

Time Series Classification Using Support Vector Machine with Gaussian Elastic Metric Kernel

Dongyu Zhang¹, Wangmeng Zuo¹, David Zhang^{2,1}, Hongzhi Zhang¹

¹School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
cswmzuo@gmail.com

²Biometrics Research Centre, Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong
csdzhang@comp.polyu.edu.hk

Abstract—Motivated by the great success of dynamic time warping (DTW) in time series matching, Gaussian DTW kernel had been developed for support vector machine (SVM)-based time series classification. Counter-examples, however, had been subsequently reported that Gaussian DTW kernel usually cannot outperform Gaussian RBF kernel in the SVM framework. In this paper, by extending the Gaussian RBF kernel, we propose one novel class of Gaussian elastic metric kernel (GEMK), and present two examples of GEMK: Gaussian time warp edit distance (GTWED) kernel and Gaussian edit distance with real penalty (GERP) kernel. Experimental results on UCR time series data sets show that, in terms of classification accuracy, SVM with GEMK is much superior to SVM with Gaussian RBF kernel and Gaussian DTW kernel, and the state-of-the-art similarity measure methods.

Keywords- time series; support vector machine; dynamic time warping; kernel method;

I. INTRODUCTION

With the increasing theoretical and practical values, time series classification has received great interest during the last decade, and has been widely applied to various disciplines, such as financial and stock data analysis, bioinformatics and biometrics etc. As a state-of-the-art classifier, support vector machine (SVM) has also been investigated and applied for time series classification in two modes. On one hand, combined with various feature extraction approaches, SVM can be adopted as a plug-in method in addressing time series classification problems. On the other hand, by designing appropriate kernel functions, SVM can also be performed based on the original time series data.

Because of the time axis distortion problem, classical kernel functions, such as Gaussian RBF (GRBF) and polynomial, generally are not suitable for SVM-based time series classification. Motivated by the success of dynamic time wrapping (DTW) distance, it has been suggested to utilize elastic measure to construct appropriate kernel. Gaussian DTW (GDTW) kernel is then proposed for SVM-

based time series classification, and has been applied to on-line handwriting recognition [1] and speech recognition [2].

Counter-examples, however, has been subsequently reported that GDTW kernel usually cannot outperform GRBF kernel in the SVM framework. Lei and Sun [3] proved that GDTW kernel is not positive definite symmetric (PDS) acceptable by SVM. Experimental results [3, 4] also showed that SVM with GDTW kernel (GDTW-SVM) cannot outperform either SVM with GRBF kernel (GRBF-SVM) or nearest neighbor classifier with DTW distance (1NN-DTW).

In this paper, we assume that elastic measures would be useful for SVM-based time series classification, and the poor performance of the GDTW kernel may be attributed to that DTW is non-metric. Motivated by recent progress in elastic measure, we propose a new class of elastic kernel, i.e. Gaussian elastic metric kernel (GEMK) by making an extension to the GRBF kernel. Using the recently developed elastic metrics, i.e. edit distance with real penalty (ERP) and time warp edit distance (TWED), we further present two examples of GEMK: Gaussian EPR (GERP) kernel and Gaussian TWED (GTWED) kernel. Our experimental results on UCR time series data sets [7] show that SVM with GEMK is significantly superior to GRBF-SVM, GDTW-SVM, and the state-of-the-art elastic measure methods.

The remainder of this paper is organized as follows: Section II first describes the definition of GEMK, and then provides two examples of GEMK. Section III presents the results of experiments using the UCR time series data sets. Finally, Section IV offers our conclusions.

II. GAUSSIAN ELASTIC METRIC KERNEL

A. Definition of Gaussian elastic metric kernel

Before the definition of GEMK, we first introduce the GRBF kernel, one of the most common kernel functions used in SVM classifier. Given two time series x and y with the same length n , the GRBF kernel is defined as

$$k_{\text{RBF}}(x, y) = \exp\left(-\frac{\|x - y\|_2^2}{2\sigma^2}\right), \quad (1)$$

where σ is the standard deviation. GRBF kernel is a PDS kernel. It can be regarded as an embedding of Euclidean distance in the form of Gaussian function. GRBF kernel requires the time series should have the same length and cannot handle the problem of time axis distortion. If the length of two time series is different, re-sampling usually is required to normalize them to the same length before further processing. Thus SVM with GRBF kernel (GRBF-SVM) usually is not suitable for time series classification.

Motivated by the effectiveness of elastic measures in handling the time axis distortion, it is interesting to embed elastic distance into SVM-based time series classification. Generally, there are two kinds of elastic distance. One is non-metric elastic distance measure, e.g. DTW, and the other is elastic metric, which is elastic distance satisfying the triangle inequality. Recently, DTW, one state-of-the-art elastic distance, has been proposed to construct the GDTW kernel [1, 2]. Subsequent studies, however, show that SVM with GDTW kernel cannot consistently outperform either GRBF-SVM or 1NN-DTW.

We assume that the poor performance of the GDTW kernel may be attributed to that DTW is non-metric, and suggest extending GRBF kernel using elastic metrics. Thus, we propose a novel class of kernel functions, Gaussian elastic metric kernel (GEMK) functions.

Definition 1 Let X be a non-empty finite set of time series and D denote the elastic metric. The GEMK function on X is defined as follows,

$$k_{\text{EM}}(x, y) = \exp\left(-\frac{D^2(x, y)}{2\sigma^2}\right), \quad \forall x, y \in X, \quad (2)$$

where σ is the standard deviation of Gaussian function.

According to Theorems 2.22 and 2.24 in [11], if the elastic distance is non-metric, the kernel function defined in (2) is not PDS and is not admissible to standard kernel machines, e.g., support vector machines. This is also the reason we utilize elastic metric rather than just elastic distance to construct kernel function in GEMK. Although we cannot guarantee the PDS property of GEMK yet, in our experiments using all the 20 UCR time series data sets, we do not observe the violation of the PDS property for the two examples of GEMK described in Section II.B. For Gaussian DTW kernel, however, the violation of the PDS property is observed in our experiments. Based on this, we suppose that GEMK may be more suitable for SVM-based time series classification.

B. Two examples of Gaussian elastic metric kernel

1) Gaussian ERP kernel.

Edit distance with real penalty (ERP) [5] is a state-of-the-art elastic metric which can be regarded as the marriage of edit distance and l_p norm. Given two time series $A = [a_1, a_2, \dots, a_m]$ with m elements and $B = [b_1, b_2, \dots, b_n]$ with n elements, the ERP distance between A and B is recursively defined as,

$$d_{\text{erp}}(A_1^m, B_1^n) = \begin{cases} \sum_{i=1}^m |a_i - g|, & \text{if } n = 0 \\ \sum_{i=1}^n |b_i - g|, & \text{if } m = 0 \\ \min \begin{cases} d_{\text{erp}}(A_2^m, B_2^n) + |a_1 - b_1|, \\ d_{\text{erp}}(A_2^m, B_1^n) + |a_1 - g|, \\ d_{\text{erp}}(A_1^m, B_2^n) + |b_1 - g|, \end{cases} & \text{otherwise} \end{cases}, \quad (3)$$

where $A_i^p = [a_i, a_{i+1}, \dots, a_p]$ denotes the subsequence of A , a_i (b_i) denotes the i th real element of the time series A (B), $|\cdot|$ denotes the l_1 -norm, and g is a constant real value [5].

It is obvious that the ERP defined in (3) is an elastic distance, and similar with DTW, it can also be calculated using the dynamic programming method. Besides, they also share the same time complexity, i.e. $O(n^2)$. But unlike DTW, ERP is a elastic metric.

Theorem 1 [5] Let Q, R, S be three time series of arbitrary length. Then it is necessary that ERP satisfies the triangle inequality $d_{\text{erp}}(Q, S) \leq d_{\text{erp}}(Q, R) + d_{\text{erp}}(R, S)$.

Corollary 2 [5] The ERP distance satisfies the triangle inequality and is a metric.

Using the ERP distance, we give the first example of GEMK, namely Gaussian ERP (GERP) kernel by substituting the elastic distance in GEMK with ERP metric.

Definition 2 Let X be a non-empty finite set, the GERP kernel on X is defined as:

$$k_{\text{GERP}}(x, y) = \exp\left(-\frac{d_{\text{erp}}^2(x, y)}{2\sigma^2}\right), \quad \forall x, y \in X, \quad (4)$$

where σ is the standard deviation of Gaussian function.

2) Gaussian TWED kernel.

Time warp edit distance (TWED) [6] is a recently developed metric by incorporating the time-stamps of time series. Suppose two time series $A = [(a_1, t_1), \dots, (a_i, t_i), \dots, (a_m, t_m)]$ with m elements, and $B = [(b_1, t_1'), \dots, (b_j, t_j'), \dots, (b_n, t_n')]$ with n elements, and t_i, t_i' are the time stamps of A and B , respectively, where $t_i < t_{i+1}$, $t_i' < t_{i'+1}'$. Then the TWED distance between A and B is recursively defined as

$$d_{\text{twed}}(A_1^m, B_1^n) = \min \begin{cases} d_{\text{twed}}(A_1^{m-1}, B_1^n) + |a_m - a_{m-1}| + v \cdot (t_m - t_{m-1}) + \lambda, \\ d_{\text{twed}}(A_1^{m-1}, B_1^{n-1}) + |a_m - b_n| + |a_{m-1} - b_{n-1}| \\ \quad + v \cdot (|t_m - t_n'| + |t_{m-1} - t_{n'-1}'|), \\ d_{\text{twed}}(A_1^m, B_1^{n-1}) + |b_n - b_{n-1}| + v \cdot (t_n - t_{n'-1}') + \lambda, \end{cases}, \quad (5)$$

where λ and v are two non-negative constant values [6].

Since it has been proved that the TWED is metric [6], we can easily define the Gaussian TWED (GTWED) kernel by substituting the elastic distance in GEMK defined in (2) with TWED metric.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, using the UCR time series data sets [7], we evaluate the effectiveness of SVM-based time series classification with the proposed GEMK functions (GEMK-SVM). First, a description of data sets and experimental setup is provided. Then, we evaluate the classification accuracies of SVM with GTWED kernel (GTWED-SVM) and SVM with GERP kernel (GERP-SVM).

A. Data set description and experimental setup

The UCR time series data sets [7] include 6 two-class problems and 14 multi-class problems for a wide variety of applications i.e., biomedical data classification, electromagnetic measurements. For each data set, a training subset is defined as well as a test subset. Table I briefly summarizes the basic information of each data set.

Following the approach suggested in [9], we adopt the 10-folder cross-validation method on the training subset to determine the hyper-parameter values. With the optimized hyper-parameter, we use the test subset to get the final classification error rate on each data set. In order to provide a comprehensive evaluation, we compare the performance of GERP-SVM and GTWED-SVM with five state-of-the-art similarity measure methods, and two other SVM-based methods. We further use the two-tailed Bonferroni-Dunn test [10] to analyze the statistical difference of multiple classification methods. We choose the Bonferroni-Dunn test because it is a nonparametric test and is suitable for comparing classifiers over multiple data sets.

TABLE I. SUMMARY OF THE UCR TIME SERIES DATA SETS

Data sets	Class	Length	Instances	
			Training	Test
Synthetic Control	6	60	300	300
Gun-Point	2	150	50	150
CBF	3	128	30	900
Face (all)	14	131	560	1,690
OSU Leaf	6	427	200	242
Swedish Leaf	15	128	500	625
50Words	50	270	450	455
Trace	4	275	100	100
Two Patterns	4	128	1,000	4,000
Wafer	2	152	1,000	6,174
Face (four)	4	350	24	88
Lightning-2	2	637	60	61
Lightning-7	7	319	70	73
ECG	2	96	100	100
Adiac	37	176	390	391
Yoga	2	426	300	3,000
Fish	7	463	175	175
Beef	5	470	30	30
Coffee	2	286	28	28
Olive Oil	4	570	30	30

B. Comparison of GEMK-SVM with the similarity measure methods

Using the error rate as the performance indicator, we compare the classification performance of GEMK-SVM with several state-of-the-art similarity measure methods, including nearest neighbor classifier with Euclidean (1NN-ED), nearest neighbor classifier with DTW (1NN-DTW),

nearest neighbor classifier with ODTW [8] (1NN-ODTW), nearest neighbor classifier with ERP (1NN-ERP) [5] and nearest neighbor classifier with OTWED (1NN-OTWED) [6]. Table II lists the classification error rates of these methods on each data set.

By using the two-tailed Bonferroni-Dunn test [10], we analyze the performance difference between these methods. The test result shows that, at the significance level $\alpha = 0.05$, the proposed GERP-SVM and GTWED-SVM methods are statistically better than 1NN-ED, 1NN-DTW, 1NN-ODTW, and 1NN-ERP. Specifically, in all the 20 data sets, the number of GTWED-SVM performs better than 1NN-OTWED is 17, the number of two methods perform the same is 1, and the number of GTWED-SVM performs worse than 1NN-OTWED is 2.

C. Comparison of SVM with GEMK, GRBF and GDTW

GEMK is an extension of the Gaussian RBF (GRBF) kernel, and thus it is interesting to verify whether GEMK could outperform GRBF kernel in the SVM framework. Moreover, GDTW kernel can be regarded as an extension of GRBF kernel by embedding non-metric elastic distance. So it is valuable to discuss the performance difference of the extensions of GRBF kernel by comparing GEMK-SVM and GDTW-SVM. Using the UCR data sets, we compare the error rates of GRBF-SVM, GDTW-SVM, GERP-SVM, and GTWED-SVM, and the results are also listed in Table II.

Compared with GRBF kernel, the proposed GEMKs are very effective in improving the classification accuracy of time series. For example, over all the 20 data sets, GTWED kernel outperforms GRBF kernel on 17 data sets and achieves equivalent error rates on 1 data set, while GRBF kernel outperforms GTWED kernel only on the “Olive Oil” and “Beef” data sets. The reason of the superiority of GEMK may be that GEMK is an elastic kernel and is more effective in handling the local time shifting in time series.

Compared with GDTW kernel, the proposed GEMKs are also more effective in terms of classification accuracy. For example, over all 20 data sets, GERP kernel is superior to GDTW kernel on 15 data sets and achieves equivalent error rates on 4 data sets, while GDTW kernel outperforms GERP kernel only on one data set “Lightning-2”.

Finally, the Bonferroni-Dunn test [10] is adopted to analyze the performance difference between these methods. The results show that, GTWED-SVM and GERP-SVM is statistically better than GRBF-SVM and GDTW-SVM at the significance level $\alpha = 0.05$.

One can see that the classification performance of GDTW kernel is quite unstable. For some data sets, i.e., “Trace” and “Lightning-2”, GDTW performs very well where the classification error rates are comparable to or even better than those of GTWED and GERP. However, on some other data sets, i.e., “Wafer” and “Adiac”, GDTW would achieve very poor classification performance. We argue that the unstable performance of GDTW could be attributed to that DTW is non-metric and GDTW is not PDS acceptable with SVM.

In our experiments, GRBF-SVM takes the least time among all above kernel methods. Because the computational

complexity of Euclidean distance in GRBF kernel is $O(n)$, while in GDTW, GERP and GTWED, the computational complexity of DTW, ERP and TWED is $O(n^2)$. Besides, the numbers of support vectors of GERP-SVM and GTWED-

SVM, which are comparable to that of GDTW-SVM, both are more than that of GRBF-SVM. Thus, compared with GRBF-SVM, it also takes more time for GERP-SVM, GTWED-SVM and GDTW-SVM.

TABLE II. COMPARATIVE STUDY USING THE UCR TIME SERIES DATA SETS: CLASSIFICATION ERROR RATES OBTAINED USING SIMILARITY MEASURE METHODS AND SVM CLASSIFIERS WITH DIFFERENT KERNELS

Datasets Names	INN-Euclidean	INN-DTW	INN-ODTW	INN-ERP	INN-OTWED	GRBF-SVM	GDTW-SVM	GERP-SVM	GTWED-SVM
Synthetic Control	0.120	0.017	0.007	0.036	0.023	0.023	0.017	0.010	0.010
Gun-Point	0.087	0.087	0.093	0.040	0.013	0.047	0.093	0.007	0.000
CBF	0.148	0.004	0.003	0.003	0.007	0.108	0.014	0.012	0.014
Face (all)	0.286	0.192	0.192	0.202	0.189	0.117	0.171	0.137	0.087
OSU Leaf	0.483	0.384	0.409	0.397	0.248	0.421	0.430	0.285	0.182
Swedish Leaf	0.213	0.157	0.210	0.120	0.102	0.112	0.141	0.056	0.053
50Words	0.369	0.242	0.310	0.281	0.187	0.328	0.319	0.253	0.196
Trace	0.240	0.010	0.000	0.170	0.050	0.190	0.000	0.000	0.000
Two Patterns	0.090	0.002	0.000	0.000	0.001	0.099	0.000	0.000	0.000
Wafer	0.005	0.005	0.020	0.009	0.004	0.004	0.025	0.003	0.003
Face (four)	0.216	0.114	0.170	0.102	0.034	0.159	0.159	0.034	0.034
Lightning-2	0.246	0.131	0.131	0.148	0.213	0.328	0.148	0.197	0.180
Lightning-7	0.425	0.288	0.274	0.301	0.247	0.343	0.233	0.192	0.151
ECG	0.120	0.120	0.230	0.130	0.100	0.080	0.160	0.090	0.070
Adiac	0.389	0.391	0.396	0.378	0.376	0.269	0.419	0.269	0.240
Yoga	0.170	0.155	0.164	0.147	0.130	0.137	0.149	0.111	0.110
Fish	0.217	0.160	0.167	0.120	0.051	0.126	0.206	0.051	0.040
Beef	0.467	0.467	0.500	0.500	0.533	0.233	0.333	0.300	0.300
Coffee	0.250	0.179	0.179	0.250	0.214	0.000	0.000	0.000	0.000
Olive Oil	0.133	0.167	0.133	0.167	0.167	0.100	0.133	0.133	0.133

IV. CONCLUSION

In this paper, we propose a novel class of elastic kernel function, GEMK, for SVM-based time series classification. GEMK is an extension of GRBF kernel by incorporating with the elastic metrics. With the help of the recently developed ERP and TWED distance measure, we further present two examples of GEMK: GERP and GTWED kernels. Using the UCR time series data sets, we evaluate the classification performance of GEMK in the SVM framework. Experimental results show that, in terms of classification accuracy, SVM with GEMK is much superior to the state-of-the-art similarity measure methods and SVM with GRBF and GDTW kernels.

ACKNOWLEDGMENT

This work is partially supported by the CERF fund from the HKSAR Government, and the NSFC/SZHK-innovation funds of China under Contract Nos. 60902099, 60871033, 60872099, and SG200810100003A.

REFERENCES

[1] C. Bahlmann, B. Haasdonk, and H. Burkhardt, "On-line handwriting recognition with support vector machines - a kernel approach," IWFHR'02, 2002, pp. 49-54.

[2] H. Shimodaira, K. Noma, M. Nakai, and S. Sagayama, "Dynamic time-alignment kernel in support vector machine," NIPS 14, 2002, pp. 921-928.

[3] H. Lei, and B. Sun, "A Study on the Dynamic Time Warping in Kernel Machines," Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007, pp. 839-845.

[4] S. Gudmundsson, T.P. Runarsson, and S. Sigurdsson, "Support vector machines and dynamic time warping for time series," IJCNN'08, 2008, pp. 2772-2776.

[5] L. Chen, and R. Ng, "On the marriage of LP-norm and edit distance," VLDB'04, 2004, pp.792-801.

[6] P.F. Marteau, "Time warp edit distance with stiffness adjustment for time series matching," IEEE Trans. PAMI, vol. 31, no.2, 2009, pp. 306-318.

[7] E.J. Keogh, X. Xi, L. Wei, C.A. Ratanamahatana (2006). The UCR Time Series Classification /Clustering, Available at: www.cs.ucr.edu/~eamonn/time_series_data/.

[8] C.A. Ratanamahatana, and E.J. Keogh, "Making time-series classification more accurate using learned constraints," SSDM '04, 2004, pp. 11-22.

[9] S.L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," Data Mining and Knowledge Discovery, vol. 1, 1997, pp. 317-327.

[10] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," JMLR, vol. 7, 2006, pp. 1-30.

[11] B. Schölkopf and A.J. Smola, Learning with Kernels, MIT Press, 2002, pp. 49 - 53.