Learning Discriminative Virtual Sequences for Time Series Classification

Abhilash Dorle[1], Fangyu Li[2], Wenzhan Song[1], Sheng Li[1]
[1]University of Georgia, Athens, Georgia, USA
[2]Kennesaw State University, Marietta, Georgia, USA
{AbhilashYeshwant.Dorle,wsong,sheng.li}@uga.edu,fli6@kennesaw.edu

ABSTRACT

Temporal data are continuously collected in a wide range of domains. The increasing availability of such data has led to significant developments of time series analysis. Time series classification, as an essential task in time series analysis, aims to assign a set of temporal sequences to different categories. Among various approaches for time series classification, the distance metric learning based ones, such as the virtual sequence metric learning (VSML), have attracted increased attention due to their remarkable performance. In VSML, virtual sequences attract samples from different classes to facilitate time series classification. However, the existing VSML methods simply employ fixed virtual sequences, which might not be optimal for the subsequent classification tasks. To address this issue, in this paper, we propose a novel time series classification method named Discriminative Virtual Sequence Learning (DVSL). Following the unified framework of sequence metric learning, our DVSL method jointly learns a set of discriminative virtual sequences that help separate time series samples in a feature space, and optimizes the temporal alignment by dynamic time warping. Extensive experiments on 15 UCR time series datasets demonstrate the efficiency of DVSL, compared with several representative baselines.

CCS CONCEPTS

• **Computing methodologies → Machine learning algorithms**; • **Information systems → Data mining**.

KEYWORDS

Time Series Classification, Metric Learning, Virtual Sequences

ACM Reference Format:
Abhilash Dorle[1], Fangyu Li[2], Wenzhan Song[1], Sheng Li[1]. 2020. Learning Discriminative Virtual Sequences for Time Series Classification. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3340531.3412099

1 INTRODUCTION

Time series data are prevalent across many domains, such as health, finance, and entertainment. To name a few examples, data readings from IoT devices could help monitor the status of electrical systems; body wearable devices can provide insightful information about human activities and behavior; surveillance cameras deployed in transit systems could help address public security concerns. The increasing availability of time series data has largely driven the research efforts on time series analysis in recent years. Some fundamental time series analysis tasks include classification, prediction, anomaly detection, clustering and visualization [7–9]. Time series classification, as the primary task in this area, aims to separate temporal sequences into some predefined categories. A large number of time series classification methods have been proposed in the past decade, which can be roughly categorized as the local feature methods, deep learning based methods, and metric learning based methods. Local feature based methods extract representative attributes from time series, such as the shapelets, which are defined as time series subsequences that maximally represent a class [12, 15]. Deep learning based methods learn a latent feature space for time series data, which have obtained impressive performance [3, 4]. Metric learning based methods try to develop metrics that bring closer time series samples from the same class and meanwhile separate out those from different classes. A commonly used metric for time series classification is the dynamic time warping (DTW), which aligns time series through dynamic programming [1]. Many variants of DTW have been proposed to improve the performance and reduce the time complexity [16, 17].

Most recently, a virtual sequence metric learning (VSML) framework for time series classification is proposed in [13]. This framework employs a set of virtual sequences that are inspired by the idea of virtual points [11]. In this framework, virtual sequences are predefined in the sequential data space, and the objectives of metric learning and temporal alignment can be jointly optimized. In particular, this framework brings samples from each class closer to a class-specific virtual sequence, such that the time series samples from different classes can be separated. Although this method has obtained quite promising results on time series classification [13], it still has some limitations. First, the performance of this method heavily relies on the quality of predefined virtual sequences. Since virtual sequence construction and time series classification are two isolated steps, such virtual sequences may not be optimal for the subsequent classification task. Second, the design of virtual sequences is very subjective, which may not fit various downstream applications in practice.

To overcome the limitations in existing work, in this paper, we propose a novel time series classification method by learning discriminative virtual sequences. Our method adaptively learns a set of virtual sequences for time series data, which can be seamlessly integrated with the unified sequence metric learning framework

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412099

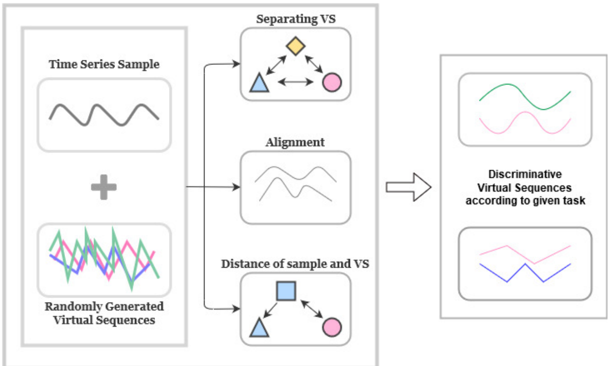

Figure 1: Framework of the proposed DVSL method. The input contains time series samples and randomly generated virtual sequences (VS). During training, our method jointly learns discriminative VS that are well separated, aligns the training samples with the VS, and optimizes the ground metric for classification purpose.

for time series classification. A new optimization algorithm is designed to jointly update the discriminative virtual sequences, the ground metrics, and the temporal alignment matrix. Experimental results on benchmark datasets demonstrate the effectiveness of our method compared with several representative baselines.

The main contributions of this paper are listed below:

(1) We propose a novel time series classification method, as illustrated in Figure 1, which jointly optimizes virtual sequences and metric learning. To the best of our knowledge, our method is the first attempt to adaptively learn virtual sequences for the task of time series classification.
(2) We design a new loss function to enhance the discriminability of virtual sequences. Time series samples from different classes could be pushed to their corresponding virtual sequences that are well separated.
(3) We conduct extensive experiments on 15 UCR benchmark datasets. Our method outperforms baselines in most cases.

2 RELATED WORK

Time series classification indicates the categorization of sequence data. A lot of methods have been proposed for time series classification [6, 9], including the local feature based methods [10], deep learning based methods [4], and metric learning based methods [1, 13]. The local feature based methods aim to extract local patterns, such as shapelets, which are predictive of time series data. The deep learning methods have obtained remarkable performance [5, 14]. For instance, a deep learning architecture based on multi-channel deep convolutional neural networks is proposed to classify multivariate time series data [18]. In [4], deep neural networks are employed for time series classification under data scarcity.

Among the metric learning based methods, the dynamic time warping (DTW) is a commonly used method [1]. DTW based nearest neighbor has produced exceptional results in time series classification. However, the time complexity of DTW is very high. Several algorithms have been proposed to mitigate the complexity brought by DTW. In [16], some learned features are used to weight time series, leading to a weighted DTW method. This method involves representing the scarce training data in an embedded space that is then used for classification. In addition, a recent study takes into consideration the local shape of the samples to improve DTW [17]. As the similarity of local shapes matches, the proposed feature encoding produces better results.

Our method is closely related to the virtual sequence metric learning methods [13]. The concept of virtual point in metric learning was first proposed in [11]. Virtual points are predefined data points that can be used to assist classification in the metric space, by bringing closer examples of the same class to a particular virtual point. This method largely reduces the number of constraints in traditional metric learning. In [13], the concept of virtual points is extended to virtual sequences, which can be used to help sequence data classification. The training time series samples are brought closer to an associated virtual sequence, which generates small values for samples from the same class than those from different classes. A unified virtual sequence metric learning framework is also introduced in [13], which jointly learns the ground metric and aligns the time series samples with virtual sequences. However, the virtual sequences in [13] are predefined and also fixed during model training. The length of the virtual sequence is set to 1 ensuring that the alignment between sequences is unique. As a result, the model performance highly depends on the construction of virtual sequences. Different from existing work, our method adaptively learns a set of discriminative virtual sequences for time series classification, which could be easily adapted to time series data in different domains.

3 OUR APPROACH

3.1 Problem Statement

In this section we provide the definitions of time series, virtual sequence and time series classification.

Definition 3.1. **Time Series.** A time series $X^n = \left[x_1, x_2, \cdots, x_L\right]$ is an ordered sequence of data, where x_t is the value of the time series at time stamp t, and L is the length of the time series. The label of the time series X^n is denoted as y^n.

Definition 3.2. **Virtual Sequence** [13]. A virtual sequence $V^n = \left[v_1, ..., v_{l^n}\right] \in \mathbb{R}^{b' \times l^n}$ is defined as a function of X^n and y^n, i.e., $V^n = f\left(X^n, z^n\right)$.

In [13], virtual sequences are used to assist sequence metric learning and sequence data classification.

Definition 3.3. **Time Series Classification.** Given a test time series X^t, the goal of time series classification is to predict the corresponding class label y^t.

3.2 Formulation

In this paper, we propose a novel time series classification method based on Discriminative Virtual Sequences Learning (DVSL). Let $\mathbf{X} = \{X^n, y^n\}_{n=1}^N$ denote a set of N time series samples. We assume that the time series belong to C different classes. In particular, a time series sample can be represented as $X^n = \left[x_1^n, \cdots, x_{L_n}^n\right]$, where L_n is the length of X^n. A Regressive Virtual Sequence Metric Learning (RVSML) framework is proposed in [13], which jointly optimizes a ground metric $D(W)$ parameterized by W and a temporal alignment matrix T. The objective function of RVSML is formulated as:

$$\min_{W,T} \Phi\big(X, D(W), T, V^*\big) + \Omega\big(T\big), \quad (1)$$

where $\Phi(\cdot)$ denotes metric learning, $\Omega(\cdot)$ denotes the regularization term on T, and V^* is a set of predefined virtual sequences.

The major limitation of RVSML is that the virtual seqeces are fixed during metric learning, which motivates us to explore adaptive and discriminative virtual sequences for time series classification. By considering virtual sequences V as variables, we formulate our method as:

$$\min_{W,T,V} \Phi\big(X, D(W), T, V\big) + \Theta\big(V\big) + \Omega\big(T\big), \quad (2)$$

where W is the parameter matrix and $\Theta\big(V\big)$ is a regularization term.

3.2.1 Distance Metric Learning. The first term of Eq. (2) aims to learn a ground metric $D(W)$ that optimizes the distance between the virtual sequence and the training time series samples, given the alignment matrix T. By using the virtual sequences, DVSL brings closer time series samples from the same class to a specific virtual sequence, and meanwhile push away samples from different classes. Specifically, $\Phi\big(X, W, T, V\big)$ is formulated as:

$$\begin{aligned}\Phi\big(X, D(W), T, V\big) &= \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \Big\langle T^n, D_I^n\Big(W\Big)\Big\rangle + \lambda\Big\|W\Big\|_F^2 \\ &= \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \sum_{i=1}^{L_n} \sum_{j=1}^{l_n} t_{ij}^n \Big\|W^T x_i^n - v_j^n\Big\|_2^2 + \lambda\Big\|W\Big\|_F^2\end{aligned} \quad (3)$$

where λ is a trade-off parameter, and l_n is the length of virtual sequence. t_{ij}^n is an element in the alignment matrix T, which captures how closely the training sample and the virtual sequence align.

3.2.2 Virtual Sequence Learning. As discussed above, the virtual sequences used in RVSML [13] are fixed, which may not be optimal for the subsequent time series classification task. Instead, our DVSL method aims to learn a set of discriminative virtual sequences that can directly benefit the classification task. In particular, the discriminative virtual sequences shall be well separated in the data space. To this end, we design the second term in Eq. (2) as follows:

$$\Theta\big(V\big) = -\sum_{n=1}^{N} \sum_{c=1,c\neq n}^{N} \sum_{j=1}^{l^n} \sum_{i=1}^{l^n} \Big\|V_j^n - V_i^c\Big\|_2^2 \quad (4)$$

where V_j and V_i denote virtual sequences. With this term, we push the virtual sequences far away from each other by maximizing their pairwise distances in the sequential data space. To align with the minimization problem in Eq. (2), a negative sign in added to Eq. (4).

3.2.3 Objective Function. Combining the distance metric learning, virtual sequence learning and the regularization term, the overall objective function of our DVSL method is written as:

$$\begin{aligned}\min_{W,T,V} \mathcal{L} = &\sum_{n=1}^{N} \sum_{i=1}^{L^n} \sum_{j=1}^{l^n} \frac{1}{N} t_{ij}^n \Big\|W^T x_i^n - v_j^n\Big\|_2^2 + \lambda\Big\|W\Big\|_F^2 \\ &- \sum_{n=1}^{N} \sum_{c=1c\neq n}^{N} \sum_{j=1}^{l^n} \sum_{i=1}^{l^n} \Big\|V_j^n - V_i^c\Big\|_2^2 + \Omega\big(T\big).\end{aligned} \quad (5)$$

Table 1: Times series classification results of the proposed DVSL method and baselines on UCR datasets. The names of the datasets have been shortened to accommodate the details

Dataset	ED	DTW	LSDTW	RVSML	DVSL
ArrowHead	**80.00**	70.29	73.14	74.86	72.00
Beef	66.67	63.33	83.33	83.33	**90.00**
Car	73.33	73.33	**86.67**	80.00	83.50
ChlConcent	65.00	64.84	72.11	60.83	**77.43**
Coffee	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
ECG200	**88.00**	77.00	85.00	85.00	83.50
ECGFiveDays	79.67	76.77	90.01	95.00	**97.36**
Herring	51.56	53.13	46.88	65.60	**65.63**
InsectWingb	56.16	35.51	50.91	**58.83**	58.19
Meat	93.33	93.33	80.00	90.00	**98.83**
MPhaOLAge	51.95	50.00	56.49	48.70	**58.18**
OliveOil	**86.67**	83.33	83.33	73.30	84.67
SonyAIBR1	69.55	72.55	78.20	**83.86**	76.16
TwoLeadECG	74.71	90.52	**93.15**	91.83	91.60
Wine	61.11	57.41	57.41	59.25	**65.00**

3.3 Optimization

Although the objective function in Eq. 5 is not jointly convex with respect to all the variables W, V and T, it is convex to each variable separately when the others are fixed. Thus, we alternatively update these variables. In particular, for the subproblem w.r.t. T, we implement it with DTW that is solved by dynamic programming. For W and V, we employ a gradient descent approach by initializing the variables $W^{(0)}$ and $V^{(0)}$ with random values and then updating them with the following rules: $W^{(t+1)} = W^{(t)} - \gamma \frac{\partial \mathcal{L}}{\partial W}$ and $V_j^{t+1} = V_j^{(t)} - \gamma \frac{\partial \mathcal{L}}{\partial V_j}$, where γ is a learning rate. The detailed derivatives with respect to W and V are written as:

$$\frac{\partial \mathcal{L}}{\partial W} = A^{-1} \sum_{n=1}^{N} \sum_{i=1}^{L_n} \sum_{j=1}^{l_n} t_{ij}^n x_i^n {v_j^n}^T, \quad (6)$$

where $A = \sum_{n=1}^{N} \sum_{i=1}^{L_n} \sum_{j=1}^{l_n} t_{ij}^n x_i^n {x_i^n}^T + \lambda N I.$

$$\frac{\partial L}{\partial V_j} = \Big(\sum_{n=1}^{N} \sum_{i=1}^{L_n} t_{ij}^n W^T x_i^n - \sum_{n=1}^{N} \sum_{c=1c\neq n}^{N} \sum_{i=1}^{l_n} V_i^c N)\Big)/C, \quad (7)$$

where $C = \sum_{n=1}^{N} \sum_{i=1}^{L_n} t_{ij} - N \sum_{n=1}^{N} \sum_{c=1c\neq n}^{N} \sum_{i=1}^{l_n} 1.$

After optimizing the virtual sequences V and the ground metric $D(W)$, we employ the one nearest neighbor (1-NN) classifer with DTW distance measure for time series classification.

4 EXPERIMENTS

In this section, we evaluate the performance of our DVSL method and compare it with baseline methods on benchmark datasets.

Datasets. We evaluate our method thoroughly on the UCR time series archive [2]. The repository contains datasets from various real word domains. It consists of time series data from various sources like sensor data, image data and spectrograph data. We perform experiments on 15 datasets to show the effectiveness of our method.

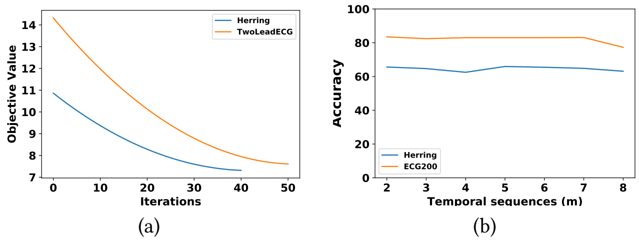

Figure 2: (a) Convergence curves. (b) Parameter sensitivity.

Table 2: Comparison between our method and deep learning methods.

Dataset	DPN	MHLF	DVSL
Coffee	**100.00**	**100.00**	**100.00**
ECGFiveDays	**100.00**	**100.00**	97.36
Herring	67.00	**68.75**	65.63
Meat	98.00	95.00	**98.83**

Baselines. We mainly compare DVSL with the following baselines. (1) *Euclidean Distance (ED)*. It's a standard metric for time series comparison. (2) *Dynamic Time Warping (DTW)* [1]. This is one of the most competitive methods that look for optimal alignments between sequences. (3) *Regressive Virtual Sequence Metric Learning (RVSML)*. [13] This method learns distances for sequences by finding a ground metric. (4) *Locally Slope-based Dynamic Time Warping for Time Series Classification (LSDTW)* [17]. It implements a weighted DTW technique by looking at regional information and pairing locally similar shapes. Moreover, we compare our method with the recently proposed deep learning methods, including the deep prototypical networks (DPN) [4] and the multi-channel MHLF [3]. We follow the standard evaluation protocols on UCR datasets and report the classification accuracy for each compared method.

Parameter Settings. There are three major hyperparameters in our method: the regularization parameter λ, the learning rate γ, and the number of temporal structures per class m. To tune the hyperparameters, we use a validation set and do a grid search. The range of values for β varies from 10^{-4} to 10^{-1}, γ varies from 10^{-4} to 10^{-2}, and m varies from 2 to 8.

Results and Analysis. We perform experiments on 15 datasets from the UCR archive and demonstrate the effectiveness of our method. Table 1 summarizes the results of our DVSL method and baselines. The results demonstrate that our method outperforms baselines in most of the cases and produces comparable results in the other cases. In addition, Table 2 shows the comparison between our method and deep learning based methods. The results from Table 2 suggest that, even without the sophisticated deep feature learning, our DVSL method obtains comparable results to the deep learning methods in many cases. Similar to RVSML, our method can be extended to learn nonlinear representations with deep neural networks, which will be explored in our future work.

Compared with RVSML, our method has more variables for optimization. However, the computational efficiency of our method is comparable to that of RVSML, owing to the efficient calculations of gradients. For instance, on the SonyAIBORobotSurface1 dataset, the training time (in seconds) of RVSML and our method are 0.55s and 0.70s, respectively.

In addition, we analyze the convergence property of our model as well as the parameter sensitivity. Figure 2(a) shows that our method converges quickly within 50 iterations. Figure 2(b) displays the parameter sensitivity to m. Our method obtains relatively stable results when m varies from 2 to 8.

5 CONCLUSION

In this paper, we proposed a novel time series classification method based on Discriminative Virtual Sequences Learning (DVSL). The goal of DVSL is to bring closer time series samples from the same class to a specific virtual sequence and meanwhile push away samples from different classes. Different from existing work, DVSL adaptively learns a set of discriminaitve virtual sequences. A new objective function is formulated and a gradient descent algorithm is designed for optimization. Experiments on 15 UCR time series datasets demonstrate that our method outperforms several representative baselines.

REFERENCES

[1] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *KDD Workshop*, Vol. 10. 359–370.
[2] Hoang Anh Dau, Eamonn Keogh, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, Yanping, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, Gustavo Batista, and Hexagon-ML. 2018. The UCR Time Series Classification Archive. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.
[3] Shuichi Hashida and Keiichi Tamura. 2019. Multi-Channel MHLF: LSTM-FCN using MACD-Histogram with Multi-Channel Input for Time Series Classification. In *IEEE IWCIA*. IEEE, 67–72.
[4] Chao Huang, Xian Wu, Xuchao Zhang, Suwen Lin, and Nitesh V Chawla. 2019. Deep Prototypical Networks for Imbalanced Time Series Classification under Data Scarcity. In *CIKM*. 2141–2144.
[5] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2017. LSTM fully convolutional networks for time series classification. *IEEE Access* 6 (2017), 1662–1669.
[6] Kang Li, Sheng Li, and Yun Fu. 2014. Early classification of ongoing observation. In *ICDM*. IEEE, 310–319.
[7] Sheng Li, Kang Li, and Yun Fu. 2015. Temporal subspace clustering for human motion segmentation. In *ICCV*. 4453–4461.
[8] Sheng Li, Kang Li, and Yun Fu. 2018. Early recognition of 3D human actions. *ACM Transactions on Multimedia Computing, Communications, and Applications* 14, 1s (2018), 1–21.
[9] Sheng Li, Yaliang Li, and Yun Fu. 2016. Multi-view time series classification: A discriminative bilinear projection approach. In *CIKM*. 989–998.
[10] Alex Nanopoulos, Rob Alcock, and Yannis Manolopoulos. 2001. Feature-based classification of time-series data. *International Journal of Computer Research* 10, 3 (2001), 49–61.
[11] Michaël Perrot and Amaury Habrard. 2015. Regressive virtual metric learning. In *NIPS*. 1810–1818.
[12] Mit Shah, Josif Grabocka, Nicolas Schilling, Martin Wistuba, and Lars Schmidt-Thieme. 2016. Learning DTW-shapelets for time-series classification. In *Proceedings of the 3rd IKDD Conference on Data Science*. 1–8.
[13] Bing Su and Ying Wu. 2019. Learning distance for sequences by learning a ground metric. In *ICML*. 6015–6025.
[14] Zhiguang Wang, Weizhong Yan, and Tim Oates. 2017. Time series classification from scratch with deep neural networks: A strong baseline. In *IJCNN*. 1578–1585.
[15] Lexiang Ye and Eamonn Keogh. 2009. Time series shapelets: a new primitive for data mining. In *KDD*. 947–956.
[16] Jidong Yuan, Ahlame Douzal-Chouakria, Saeed Varasteh Yazdi, and Zhihai Wang. 2019. A large margin time series nearest neighbour classification under locally weighted time warps. *Knowledge and Information Systems* 59, 1 (2019), 117–135.
[17] Jidong Yuan, Qianhong Lin, Wei Zhang, and Zhihai Wang. 2019. Locally Slope-based Dynamic Time Warping for Time Series Classification. In *CIKM*. 1713–1722.
[18] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J Leon Zhao. 2016. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science* 10, 1 (2016), 96–112.