in a gas turbine. ISA Transactions **71** (2017) 513 – 529
6. Hu, M., Feng, X., Ji, Z., Yan, K., Zhou, S.: A novel computational approach for discord search with local recurrence rates in multivariate time series. Information Sciences **477** (2019) 220 – 233
7. Yu, C., Luo, L., Chan, L.L.H., Rakthanmanon, T., Nutanong, S.: A fast lsh-based similarity search method for multivariate time series. Information Sciences **476** (2019) 337 – 356
8. Mikalsen, K.Ø., Bianchi, F.M., Soguero-Ruiz, C., Jenssen, R.: Time series cluster kernel for learning similarities between multivariate time series with missing data. Pattern Recognition **76** (2018) 569 – 581
9. Vázquez, I., Villar, J.R., Sedano, J., Simic, S.: A preliminary study on multivariate time series clustering. In: 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019) - Seville, Spain, May 13-15, 2019, Proceedings. (2019) 473–480
10. Vázquez, I., Villar, J.R., Sedano, J., Simic, S., de la Cal, E.A.: A proof of concept in multivariate time series clustering using recurrent neural networks and sp-lines. In: Hybrid Artificial Intelligent Systems - 14th International Conference, HAIS 2019, León, Spain, September 4-6, 2019, Proceedings. (2019) 346–357
11. Ferreira, A.M.S., de Oliveira Fontes, C.H., Cavalcante, C.A.M.T., Marambio, J.E.S.: Pattern recognition as a tool to support decision making in the management of the electric sector. part ii: A new method based on clustering of multivariate time series. International Journal of Electrical Power and Energy Systems **67** (2015) 613 – 626

12. Salvo, R.D., Montalto, P., Nunnari, G., Neri, M., Puglisi, G.: Multivariate time series clustering on geophysical data recorded at mt. etna from 1996 to 2003. Journal of Volcanology and Geothermal Research **251** (2013) 65 – 74 Flank instability at Mt. Etna.
13. Li, J., Pedrycz, W., Jamal, I.: Multivariate time series anomaly detection: A framework of hidden markov models. Applied Soft Computing **60** (2017) 229 – 240
14. Duan, L., Yu, F., Pedrycz, W., Wang, X., Yang, X.: Time-series clustering based on linear fuzzy information granules. Applied Soft Computing **73** (2018) 1053 – 1067
15. Bode, G., Schreiber, T., Baranski, M., Müller, D.: A time series clustering approach for building automation and control systems. Applied Energy **238** (2019) 1337 – 1345
16. Anstey, J., Peters, D., Dawson, C.: An improved feature extraction technique for high volume time series data. Proceedings of the Fourth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (01 2007) 74–81
17. Keogh, E., Lonardi, S., Chiu, B.Y.c.: Finding surprising patterns in a time series database in linear time and space. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2002) 550–556
18. Duan, L., Yu, F., Pedrycz, W., Wang, X., Yang, X.: Time-series clustering based on linear fuzzy information granules. Applied Soft Computing **73** (12 2018) 1053–1067
19. Chakrabarti, K., Keogh, E., Mehrotra, S., Pazzani, M.: Locally adaptive dimensionality reduction for indexing large time series databases. ACM Transactions on Database Systems (TODS) (January 2002)
20. Chan, K.p., Fu, A.W.C.: Efficient time series matching by wavelets. Proceedings of the 15th International Conference on Data Engineering (1999) 126
21. Bellman, R.: Adaptive Control Processes. Princeton University Press (1961)
22. Singleton, R.: An algorithm for computing the mixed radix fast fourier transform. IEEE Transactions on Audio and Electroacoustics **17**(2) (June 1969) 93–103
23. Keogh, E., Lonardi, S., Ratanamahatana, C., Wei, L., Lee, S.H., Handley, J.: Compression-based data mining of sequential data. Data Mining and Knowledge Discovery **14** (02 2007) 99–129
24. Öztürk, A., Lallich, S., Darmont, J.: A visual quality index for fuzzy c-means. Artificial Intelligence Applications and Innovations (06 2018)
25. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The Great Time Series Classification Bake Off: a Review and Experimental Evaluation of Recent Algorithmic Advances. Data Mining and Knowledge Discovery **31**(3) (2017) 606–660
26. Wang, J., Balasubramanian, A., de la Vega, L.M., Green, J.R., Samal, A., Prabhakaran, B.: Word recognition from continuous articulatory movement time-series data using symbolic representations. In: ACL/ISCA interspeech workshop on speech and language processing for assistive technologies. (2013) 119–127
27. Shokoohi-Yekta, M., HuHongxia, B., Wang, J., Keogh, E.: Generalizing DTW to the multi-dimensional case requires an adaptive approach. Data Mining and Knowledge Discovery **31**(1) (2017) 1–31
28. Ko, M., West, G., Venkatesh, S., Kumar, M.: Online context recognition in multisensor systems using dynamic time warping. In: Proceedings of the IEEE international conference on intelligent sensors, sensor networks and information processing (ISSNIP). (2005) 283–288

29. Villar, J.R., Vergara, P., Menéndez, M., de la Cal, E., González, V.M., Sedano, J.: Generalized Models for the Classification of Abnormal Movements in Daily Life and its Applicability to Epilepsy Convulsion Recognition. International Journal of Neural Systems **26**(06) (2016) 1650037
30. Blankertz, B., Curio, G., Muller, K.R.: No Title. Advances in Neural Information Processing Systems 14 (NIPS 2001) (2011)
31. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: hysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. Circulation **101**(23) (2000)
32. Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E.W., Syed, Z., Schmidt, S.E., Papadaniil, C.D., Hadjileontiadis, L., Naseri, H., Moukadem, A., Dieterlen, A., Brandt, C., Tang1, H., Samieinasab, M., Samieinasab, M.R., SameniRoger, R., Mark, G., Clifford, G.D.: An open access database for the evaluation of heart sound algorithms. Physiological Measurement **37**(12) (2016)
33. Zakaria, J., Mueen, A., Keogh, E.: Clustering time series using unsupervised-shapelets. Proceedings of the 2012 IEEE 12th International Conference on Data Mining (2012) 785–794