Bayesian Transfer Learning

Apoorv Dwivedi - IIT2014142
Nishit Gupta - IIT2014502
Sahil Prakash - IIT2014504

INTRODUCTION

- Machine Learning algorithms have already achieved great heights in knowledge engineering areas including classification, clustering etc.
- However, these traditional algorithms are still not able to mimic the human tendency of applying previously gained knowledge to a different but related problem.
- To efficiently utilize the gained knowledge to a problem of interest in another area, the concept of 'Transfer Learning' is introduced.

MOTIVATION

- Traditional ML Algorithms work well under a common assumption - the training and test data are drawn from the same domain.
- But sometimes, training data may be -
  - **Expensive.**
  - **Difficult to obtain.**
  - **Easily outdated.**
  - **Not available.**
- In such a scenario, the traditional ML algorithms will not produce accurate results.
- Hence, there is a need to transfer knowledge from one domain to another - a concept known as ‘Transfer Learning’.

PROBLEM DEFINITION AND OBJECTIVE

- Given a source domain D_S, corresponding source task T_S as well as a target domain D_T and corresponding target task T_T , the objective of transfer learning is to enable us to learn the target conditional probability $P(Y_T|X_T)$ in D_T with the information gained from D_S and T_S where $D_S \neq D_T$ or $T_S \neq T_T$.
- The first part of the project aims to **implement the Kernelized Bayesian Transfer Learning approach for supervised learning**, i.e. only labelled data of both the domains are available.
- We also extend the idea of transfer learning in a scenario when we have a mixture of Labeled and Unlabeled data in the training data-set i.e **Semi-Supervised Learning** and develop a model for the problem defined.

Our Contribution

- Though a little amount of research work has been done in the field of Supervised Transfer Learning but no work has been proposed in field of Transfer Learning when the data is Semi-supervised.
- No proposed approach as of now.
- Since we are using **Bayesian approach**, there must be some way to sanitize such data-set to obtain some interesting knowledge pertaining to the problem statement.
  - **Semi-Supervised Clustering** can yield the labels of Unknown Data using the information present in the Labeled Data and constraints existing between both of them.
- We therefore propose such an algorithm which could perform Semi-Supervised Transfer Learning.

LITERATURE SURVEY

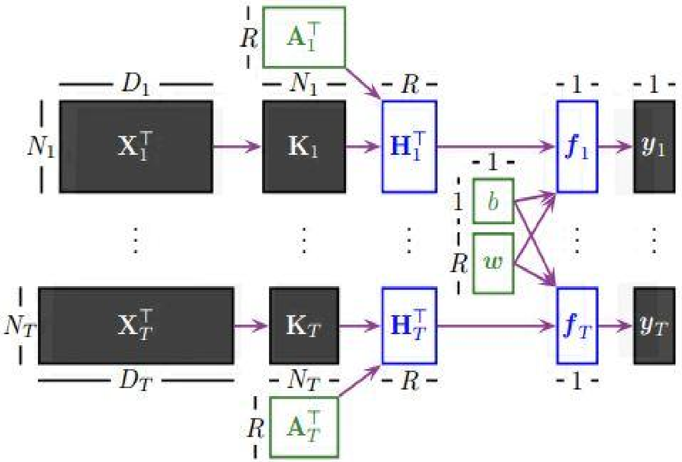
S No.	Author	Paper Title	Year	Crux	Venue
1.	Mehmet Gonen Adam A. Margolin	*Kernelized Bayesian Transfer Learning* [1]	2014	*This paper provides an insight on Transfer Learning. Firstly it defines Transfer Learning Mechanism and then a Bayesian Model is proposed to learn in One Domain and use this knowledge to the problem of interest in Another Domain. The concept of Kernel trick, priors and posterior probability distribution are used to achieve this.*	Twenty-eighth AAAI Conference on Artificial Intelligence.
2.	Kevin P. Murphy	*Book : Machine Learning – A probabilistic perspective.* [4] *Chapter 3 and 5*	2012	*Chapter 3 – Generative Models for Discrete Data define the concepts of Prior, Likelihood, Posterior and Posterior Predictive Distribution.* *Chapter 5- Bayesian Statistics gives advance details on summarizing posterior distributions and Bayesian Model Selection using Priors and Hierarchical Bayesian Approach.*	Massachusetts Institute of Technology.
3.	Mehmet Gonen	*Bayesian Efficient Multiple Kernel Learning.* [5]	2012	*This Paper provides the conjugate distribution theory for approximation of hper-parameters during training.*	Helsinki Institute of Information Technology HIIT, Dept of Information and Computer Science, Aalto University School of Science
4.	E. Amid, A. Gionis and A. Ukkonen	*A Kernel-learning approach to Semi-Supervised clustering with relative distance comparisons.* [7]	2015	*This Paper provides a Kernel-based approach to minimize the relative distance between data points to obtain clusters in a Semi-Supervised environment.*	Helsinki Institute for Information Technology, Department of Computer Science, Aalto University

PROPOSED METHODOLOGY

Supervised Learning : Kernelized Bayesian Transfer Learning

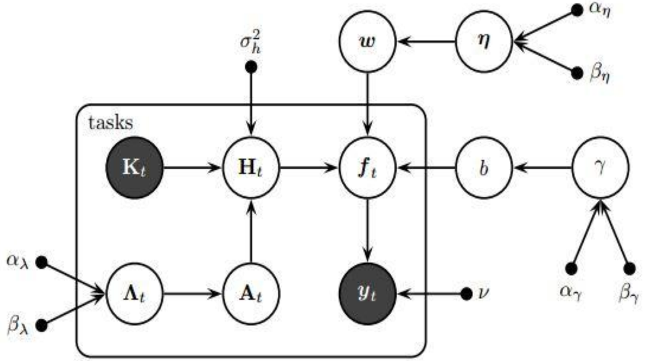
The model uses a two-step process for classification:

- **Subspace Reduction**: The data points from the different tasks having different dimensionalities are brought into a common subspace. This is done using the kernel-based dimensionality reduction technique.
- **Classification**: After bringing the data points into a common subspace, classification is done using the parameters learned during the training.



Bayesian Approach

- The Bayesian approach to Machine learning uses the probabilistic methods to learn from the available data.
- For the learning purpose, a model is defined which will have some unknown parameters.
- For the unknown parameters, we assume that they follow some known probability distribution, for example, Normal or Gaussian distribution.
- On the basis of the training data available, we calculate the posterior probability for the unknown parameters given the available training data.
- The posterior distribution thus obtained is used to determine the optimal value for the parameters such that it produces the best possible results.



*But how to handle **mixture of Labeled and Unlabeled Data**?*

- Since Unlabeled data doesn't have Output Labels, KBTL Model hence can not be applied. But if data is pre-processed in some manner to have labels, then Good to Go!

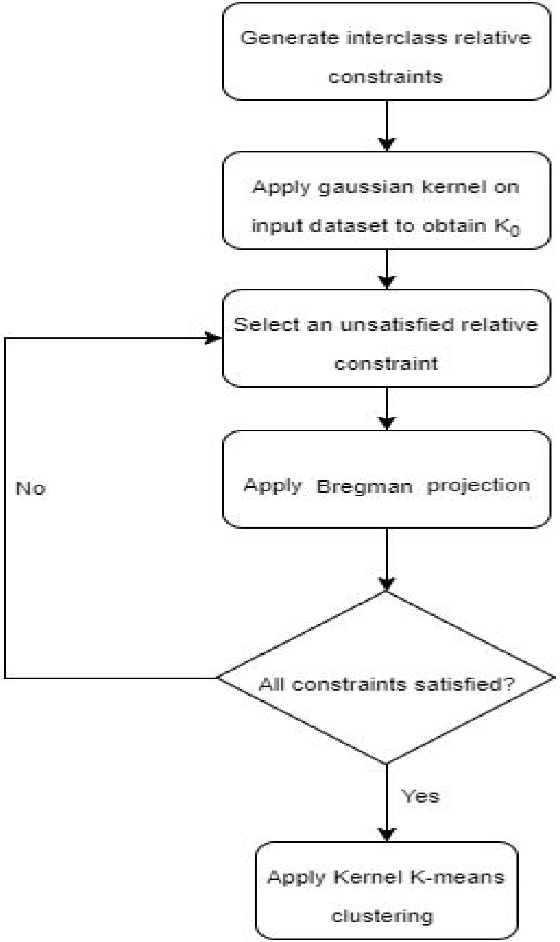
We apply a preprocessing step which uses clustering to assign output labels to each unlabeled data point. KBTL model can now be used as usual.

Semi-Supervised Clustering : <u>Kernel Learning using Relative Constraints</u>

➢ Cluster a given data-set into k-clusters subject to an additional set of constraints based on relative distance comparisons between the data points.

➢ Additional constraints present some background/side information which is specifically not present in the feature vectors directly but is provided by the few labeled data points.

- ➢ This approach uses a Log Determinant Divergence for learning the Kernel matrix subject to relative distance constraints.
- ➢ Now using this learned Kernel matrix, clustering can now be obtained using any general algorithm like kernel k-means.
    - ○ uses Kernel Matrix as the Distance Measure.



EXPERIMENTS

- Dataset Used = "*Office*" Dataset containing images from different categories like Backpack, Bottle, Helmet, Mouse, Printer etc. under 3 domains wiz. Amazon, DSLR and Webcam. Categories used for Binary Classification - Backpack and Bottle.
- Source Domain = Amazon Images
- Target Domain = Webcam Images
- Labels : The output labels assigned to the images are
  - **y = +1 for Backpack and**
  - **y = −1 for Bottle.**

Supervised Transfer Learning

Experiment 1:

- Training Data :
  - Source Domain (Amazon images): 20 Images from each category (Backpack and Bottle).
  - Target Domain (Webcam images): 3 Images from each category (Backpack and Bottle).
- Testing Data:
  - Target Domain (Webcam images): 26 Images of Backpack and 16 images of Bottle.
- **Result:**
  - Accuracy: 80.95%

Experiment 2:

- Training Data :
  - Source Domain (Amazon images): 92 images of Backpack, 36 images of Bottle.
  - Target Domain (Webcam images): 5 Images of Backpack and 3 images of Bottle.
- Testing Data:
  - Target Domain (Webcam images): 29 Images of Backpack and 16 images of Bottle.
- **Result:** On running the algorithm for variable number of iterations and also different subspace dimensionalities, the
  - Maximum Accuracy: 95%
  - Average Accuracy: 93.3%

Experiment 3:

- Training Data :
    - Source Domain (Amazon images):
        - Backpack: 92 images.
        - Bottle: 36 images.
        - Bookcase: 82 images.
        - File cabinet: 81 images.
    - Target Domain (Webcam images):
        - Backpack: 5 images.
- Testing Data:
    - Target Domain (Webcam images):
        - Backpack: 29 images.
        - Bottle: 16 images.
        - Headphones: 5 images.
- **Result:** On running the algorithm for variable number of iterations and also different subspace dimensionalities, the model predicted the image to be a Backpack or not by a:
    - Maximum Accuracy: 92%
    - Average Accuracy: 90%

Semi-Supervised Transfer Learning

Dataset Used

- Training Data-set
  - Source Images
    - Category 1 : Backpacks, 80 Data Points
    - Category 2 : Bike, 80 Data Points
    - Category 3 : Bookcase, 80 Data Points
    - Category 4 : Calculator, 80 Data Points
    - Category 5 : Desk-chair, 80 Data Points
    - Here 10 data points from each category are LABELED whereas 70 are UNLABELED.
  - Target Images
    - Category 1 : Backpacks, 5 Data Points
    - Category 2 : Bottles, 3 Data Points
- Testing Data-set
  - Target Images
    - Category 1 : Backpacks, 29 Data Points
    - Category 2 : Bottles, 16 Data Points

Experiment 1:

- Applied clustering to obtain labels for the unlabeled data points in the training data-set.
- Dimensionality of data points : 800
- **Result:** Accuracy of clustering : 70.3%

Experiment 2:

- Applied PCA to reduce the dimensionality of the data points before applying clustering.
- After obtaining the labels, applied KBTL for classification.
- Dimensionality of data points : 5
  - Accuracy of Clustering : 72.5%
  - Accuracy of Classification : 84.4%
- Dimensionality of data points : 10
  - Accuracy of Clustering : 83.25%
  - Accuracy of Classification : 86.67%

Experiment 3:

- Using the same set-up as Experiment 2, we increased the number of relative constraints between the data points.
- After iteratively increasing the number of constraints from 2 to 5, the following results were obtained for the mentioned values of dimensionalities
- Dimensionality : 5
  - Accuracy of Clustering : 69.5%
  - Accuracy of Classification : 93.3%
- Dimensionality : 10
  - Accuracy of Clustering : 82.0%
  - Accuracy of Classification : 93.3%

CONCLUSIONS

- KBTL model is able to perform classification of images in Target Domain with good accuracy. Hence, apt for Transfer Learning.
- Due to curse of high dimensionality, the distance between data points is not accurate, hence we use PCA to reduce dimensionality and then apply SKLR. Reducing dimensionality enhances the clustering capability.
- The accuracy of clustering using SKLR approach decreases as we increase the number of constraints for each class combination due to tight clustering.

FUTURE WORK

- Firstly the KBTL Model can be extended to perform multi-label classification apart from Binary Classification. This would in turn increase the overall scope of Transfer Learning to a variety of object classification rather than just 2.
- Secondly, rather than pre-processing the data to obtain output labels, a model can be developed which could incorporate SKLR and KBTL in one single unit which would thus eliminate the computation complexity and effort.
- This model can be further extended to be used for unsupervised learning, i.e. when none of the data points in the training data set are labeled.

Thank You!