

Landmarks-based Kernelized Subspace Alignment for Unsupervised Domain Adaptation

Qilei Zhang

Jul 16 2018

Abstract

Domain adaptation has achieved great success in computer vision in recent years, in order to process the learning process to transfer knowledge from the source to the target domain. In this article, we introduce a new unsupervised DA method based on subspace alignment and selection of landmarks, which are also distributed between two domains. These landmarks are selected to reduce the differences between domains, and then used to projecting data in the same space, in which the effective subspace alignment is performed in a closed form.

1. Introduction

While the standard machine learning setting assumes that training and test data come from the same statistical distribution, it turns out that many real world applications, in Computer Vision or in Natural Language Processing, challenge this assumption [4]. Indeed, the data from which the classifier is supposed to be learned often differ from the target data the classifier will be deployed on Figure 1. To deal with such situations, the learning system has to take into account the distribution shift between the two domains in order to adapt well from the source to the target. In this framework, two main categories of domain adaptation methods are available [3].

2. Related Work

Domain adaptation is one of the possible settings for transfer learning. In the article, we focus on the most complex case, called unsupervised DA, where the label data can only be from the source domain. In this case, in order to reduce the offset between the source and the target distribution, a strategy includes weighting each source data based on the similarity between the source data and the target domain [2].



Figure 1. Example of distribution shift between images from two datasets. First row, some bike helmets from the Amazon subset and second row, bike helmets from the webcam subset. These 2 subsets are from the Office dataset.

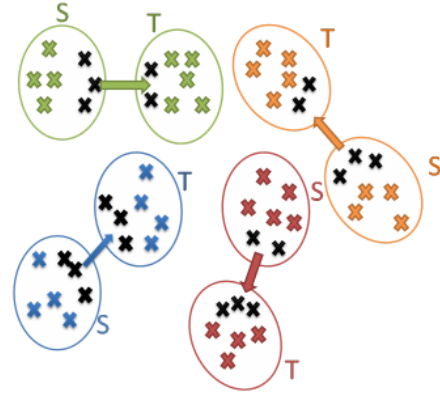


Figure 2. (Best shown in color) Domain adaptation problem that needs to adapt a source (S) domain to a target (T) domain w.r.t. four classes of examples (green, orange, blue and red). We can see that no linear transformation will be able to entirely bridge the source to the target. On the other hand, we can note that, for each class, some subsets of source and target examples seem to be similarly distributed (in black). The objective is to automatically discover those so-called landmarks that will be used to map both source and target data onto a common space using a kernel.

3. Conclusion

The DA method based on subspace alignment has attracted much interest recently. In this framework, the most accurate method assumes that the offset between source and target distributions can be corrected by linear functions [5]. However, experiments show that this assumption is challenged in most real-world applications. Therefore, nonlinear mapping functions should be optimized for homogeneous sources and target subspaces. Moreover, there is no reason to prove that the constrained DA algorithm adapts all training source data to the target domain [1].

References

- [1] R. Fabbri and B. B. Kimia. Multiview differential geometry of curves. *International Journal of Computer Vision*, 120(3):324–346, 2016. 2
- [2] S. Khuller, B. Raghavachari, and A. Rosenfeld. Landmarks in graphs. *Discrete Applied Mathematics*, 70(3):217–229, 1996. 1
- [3] H. Sun, S. Liu, and S. Zhou. Discriminative subspace alignment for unsupervised visual domain adaptation. *Neural Processing Letters*, 44(3):1–15, 2016. 1
- [4] S. Wu, X. Jing, Z. Wei, J. Yang, and J. Yang. Learning image manifold via local tensor subspace alignment. *Neurocomputing*, 139(139):22–33, 2014. 1
- [5] J. Yu, F. Lin, H. S. Seah, C. Li, and Z. Lin. Image classification by multimodal subspace learning. *Pattern Recognition Letters*, 33(9):1196–1204, 2012. 2