

# Adversarial Spectral Kernel Matching for Unsupervised Time Series Domain Adaptation

Qiao Liu<sup>1,2</sup>, Hui Xue<sup>1,2\*</sup>

<sup>1</sup>School of Computer Science and Engineering, Southeast University, Nanjing, 210096, China

<sup>2</sup>MOE Key Laboratory of Computer Network and Information Integration (Southeast University), China  
 {qiaoliu, hxue}@seu.edu.cn

## Abstract

Unsupervised domain adaptation (UDA) has been received increasing attention since it does not require labels in target domain. Most existing UDA methods learn domain-invariant features by minimizing discrepancy distance computed by a certain metric between domains. However, these discrepancy-based methods cannot be robustly applied to unsupervised time series domain adaptation (UTSDA). That is because discrepancy metrics in these methods contain only low-order and local statistics, which have limited expression for time series distributions and therefore result in failure of domain matching. Actually, the real-world time series are always non-local distributions, i.e., with non-stationary and non-monotonic statistics. In this paper, we propose an Adversarial Spectral Kernel Matching (AdvSKM) method, where a hybrid spectral kernel network is specifically designed as inner kernel to reform the Maximum Mean Discrepancy (MMD) metric for UTSDA. The hybrid spectral kernel network can precisely characterize non-stationary and non-monotonic statistics in time series distributions. Embedding hybrid spectral kernel network to MMD guarantees precise discrepancy metric and benefits domain matching. Besides, the differentiable architecture of the spectral kernel network enables adversarial kernel learning, which brings more discriminatory expression for discrepancy matching. The results of extensive experiments on several real-world UTSDA tasks verify the effectiveness of our proposed method.

## 1 Introduction

Unsupervised domain adaptation (UDA) is a machine learning strategy that transfers knowledge from one domain to another [Pan and Yang, 2009]. The unsupervised adaptation progress utilizes labeled data from the source domain but unlabeled data from the target domain to train the model. Many successful cases have been achieved by methods in traditional research fields as computer vision [Long *et al.*, 2015] and

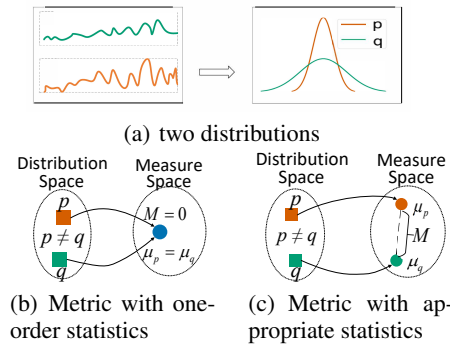


Figure 1: Illumination of the importance of suitable statistics. (a): Two different example distributions from time series. (b): Unsuitable statistics fails to capture all information in distribution, leading to non-injective embedding and an incorrect metric value. (c): Appropriate statistics can represent sufficient information in distribution, resulting in injective embedding and an appropriate metric value.

text categorization [Dai *et al.*, 2009]. These methods usually learn a domain-invariant space by minimizing discrepancy distance computed by a certain metric between source and target domains. Maximum Mean Discrepancy (MMD) metric [Gretton *et al.*, 2007] is the which utilizes a kernel function to extract statistics from samples to the Reproducing Kernel Hilbert Space (RKHS) and measures the domain discrepancy on RKHS. For example, DDC [Tzeng *et al.*, 2014] applies MMD with a linear kernel which captures first-order statistics (Mean). DAN [Long *et al.*, 2015] enhances MMD metric with multiple Gaussian kernels. Besides, HoMM [Chao *et al.*, 2020] proposed an innovate high-order metric, which approximates to MMD with third or fourth order polynomial kernel. UDA methods with MMDs embedded by distinct kernels have several advantages over methods with classical metrics, namely easy convergence and low bias of finite sample estimates. However, when met with the unsupervised time series domain adaptation (UTSDA) task, a challenging and emerging area of UDA, these methods cannot be robustly applied to.

For the real-world time series, the data usually change through time (i.e., non-stationary) and continuously reverse derivatives (i.e., non-monotonic). Based on Fourier spectral theory, these non-stationary and non-monotonic statis-

\*Contact Author

tics are non-local in the energy-frequency-time distributions [Clements and Hendry, 2000]. However, existing methods based on MMD metrics only capture the low-order or local statistics of distributions on structured objects such as images, instead of appropriate statistics for time series, leading to rough-grained domain matching. The linear-kernel MMD in DDC capturing low-order statistics degrades to Euclidean distance metric. The MMD with multiple Gaussian kernel in DAN reveals local statistics, i.e., only capturing the identical similarity of structured data [Xue *et al.*, 2019]. Even the high-order MMD in HoMM is only suitable for traditional images with translation invariance and still suffers from locality. Visual proof of the shortcomings of these local statistics are shown in Fig.1, where two time series distributions with same mean are given in Fig.1(a). In Fig.1(b), the unsuitable statistics maps two distributions into the same point in the measure space and leads to incorrect distance metric. But the appropriate statistics in Fig.1(c) can capture sufficient information of distributions and make method effectively distinguish between domains.

To solve the above problems, we propose a Adversarial Spectral Kernel Matching (AdvSKM) method for UTSDA. Our method perform time series domain matching by minimizing a reinforced MMD metric embedded by a well-designed hybrid spectral kernel network. The hybrid spectral kernel network interpretably integrates spectral kernel and arc-cosine kernel into the deep architecture via spectral theorems and kernel approximation, which exposes non-stationary and non-monotonic properties through compact representation of complex time series and boosts computational efficiency. Adopting the hybrid spectral kernel network as inner kernel for MMD not only facilitates time series feature mapping to RKHS, but also reveals non-local properties in time series distributions, and therefore ensures injective embedding to RKHS and guarantees fine-grained domain matching. Besides, the differentiable architecture of the spectral kernel network enables adversarial kernel learning, thus bringing more discriminatory expression for discrepancy matching. Extensive experimental results demonstrate the superiority of our AdvSKM to several state-of-the-art adaptation methods in UTSDA task.

## 2 Related Works

Several researches have been devoted to time series classification (TSC) problems. Wang *et al.* [2017] designed a Fully Convolutional Neural (FCN) Network for the TSC problem. The FCN consists of three convolutional layers without any subsampling, followed by a global average pooling layer, and the last layer is a softmax classifier. Although recurrent neural networks (RNN) propagating through time are intuitively more suitable for TSC problems, experimental results in [Bai *et al.*, 2018] show that the simple FCN outperforms canonical RNNs such as LSTMs across a diverse range of tasks. However, the performance of backbone FCN trained on specific domain decreases abruptly when applied to the other domain. This problem can be solved with transfer learning strategy.

Unsupervised domain adaptation (UDA) is a rising field in transfer learning since it does not require labels in tar-

get domains. Most UDA methods generally learn a domain-invariant space by minimizing domain discrepancy between domains. These discrepancy-based methods can be categorized into two groups: (1) Metric-based: Learning domain-invariant features between domains by reducing the discrepancy distance computed by a certain metric. (2) Domain-discrimination-based: Blending domains with a domain adversarial loss, specifically by training discriminators to distinguish between source and target domains.

Most of the metric-based methods are based on Maximum Mean Discrepancy (MMD) metric. MMD has several advantages over classical metrics, since it compares two distributions without density estimation. Specifically, Tzeng *et al.* [2014] proposed DDC which matches one-order statistic at last layer of the model with linear-kernel MMD. Further, Long *et al.* [2015] presented DAN which performs domain matching at multiple layers via multiple Gaussian kernel MMD. Sun and Saenko [2016] proposed Correlation alignment (CORAL) which is equivalent to MMD with second-order polynomial kernel. Besides, Central Moment Discrepancy (CMD) [Zellinger *et al.*, 2017] and Higher-order Moment Matching (HoMM) [Chao *et al.*, 2020] further considered higher-order metrics, which both approximate to the MMD using third or fourth order polynomial kernel. However, these methods capture low-order or local statistics that are only suitable for structured data instead of non-local statistics for time series, leading to rough-grained domain matching in unsupervised time series domain adaptation (UTSDA) task.

The second category of domain-discrimination-based approaches such as Domain-adversarial Neural Network (DANN) [Ganin *et al.*, 2016] attempt to minimize domain discrepancy by distinguishing domains. However, the backbone architectures in these methods are designed for image data and is not adapted to the time series data. Therefore, they cannot be applied robustly to UTSDA.

Although less numerous, a few methods have been introduced to UTSDA. These methods combine different TSC models as backbones with an domain adversary following DANN. Specifically, variational recurrent adversarial deep domain adaptation (VRADA) [Purushotham *et al.*, 2017] and recurrent domain adversarial neural network (R-DANN) explore long short-term memory (LSTM) network and variational RNN as backbone, respectively. While Convolutional Domain Adaptation model for Time Series data (CoDATS) [Wilson *et al.*, 2020] utilizes FCN as backbone network. However, in terms of domain matching methodology, these UTSDA methods all follow domain-discrimination rule, where domain and category classification losses are equally important and a slight imbalance between them can cause inter-class confusion and leads to accuracy degradation.

Unlike the UTSDA methods above, our approach minimize the domain divergence directly on a special MMD metric for time series data. The special MMD metric is embedded by a well-designed hybrid spectral kernel network as inner kernel. The hybrid spectral kernel network can precisely characterize non-stationary and non-monotonic statistics instead of local statistics in time series distributions. Embedding the hybrid spectral kernel network to MMD guarantees precise discrep-

ancy metric and benefits domain matching in UTSDA.

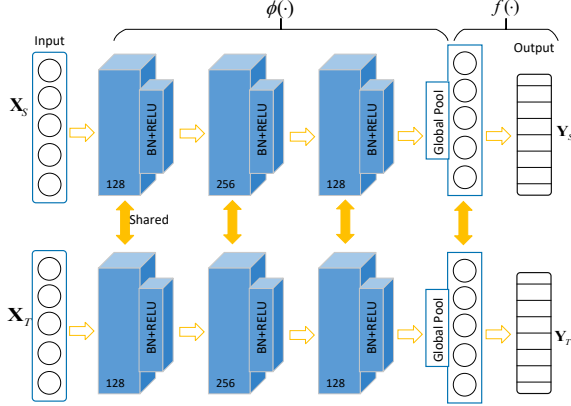


Figure 2: The architecture of the two-stream FCN  $\phi(\cdot)$  with shared parameters and the classifier  $f(\cdot)$ . The up-stream operates the source samples and the down-stream operates the target samples.

### 3 Our Method

Our method aim to solve the unsupervised domain adaptation task where a significant amount of labeled source instances and several unlabeled target instances are given. We first denote the instances and labels from the source and target domains with  $\mathbf{X}_S = [\mathbf{x}_1^s, \dots, \mathbf{x}_{n_s}^s]^\top \in \mathbb{R}^{d_s \times n_s}$ ,  $\mathbf{Y}_S = [\mathbf{y}_1^s, \dots, \mathbf{y}_{n_s}^s] \in \mathbb{R}^{n_s \times 1}$  and  $\mathbf{X}_T = [\mathbf{x}_1^t, \dots, \mathbf{x}_{n_t}^t]^\top \in \mathbb{R}^{d_t \times n_t}$ ,  $\mathbf{Y}_T = [\mathbf{y}_1^t, \dots, \mathbf{y}_{n_t}^t] \in \mathbb{R}^{n_t \times 1}$ , respectively. The source domain provides  $N_S$  labeled instances and the target domain data contains  $N_T$  unlabeled instances. In this paper, we focus on the multivariate time series classification task. Time series include both univariate and multivariate data. If univariate, the sample  $\mathbf{x} = [x_1, x_2, \dots, x_H]$  consists of  $H$  real values through time. If multivariate, the sample  $\mathbf{x} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^K]$  contains  $K$  univariate time series, each of which consists of  $H$  real values through time.

Given source and target samples, we aim to train a cross-domain FCN backbone network  $\phi(\cdot)$  and a classifier  $f(\cdot)$  which together map cross-domains samples to a unified space and identify unlabeled target samples on the new space.  $\phi(x)$  denotes the outputs of the deep FCN and  $f(\phi(x))$  denotes the predicted labels from the classifier. Similar to the experimental framework in [Long *et al.*, 2015], a two-stream FCN framework is adopted as in Fig.2. We aim to minimize domain discrepancy at the last global pooling layer. The framework is trained under a basic domain adaptation loss including source domain classification loss and domain discrepancy loss:

$$\mathcal{L} = \underbrace{\mathcal{L}_c(\phi, f, \mathbf{X}_S, \mathbf{Y}_S)}_{\text{classification}} + \lambda \cdot \underbrace{\mathcal{L}_d(\phi, \mathbf{X}_S, \mathbf{X}_T)}_{\text{discrepancy}} \quad (1)$$

$$\mathcal{L}_c(\phi, f, \mathbf{X}_S, \mathbf{Y}_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} J(f(\phi(\mathbf{X}_S)), \mathbf{Y}_S) \quad (2)$$

where  $\mathcal{L}_c$  is the classification loss for source domain,  $\mathcal{L}_d$  is the domain discrepancy loss, and  $\lambda$  is the factor. Specifically,

the classification loss function is defined as (2) and  $J$  is the cross-entropy function.

#### 3.1 Maximum Mean Discrepancy Matching

The two-stream FCN with shared parameters  $\phi(\cdot)$  transform instances from two distinct domains into the unified space. The new representations of the two domains on the new space often obey distinct distributions. We utilized kernel-based MMD metric as a loss to minimize this divergence.

And to go deeper, MMD compares two distributions by embedding each distribution in to RKHS  $\mathcal{H}$  with a kernel function  $k$ . When given two distributions  $\mathbb{P}_X, \mathbb{Q}_X$  and the MMD can be expressed as:

$$\begin{aligned} \text{MMD}[\mathcal{H}, \mathbb{P}, \mathbb{Q}] &= \|\mu_{\mathbb{P}_X} - \mu_{\mathbb{Q}_Y}\|_{\mathcal{H}} \\ &= \left\| \mathbb{E}_{\mathbb{P}_X} [k(X, \tilde{X})] - \mathbb{E}_{\mathbb{Q}_Y} [k(Y, \tilde{Y})] \right\|_{\mathcal{H}} \end{aligned} \quad (3)$$

The kernel function serves to extract statistics from samples to RKHS. However, when met with time series data, the commonly used shallow kernels cannot effectively capture appropriate statistics for MMD. For instance, MMD with a linear kernel captures only one-order statistics. MMD with Gaussian kernel cannot reveal non-local statistics except for the identical similarity of structured data [Xue *et al.*, 2019]. Besides, these shallow kernels are also restricted by the computational inefficiency as the learning problems become more complex.

#### 3.2 Hybrid Spectral Kernel Network

To overcome the shortcoming of these kernels, we design a deep hybrid spectral kernel network specifically for time series data. We first interpretably integrate non-stationary and non-monotonic spectral kernel into the deep architectures via several spectral theorems [Xue *et al.*, 2019]. Based on the formula of deep spectral kernel network, the spectroscopic kernel function is represented by the spectral kernel mapping:

$$\Phi_s(\mathbf{x}) = \sqrt{\frac{1}{2M}} [\cos(\Omega^T \mathbf{x} + \varphi) + \cos(\Omega'^T \mathbf{x} + \varphi')] \quad (4)$$

$$k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle \quad (5)$$

where (4) is the spectral kernel mapping. Through this spectral kernel mapping, the spectroscopic kernel function can be represented as (5) and  $\Omega = [\omega_1, \dots, \omega_M]$ ,  $\Omega' = [\omega'_1, \dots, \omega'_M]$  denotes the frequency matrices in inverse Fourier transform. In terms of structural form, the kernel function (5) represented in spectral way is similar to a one-layer neural network. Except for non-stationary and non-monotonic property, this deep kernel with cosine activation function is more likely to depict the periodic and long-term characteristics of a distribution.

In order to preserve locality and general information in a distribution, we further introduce arc-cos kernel [Cho and Saul, 2009]:

$$k(x, x') = \frac{1}{\pi} \|x\| \|x'\| J(\theta) \quad (6)$$

where  $J(\theta) = (\sin \theta + (\pi - \theta) \cos \theta)$  and  $\theta = \cos^{-1} \left( \frac{x \cdot x'}{\|x\| \|x'\|} \right)$ .  $\theta$  is the angle between  $x$  and  $x'$ . Through kernel approximation, the feature map of arc-cos kernel is related to ReLUs in deep neural networks as:

$$\Phi_a(x) = \sqrt{\frac{1}{D}} \max(0, W^T x) \quad (7)$$

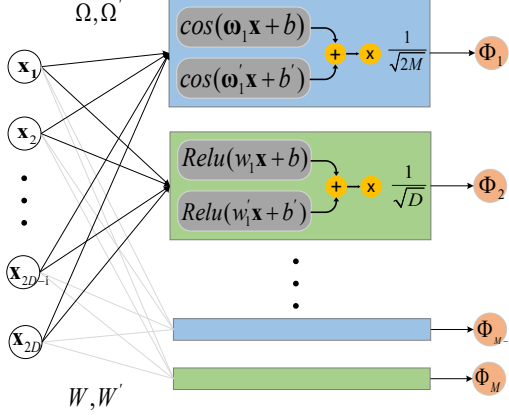


Figure 3: The structure of the hybrid spectral kernel mapping.

Based on the above two deep kernel mappings, we design a hybrid spectral kernel mapping as in Fig. 3. In the hybrid spectral kernel mapping, spectral and arc-cos kernels are parallel and intertwined as  $\Phi_H(\mathbf{x}) = [\Phi_s^1(\mathbf{x}), \Phi_a^1(\mathbf{x})]$ . Then, we use the unified spectral transform technique [Xue *et al.*, 2019] to stack an extra layer-wise kernel mapping in a hierarchical composite way to construct the hybrid spectral kernel network. The notation of the final hybrid kernel network is:

$$\mathcal{K}^{(2)}(\mathbf{x}, \mathbf{x}') = \langle \Xi^{(2)}(\mathbf{x}), \Xi^{(2)}(\mathbf{x}') \rangle \quad (8)$$

where the composite mapping is  $\Xi^{(2)}(\mathbf{x}) = \Phi_H^2(\Phi_H^1(\mathbf{x}))$ . Combining deep spectral and arc-cos kernels provides a complete picture of the distribution and therefore circumvents drawbacks of the classical kernels. Besides, the deep architecture can also avoid the tricky kernel training with specific predefined kernel.

With the well-designed hybrid kernel network  $\mathcal{K} = \mathcal{K}^{(2)}$ , we reinforce the biased empirical estimation of MMD metric and design the distribution matching loss function as:

$$\begin{aligned} \mathcal{L}_d(\phi, \mathbf{X}_S, \mathbf{X}_T, \mathcal{K}) &= \sum_{i,j} \frac{\mathcal{K}(\phi(\mathbf{x}_i^s), \phi(\mathbf{x}_j^s))}{N_S^2} \\ &+ \sum_{i,j} \frac{\mathcal{K}(\phi(\mathbf{x}_i^t), \phi(\mathbf{x}_j^t))}{N_T^2} - 2 \sum_{i,j} \frac{\mathcal{K}(\phi(\mathbf{x}_j^s), \phi(\mathbf{x}_i^t))}{N_S N_T} \end{aligned} \quad (9)$$

### 3.3 Adversarial Kernel Training

The differentiable architecture of hybrid spectral kernel network makes adversarial kernel learning possible and therefore results in a more precise distribution alignment. Unlike domain-discrimination-based methods, our method directly applies adversarial learning to the metric computation,

thus bringing more discriminatory expression for discrepancy metric. The adversarial training procedure in our AdvSKM can be represented by reformulating (9) as:

$$\min_{\phi, f} \max_{\mathcal{K}} \mathcal{L}_d(\phi, \mathbf{X}_S, \mathbf{X}_T, \mathcal{K}) \quad (10)$$

The procedure in (10) is associated with Generative Adversarial Network (GAN), where a generative model  $G$  and a discriminative model  $D$  are constructed to generate and discriminate instances, respectively. In our paper, the backbone network  $\phi(\cdot)$  is similar to  $G$  and strive to narrow the gap between domains, while  $\mathcal{K}$  is similar to  $D$  and enlarges the gap. These two competitive components are trained separately.  $\mathcal{K}$  is trained by maximizing the distribution matching loss (9), while  $\phi(\cdot)$  and  $f$  are trained when minimizing the overall loss function (11) with fixed  $\mathcal{K}$  as input. Through adversarial training, the MMD with backbone network  $\phi$  and classifier  $f$  can achieve more accurate results and better generalization capability.

### 3.4 Optimization

The components above construct the final objective function:

$$\mathcal{L} = \mathcal{L}_c(\phi, f, \mathbf{X}_L, \mathbf{Y}_L) + \lambda \cdot \mathcal{L}_d(\phi, \mathbf{X}_S, \mathbf{X}_T, \mathcal{K}) \quad (11)$$

Based on the designed objective, the main steps of our proposed method are illustrated in Algorithm 1 where the back propagation optimization is based on the adaptive moment estimation (ADAM) algorithm. At first, the networks  $f, \phi, \mathcal{K}$  are initialized by truncated Gaussian distribution. Secondly, source and target instances are transformed into the invariant space with the FCN  $\phi$ . Then our model is trained in a max-mini adversarial manner. At the maximization stage, the hybrid spectral kernel network  $\mathcal{K}$  is trained with iteration  $L = 3$ . Correspondingly at the minimization phase, the projection networks and the classifier are jointly optimized with kernel network  $\mathcal{K}$  as input. With tighter distribution distance, models can be fitted before the max iterations  $M = 30,000$ .

---

#### Algorithm 1 AdvSKM

---

**Input:** Instances and labels  $\mathbf{X}_S, \mathbf{X}_T$  and  $\mathbf{Y}_S$ ; Initialized  $\phi, f$  and  $\mathcal{K}$ ; Hyperparameters  $\lambda, L$  and  $M$ .

**Output:** Labels of target test data,  $\mathbf{Y}_T$ .

**while** not convergence or max iteration **do**

    Project  $\mathbf{X}_S$  and  $\mathbf{X}_T$  to the common space with  $\phi$ .

**for**  $i=1:L$  **do**

        Compute and gradient of (9) w.r.t  $\mathcal{K}$ .

        Update  $\mathcal{K}$  to maximize (9) with gradient.

**end for**

    Compute the loss (9) with updated  $\mathcal{K}$ .

    Compute gradient of (11) w.r.t  $\phi$  and  $f$ .

    Update  $\phi$  and  $f$  to minimize (11) with gradient.

**end while**

Predict  $\mathbf{Y}_T$  with the optimized model.

---

## 4 Experiments

### 4.1 Datasets

In the experimental settings, four multivariate time series datasets collected by sensors are selected following [Wilson *et al.*, 2020]. The first dataset is the Human Activity Recognition (HAR) dataset [Anguita *et al.*, 2013] which contains accelerometer, gyroscope, and estimated body acceleration data from 30 participants. The second is the Heterogeneity Human Activity Recognition (HHAR) dataset [Stisen *et al.*, 2015], including acceleration sensor data from 31 smartphones. Next is WISDM activity recognition (WS AR) dataset [Kwapisz *et al.*, 2011] including 33 participants’ accelerometer data. Finally, hand gesture accelerometer data from 8 participants is provided from the gesture recognition dataset (uWave) [Liu *et al.*, 2009]. In each dataset, certain sensor device is used to capture different activity behaviors of the participants. Each dataset includes a participant identifier, and we use this property to delineate the domains.

### 4.2 Comparing Methods

We compare our method with the following methods: Deep Domain Confusion (DDC) [Tzeng *et al.*, 2014], Deep Adaptation Network (DAN) [Long *et al.*, 2015], Deep Correlation Alignment (CORAL) [Sun and Saenko, 2016], Domain-adversarial Neural Network (DANN) [Ganin *et al.*, 2016], Central Moment Discrepancy (CMD) [Zellinger *et al.*, 2017], Higher-order Moment Matching (HoMM) [Chao *et al.*, 2020], Convolutional Domain Adaptation for Time Series (CoDATS). Specially, DDC, DAN, CORAL and CMD are the representative metric-based methods, while DANN and CoDATS are the representative domain-discrimination-based

methods. Among them, CoDATS is the latest state-of-the-art adaptation method for time series data.

### 4.3 Experimental Settings

All the comparison methods are based on the same FCN backbone network for a fair comparison. We implemented our experimental framework by following the official codes of CoDATS. Following CoDATS’s released codes, the difference between CoDATS and DANN is the domain/task classifier with different layers. We split the data in each dataset into training, validation and test. The training-test were split at 80% and 20%, respectively, and the training data were further split into training-validation with the same proportions. We randomly selected 5 adaptation problems between two domains. For each data point, we average over three different randomly initialized network weights on test set. During the training, we performed model selection by selecting the model that performed best on the validation set. We evaluate the models every 4,000 iterations in addition to the end of the training. The framework was trained for 30,000 iterations, which means that the best model was selected from nine models. Our codes <sup>1</sup> are published online.

### 4.4 Results

Table 1 tabulate the experimental results on four datasets. Boldface is used to highlight the best result of each case. "FCN" in the table represents the no adaptation model. From the results, we can observe that our method achieves the highest accuracy in all cases, which demonstrates the superiority of our method in various cross-domain scenarios. The poor

<sup>1</sup><https://github.com/jarheadjoe/Adv-spec-ker-matching>

Problems	FCN	DDC	DAN	CORAL	DANN	CMD	HoMM	CoDATS	Ours
HAR 14→19	13.9±0.0	20.4±0.7	76.4±6.0	79.6±8.8	85.6±9.4	54.6±3.3	88.9±11.2	66.2±4.3	<b>98.1±0.7</b>
HAR 17→25	22.0±2.0	28.9±8.3	91.6±3.4	96.3±1.0	90.7±2.1	87.8±3.6	92.3±7.1	93.5±1.5	<b>97.6±2.6</b>
HAR 7→24	23.4±9.2	34.6±2.2	99.6±0.6	100.0±0.0	97.4±0.0	98.7±1.1	97.8±0.6	99.1±1.2	<b>100.0±0.0</b>
HAR 9→18	16.0±1.3	21.5±5.6	74.0±1.1	84.9±1.1	71.2±5.8	75.8±3.2	86.8±3.9	89.5±7.6	<b>93.2±3.9</b>
HAR 12→18	16.9±2.6	77.6±2.8	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	99.5±0.6	<b>100.0±0.0</b>
HAR Avg	18.44	36.6	88.32	92.16	88.98	83.38	93.16	89.56	<b>97.78</b>
uWave 2→6	11.3±3.0	73.5±4.5	92.6±0.4	91.4±0.4	83.0±0.7	90.5±1.1	92.3±2.2	92.3±4.4	<b>99.1±0.7</b>
uWave 4→8	14.0±2.1	92.3±0.8	100.0±0.0	100.0±0.0	98.2±1.3	100.0±0.0	100.0±0.0	100.0±0.0	<b>100.0±0.0</b>
uWave 7→8	13.4±1.3	85.7±2.9	99.4±0.4	96.1±1.5	87.2±3.0	97.9±0.4	97.6±1.1	100.0±0.0	<b>100.0±0.0</b>
uWave 3→5	14.9±3.4	79.2±3.4	99.4±0.4	94.9±0.4	81.5±4.1	91.4±1.5	91.7±2.6	98.8±1.7	<b>100.0±0.0</b>
uWave 2→5	11.6±1.9	75.6±2.1	100.0±0.0	98.8±0.4	83.9±6.0	92.6±1.5	100.0±0.0	100.0±0.0	<b>100.0±0.0</b>
uWave Avg	13.04	81.26	98.28	96.24	86.76	94.48	96.32	98.22	<b>99.82</b>
HHAR 2→7	17.5±1.8	45.2±1.3	56.0±7.5	43.0±0.6	50.1±1.9	49.1±1.9	46.7±0.4	59.4±0.1	<b>73.0±18.0</b>
HHAR 0→6	17.9±2.6	42.7±5.0	48.2±6.6	44.7±3.0	41.9±3.8	43.0±1.8	36.9±1.1	71.0±9.3	<b>81.9±3.8</b>
HHAR 1→6	20.7±4.3	62.0±1.8	89.9±1.0	85.5±1.4	85.7±0.7	88.7±1.0	82.7±4.8	90.1±1.7	<b>93.1±0.9</b>
HHAR 3→8	21.1±3.1	74.4±2.8	92.0±1.4	83.8±0.2	78.7±1.9	85.6±1.6	77.7±3.5	92.7±0.8	<b>94.2±0.7</b>
HHAR 4→5	18.9±0.3	59.1±4.6	91.5±0.7	79.6±1.3	76.5±3.4	81.5±0.9	70.5±1.4	94.2±0.6	<b>95.1±0.2</b>
HHAR Avg	19.22	56.68	75.52	67.32	66.58	69.58	62.9	81.48	<b>87.46</b>
WS AR 21→31	31.4±0.0	57.1±0.0	76.2±2.7	70.5±11.5	79.0±1.3	81.9±3.6	74.3±13.0	73.3±2.7	<b>86.7±1.3</b>
WS AR 2→25	37.2±2.8	16.7±4.7	93.3±0.0	87.8±0.8	87.8±2.1	83.9±1.6	73.3±15.3	92.8±1.6	<b>98.9±0.8</b>
WS AR 2→32	19.6±0.0	38.6±9.1	75.2±0.9	69.9±4.6	70.6±1.6	54.9±2.8	52.3±8.8	66.0±0.9	<b>79.1±8.2</b>
WS AR 4→15	82.7±1.6	27.6±12.7	82.7±5.4	86.5±0.0	82.7±1.6	78.8±0.0	77.6±5.0	86.5±2.7	<b>89.7±0.9</b>
WS AR 7→30	37.5±0.0	44.6±5.3	83.3±16.0	67.9±0.0	62.5±5.3	82.1±5.3	70.2±3.4	55.4±0.0	<b>93.5±1.7</b>
WS AR Avg	41.68	36.92	82.14	76.52	76.52	76.32	69.54	74.8	<b>89.58</b>

Table 1: classification accuracy(source→target, mean±std%) on 5 randomly chosen problems for 5 datasets.

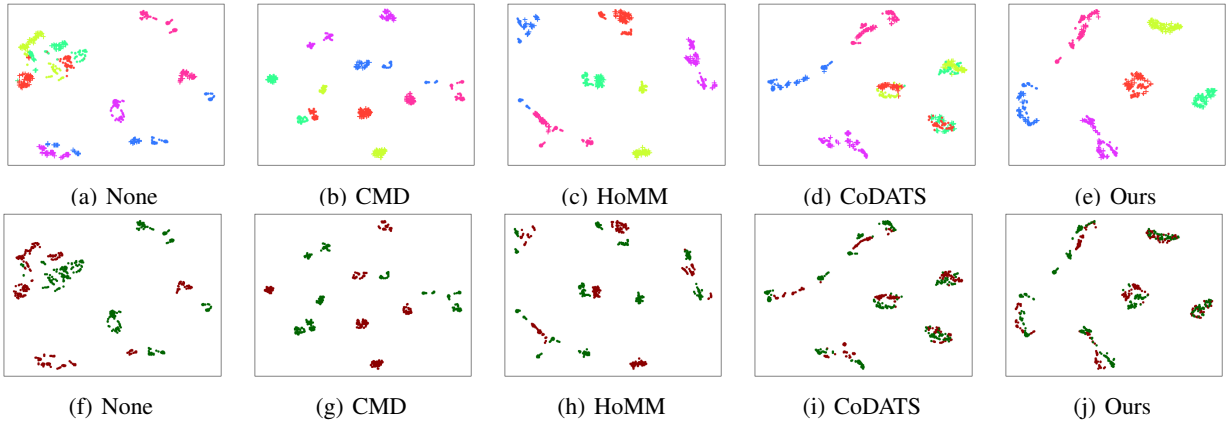


Figure 4: Feature visualization on the HAR 14→19 task. The first row illustrates the t-SNE embedding features where each color symbolizes a category. The second row illustrates the t-SNE embedding features which are marked by domain information, red and green points represent the samples drawn from the source and target domains.

performance of the no adaptation model indicates the significant difference between the source and target domains and the necessity for domain adaptation. The DDC model is much worse than our method on the HAR and WISDM AR datasets, indicating that the first-order statistic cannot characterize the samples between domains. Our method surpasses the second-order method CORAL by a large margin (20%) on the uWave and HHAR. Compared to the third and fourth order methods CMD and HoMM, our method outperforms them by 3-10% on the HAR dataset and far exceeds them on WS AR dataset. Our approach surpasses above metric-based approaches, which shows the importance of non-stationary, non-monotonic properties captured by our hybrid spectral kernel network. Besides, our method outperforms the domain-discrimination-based methods (DANN, CoDATS). Besides, from the feature visualization result in Fig. 4, we can see that these methods can confuse domains by domain-oriented discrimination. But they may eventually lead to the failure of the classification task due to the inter-class confusion during domain matching.

#### 4.5 Feature Visualization

We also used the visualization tool t-SNE to help prove the validity of the AdvSKM. We visualize the final features obtained by different methods on the HAR 14→19 task. T-SNE embedding results are in Fig. 4(a)-4(j). We can distill the following observations: The shift between source and target domains is significant. The third and fourth order metric-based methods CMD and HoMM cannot effectively align domain despite the relatively higher classification accuracy. The latest domain-discrimination-based method CoDATS neglects category aggregation while conducting relatively effective domain confusion. With our proposed method, the source and target samples are matched better and categories are discriminated better as well.

#### 4.6 Sensitivity Analysis

Two hyperparameters including  $\lambda$  and  $L$  are prepared for sensitivity analysis on four tasks. The influence of parameter variations on accuracy is shown in Fig. 5(a)-5(b). As shown

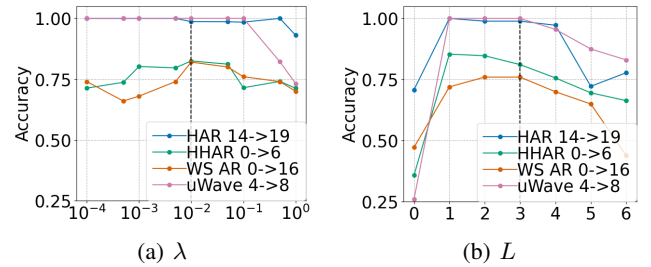


Figure 5: Sensitivity analysis on parameters (a)  $\lambda$ , (b)  $L$ . The dashed line parallel to the y-axis represents the default parameters.

in Fig. 5(a), since the accuracy rate does not fluctuate much, our model is not sensitive to the change of  $\lambda$ . Fig. 5(b) shows that our model is relatively insensitive within the appropriate range of adversarial iteration  $L$ . Specifically,  $L = 0$  in Fig. 5(b) indicates that the distance metric is calculated using the initialized kernel mapping without updating the weights in the maximization phase.

## 5 Conclusion

In this paper, a novel metric-based domain matching method AdvSKM is proposed for UTSDA task by reform MMD metric with a well-designed hybrid spectral kernel network. Specially, the hybrid spectral kernel network maps the time series samples to RKHS and preserves the non-stationary and non-monotonic statistics in the distributions, and thus bringing a more accurate discrepancy metric and ensuring the effective domain matching for UTSDA. Experimental results on distinct datasets demonstrate the superiority of our method to several state-of-the-art adaptation methods.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62076062) and National Key RD Program of China (Grant No. 2017YFB1002801). Furthermore, the work was also supported by Collaborative Innovation Center of Wireless Communications Technology.

## References

- [Anguita *et al.*, 2013] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, page 3, 2013.
- [Bai *et al.*, 2018] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
- [Chao *et al.*, 2020] Chen Chao, Fu Zhihang, Chen Zhihong, Jin Sheng, Cheng Zhaowei, Jin Xinyu, and Hua Xian-Sheng. Homm: Higher-order moment matching for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2020.
- [Cho and Saul, 2009] Youngmin Cho and Lawrence K. Saul. Kernel methods for deep learning. In *Advances in Neural Information Processing Systems 22*, pages 342–350, 2009.
- [Clements and Hendry, 2000] Michael P Clements and David F Hendry. *Forecasting non-stationary economic time series*. mit Press, 2000.
- [Dai *et al.*, 2009] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu. Translated learning: Transfer learning across different feature spaces. In *Advances in neural information processing systems*, pages 353–360, 2009.
- [Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.
- [Gretton *et al.*, 2007] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex J Smola. A kernel method for the two-sample-problem. In *Advances in neural information processing systems*, pages 513–520, 2007.
- [Kwapisz *et al.*, 2011] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.
- [Liu *et al.*, 2009] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675, 2009.
- [Long *et al.*, 2015] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.
- [Pan and Yang, 2009] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [Purushotham *et al.*, 2017] Sanjay Purushotham, Wilka Carvalho, Tanachat Nilanon, and Yan Liu. Variational recurrent adversarial deep domain adaptation. In *ICLR (Poster)*, 2017.
- [Stisen *et al.*, 2015] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 127–140, 2015.
- [Sun and Saenko, 2016] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016.
- [Tzeng *et al.*, 2014] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014.
- [Wang *et al.*, 2017] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.
- [Wilson *et al.*, 2020] Garrett Wilson, Janardhan Rao Doppa, and Diane J. Cook. Multi-source deep domain adaptation with weak supervision for time-series sensor data. In *KDD*, 2020.
- [Xue *et al.*, 2019] Hui Xue, Zheng-Fan Wu, and Wei-Xiang Sun. Deep spectral kernel learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4019–4025. AAAI Press, 2019.
- [Zellinger *et al.*, 2017] Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, and Susanne Saminger-Platz. Central moment discrepancy (cmd) for domain-invariant representation learning. *arXiv preprint arXiv:1702.08811*, 2017.