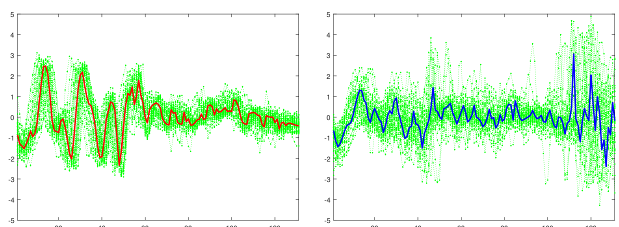
Order-preserving Wasserstein Discriminant Analysis: Supplementary Material

Bing Su[1], Jiahuan Zhou[2], Ying Wu[2]
[1]Science & Technology on Integrated Information System Laboratory,
Institute of Software, Chinese Academy of Sciences, Beijing, 100190, China
[2]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, 60208
subingats@gmail.com, {jzt011, yingwu}@eecs.northwestern.edu



(a) (b)

Figure 1. The learned barycenters of two sequence classes in the Face (all) dataset from the UCR Time Series Archive [1].

1. Deduction of Eq. (4)

$$\begin{aligned}
&\langle \boldsymbol{T}_k, \boldsymbol{D}_k\rangle - \lambda_1 I(\boldsymbol{T}_k) + \lambda_2 KL(\boldsymbol{T}_k||\boldsymbol{P}) \\
&= \sum_{i=1}^{L}\sum_{j=1}^{N_k} t_{ij}^k d_k(\boldsymbol{\mu}_i, \boldsymbol{x}_j^k) - s_{ij}^{\lambda_1} t_{ij}^k + \lambda_2 t_{ij}^k \log \frac{t_{ij}^k}{p_{ij}} \\
&= \lambda_2 \sum_{i=1}^{L}\sum_{j=1}^{N_k} t_{ij}^k (\log t_{ij}^k - \log p_{ij} - \tfrac{1}{\lambda_2}(s_{ij}^{\lambda_1} - d_{ij}^k)) \\
&= \lambda_2 \sum_{i=1}^{L}\sum_{j=1}^{N_k} t_{ij}^k (\log t_{ij}^k - \log p_{ij} e^{\frac{1}{\lambda_2}(s_{ij}^{\lambda_1} - d_{ij}^k)}) \\
&= \lambda_2 \sum_{i=1}^{L}\sum_{j=1}^{N_k} t_{ij}^k \log \frac{t_{ij}^k}{\boldsymbol{K}_k(i,j)} = \lambda_2 KL(\boldsymbol{T}_k||\boldsymbol{K}_k)
\end{aligned}, \tag{1}$$

2. Illustration of the alignment between barycenters

Fig. 1 illustrates the learned barycenters for two sequence classes from the UCR Time Series Archive [1]. Note that the sequences are univariate sequences for illustration. In this paper, we tackle multivariate sequences. We can observe that each barycenter reflects the average evolution and acts as the mean sequence of all training sequences.

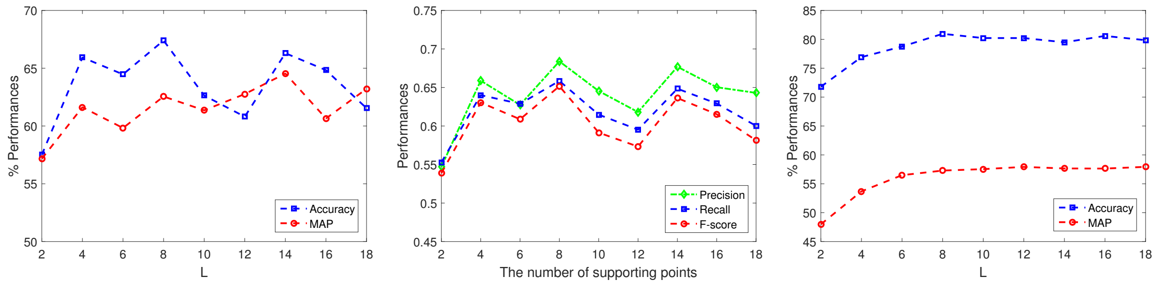
The distance between two classes is defined as the OPW distance between the corresponding barycenters of the two classes. The inter-sequence-class scatter is based on the OPW transport between the two barycenters. In Fig. 2, we illustrate the transport or soft alignment between the two barycenters in Fig. 1.

We can observe that more pairwise differences between elements in all barycenters contribute to the overall inter-class scatter according to different weights. The weight $t_{ij}^{cc'*}$ of a pair $(\boldsymbol{\mu}_i^c, \boldsymbol{\mu}_j^{c'})$ encodes the local relations of the two elements and indicates their joint probability. $\boldsymbol{\Gamma}_b$ concentrates more on the differences between the pairs with large joint probabilities. Such differences reflect the essential distinctions of two classes, because the matched pairs represent the homologous temporal structures and thus are distinctive for discriminating the two classes. Different from the alignments by DTW, where the weights are 1 for a small portion of aligned pairs and 0 for other pairs, the weights by OPW are soft probability values and hence $\boldsymbol{\Gamma}_b$ also incorporates the differences between the pairs with smaller weights. This compromises more information and is more robust to incorrect or ambiguous alignments caused by noises.

We can also observe that large transports appear near the diagonal. This shows that the inter-class scatter focuses more on the temporal differences between the two barycenters.

3. Influence of hyper-parameters

We set the values of the hyper-parameters λ_1, λ_2, and δ of OPW as suggested in [3] on the three datasets. On the MSR Action3D dataset, these hyper-parameters are suggested for the nearest mean (NM) classifier, but the optimal hyper-parameters for the NM classifier and the NN classifier are different. Also, on this dataset, the pairwise distance matrix $\mathbf{D}$ is normalized in [3]. We do not perform this preprocessing. These lead to worse results with the NN classifier than those obtained using the optimal hyper-parameters for this classifier. In [3], λ_2 was fixed to 0.1 for all datasets, and OPW is not sensitive to λ_1. Since λ_1, λ_2, and δ influ-



(a) (b) (c)

Figure 3. Performances as functions of L by (a) (b) the SVM classifier and (c) the NN classifier on the MSR Action3D dataset.

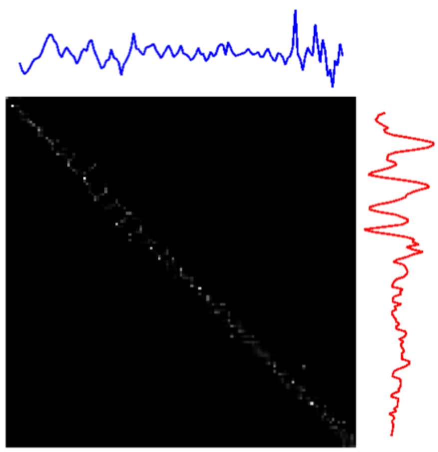


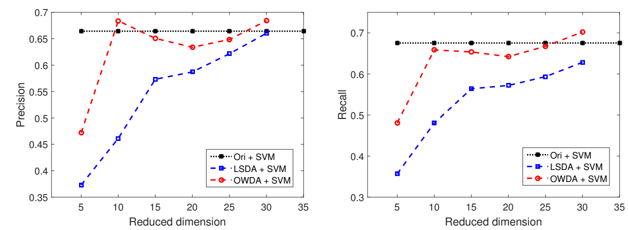
Figure 2. The OPW alignment between two barycenters.

ence our method through OPW distance, our method should share similar sensitivities to them. We simply fix L to 8 in our experiments. In fact, Fig. 3 shows the influence of L on the Action3D dataset. In most cases, our method is not very sensitive to L.

For the competitive LSDA, we directly compare with the reported results in [2] on the ChaLearn dataset. On other datasets, we fix the hyper-parameter of LSDA, the number of states for HMM per class, to the suggested value in [2]. Comparisons of LSDA and the proposed OWDA with different hyper-parameters are shown in Fig. 4. We can observe that the proposed OWDA consistently outperforms LSDA with different hyper-parameters using different classifiers and performance measures.

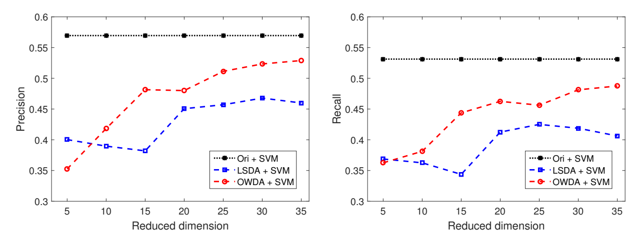
4. Comparison with LSDA on multi-class precision and recall

LSDA is the major competitor of the proposed OWDA. We further evaluate the multi-class precision and recall by the two DRS methods with the SVM classifier on the MSR Action3D dataset. The results are shown in Fig. 5. We can observe that OWDA outperforms LSDA consistently for all subspace dimensions. Again, OWDA improves the precision and recall of classifying the original sequences with the SVM classifier when reducing the dimension to 30.



(a) (b)

Figure 5. (a) Precisions and (b) Recalls with the SVM classifier as functions of the dimensionality of the subspace on the MSR Action3D dataset.



(a) (b)

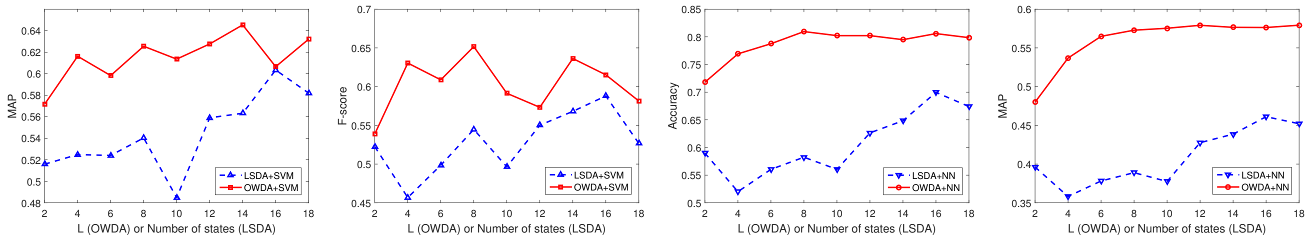
Figure 6. (a) Precisions and (b) Recalls with the SVM classifier as functions of the dimensionality of the subspace on the MSR Daily Activity3D dataset.

Fig. 6 compares the multi-class precision and recall by OWDA and LSDA with the SVM classifier on the MSR Activity3D dataset. We can observe similar trends.

We again evaluate the multi-class precision and recall by OWDA and the major competitor, LSDA, with the SVM classifier on the ChaLearn dataset. As shown in Fig. 7, OWDA outperforms LSDA significantly on all subspace dimensions, and also outperforms the original sequences when more than 25 dimensions are preserved.

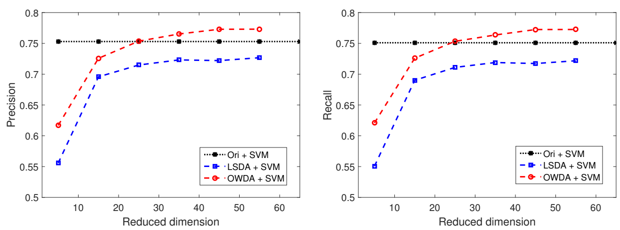
References

[1] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, and Gustavo



(a) (b) (c) (d)

Figure 4. Performances of OWDA as functions of L and performances of LSDA as functions of the number of states by (a) (b) the SVM classifier and (c) (d) the NN classifier on the MSR Action3D dataset.



(a) (b)

Figure 7. (a) Precisions and (b) Recalls with the SVM classifier as functions of the dimensionality of the subspace on the ChaLearn Gesture dataset.

Batista. The ucr time series classification archive, July 2015. `www.cs.ucr.edu/~eamonn/time_series_data/`.

[2] Bing Su, Xiaoqing Ding, Hao Wang, and Ying Wu. Discriminative dimensionality reduction for multi-dimensional sequences. *TPAMI*, 2018.

[3] Bing Su and Gang Hua. Order-preserving optimal transport for distances between sequences. *TPAMI*, 2018.