

Interaction representation-based subspace learning for domain adaptation

Jijia Chen¹, Ao Li², Deyun Chen³, Kezheng Lin⁴

1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China
2. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China

Email: {544953065@qq.com(Jijia Chen), dargonboy@126.com (Ao Li)}

Abstract Generally, Transfer Learning (TL) or Domain Adaptation (DA) are used to solve domain inconsistency, but the conventional domain adaptation methods mostly only consider local information and ignore global information, and only consider one-way data and ignore the possibility of two-way data. Therefore, in this paper, we proposed an interactive representation-based framework for domain adaptation. In the novel framework, two low-rank based interactive representation models are built on both of source and target domains, which can be used to better align distribution discrepancy. Then, a distance constraint is designed to model the subspace relationship between source and target domain. Finally, the label-based regression is jointly used to earn extra discrimination for classification. Experiments on a number of public databases demonstrate that our method has competitive performance among comparison methods.

Keywords: Subspace learning, Domain adaptation, Interaction representation

1 Introduction

In most machine learning methods, training set and test set are generally considered to have similar distribution. However, in the real world, the training and test set are affected by light, Angle, background, Fig.1 show a few of pictures of different distributions, Therefore, the question of domain inconsistencies arises, that is, training sets and test sets cannot have similar distributions. However, domain inconsistency will lead to the reduction of classification accuracy when classifying images^[2]. Usually, the solution to the problem of domain inconsistency is Transfer Learning or Domain Adaptation^[3].

Supervised^[4], semi supervised^[5] and unsupervised^[6] are three DA strategies. Domain adaptation can be used to solve the question of data coming from two different domains but related to each other. The method of domain adaptation is to explore the invariant structure in the two domains, through which a subspace can be found to connect the source and target domain^[12]. In this paper, we think that the domain adaptive method only considers the unidirectional alignment of data sets and ignores the possibility of bidirectional alignment. Therefore, we advance a subspace learning method based on interactive representation. The main contents of this method are as follows:

(1) Inspired by the alignment of two domains, we propose an interactive representation model based on two low rank. In this model, we make the two domains align with each other. In addition to making the target domain use the linear representation of the source domain, we can

also make the source domain use the linear representation of the target domain.

(2) A distance constraint is added to the model, which makes the two domains approach each other infinitely, and further achieves the effect of mutual representation.

(3) In order to obtain extra discrimination information, we explore a great of label information in source domain to strengthen the robustness of the model.



Fig. 1. Examples of different distributions in dataset CMU

2 Related Works

We will introduce two topics related to research work, namely low rank and subspace learning.

2.1 Subspace Learning

In the field of pattern recognition and image classification, Subspace learning has been used. It is expected to find a subspace in which all required data are kept. Subspace level method can align the features of both domains. For example, PCA will project data along the direction of the maximum variance. PCA pursues to maximize the internal information of data after dimensionality reduction. It does not consider classification information. LDA seeks the most effective recognition direction by minimizing the ratio between in-class and inter-class scatterers. What LDA pursues is that data points after dimension reduction can be distinguished as easily as possible. LDA pursues that the data points after dimensionality reduction can be distinguished as easily as possible, that is, the data has better separability in the low dimensional space, so that the data samples after dimensionality reduction have not only the largest distance between classes, but also the smallest intra class variance in this new dimensional space. SA^[1] mainly searches for a linear mapping by aligning the subspaces covered by vector features, which can find a domain invariant feature space. The ULRA method proposes a special logarithmic determinant function to approximate the rank function. It reduces the contribution of large singular value to kernel norm while keeping the contribution of small singular value close to zero^[15].

2.2 Low-Rank

During the old days, low-rank learning has become a study hotspot and has been widely used in many applications^[7], such as image clustering, discriminant subspace learning and outlier test. When data is affected by noise, Wright^[8] proposed a robust principal component analysis method for recovering the low-rank property of data structure. RPCA is a typical low order method of study. The principle of the observation matrix can be broken down into low rank matrix and the sum of sparse matrix. The purpose of RPCA is to recover from low rank matrix to the observation matrix. It can correspond to a rank minimum optimization problem, in the shadow removal and background modeling has a good effect in actual application. When images, high-dimensional data and some unbalanced data are encountered, Peng proposes a classification method based on Discriminant ridge regression (DRM). DRM does not need to find binary similarity, but estimates

the representation of a new example as a soft similarity vector by minimizing the fitting error in the regression framework, so that the model has a continuous neural network supervision label^[14].

Low rank said it would RPCA from a single subspace extends to more subspace, but most of the low rank representation may encounter the problem of insufficient samples. if only to make a study of domain data information for the final model will be no good robustness. Therefore, the potential of low rank representation method was used to solve this problem are put forward^{[9][10]}.

Recently, a few studies have proposed the introduction of low rank constraints in transfer learning.^[11] The learning method of low-order transfer subspace is to constrain the Shared subspace of two domains, while the adaptive method of low-rank domains is to use low-rank representations to reduce the difference in domain distribution^[13]. Our method is different from the method above, which only adopts one-way domain alignment in subspace learning, while our proposed IRSL method is the interactive representation of two domains, not only the training set alignment test set, but also the test set alignment training set.

3 IRSL:Interaction representation-based subspace learning

3.1 Mathematical Notation

We define the characters required below. The identifier of the source domain is $S = \{X_s, y_s\}$, The identifier of the source domain is $T = \{X_t, y_t\}$, where $X_s \in R^{D \times n_s}$ and $X_t \in R^{D \times n_t}$ are samples, y_s and y_t are labels. In the source domain, n_s represents the amount of samples and $P_s \in R^{D \times d}$ represents the projection. In the target domain, where $P_t \in R^{D \times d}$ represents the projection, n_t represent the amount of samples. D represents the dimension of the primitive samples. d is the dimension of the subspace. $Z \in R^{n_s \times n_t}$ represents the reconstruction matrix.

3.2 Formulation

In this section, the IRSP method can solve the domain inconsistency problem by learning a subspace. In IRSP, we introduce a subspace learning model based on interactive representation. First of all, we can learn P_t from the invariable information in the data. Then, we expect to find an invariant subspace. In this subspace, the target domain and the source domain are infinitely close. Therefore, the target domain and the source domain can be expressed linearly with each other. In this way, the two domains achieve the interactive effect, and the two domains achieve a better effect in alignment with each other. Mathematically, A low-order model can be formed, that is, a low rank constraint is used to realize the reconstruction matrix, and its expression is as follows:

$$\min_{P_s, P_t, Z} \left\| P_t^T X_t - P_s^T X_s Z_1 \right\|_2^2 + \lambda \|Z_1\|_* + \left\| P_s^T X_s - P_t^T X_t Z_2 \right\|_2^2 + \eta \|Z_2\|_* \quad (1)$$

Then, we hope to settle the question of domain inconsistency by aligning the two domains.

Therefore, we set the distance constraint on the subspace of the two domains so that the two fields are close. This distance constraint can be achieved by minimizing L2 norm:

$$\min_{P_t, P_s} \|P_s - P_t\|_2^2 \quad (2)$$

This formula can retain the useful information of two datasets to a great extent, meanwhile, adjust the subspaces of two domains. Therefore, a better P_t can be obtained.

For the final classification model, it is not enough to use only subspace to distinguish, and extensive label information in the source domain is ignored. Therefore, we suggest labels in the two domains into the classification model to obtain additional discrimination information and improve the resolution of the model. In order to make the best of the known label information in the source domain, we define the constructed label matrix Y as:

$$Y_{\{i, j\}} = \begin{cases} 1, & \text{if } x_j \in c_i \\ -1, & \text{otherwise} \end{cases} \quad (3)$$

The purpose of introducing labels into the model is to find a discriminative P_t , which is similar to a subspace between domains. The formula can be expressed as follows:

$$\min_{P_t} \|P_t^T X - Y\|_2^2 \quad (4)$$

Where $X = [X_s, X_t]$.

We can get the final formula by Eqs. 1, 2 and 4, as follows:

$$\min_{P_t, P_s, Z_1, Z_2} \|P_t^T X_t - P_s^T X_s Z_1\|_2^2 + \lambda \|Z_1\|_* + \|P_s^T X_s - P_t^T X_t Z_2\|_2^2 + \eta \|Z_2\|_* + \partial \|P_s - P_t\|_2^2 + \frac{1}{2} \|P_t X - Y\|_2^2 \quad (5)$$

Where ∂ , η and λ are trade-off parameters to balance the constraints.

3.3 Optimization

It can be seen from formula (5) that when Y is fixed, four variables will be involved. Therefore, solving formula (5) by the method of inexact augmented Lagrange multiplier method (IALM), in which two auxiliary variables L1 and L2 are involved. Formula (5) is converted into:

$$\min_{P_t, P_s, Z_1, Z_2} \|P_t^T X_t - P_s^T X_s Z_1\|_2^2 + \lambda \|L_1\|_* + \|P_s^T X_s - P_t^T X_t Z_2\|_2^2 + \eta \|L_2\|_* + \partial \|P_s - P_t\|_2^2 + \frac{1}{2} \|P_t X - Y\|_2^2 \quad (6)$$

s.t. $Z_1 = L_1, Z_2 = L_2$

Using the variable alternation strategy, we can get the following results by derivation:

$$P_t = (2\partial + 2X_t X_t^T - 2X_t Z_2 Z_2^T X_t^T + X X^T)^{-1} (2\partial P_s + 2X_t Z_1^T P_s X_s^T - 2X_t Z_2 X_s^T P_s + X Y) \quad (7)$$

$$P_s = (2\partial + 2X_s Z_1 Z_1^T X_s^T + 2X_s X_s^T)^{-1} (2\partial P_t - 2X_s Z_1 X_t^T P_t + 2X_s Z_2^T X_t^T P_t) \quad (8)$$

$$Z_1 = (2P_s X_s^T P_s^T X_s + \mu)^{-1} \left(2P_t X_s^T P_t^T X_t + \mu \left(L_1 - \frac{Y_1}{\mu} \right) \right) \quad (9)$$

$$Z_2 = \left(2P_t X_t^T P_t^T X_t + \mu\right)^{-1} \left(2P_t X_t^T P_s^T X_s + \mu \left(L_2 - \frac{Y_2}{\mu}\right)\right) \quad (10)$$

$$L_1 = \underset{L_1}{\operatorname{argmin}} \left\|L_1\right\|_* + \frac{\mu}{2} \left\|Z_1 - L_1 + \frac{Y_1}{\mu}\right\|_2^2 \quad (11)$$

$$L_2 = \underset{L_2}{\operatorname{argmin}} \left\|L_2\right\|_* + \frac{\mu}{2} \left\|Z_2 - L_2 + \frac{Y_2}{\mu}\right\|_2^2 \quad (12)$$

where Y_1 and Y_2 are Langrange multiplier, $\mu > 0$ is a penalty parameter. P_t can be obtained by IALM algorithm.

4 Experiment

A. Data set

4DA data set: There are four domains in 4da: Amazon (A), Webcam (W), DSLR (D) and Caletch(C). There are 10 shared categories in these four domains. We experiment with these four domains alternately as source and target domain, that is to say, we have carried out cross domain experiments of 12 tasks.

CMU PIE Face data set: PIE includes 41368 pictures taken by 68 people in 13 different postures and 21 different kinds of light. The size of these 41368 pictures is 32x32. We divide these pictures into 5 different postures, P1 is left postures, P2 is up postures, P3 is down postures, P4 is positive postures, P5 is right postures.

B. Experimental setting

The dimension d of the learning subspace is set to the category number c in each data set. According to a large number of experiments, the optimal accuracy can be obtained by setting the trade-off parameter α to 25, as shown in Figure 2.

C. Comparisons With Other Approaches

In this section, we compare with SA, JDA and TSL in the above two datasets. Table 1 and Table 2 shows the results of the above method comparison. From the table, we can see that the classification performance of the proposed IRSP method is the best.

Table 1. Accuracy(%) of 4DA data set

Data Set	SA ^[1]	JDA	TSL ^[12]	IRSP
C→A(1)	48.02	51.46	52.30	57.10
C→W(2)	31.86	41.36	40.34	57.97
C→D(3)	42.68	46.50	49.04	50.32
A→C(4)	34.37	43.90	43.28	44.70
A→W(5)	33.90	33.90	34.58	47.80
A→D(6)	38.85	33.76	38.85	47.13
W→C(7)	30.01	31.17	31.43	38.47
W→A(8)	32.15	36.33	34.66	41.34
W→D(9)	83.44	77.71	79.62	83.35

D→C(10)	32.24	31.43	33.13	33.57
D→A(11)	33.40	38.41	32.57	39.46
D→W(12)	70.51	75.59	72.54	84.07
Average	42.61	45.13	45.19	52.11

Table 2. Accuracy(%) of CMU data set

Data Set	SA ^[1]	JDA	TSL ^[12]	IRSP
P1→P4	42.75	25.14	46.68	87.01
P4→P4	51.41	33.76	59.15	86.94
P4→P5	47.92	29.47	45.22	73.24
P5→P4	43.11	25.38	53.08	65.98
Average	46.29	28.43	51.03	78.29

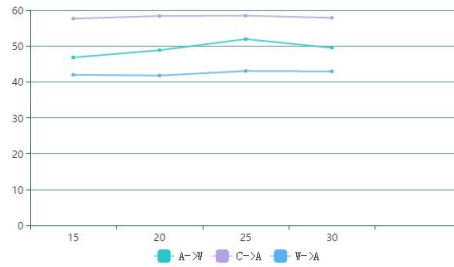


Fig. 2. Parament sensitivity analysis

5 Conclusion

In this paper, We propose a subspace learning based on interactive representation framework to solve the question of domain inconsistency. In this framework, We align the two fields and make the two fields as close as possible. In order to achieve this better, We further constrain the distance of the model to make the two domains closer. Finally, In the source domain, we take full advantage of the known label information to obtain additional classification information, so as to increase the robustness of the model. A great of experiments show that our way is superior to the existing method.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grant 61501147, in part by the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant UNPYSCT-2018203, in part by the Natural Science Foundation of Heilongjiang Province under Grant YQ2019F011, in part by the Fundamental Research Foundation for University of Heilongjiang Province under Grant LGYC2018JQ013, and in part by the Postdoctoral Foundation of Heilongjiang Province under Grant LBH-Q19112.

References

- [1] Fernando, B., Habrard, A, Sebban, M, Tuytelaars, T: Unsupervised visual domain adaptation using subspace alignment. In: IEEE ICCV, pp. 2960–2967 ,2014.
- [2] Kan, M, Junting, W, Shan, S, Chen, X: Domain adaptation for face recognition: targetize source domain bridged by common subspace. IJCV **109**(1–2), 94–109 ,2014.

- [3] Gong, B., Shi, Y., Sha, F, Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: IEEE CVPR, pp. 2066–2073,2012.
- [4] L. Duan, D. Xu, and I. Tsang, “Learning with augmented features for heterogeneous domain adaptation,” *arXiv*, 2012.
- [5] T. Yao, Y. Pan, C.-W. Ngo, H. Li, and T. Mei, “Semi-supervised domain adaptation with subspace learning for visual recognition,” in ICCV, pp. 2142–2150,2015.
- [6] M. Long, H. Zhu, J. Wang, and M. I. Jordan, “Unsupervised domain adaptation with residual transfer networks,” in *NIPS*, pp. 136–144,2016.
- [7] Z. Jiang, P. Guo, and L. Peng. Locality-constrained low-rank coding for image classification. In *AAAI*, pages 2780–2786, 2014.
- [8] E. J. Candes, X. D. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11, 2011.
- [9] Y. Y. Liu, L. C. Jiao, and F. H. Shang. An efficient matrix factorization based low-rank representation for subspace clustering. *Pattern Recognition*,46(1):284–292, 2013.
- [10] Z. Xu, W. Li, L. Niu, and D. Xu. Exploiting low-rank structure from latent domains for domain generalization. In *ECCV*, pages 628–643,2014.
- [11] I. H. Jhuo, D. Liu, D. T. Lee, and S. F. Chang. Robust visual domain adaptation with low-rank reconstruction. In *CVPR*, pages 2168–2175,2012.
- [12] F.Jingru,Z,Lei,Z.Bob,J,Wei.Guided Learning:A New Paradigm for Multi-task Classification.In *CCBR*,pp. 239-246,2018
- [13] L.Sheng,F.Yun.Unsupervised transfer learning via low-rank coding for image clustering.In *IEEE*,pp. 1795-1802,2016
- [14] P.Chang, C.Qiang. Discriminative Ridge Machine: A Classifier for High-Dimensional Data or Imbalanced Data[J].*arXiv*, 2019.
- [15] P.Chong,K.Zhao,L.Huiqing,and C.Qiang.Subspace Clustering Using Log-determinant Rank Approximation,pp.925-934,2015