A SURVEY ON TRANSFER LEARNING FRAMEWORK FOR DATA SETS USING SEMI SUPERVISED LEARNING

[1]S. Balaji [2]Dr. G. Vasanth [3]Dr. D. Sumathi
1,3 : Associate Professors, Kuppam Engineering College, Kuppam, AP., India
2 : HoD & Professor, Govt. Engineering College, K.R.Pet, Karnataka, India.

Abstract: An assumption in machine learning and data mining algorithms is training data and future data are same. When it comes to real word applications this assumption may be failed because the future data may be from different sources with different domains. Here the importance of transfer learning comes into the picture. Transfer learning deals with data like labeled, unlabeled data and semi labeled data. It is important to design effective transfer learner to deal with above said data. This article focused on Survey on different transfer learning methods like Domain adaptation, Semi-Supervised learning, Transfer learning, Self-taught learning.

Key words: Transfer learning, Self- taught learning, Domain Adaptation, Data Mining, Machine learning, Translated Learning.

I. INTRODUCTION

Labeled data for machine learning is often very difficult and expensive to obtain, and thus the ability to use unlabeled data holds significant promise in terms of vastly expanding the applicability of learning methods [1]. Automatically organizing and indexing multimedia content becomes increasingly important as the online images and videos continue to be a vital resource in everyday life. Consequently, devising effective visual category models has attracted considerable attention in computer vision area in recent years. When sufficient labeled training images are available, traditional classification methods usually work well. However, because image labeling requires expensive and time-consuming human labors, it is not likely that we always have enough training data to achieve satisfactory performance for new computer vision tasks [7].

A fundamental assumption in classification learning is that the data distributions of training and test sets should be identical. When the assumption does not hold, traditional classification methods might perform worse. However, in practice, this assumption may not always hold. For example, in Web mining, the Web data used in training a Web-page classification model can be easily out-dated when applied to the Web sometime later, because the topics on the web change frequently. Often, new data are expensive to label and thus their quantities are limited due to cost issues. How to accurately classify the new test data by making the maximum use of the old data becomes a critical problem [2]. Facing the information flood in our daily lives, search engines mainly respond to our submitted queries passively, while recommender systems aim to discover and meet our needs in a more active way. Collaborative filtering techniques have been applied in various recommendation-embedded applications. However, lack of users' accurate preference data, for example, five-star numerical ratings might limit this approach's applicability in real deployment. On the other side, a real recommender system can usually make use of additional types of user feedback [6]

Domain adaptation aims at solving a learning problem in the target domain by utilizing training data in the source domain, even when these domains may have different distributions. This is an important learning problem because labeled data are often difficult to come by, making it desirable to make the best use of any related data available [3]. Traditional machine learning relies on the availability of a large amount of labeled data to train a model in the same feature space. However, labeled data are often scarce and expensive to obtain. In order to save much labeling work, various machine learning strategies have been proposed, including semi-supervised learning, transfer learning, self-taught learning etc. One commonality among these strategies is they all require the training data and test data to be in the same feature space [4]. A commonality among these methods is that they all require the training data and test data to be in the same feature space. In addition, most of them are designed for supervised learning. However, in practice, we often face the problem where the labeled data are scarce in their own feature space, whereas there may be a large amount of labeled heterogeneous data in another feature space. In such situations, it would be desirable to transfer the knowledge from heterogeneous data to domains where we have relatively little training data available [5].

II. RELATED WORKS

Most machine learning tasks in data classification and information retrieval require manually labeled data examples in the training stage. The goal of active learning was to select the most informative examples for manual labeling in these learning tasks. Most of the previous studies in active learning have focused on selecting a single unlabeled example in each iteration. This could be inefficient, since the classification model has to be retrained for every acquired labeled example. It was also inappropriate for the setup of information retrieval tasks where the user's relevance feedback was often provided for the top K retrieved items. Hoi *et al.* [8] have presented a framework for batch mode active learning, which selects a number of informative examples for manual labeling in each iteration. The key feature of batch mode active learning was to reduce the redundancy among the selected examples such that each example provides unique information for model updating. To this end, they employed the Fisher information matrix as the measurement of model uncertainty, and choose the set of unlabeled examples that could efficiently reduce the Fisher information of the classification model. They applied their batch mode active learning framework to both text categorization and image retrieval.

Transfer learning was established as an effective technology to leverage rich labeled data from some source domain to build an accurate classifier for the target domain. The basic assumption was that the input domains may share certain knowledge structure, which could be encoded into common latent factors and extracted by preserving important property of original data, e.g., statistical property and geometric structure. Long *et al.* [9] have showed that different properties of input data could be complementary to each other and exploring them simultaneously could make the learning model robust to the domain difference. They proposed a general framework, referred to as Graph Co-Regularized Transfer Learning (GTL), where various matrix factorization models could be incorporated. Specifically, GTL aims to extract common latent factors for knowledge transfer by preserving the statistical property across domains, and simultaneously, refine the latent factors to alleviate negative transfer by preserving the geometric structure in each domain.

A major assumption in many machine learning and data mining algorithms was that the training and future data must be in the same feature space and have the same distribution. However, in many real-world applications, this assumption may not hold. For example, sometimes have a classification task in one domain of interest, but could only have sufficient training data in another domain of interest, where the latter data might be in a different feature space or follow a different data distribution. In such cases, knowledge transfer, if done successfully, would greatly improve the performance of learning by avoiding much expensive data-labeling efforts. In recent years, transfers learning have emerged as a learning framework to address this problem. Pan and Yang [10] have surveyed focuses on categorizing and reviewing the current progress on transfer learning for classification, regression, and clustering problems. They have discussed the relationship between transfer learning and other related machine learning techniques such as domain adaptation, multitask learning and sample selection bias, as well as covariate shift.

Regular machine learning and data mining techniques study the training data for future inferences under a major assumption that the future data are within the same feature space or have the same distribution as the training data. However, due to the limited availability of human labeled training data, training data that stay in the same feature space or have the same distribution as the future data cannot be guaranteed to be sufficient enough to avoid the over-fitting problem. In real-world applications, apart from data in the target domain, related data in a different domain could also be included to expand the availability of the prior knowledge about the target future data. Transfer learning addresses such crossdomain learning problems by extracting useful information from data in a related domain and transferring them for being used in target tasks. In recent years, with transfer learning being applied to visual categorization, some typical problems, e.g., view divergence in action recognition tasks and concept drifting in image classification tasks, could be efficiently solved. Shao *et al.* [11] have surveyed state-of-the-art transfer learning algorithms in visual categorization applications, such as object recognition, image classification, and human action recognition.

Tao and Geng [12] have presented a family of subspace learning algorithms based on a form of regularization, which transfers the knowledge gained in training samples to testing samples. In particular, the regularization minimizes the Bregman divergence between the distribution of training samples and that of testing samples in the selected subspace, so it boosts the performance when training and testing samples are not independent and identically distributed. To test the effectiveness of the proposed regularization, they introduced it to popular subspace learning algorithms, e.g., principal components analysis (PCA) for cross-domain face modeling; and Fisher's linear discriminant analysis (FLDA), locality preserving projections (LPP), marginal Fisher's analysis (MFA), and discriminative locality alignment (DLA) for cross-domain face recognition and text categorization. Finally, they presented experimental evidence on both face image data sets and text data sets, suggesting that the proposed Bregman divergence-based regularization was effective to deal with cross-domain learning problems.

2.1 Homogeneous transfer learning:

This section presents surveyed papers covering homogeneous transfer learning solutions and is divided into subsections that correspond to the transfer categories of instance based, feature-based (both asymmetric and symmetric), parameter-based, and relational-based. Recall that homogeneous transfer learning is the case where XS = XT. The algorithms surveyed the methodology of homogeneous transfer learning is directly applicable to a big data environment. As repositories of big data become more available, there is a desire to use this abundant resource for machine learning tasks, avoiding the timely and potentially costly collection of new data. If there is an available dataset that is drawn from a domain that is related to, but does not an exactly match a target domain of interest, then homogeneous transfer learning can be used to build a predictive model for the target domain as long as the input feature space is the same.

2.2 Instance-based transfer learning:

The paper by Chattopadhyay proposes two separate solutions both using multiple labeled source domains. The first solution is the conditional probability based multisource domain adaptation (CP-MDA) approach, which is a domain adaptation process based on correcting the conditional distribution differences between the source and target domains. The CP-MDA approach assumes a limited amount of labeled target data is available. The main idea is to use a combination of source domain classifiers to label the unlabeled target data. This is accomplished by first building a classifier for each separate source domain. Then a weight value is found for each classifier as a function of the closeness in conditional distribution between each source and the target domain. The weighted source classifiers are summed together to create a learning task that will find the pseudo labels (estimated labels later used for training) for the unlabeled target data. Finally, the target learner is built from the labeled and pseudo labeled target data. The second proposed solution is the two stage weighting framework for multi-source domainadaptation (2SW-MDA) which addresses both marginal and conditional distribution differences between the source and target domains.

Labeled target data is not required for the 2SW-MDA approach; however, it can be used if available. In this approach, a weight for each source domain is computed based on the marginal distribution differences between the source and target domains. In the second step, the source domain weights are modified as a function of the difference in the conditional distribution as performed in the CP-MDA approach previously described. Finally, a target classifier is learned based on the reweighted source instances and any labeled target instances that are available. The work presented in Chattopadhyay is an extension of Duan where the novelty is in calculating the source weights as a function of conditional probability. Note, the 2SW-MDA approach is an example of an instance-based Transfer Category, but the CP-MDA approach is more appropriately classified as a parameter-based Transfer Category (see "Parameter-based transfer learning" section). Experiments are performed for muscle fatigue classification using surface electromyography data where classification accuracy is measured as the performance metric. Each source domainrepresents one person's surface electromyography measurements. A baseline approach is constructed using a support vector machine (SVM) classifier trained on the combination of seven sources used for this test. The transfer learning approaches that are tested against include an approach proposed by Huang, Pan, Zhong, Gao, and Duan. The order of performance from best to worst is 2SW-MDA, CP-MDA, Duan, Zhong, Gao, Pan, Huang, and the baseline approach. All the transfer learning approaches performed better than the baseline approach.

2.3 Asymmetric feature-based transfer learning:

In an early and often cited work, Daumé proposes a simple domain adaptation algorithm, referred to as the feature augmentation method (FAM), requiring only ten lines of Perl script that uses labeled source data and limited labeled target data. In a transfer learning environment, there are scenarios where a feature in the source domain may have a different meaning in the target domain. The issue is referred to as context feature bias, which causes the conditional distributions between the source and target domains to be different. To resolve context feature bias, a method to augment the source and target feature space with three duplicate copies of the original feature set is proposed. More specifically, the three duplicate copies of the original feature set in the augmented source feature space represent a common feature set, a source specific feature set, and a target specific feature set which is always set to zero. In a similar way, the three duplicate copies of the original feature set in the augmented target feature space represent a common feature set, a source specific feature set which is always set to zero, and a target specific feature set. By performing this feature augmentation, the feature space is duplicated three times. From the feature augmentation structure, a classifier learns the individual feature weights for the augmented feature set, which will help correct for any feature bias issues. Using a text document example where features are modeled as a bagof-words, a common word like "the" would be assigned (through the learning process) a high weight for the common feature set, and a word that is different between the source and target like "monitor" would be assigned a high weight for the corresponding domain feature set. The duplication of features creates feature separation between the source and target domains, and allows the final classifier to learn the optimal feature weights. For the experiments, a number of different natural language processing applications are tested and in each case the classification error rate is measured as the performance metric.

An SVM learner is used to implement the Daumé approach. A number of baseline approaches with no transfer learning techniques are measured along with a method by Chelba. The test results show the Daumé method is able to outperform the other methods tested. However, when the source and target domains are very similar, the Daumé approach tends to underperform. The reason for the underperformance is the duplication of feature sets represents irrelevant and noisy information when the source and target domains are very similar. Multiple kernel learning is a technique used in traditional machine learning algorithms as demonstrated in the works of Wu and Vedaldi. Multiple kernel learning allows for an optimal kernel function to be learned in a computationally efficient manner. The paper by Duan proposes to implement a multiple kernel learning framework for a transfer learning environment called the domain transfer multiple kernel learning Weiss et al. J Big Data (2016) 3:9 Page 10 of 40 (DTMKL). Instead of learning one kernel, multiple kernel learning assumes the kernel is comprised of a linear combination of multiple predefined base kernels. The final classifier and the kernel function are learned simultaneously which has the advantage of using labeled data during the kernel learning process. This is an improvement over Pan and Huang where a two-stage approach is used. The final classifier learning process minimizes the structural risk functional and the marginal distribution between domains using the maximum mean discrepancy measure. Pseudo labels are found for the unlabeled target data to take advantage of this information during the learning process.

The pseudo labels are found as a weighted combination of base classifiers (one for each feature) trained from the labeled source data. A regularization term is added to the optimization problem to ensure the predicted values from the final target classifier and the base classifiers are similar for the unlabeled target data. Experiments are performed on the applications of video concept detection, text classification, and email spam detection. The methods tested against include a baseline approach using an SVM classifier trained on the labeled source data, the feature replication method from Daumé, an adaptive SVM method from Yang, a cross-domain SVM method proposed by Jiang, and a kernel mean matching method by Huang. The DTMKL approach uses an SVM learner for the experiments. Average precision and classification accuracy are measured as the performance metrics. The DTMKL method performed the best for all applications, and the baseline approach is consistently the worst performing. The other methods showed better performance over the baseline which demonstrated a positive transfer learning effect. The work by Long is a joint domain adaptation (JDA) solution that aims to simultaneously correct for the marginal and conditional distribution differences between the labeled source domain and the unlabeled target domain. Principal component analysis (PCA) is used for optimization and dimensionality reduction. To address the difference in marginal distribution between the domains, the maximum mean discrepancy distance measure is used to compute the marginal distribution differences and is integrated into the PCA optimization algorithm. The next part of the solution requires a process to correct the conditional distribution differences, which requires labeled target data. Since the target data is unlabeled, pseudo labels (estimated target labels) are found by learning a classifier from the labeled source data. The maximum mean discrepancy distance measure is modified to measure the distance between the conditional distributions and is integrated into the PCA optimization algorithm to minimize the conditional distributions. Finally, the features identified by the modified PCA algorithm are used to train the final target classifier. Experiments are performed for the application of image recognition and classification accuracy is measured as the performance metric. Two baseline approaches of a 1-nearest neighbor classifier and a PCA approach trained on the source data are tested. Transfer learning approaches tested for this experiment include the approach by Pan, Gong, and Si. These transfer learning approaches only attempt to correct for marginal distribution differences between domains. The Long approach is the best performing, followed by the Pan and Si approaches (a tie), then the Gong approach, and finally the baseline approaches. All transfer learning approaches perform better than the baseline approaches. The possible reason behind theunderperformance of the Gong approach is the data smoothness assumption that is made for the Gong solution may not be intact for the data sets tested. The paper by Long proposes an Adaptation Regularization based transfer learning (ARTL) framework for scenarios of labeled source data and unlabeled target data. This transfer learning framework proposes to correct the difference in marginal distribution between the source and target domains, correct the difference in conditional distribution between the domains, and improve classification performance through a manifold regularization process (which optimally shifts the hyperplane of an SVM learner).

This complete framework process is depicted. The proposed ARTL framework will learn a classifier by simultaneously performing structural risk minimization, reducing the marginal and conditional distributions between the domains, and optimizing the manifold consistency of the marginal distribution. To resolve the conditional distribution differences, pseudo labels are found for the target data in the same way as proposed by Long. A difference between the ARTL approach and Long is ARTL learns the final classifier simultaneously while minimizing the domain distribution differences, which is claimed by Long to be a more optimal solution. Unfortunately, the solution by Long is not included in the experiments. Experiments are performed on the applications of text classification and image classification where classification accuracy is measured as the performance metric. There are three baseline methods tested where different classifiers are trained with the labeled source data. There are five transfer learning methods tested against, which include methods by Ling, Pan, Pan, Quanz, and Xiao. The order of performance from best to worst is ARTL, Xiao, Pan, Pan, Quanz and Ling (tie), and the baseline approaches. The baseline methods underperformed all other transfer learning approaches tested.

III. PROBLEM DEFINITION:

Transfer learning focuses on the learning scenarios when the test data from target domains and the training data from source domains are drawn from similar but different data distributions with respect to the raw features. Data classification has been an active research topic in the machine learning community for many years. The goal of data classification is to automatically assign data examples to a set of predefined categories. The major computational problems of transfer learning are

- To explore the shared knowledge structure underlying input domains as the bridge to propagate supervision information from the source domains to the target domains.
- Data collected at different periods may suffer from the concept drifting.
- Sufficient training data belonging to the same feature space or the same distribution as the testing data may not always be available.

IV. PROPOSED METHODOLOGY:

Transfer learning has been identified to be an effective tool to address this problem by storing knowledge gained from training samples and applying the knowledge to testing samples. Conventional algorithms for image and video data analysis and organization perform well under the assumption that training and testing samples are independent and identically distributed. Regular machine learning approaches have achieved promising results under the major assumption that the training and testing data stay in the same feature space or share the same distribution. Transfer learning addresses cross domain learning problems by extracting useful information from data in a related domain and transferring them for being used in target tasks. In this work we have proposed an efficient technique for transfer learning framework for text and image data sets using semi supervised learning methods. The proposed method can be categorized into different steps of knowledge input, annotation of learning materials, creation of knowledge space and classification. The data set are subjected to supervised learning and the classification done in the neural network which is modified by incorporating the optimization algorithm for weight selection. The optimization algorithm utilized is the Modified cuckoo search algorithm. The data are trained in neural network based on supervised learning scenario and the classification is done based on this trained data's. The proposed approach will be implemented in the working platform of JAVA and effectiveness of the method is verified by comparing the parameters like precision, recall and F-measure with some existing techniques

REFERENCE:

1. Rajat Raina,Alexis Battle,Honglak Lee,Benjamin Packer and Andrew Y. Ng, "Self-taught Learning: Transfer Learning from Unlabeled Data",In.Proc.of the 24 th International Conference on Machine Learning, Corvallis,2007.
2. Wenyuan Dai, Qiang Yang, Gui-Rong Xue and Yong Yu, "Boosting for Transfer Learning", In.Proc.of the 24 th International Conference on Machine Learning, Corvallis,2007.
3. Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang, "Domain Adaptation via Transfer Component Analysis",IEEE transactions on neural networks, vol. 22, no. 2, 2011.
4. Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang and Yong Yu, "Translated Learning: Transfer Learning across Different Feature Spaces", Advances in Neural Information Processing Systems,pp.353--360, 2008
5. Qiang Yang, "Heterogeneous Transfer Learning with Real-world Applications" In,proc.of 4th International Joint Conference on Natural Language Processing of the AFNLP,2009.
6. Weike Pan and Zhong Ming, "Interaction-Rich Transfer Learning for Collaborative Filtering with Heterogeneous User Feedback",IEEE intelligent systems,2014
7. Hua Wang, Feiping Nie, Heng Huang and Chris Ding, "Dyadic Transfer Learning for Cross-Domain Image Classification",In.proc.of IEEE International Conference on Computer Vision,2011.
8. Steven C.H. Hoi, Rong Jin and Michael R. Lyu, "Batch Mode Active Learning with Applications to Text Categorization and Image Retrieval",IEEE transactions on knowledge and data engineering, vol.21, no.9,2009.
9. Mingsheng Long, Jianmin Wang, Guiguang Ding, Dou Shen, and Qiang Yang, "Transfer Learning with Graph Co-Regularization",IEEE transactions on knowledge and data engineering, vol. 99,2011
10. Sinno Jialin Pan and Qiang Yang, "A Survey on Transfer Learning", IEEE transactions on knowledge and data engineering, vol. 22, no. 10,2010.
11. Ling Shao, Fan Zhu, and Xuelong Li, "Transfer Learning for Visual Categorization: A Survey",IEEE Transactions On Neural Networks And Learning Systems,2014.
12. Si Si, Dacheng Tao and Bo Geng, "Bregman Divergence-Based Regularization for Transfer Subspace Learning",IEEE Transactions On Knowledge And Data Engineering, Vol.22, No.7,2010
13. Karl Weiss,Taghi M.Khoshgoftaar And DingDing wang, “A Survey of Transfer Learning”,Journal of Bigdata,Vol:3,Issue:9,2016.