IEICE TRANS. INF. & SYST., VOL.E102–D, NO.12 DECEMBER 2019

LETTER

Target-Adapted Subspace Learning for Cross-Corpus Speech Emotion Recognition

Xiuzhen CHEN[†], **Xiaoyan ZHOU**[†a)], **Cheng LU**[††], **Yuan ZONG**[††], **Wenming ZHENG**[††], ***Nonmembers***, ***and*** **Chuangao TANG**[††], ***Member***

SUMMARY For cross-corpus speech emotion recognition (SER), how to obtain effective feature representation for the discrepancy elimination of feature distributions between source and target domains is a crucial issue. In this paper, we propose a Target-adapted Subspace Learning (TaSL) method for cross-corpus SER. The TaSL method trys to find a projection subspace, where the feature regress the label more accurately and the gap of feature distributions in target and source domains is bridged effectively. Then, in order to obtain more optimal projection matrix, ℓ_1 norm and $\ell_{2,1}$ norm penalty terms are added to different regularization terms, respectively. Finally, we conduct extensive experiments on three public corpuses, EmoDB, eNTERFACE and AFEW 4.0. The experimental results show that our proposed method can achieve better performance compared with the state-of-the-art methods in the cross-corpus SER tasks.

key words: *cross-corpus speech emotion recognition, transfer learning, domain adaptation, subspace learning*

1. Introduction

Emotion recognition is a hot topic and widely used in the fileds of human-machine interaction and signal processing, including multiple modalities, such as speech [1], facial expression [2], physiological signals [3], [4] and so on. SER utilizes the labeled samples to train a robust model for the label prediction of unlabeled samples, all of the samples come from the same database and their corresponding conditional distributions and marginal distributions are similar in general, which can boost the good performance. However, in real scenario, the labeled training data and unlabeled testing data are always from different datasets, called cross-corpus SER, thus, the general methods based on the same dataset do not work well.

Cross-corpus SER aims to eliminate the discrepancy of feature distributions between source and target domains. For source domain (labeled training data) and target domain (unlabeled testing data) come from different datasets, they have different conditional distributions and marginal distributions such that the traditional methods of SER are no longer applicable to the cross-corpus SER. To address these cross-corpus issues, the popular approach is to obtain an invariant feature space employing two domains to narrow the discrepancy. Specially, various normalization schemes [5] are firstly utilized to conduct cross-corpus SER on different databases. With the rise of transfer learning, domain adaptation methods are widely used on cross-corpus issues [6]. In [7], an importance-weighted support vector machine (IW-SVM) by incorporating three domation adaptation is proposed for the evaluation of cross-corpus SER. Meanwhile, Deng et al. [8] combined auto-encoder with domain adaptation to propose an unsupervised framework in order to obtain a common representation of source domain and target domain.

Despite these domain adaptation methods are widely used, Least Squares Regression (LSR) method is also a popular subspace learning method to bridge the feature space and the label space, which is benefitting to the label prediction of target data. Thus, to improve the performance of the model, the basic idea of transfer learning is borrowed into LSR model. In [9], an Incomplete Sparse Least Squares Regression (ISLSR) model was proposed by Zheng et al. to alleviate the discrepancy in the data distribution between the training corpus and testing corpus, and the result shows that this model has an effective performance. Benefitting to this method, large domain-adaptation researches based on LSR emerge continually. Zong et al. [10] utilized a Domain-adaptive Least-Squares Regression (DaLSR) model to handle the mismatch problems between source and target speech corpuses, which used an additional unlabeled dataset from target speech corpus as an auxiliary dataset and combined with the labeled training dataset from source speech corpus for jointly training the DaLSR model. Although these methods can achieve good results, how to reduce the discrepancy of feature distributions between the source and target domains in the research of cross-corpus SER is still a challenge.

Recently, Liu et al. [11] proposed a Domain-adaptive Subspace Learning (DoSL) method for learning a projection matrix to transform the source and target speech samples from the original feature space to the label space so that every target sample can be represented by source samples. In detail, it not only bridges the feature space and label space by LSR, but also selects effective features contributing to regress the label of source domain by $\ell_{2,1}$ norm, which are benefitting to describe the relationship between the source feature space and label space. Furthermore, it also contains

Manuscript received February 18, 2019.
Manuscript revised July 10, 2019.
Manuscript publicized August 26, 2019.

[†]The authors are with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.
[††]The authors are with the Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China.
a) E-mail: xiaoyan_zhou@nuist.edu.cn

DOI: 10.1587/transinf.2019EDL8038

a LSR term between source and target domains to eliminate the data distributions discrepancy between source and target domains. Unfortunately, this method can not reconstruct the target samples well when the speech samples are transformed from the original feature space to the label space. In order to overcome the shortcoming of this method, we will extend the DoSL method and then propose the Target-adapted Subspace Learning (TaSL) method for cross-corpus SER in this paper, in which another projection matrix and ℓ_1 norm are added to get sparse features of source domain to represent target domain features better.

2. Proposed Method

2.1 TaSL Model

In this section, some notations are given for better illustration of TaSL model in this paper. Suppose that we have two different speech emotion corpuses, in which source speech feature matrix and target speech feature matrix are described as $\mathbf{D}^s = [d_1^s, \ldots, d_M^s] \in \mathbb{R}^{k\times M}$ and $\mathbf{D}^t = [d_1^t, \ldots, d_N^t] \in \mathbb{R}^{k\times N}$, where k is the dimension of the speech feature in two domains. Specially, M and N are the numbers of source and target speech samples, respectively. In particular, source speech samples have labels, denoted by $\mathbf{L}^s = [l_1, \ldots, l_M] \in \mathbb{R}^{c\times M}$, where c is the number of speech emotion states, while target samples are unlabeled. For $\mathbf{L}^s$, the values of the l_j are defined as 1 when d_j^s belongs to the ith speech emotion class, otherwise the values are 0.

The main purpose of our proposed method is to find a projection subspace to make the target speech samples be represented by the source speech samples. Meanwhile, it can bridge the discrepancy of feature distributions between source and target samples and then learn a regression projection matrix to predict the emotion categories of target samples. Based on this method, the objective function can be formulated as:

$$\min_{\mathbf{C},\mathbf{Z}} \|\mathbf{L}^s - \mathbf{C}^T\mathbf{D}^s\|_F^2 + \lambda \sum_{i=1}^{N} \|\mathbf{C}^T\mathbf{d}_i^t - \mathbf{C}^T\mathbf{D}^s\mathbf{Z}_i\|_2^2 + \mu \sum_{i=1}^{N} \|\mathbf{Z}_i\|_1 + \tau\|\mathbf{C}^T\|_{2,1} \quad (1)$$

$\mathbf{C}$ and $\mathbf{Z}$ which are the regression coefficient matrix and the projection matrix, used to predict the emotion categories of target speech samples well. Meanwhile, the λ, μ, τ are the trade-off parameters in this function, all of them are defined as positive values.

For the Eq. (1), it contains four terms, respectively. The first term is least-squares loss function where $\|\cdot\|_F$ denotes Frobenius norm, and it aims to bridge the relationship between labels and features in source domain. Then, a squared ℓ_2 norm term is used to reflect the relationship of features between source and target domains onto the second term. Meanwhile, we added a sparse ℓ_1 norm penalty in $\mathbf{Z}_i$ onto the third term for better reconstruction of target domain features. Therefore, the target-adapted term will more effective to reduce the discrepancy between the features of source and target domains. Finally, for the fourth term, it is a $\ell_{2,1}$ norm penalty of the regression coefficient $\mathbf{C}$, the purpose of $\ell_{2,1}$ norm penalty is to transfer projection matrix to a group sparse matrix by setting all elements of rows in projection matrix to 0, since the sparse rows of projection matrix will correspond to the speech features that contribute less to the regress labels, namely feature selection.

Specially, for the optimation of the Eq. (1), the corresponding equivalent equation is follows:

$$\min_{\mathbf{C},\mathbf{Z}} \|\mathbf{L}^s - \mathbf{C}^T\mathbf{D}^s\|_F^2 + \lambda\|\mathbf{C}^T\mathbf{D}^t - \mathbf{C}^T\mathbf{D}^s\mathbf{Z}\|_F^2 + \mu\|\mathbf{Z}\|_1 + \tau\|\mathbf{C}^T\|_{2,1} \quad (2)$$

2.2 Optimization of TaSL

For the optimization of proposed TaSL model, Eq. (1) can be effectively solved by Alternative Direction Method (ADM) [12] and the Augmented Lagrangian Multiplier (ALM) method [13]. Specifically, for this cross-corpus SER problem, it can be easily solved by dividing it into two steps and then update them alternately. The updating rule can be summarized as follows:

(1) Fix $\mathbf{Z}$ and optimize $\mathbf{C}$: in this step, the regression coefficient matrix C will be computed, and the optimization problem can be written as follows:

$$\min_{\mathbf{C}} \|\mathbf{L}^s - \mathbf{C}^T\mathbf{D}^s\|_F^2 + \lambda\|\mathbf{C}^T(\mathbf{D}^t - \mathbf{D}^s\mathbf{Z})\|_F^2 + \tau\|\mathbf{C}^T\|_{2,1} \quad (3)$$

The optimization problem of (3) can be solved by using inexact ALM. To solve this problem, we can also rewrite the optimization problem as a simple form, following is the equivalent equation:

$$\arg\min_{\mathbf{Q},\mathbf{C}} \|[\mathbf{L}^s, \mathbf{0}] - \mathbf{Q}^T[\mathbf{D}^s, \sqrt{\lambda}\tilde{\mathbf{D}^{ts}}]\|_F^2 + \tau\|\mathbf{C}^T\|_{2,1}, \ s.t.\ \mathbf{Q} = \mathbf{C}. \quad (4)$$

where $\tilde{\mathbf{D}^{ts}} = \mathbf{D}^t - \mathbf{D}^s\mathbf{Z}$ and $\mathbf{0}$ is the matrix of zero, the corresponding augmented Lagrangian function of this equation can be described as:

$$\|[\mathbf{L}^s, \mathbf{0}] - \mathbf{Q}^T[\mathbf{D}^s, \sqrt{\lambda}\tilde{\mathbf{D}^{ts}}]\|_F^2 + tr[\mathbf{T}^T(\mathbf{Q} - \mathbf{C})] + \frac{l}{2}\|\mathbf{Q} - \mathbf{C}\|_F^2 + \tau\|\mathbf{C}^T\|_{2,1} \quad (5)$$

where $\mathbf{T}$ and l are the Lagrangian multiplier and regularization constant in this equation, furthermore, for the parameter l, it is defined as $l > 0$.

For the optimation of this step, it is similar to the optimation problem of (5) in [9]. Firstly, fix other parameters and update $\mathbf{Q}$, then update $\mathbf{C}$, finally, check the convergence conditions.

(2) Fix $\mathbf{C}$ and update $\mathbf{Z}$: in this step, the optimation function of $\mathbf{Z}$ is follows:

$$\min_{\mathbf{Z}} \lambda \|\mathbf{C}^T\mathbf{D}^t - \mathbf{C}^T\mathbf{D}^s\mathbf{Z}\|_2^2 + \mu\|\mathbf{Z}\|_1 \quad (6)$$

this problem is the Least Absolute Shrinkage and Selection Operator (LASSO) problem, and it can be easily solved by many optimization methods [14], same like one step of the optimization problem in [15].

2.3 Cross-Corpus SER Based on TaSL

The task of cross-corpus SER mainly is using data from two databases as training set and testing set to train models for the classification of speech emotion, then the features and labels of source domain are used to predict the labels of the target domain by the trained models. Based on the idea of our proposed method, the speech label matrix $\mathbf{L}^s$, speech feature matrix $\mathbf{D}^s$ of the source domain and the speech feature matrix $\mathbf{D}^t$ of the target domain are given while the speech label matrix $\mathbf{L}^t$ of target domain is unknown, then we use the proposed method to get the learned optimal regression coefficient matrix. Therefore, if we learn the optimal regression coefficient matrix $\hat{\mathbf{C}}$, then the optimal common feature space is obtained and the speech emotion states of speech samples can be judged by $\mathbf{l}^t = \hat{\mathbf{C}}^T\mathbf{d}^t$, where $\mathbf{d}^t$ denotes the feature vector of testing speech samples from the target domain, and $\mathbf{l}^t$ represents the column of $\mathbf{L}^t$. Furthermore, the speech emotion class of target speech samples is defined as i^{th} when the maximum value of the $\mathbf{l}^t$ belongs i^{th} row.

3. Experiments

3.1 Datasets and Protocols

In order to evaluate the performance of proposed TaSL method, we use three public speech emotion corpuses for cross-corpus SER, which are EmoDB [16], eNTERFACE [17], and AFEW4.0 [18]. Specially, EmoDB is a German database which consists of 800 utterances and that is recorded by professional actors, and it contains seven basic emotions (angry, disgust, fear, joy, sadness, neutral and boredom). eNTERFACE is recorded in English and consists of 1287 emotion videos from 43 subjects with six basic emotions, such as anger, disgust, fear, happiness, sadness and surprise. Besides, AFEW4.0 has 2577 clips with seven basic emotions, it contains anger, disgust, fear, happiness, neutral, sadness and surprise.

To evaluate our proposed method, we conduct our experiments on above three public databases by using the leave-one-corpus-out (LOCO) experimental protocol. In order to conduct the cross-corpus experiments, the common emotion states are choosen between two databases. Specially, in the experiments of EmoDB and eNTERFACE, we choose five emotions (angry, disgust, fear, joy/happy and sad) for training and the same five emotions for testing. Similarly, for AFEW4.0 and EmoDB, six common emotions are used, which are angry, disgust, fear, joy/happy, neutral and sad. Meanwhile, these common emotions (angry, disgust, fear, happy, sad and surprise) are used in the experiments of AFEW4.0 and eNTERFACE.

3.2 Experimental Setup

In the experiments of cross-corpus SER, we extract 384 dimensions features as input by the openSMILE toolkit [19], [20]. openSMILE features are the most popular feature set in cross-corpus SER, which contains lots of information such as Mel-frequency cepstral coefficient (MFCC), zero-crossing-rate (ZCR), fundamental frequency (F0) and so on. Furthermore, all emotion features are normalized to range of [0, 1] for better feature selection.

For this cross-corpus SER task, we designed six types of experiments based on the EmoDB, eNTERFACE and AFEW4.0 corpuses to evaluate our method, and the details are as follow:

- **b to e**: The source corpus is EmoDB and the target corpus is eNTERFACE.
- **e to b**: The source corpus is eNTERFACE and the target corpus is EmoDB.
- **b to a**: The source corpus is EmoDB and the target corpus is AFEW4.0.
- **a to b**: The source corpus is AFEW4.0 and the target corpus is EmoDB.
- **e to a**: The source corpus is eNTERFACE and the target corpus is AFEW4.0.
- **a to e**: The source corpus is AFEW4.0 and the target corpus is eNTERFACE.

Subsequently, to evaluate the performance of our proposed method, we use two different accuracy evaluation methods, which contain the weighted average recall (WAR) and the unweighted average recall (UAR). WAR is denoted as $WAR = \frac{te}{n}$, while UAR is $UAR = \frac{tc}{c}$, where te, tc, n and c represent the number of correctly predicted test samples, the sum of accuracy per class, the total number of test samples and the emotion states, respectively.

In order to verify the effectiveness of our proposed method, several popular methods are used to compare, which are KMM+SVM [21], KLIEP+SVM [22], uLSIF+SVM [23], ISLSR [9], DaLSR [10], DoSL [11]. As the baseline method, SVM is used without any domain adaptation methods. Furthermore, in order to obtain better performance, we finally set the optimal trade-off parameters (λ, μ and τ) in six experiments as (0.01, 1, 1), (15, 90, 76), (1, 0.001, 10), (0.42, 10, 9.6), (0.001, 5, 14) and (1, 0.01, 10), respectively.

3.3 Experimental Results

For the tasks of cross-corpus SER in this paper, the final experimental results about UAR and WAR are shown on Table 1. In particular, we can see that the proposed TaSL method can achieve better accuracies than any other methods in most cases and it is better than baseline method SVM in all of the experiments.

In contrast to DoSL, TaSL achieves the same accu-

Table 1 Results (UAR / WAR) of all the methods in the cross-corpus SER experiments

Method	**b to e**	**e to b**	**b to a**	**a to b**	**e to a**	**a to e**
SVM	30.06 / 30.08	27.83 / 24.27	26.07 / 25.99	29.87 / 35.02	20.80 / 18.39	18.68 / 18.72
KMM+SVM	23.08 / 23.14	40.18 / 44.69	**30.39** / 29.78	38.17 / 46.81	23.79 / 25.72	19.79 / 19.75
KLIEP+SVM	21.79 / 21.82	28.58 / 27.01	25.47 / 25.57	27.41 / 31.37	18.66 / 18.60	17.48 / 17.47
uLSIF+SVM	25.75 / 25.75	40.42 / 42.27	25.75 / 25.93	36.25 / 44.38	22.61 / 21.21	18.10 / 18.11
ISLSR	32.52 / 32.59	42.11 / 50.93	27.73 / 30.19	36.13 / 46.70	24.24 / 26.20	22.55 / 22.58
DaLSR	36.36 / 36.40	**44.41** / **52.27**	27.51 / 30.19	37.33 / 47.80	24.67 / **26.70**	21.93 / 21.96
DoSL	**37.49** / **37.51**	44.25 / 52.00	29.10 / 31.00	39.66 / 50.00	**24.83** / 26.20	21.64 / 21.66
TaSL	**37.49** / **37.51**	42.64 / 45.60	29.22 / **31.70**	**43.69** / **51.76**	24.67 / 26.07	**25.19** / **25.21**

racy in the experiment **b to e**, UAR and WAR of them are (37.49%, 37.51%), better than DoSL in the experiments **a to b** and **a to e**. It reflects that TaSL can get better results than DoSL, so the added projection matrix and ℓ_1 norm penalty term can represent target domain features better by getting sparse features of source domain, and TaSL is more effective than DoSL. However, our method TaSL is worse than DaLSR in the experiments **e to b** and **e to a**, where DaLSR achieves 44.41% (UAR) and 52.27% (WAR) in **e to b**, and 26.70% (WAR) in the experiment **e to a**, while DoSL method gets 24.83% (UAR) in this experiment, this may due to the imbalance of class samples in these cases. Nevertheless, in comparison of ISLSR, DaLSR and DoSL, TaSL is better than ISLSR in experiment **b to e**, worse than DaLSR and DoSL. Specifically, these four methods get similar accuracy in experiment **e to a**, the TaSL method performances well in experiment **a to b** and **a to e**, respectively. Based on these results, it is obvious that transfer subspace learning methods are effective to cope with the cross-corpus SER problems, especially for the method we proposed.

4. Conclusion

In this paper, we proposed a Target-adaption subspace learning (TaSL) method to cope with the cross-corpus SER tasks. The proposed TaSL model contains the least-squares regression term with ℓ_1 norm for the label regression and the target-adapted term with $\ell_{2,1}$ norm, which can not only bridge the relationship between feature space and label space in source domina better, but also effectively reduce the discrepancy of feature distributions in source and target domains. Furthermore, a optimal projection common space is obtained to regress the labels from features of speech samples effectively and the features of target samples are reconstructed by source samples features better. Finally, extensive experiments on three speech emotion corpuses (EmoDB, eNTERFACE, and AFEW4.0) show that compared with other domain adaptation methods, the proposed TaSL method achieves the better performance in dealing with the cross-corpus SER tasks. Although the proposed method is effective for cross-corpus SER, the transfer performance of different unbalanced corpuses is worse. In the future, we will focus on more robust model for the problem of imbalance sample.

References

[1] M.E. Ayadi, M.S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognition, vol.44, no.3, pp.572–587, 2011.

[2] L. Li, X. Zhou, Y. Zong, W. Zheng, X. Chen, J. Shi, and P. Song, "Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression," IEICE Trans. Inf. & Syst., vol.E102-D, no.7, pp.1417–1421, 2019.

[3] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "Mped: A multi-modal physiological emotion database for discrete emotion recognition," IEEE Access, vol.7, pp.12177–12191, 2019.

[4] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," IEEE Transactions on Affective Computing, p.1, 2018.

[5] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies," IEEE Transactions on Affective Computing, vol.1, no.2, pp.119–131, 2010.

[6] S.J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol.22, no.10, pp.1345–1359, 2010.

[7] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: Compensating for covariate shift," IEEE Transactions on Audio Speech & Language Processing, vol.21, no.7, pp.1458–1468, 2013.

[8] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse autoencoder-based feature transfer learning for speech emotion recognition," Affective Computing and Intelligent Interaction, 2013.

[9] W. Zheng, M. Xin, X. Wang, and B. Wang, "A novel speech emotion recognition method via incomplete sparse least square regression," IEEE Signal Process. Lett., vol.21, no.5, pp.569–572, 2014.

[10] Y. Zong, W. Zheng, T. Zhang, and X. Huang, "Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression," IEEE Signal Process. Lett., vol.23, no.5, pp.585–589, 2016.

[11] N. Liu, Y. Zong, B. Zhang, L. Liu, J. Chen, G. Zhao, and J. Zhu, "Unsupervised cross-corpus speech emotion recognition using domain-adaptive subspace learning," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.5144–5148, IEEE, 2018.

[12] Z. Qin and D. Goldfarb, "Structured sparsity via alternating direction methods," Journal of Machine Learning Research, vol.13, no.1, pp.1435–1468, 2012.

[13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.35, no.1, pp.171–184, 2013.

[14] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July, pp.547–556, 2009.

[15] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning from hierarchical spatiotemporal descriptors for micro-expression recognition," IEEE Trans. Multimedia, vol.20, no.11, pp.3160–3172, 2018.

[16] F. Burkhardt, A. Paeschke, M. Rolfes, W.F. Sendlmeier, and B. Weiss, "A database of german emotional speech," INTERSPEECH 2005 - Eurospeech, European Conference on Speech Communication and Technology, Lisbon, Portugal, September, pp.1517–1520, 2005.

[17] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The enterface'05 audio-visual emotion database," International Conference on Data Engineering Workshops, p.8, 2006.
[18] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion recognition in the wild challenge 2014: baseline, data and protocol," ACM on International Conference on Multimodal Interaction, pp.461–466, 2014.
[19] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," ACM International Conference on Multimedia, pp.835–838, 2013.
[20] F. Eyben, openSMILE:): the Munich open-source large-scale multimedia feature extractor, ACM, 2015.
[21] J. Huang, A.J. Smola, A. Gretton, K.M. Borgwardt, and B. Scholkopf, "Correcting sample selection bias by unlabeled data," International Conference on Neural Information Processing Systems, 2006.
[22] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P.V. Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," Annals of the Institute of Statistical Mathematics, vol.60, no.4, pp.699–746, 2008.
[23] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," Journal of Machine Learning Research, vol.10, no.Jul, pp.1391–1445, 2009.