

Weakly Supervised Metric Learning towards Signer Adaptation for Sign Language Recognition

Fang Yin^{1,2}

fang.yin@vipl.ict.ac.cn

Xiujuan Chai^{1,2}

xiujuan.chai@vipl.ict.ac.cn

Yu Zhou³

zhouyu@iie.ac.cn

Xilin Chen^{1,2}

xlchen@ict.ac.cn

¹ Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS
Beijing, 100190, China

² Cooperative Medianet Innovation Center, China

³ Institute of Information Engineering, CAS
Beijing, 100093, China

In this paper, we introduce metric learning into Sign Language Recognition (SLR) for the first time and propose a signer adaption framework to address signer-independent SLR. For adapting the general model to the new signer, both clustering and manifold constraints are considered in the adaptive distance metric optimization.

The contribution of our work mainly lies in three-folds. Firstly, a Weakly Supervised Metric Learning (WSML) framework is proposed, which combines the clustering and manifold constraints simultaneously. Secondly, the general framework is applied to signer adaptation and achieves good performance. Thirdly, a fragment based feature is designed for sign language representation and the effectiveness is verified in large vocabulary datasets.

Our proposed WSML framework can be decomposed into two key steps. The first one is to learn a generic metric from the given labeled data. Then the second step is to realize the distance metric adaptation by considering the clustering and manifold constraints with the unlabeled data.

To learn a generic distance metric, the labeled data are used under clustering assumption with classical large margin hinge loss. Specifically, the distances between data points within the same cluster (with same label) should be minimized and the distances between data points from different clusters (with different labels) should be maximized. Here we define the index set with same labels as $S_g = \{(i, j) | y_i = y_j, x_i, x_j \in X_l\}$ and the index triplet $B_g = \{(i, j, k) | y_i = y_j, y_i \neq y_k, x_i, x_j, x_k \in X_l\}$. The objective function is

$$\min_{M_g} J_g = \min_{M_g} \sum_{(i,j) \in S_g} D_g(x_i, x_j) + \sum_{(i,j,k) \in B_g} [1 + D_g(x_i, x_k) - D_g(x_i, x_j)]_+ \quad (1)$$

s.t. $M_g \geq 0$,

where the term $[z]_+ = \max(z, 0)$ denotes the standard hinge loss and M_g is the metric matrix for distance metric D_g .

After learning the generic distance metric, clustering constrain on both the labeled and unlabeled data can be considered. First of all, with the generic distance metric D_g determined by matrix M_g and the labeled data set X_l , a classifier is trained. Thus the labels of the unlabeled data can be predicted. We hope that the generic distance metric can be adaptive to the data from a different distribution through the predicted labels, although they are uncertain. For each unlabeled data x_i , the classifier gives a predicted label y_i and belief b_i , where $0 \leq b_i \leq 1$. Here the label y_i is not precise so the metric learning method with these uncertain labels is weakly supervised. The objective function of clustering constraint is

$$J_c = \sum_{(i,j) \in S_c} b_i D(x_i, x_j) + \sum_{(i,j,k) \in B_c} b_i [1 + D(x_i, x_k) - D(x_i, x_j)]_+ \quad (2)$$

where S_c is a index set defined as $S_c = \{(i, j) | y_i = y_j, x_i \in X_u, x_j \in X_l\}$. While B_c is a triplet set $B_c = \{(i, j, k) | y_i = y_j, y_i \neq y_k, x_i \in X_u, x_j, x_k \in X_l\}$.

To further investigate the topological structure of the unlabeled data, manifold assumption is considered. Specifically, if two data points x_i and x_j are close in the intrinsic geometry of Euclidean distance, they should also be close to each other in the Mahalanobis distance. A straightforward expression for manifold assumption is

$$J_m = \sum_{i=l+1}^u \sum_{j=l+1}^u D(x_i, x_j) W_{ij}, \quad (3)$$

where W is a weight matrix. $W_{ij} = 1$ if x_i is among the k -nearest neighbors of x_j or x_j is among the k -nearest neighbors of x_i . Otherwise, $W_{ij} = 0$.

The final objective function of metric adaptation can be formulated by incorporating expressions (2) and (3) into (1).

$$\min J = \min(J_g + \alpha J_c + \beta J_m), \quad (4)$$

where α and β are the weights corresponding to clustering constraint and manifold constraint respectively.

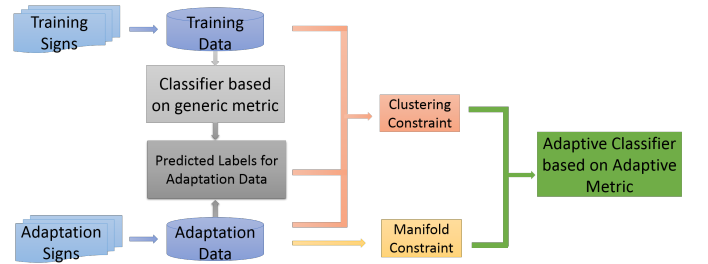


Figure 1: The pipeline of the signer adaptation.

Another main contribution of our work is applying WSML to signer-independent SLR. Figure 1 illustrates the pipeline of the signer adaptation using WSML. Concretely speaking, the features are extracted from the labeled signs and the unlabeled adaptation signs. With the features of labeled training data, a generic distance metric can be learnt, with which a classifier is trained. Then the labels of unlabeled data are predicted with believes. In the adaptation stage, both the labeled and unlabeled training data are used for adaptive metric learning. Specifically, the labeled data and the unlabeled data with their predicted labels are used together for clustering constraint, while the unlabeled data are used for manifold constraint. The classifier based on the adaptive metric is suitable for the new signer and is used for recognition.

WSML is compared with HMM[1], DTW, ARMA[2] and generic metric learning (ML) on our own dataset in the experiments. The accuracies are shown in Figure 2(a). Figure 2(b) illustrates the accuracy trend with the increasing of adaptation data size.

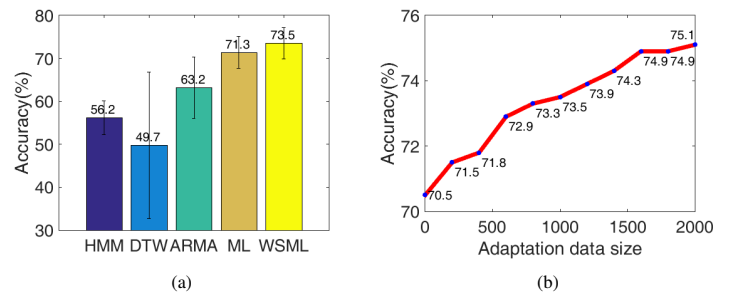


Figure 2: The accuracies of our proposed method and the baseline methods are given in (a). (b) shows the accuracies of our proposed method with different adaptation data sizes.

- [1] Chunli Wang, Wen Gao, and Shiguang Shan. An approach based on phonemes to large vocabulary chinese sign language recognition. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 411–416. IEEE, 2002.
- [2] Chunyan Xu, Tianjiang Wang, Junbin Gao, Shougang Cao, Wenbing Tao, and Fang Liu. An ordered-patch-based image classification approach on the image grassmannian manifold. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(4):728–737, 2014.