Impact of the Sakoe-Chiba Band on the DTW Time-Series Distance Measure for *k*NN Classification

Zoltan Geler[1], Vladimir Kurbalija[2], Miloš Radovanović[2], Mirjana Ivanović[2]

[1] Faculty of Philosophy, University of Novi Sad
Dr Zorana Đinđića 2, 21000 Novi Sad, Serbia
gellerz@gmail.com
[2] Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad
Trg D. Obradovića 4, 21000 Novi Sad, Serbia
{kurba,radacha,mira}@dmi.uns.ac.rs

Abstract. For classification of time series, the simple 1-nearest neighbor (1NN) classifier in combination with an elastic distance measure such as Dynamic Time Warping (DTW) distance is considered superior in terms of classification accuracy to many other more elaborate methods, including *k*-nearest neighbor (*k*NN) with neighborhood size $k > 1$. In this paper we revisit this apparently peculiar relationship and investigate the differences between 1NN and *k*NN classifiers in the context of time-series data and constrained DTW distance. By varying neighborhood size *k*, constraint width *r*, and evaluating 1NN and *k*NN with and without distance-based weighting in different schemes of cross-validation, we show that the first nearest neighbor indeed has special significance in labeled time-series data, but also that weighting can drastically improve the accuracy of *k*NN. This improvement is manifested by better accuracy of weighted *k*NN than 1NN for small values of *k* (3–4), better accuracy of weighted *k*NN than unweighted *k*NN in general, and reduced need to use large values of constraint *r* with weighted *k*NN.

Keywords: Time series, Dynamic Time Warping, global constraints, classification, *k*-nearest neighbor.

1 Introduction

A time series represents a series of numerical data points in successive order, usually with uniform intervals between them. This form of data can appear in almost every aspect of human activity including: representing social, economic and natural phenomena, medical observations, results of scientific and engineering experiments, etc. Time-series mining is the subfield of artificial intelligence where different data mining methods are applied on time-series data in order to understand the phenomenon which generated those time series. These methods include classification, clustering, anomaly detection, prediction, and indexing.

The choice of appropriate distance/similarity measure is a crucial aspect of time-series mining since all mentioned methods explicitly or implicitly use distance

adfa, p. 1, 2011.

measures. These measures should be carefully defined in order to reflect the essential similarities between time series which are commonly based on shapes and trends. Research in this field yielded several distance measures – from Euclidean distance [1] as the most simple and intuitive to the more sophisticated distance measures such as Dynamic Time Warping (DTW) [2], Longest Common Subsequence (LCS) [3], Edit Distance with Real Penalty (ERP) [4] and Edit Distance on Real sequence (EDR) [5].

Unfortunately, the quality of distance measures is usually hard to evaluate since the notion of similarity is a very subjective and data-dependent issue. The most common approach to the assessment of distance measures in the literature [6,7,8] is through evaluation of classification accuracies of distance-based classifiers. The quality of the nearest-neighbor based techniques strongly depends on the quality of the used distance measures, which makes the NN classifier very suitable for distance-measure assessment. Furthermore, the simple 1NN classifier is selected in several works [7, 9], as one of the most accurate classifiers for time-series data, demonstrating comparable and even superior performance than many more complex classification approaches, including the k-nearest neighbor classifier with $k > 1$.

The main goal or this paper is to provide a more detailed investigation of differences between 1NN and kNN classifiers in the context of time-series data and DTW distance. We will show that the accuracy of kNN can be improved and made superior to 1NN when the importance of the first neighbor is taken into account. The rest of the paper is organized as follows: next section gives some basic facts and an overview of the recent work in this area. Section 3 presents the detailed results of our experiments which are conducted on 46 datasets available from [10]. The final section contains conclusions drawn from the experiments, as well as possibilities for future work.

2 Background and Related Work

The advantages of Euclidean distance (easily implementable, fast to compute and represents a distance metric) have made it, over time, probably one of the most commonly used similarity measure for time series [11,12,13,14]. However, due to the linear aligning of the points of the time series it is sensitive to distortions and shifting along the time axis [15, 16]. To address this shortcoming, many different elastic similarity measures were proposed. Among them, some of the most widely used and studied are Dynamic Time Warping (DTW) and Longest Common Subsequence (LCS), and their extensions, Edit Distance with Real Penalty (ERP) and Edit Distance on Real sequence (EDR).

Implementations of these elastic similarity measures are based on dynamic programming: in order to determine the similarity between two time series we need to compare each point of one time series with each point of the other one. This can lead to pathological non-linear aligning of the points (where a relatively small part of one time series maps onto a large section of the other time series) and slow down the computations. One way to avoid these adverse effects is to constrain the warping path using the Sakoe-Chiba band [17].

It is reported that the elastic measures can have better classification accuracy than Euclidean distance and that constraining the warping window can further improve the accuracy of these measures [7, 9]. In [18] and [19] we have shown that when the constraint parameter is tight enough (less than 15%-10% of the length of the time series), constrained versions of the elastic measures (DTW, LCS, ERP and EDR) become qualitatively different from their unconstrained counterparts (in the sense of producing significantly different 1-nearest neighbor graphs). In [9] and [15], based on experiments using a limited number of datasets it is reported that narrow constraints (less than 10% of the length of time series) are necessary for accurate DTW and that a warping window which is too large may actually deteriorate classification accuracy.

All mentioned experiments for distance-measure assessment were conducted with 1NN classifier as it was shown that it gives among the best results (compared to many not only distance-based classifiers) with time-series data [7, 9]. This fact strongly indicates that the first neighbor has particularly important meaning in the time-series datasets. In [20], the reasons and origins of this special behavior of the first neighbor are investigated, and related with the observed diversity of class labels in *k*-neighborhoods. In this paper, we will compare the accuracies of 1NN and *k*NN classifiers when using the DTW time-series distance measure in order to understand the special meaning of the first neighbor. Furthermore, we will attempt to improve the accuracy of *k*NN by favoring the first (few) neighbors.

3 Experimental Results

Through extensive experiments in this section we will investigate the suggestions and findings regarding the influence of the Sakoe-Chiba band on the Dynamic Time Warping similarity measure, 1NN and *k*NN classifiers discussed above. We will observe the following widths of the warping window: 100% (the unconstrained similarity measure), 90%, 80%, 70%, 60%, 50%, 45%, 40%, 35%, 30%, and all values from 25% to 0% in steps of 1%. These values were chosen based on reports that the measures with larger constraints behave similarly to the unconstrained ones, while the smaller constraints show more apparent discrepancies [7, 9, 15, 18, 19].

We are going to report the minimal value of the warping window that maximizes the classification accuracy of the *k*-nearest neighbor classifier for a large number of datasets. This classifier is chosen taking into account that among many classification methods (decision trees, neural networks, Bayesian networks, support vector machines, etc.) simple nearest-neighbor methods often give the best results when working with time series [7, 9]. In addition to that, the quality of distance/similarity measure directly influences the accuracy of the NN classifier, which makes it appropriate for distance/similarity measure assessment.

To obtain a better insight into the impact of constraining the warping window our experiments encompass five different evaluation methods of classification accuracy: leave-one-out (LOO), stratified 9-fold cross-validation (SCV1x9), 5 times repeated stratified 2-fold cross-validation (SCV5x2), 10 times repeated stratified 10-fold cross validation (SCV10x10) and 10 times repeated stratified holdout method (SHO10x) using two-thirds of available time series for training and one third for testing. The datasets are randomly shuffled in each run. Furthermore, we observe the unweighted and the weighted *k*NN classifier with the values of parameter *k* in range from 1 to 30. Weights are calculated by the formula $1/\mathrm{d}(q,c)^2$ where $\mathrm{d}(q,c)$ denotes the distance between the time series *q* and *c* [21].

The experiments were conducted on 46 datasets from [10], which includes the majority of all publicly available, labeled time-series datasets in the world. In addition to that, this collection of datasets is most commonly used for validation of different time-series mining concepts. The length of time series varies from 24 to 1882 depending of the dataset. The number of time series per dataset varies from 56 to 9236 and the number of classes varies from 2 to 50.

The unweighted *k*NN classifier. In Fig. 1 we can clearly notice that the relationship between the parameter *k* and the average smallest error rate is almost linear – the growth of parameter *k* leads to the decline of classification accuracy. The highest average classification accuracy (88.772%) was achieved with the 1NN classifier and the LOO evaluation method and the lowest one (74.536%) with the 30NN classifier and the SCV5x2 evaluation method (Table 1).

In case of the unweighted *k*NN classifier the average width of the smallest warping window which gives the lowest error rate for DTW varies in the range from 3.783 to 10.087. We can see that the increase of the parameter *k* implies the growth of the average warping window widths (Fig. 2): we need wider and wider windows to get the best accuracy. The smallest average warping window (3.783) was obtained using the LOO evaluation method and the 1NN classifier and the largest one (10.087) with the SHO10x evaluation method and the 24NN classifier (Table 2).

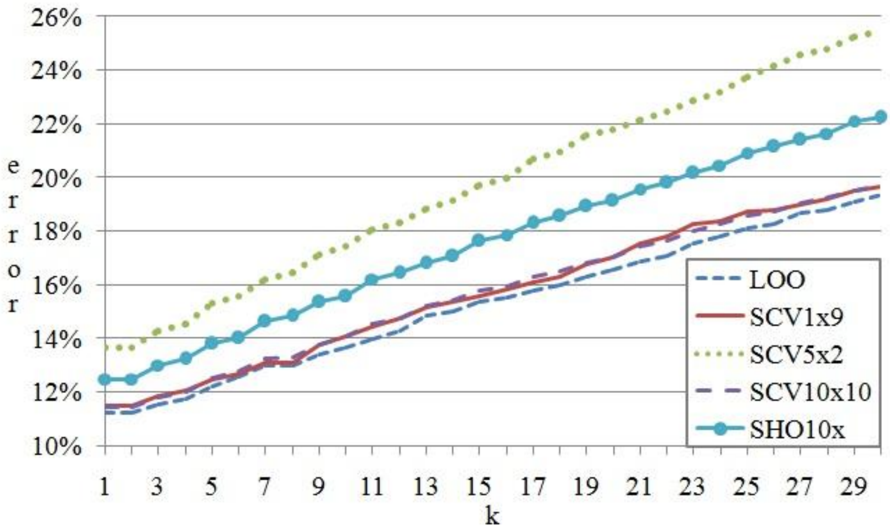

Fig. 1. Average lowest error rates for DTW with unweighted *k*NN

Table 1. Minimum and maximum of the average lowest error rates for DTW with unweighted *k*NN

	MIN error	*k*	MAX error	*k*	MAX-MIN
LOO	**11.228%**	**1**	19.317%	30	8.089
SCV1x9	11.494%	1	19.636%	30	8.142
SCV5x2	13.628%	1	**25.464%**	**30**	**11.836**
SCV10x10	11.410%	1	19.701%	30	8.291
SHO10x	12.471%	1	22.223%	30	9.752

Fig. 2. Average smallest warping window widths for DTW with unweighted *k*NN

Table 2. Minimum and maximum of the average smallest warping window widths for DTW with unweighted *k*NN

	MIN *r*	*k*	MAX *r*	*k*	MAX-MIN
LOO	**3.783**	**1**	7.652	30	3.870
SCV1x9	4.065	1	8.587	30	4.522
SCV5x2	4.913	1	9.935	24	5.022
SCV10x10	4.261	1	8.565	21	4.304
SHO10x	4.000	1	**10.087**	**24**	**6.087**

The weighted *k*NN classifier. Looking at the chart in Fig. 3 we can see that in the case of DTW the use of weights changes the influence of the parameter *k* on the accuracy of classification: instead of 1NN the smallest average error rates were achieved with 3NN (or 4NN in the case of SCV5x2 and SHO10x). After a brief decline and reaching the minimum value, the error rates begin to grow again, similarly as in the case of the unweighted *k*NN classifier but visibly slower. The attained maximum values of the classification errors are more than 1.5 times less than without weights (Table 3). The highest average classification accuracy was achieved by LOO and the lowest one by SCV5x2.

Fig. 4 shows that the introduction of weights into the *k*NN classifier noticeably alleviates the growth of the average warping window widths. In this case the largest average warping window (6.848) was achieved by the combination of the 8NN classifier and the SCV5x2 evaluation method (Table 4). The smallest average warping window (3.783) was obtained using the 1NN classifier and the LOO evaluation method. The differences between the minimum and maximum average *r* values are about two times smaller than in the case of the unweighted *k*NN classifier.

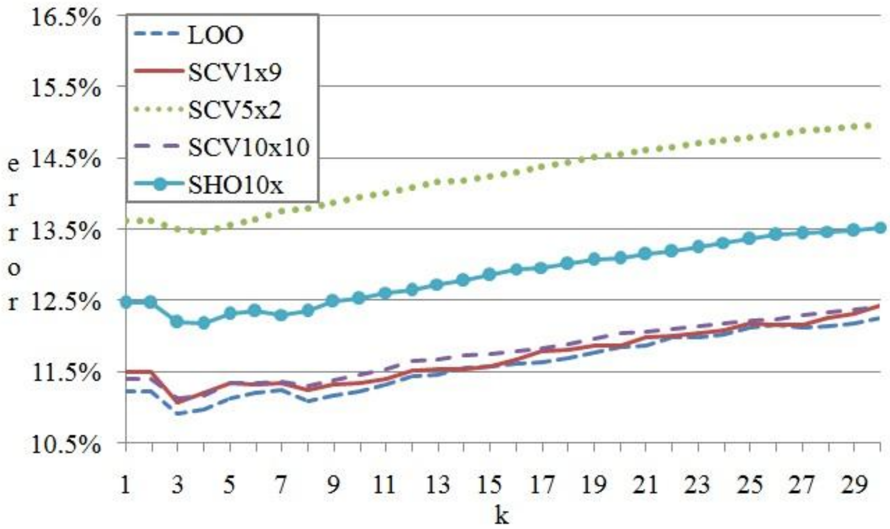

Fig. 3. Average lowest error rates for DTW with weighted *k*NN

Table 3. Minimum and maximum of the average lowest error rates for DTW with weighted *k*NN

	MIN error	*k*	MAX error	*k*	MAX-MIN
LOO	**10.923%**	**3**	12.256%	30	1.333
SCV1x9	11.072%	3	12.426%	30	1.354
SCV5x2	13.468%	4	**14.970%**	**30**	**1.502**
SCV10x10	11.134%	3	12.412%	30	1.278
SHO10x	12.177%	4	13.527%	30	1.350

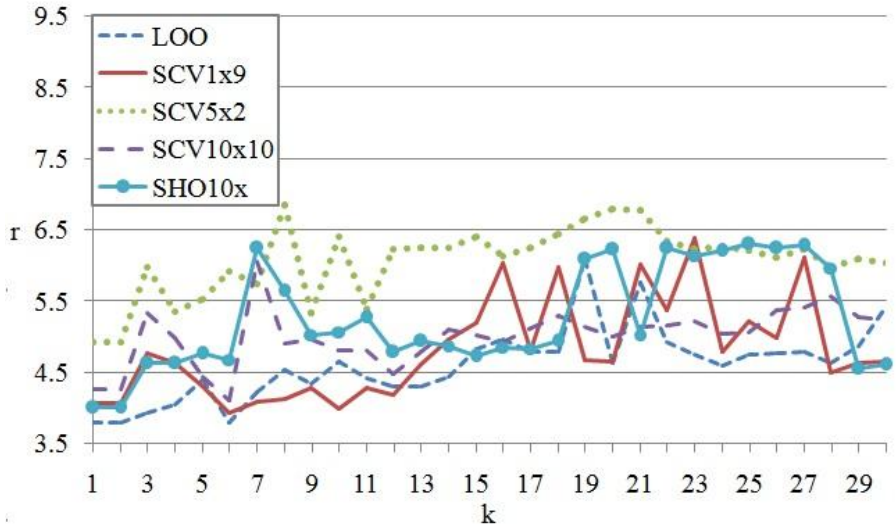

Fig. 4. Average smallest warping window widths for DTW with weighted *k*NN

Table 4. Minimum and maximum of the average smallest warping window widths for DTW with weighted *k*NN

	MIN *r*	*k*	MAX *r*	*k*	MAX-MIN
LOO	**3.783**	**1**	6.087	19	2.304
SCV1x9	3.935	6	6.370	23	**2.435**
SCV5x2	4.913	1	**6.848**	**8**	1.935
SCV10x10	4.109	6	6.043	7	1.935
SHO10x	4.000	1	6.304	25	2.304

4 Conclusions and Future Work

The results of experiments clearly confirmed the special importance of the first neighbor in time-series data. As seen in Fig. 1, the error rate of the unweighted *k*NN classifier almost linearly grows as the number of neighbors *k* grows. The *k*NN classifier actually gives the best results for the value $k = 1$ when consider *k* neighbors without a weighting scheme. On the other hand, when the weighting scheme is introduced (Fig. 3) the situation is changed to some extent. The best results are obtained for the value $k = 3$. Furthermore, the weighting scheme which favors the first neighbor significantly improved the accuracy for all values ok *k*.

When observing the value of constraint (Fig. 2 and 4) the introduction of the weighting scheme has an important impact. For unweighted *k*NN, the value of the constraint grows as *k* grows. On the other hand, with the weighting scheme the value of the constrains remains approximately the same for all values of *k*. In addition, the difference between minimum and maximum values of constraints is about two times smaller with the weighting scheme.

All these observations indicate that favoring the first neighbor with a weighting scheme improves the quality and stability of *k*NN. The first neighbor has a special meaning in time-series data and taking this fact into consideration can significantly improve the quality of *k*NN for all values of *k*, by making it even more accurate than 1NN for some small values of *k*.

In future work, it would be interesting to investigate the influence of weighting on other popular time-series distance measures like Euclidian distance, LCS, EDR, ERP, etc. In addition, the behavior of other weighting schemes [21, 22, 23, 24] we believe also warrants further investigation.

Acknowledgments

The authors would like to thank Eamonn Keogh for collecting and making available the UCR time series data sets, as well as everyone who contributed data to the collection, without whom the presented work would not have been possible. V. Kurbalija, M. Radovanović and M. Ivanović thank the Serbian Ministry of Education, Science and Technological Development for support through Project no. OI174023, "Intelligent Techniques and their Integration into Wide-Spectrum Decision Support."

References

1. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast subsequence matching in time-series databases. ACM SIGMOD Rec. 23, 419–429 (1994).
2. Berndt, D., Clifford, J.: Using dynamic time warping to find patterns in time series. KDD Workshop, pp. 359–370. Seattle, WA (1994).
3. Vlachos, M., Kollios, G., Gunopulos, D.: Discovering similar multidimensional trajectories. In Proceedings of the 18th International Conference on Data Engineering (ICDE), pp. 673–684. IEEE Comput. Soc. (2002).

4. Chen, L., Ng, R.: On The Marriage of Lp-norms and Edit Distance. In Proceedings of the 30th International Conference on Very Large Data Bases (VLDB), pp. 792–803. VLDB Endowment (2004).
5. Chen, L., Özsu, M.T., Oria, V.: Robust and fast similarity search for moving object trajectories. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 491–502. ACM, New York, NY, USA (2005).
6. Keogh, E., Kasetty, S.: On the need for time series data mining benchmarks: A survey and empirical demonstration. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 102–111. ACM, New York, NY, USA (2002).
7. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: Experimental comparison of representations and distance measures. Proc. VLDB Endow. 1, 1542–1552 (2008).
8. Esling, P., Agon, C.: Time-series Data Mining. ACM Comput. Surv. 45, 12:1–12:34 (2012).
9. Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C.A.: Fast time series classification using numerosity reduction. In Proceedings of the 23rd International conference on Machine Learning (ICML), pp. 1033–1040. ACM, New York, NY, USA (2006).
10. Keogh, E., Zhu, Q., Hu, B., Y., H., Xi, X., Wei, L., Ratanamahatana, C.A.: The UCR Time Series Classification/Clustering Homepage: www.cs.ucr.edu/~eamonn/time_series_data/, citeulike-article-id:2139261, (2011).
11. Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. In Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms (FODO), pp. 69–84. Springer Berlin Heidelberg (1993).
12. Chan, K.-P., Fu, A.W.-C.: Efficient time series matching by wavelets. In Proceedings of the 15th International Conference on Data Engineering (ICDE). pp. 126–133. IEEE Comput. Soc. (1999).
13. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Dimensionality reduction for fast similarity search in large time series databases. Knowl. Inf. Syst. 3, 263–286 (2001).
14. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Locally adaptive dimensionality reduction for indexing large time series databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 151–162. ACM, New York, NY, USA (2001).
15. Ratanamahatana, C.A., Keogh, E.: Three myths about dynamic time warping data mining. In Proceedings of the 5th SIAM International Conference on Data Mining (SDM), pp. 506–510 (2005).
16. Keogh, E.: Exact indexing of dynamic time warping. In Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), pp. 406–417. VLDB Endowment (2002).
17. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE T. Acoust. Speech Signal Process. 26, 43–49 (1978).
18. Kurbalija, V., Radovanović, M., Geler, Z., Ivanović, M.: The influence of global constraints on DTW and LCS similarity measures for time-series databases. In: Dicheva, D., Markov, Z., and Stefanova, E. (eds.) Third International Conference on Software, Services and Semantic Technologies (S3T) SE - 10. pp. 67–74. Springer Berlin Heidelberg (2011).

19. Kurbalija, V., Radovanović, M., Geler, Z., Ivanović, M.: The influence of global constraints on similarity measures for time-series databases. Knowl-Based Syst. 56, 49–67 (2014).
20. Radovanović, M., Nanopoulos, A., Ivanović, M.: Time-series classification in many intrinsic dimensions. In Proceedings of the 10th SIAM International Conference on Data Mining (SDM), pages 677–688, Columbus, Ohio, USA, 2010
21. Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. IEEE T. Syst. Man Cy. 6, 325–327 (1976).
22. Macleod, J.E.S., Luk, A., Titterington, D.M.: A re-examination of the distance-weighted k-nearest neighbor classification rule. IEEE T. Syst. Man Cy. 17, 689–696 (1987).
23. Gou, J., Xiong, T., Kuang, Y.: A novel weighted voting for k-nearest neighbor rule. J. Comput. 6, 833–840 (2011).
24. Gou, J., Du, L., Zhang, Y., Xiong, T.: A new distance-weighted k-nearest neighbor classifier. J. Inf. Comput. Sci. 9, 1429–1436 (2012).