Draft

Kris Sankaran, Xinran Miao
ksankaran@wisc.edu, xmiao27@wisc.edu

1 Introduction

1.1 Motivation

1. The lack of well-labelled data has been an obstacle of deep learning problems in the remote sensing context, as well as other scenarios. As a way of borrowing extra knowledge from out-of-distribution data, transfer learning shows great performance in practice, but its mysterious mechanism makes it difficult to choose a suitable dataset to pre-train a model on, especially under a computational budget.

2. To make the transfer learning more efficient, a natural thought is to make the source and target tasks as similar as possible. Then questions become (1) how to measure the similarity between the target and different source data, and (2) to what extent can this similarity between datasets help. For the former, choices of information that differentiates data include auxiliary information like geographical locations for earth observations, and predictors like VGG features with reduced dimensions. For the latter, there is an exploration-exploitation trade-off between exploring unseen source data and exploiting the most useful data at hand.

2 Literature Review

2.1 Thompson Sampling

In the multi-armed bandit problems, a player in front of a set of slot machines (also called arms) need to decide which machine to play at each time, with the goal of maximizing cumulative rewards after a sequence of choices. The inherent exploration - exploitation dilemma is addressed in Thompson Sampling by selecting the arm optimally after sampling parameters from posterior distributions ([1]). This can be formulated to the sample selection problem where all training data can be clustered into multiple arms and then selected sequentially ([2, 3]).

2.2 Transfer Learning

3 Models and Algorithms

3.1 Thompson Sampling in the Beta-Bernoulli Bandit Model

3.1.1 Multi-armed Bandit Problems

A multi-armed bandit consists of n unknown distributions, each of which is associated rewards with unknown mean $\theta_i,\ i = 1, \cdots, n$. The player chooses arm a_t and observes reward $\hat{r}_t$ at time t in T rounds, with the purpose of maximizing the cumulative reward

$$R_T = T\theta^* - \sum_{i=1}^{T} \hat{r}_t, \tag{1}$$

where $\theta^* = \max_{i \in \{1,2,\cdots,n\}} \theta_i$.

3.1.2 Thompson Sampling

Thompson sampling addresses the trade-off between exploiting what is known and exploring new information that may make an improvement by posterior sampling. Let $\mathcal{D}_t = \{(a_k, \hat{r}_k)\}_{k=1}^{t}$ be the observations at time t. Suppose the likelihood of rewards at time t with arm i can be modeled as $P(r|\mathcal{D}_t, \theta_i)$, then its posterior can be obtained since $P(\theta_i|\mathcal{D}) \propto P(\theta_i)P(\mathcal{D}|\theta_i)$. At each round, $\hat{\theta}_i$'s are sampled from posteriors and the action that maximizes the expected reward, i.e.,

$$a_t = \max_a \mathbb{E}[r|\hat{\theta}_a]. \tag{2}$$

For 0-1 rewards with Bernoulli likelihoods, if we assume Beta priors for θ_i's with parameters α_i's and β_i's, then the posterior for the arm I_t at time t can be updated to $\alpha_{I_t} + \mathbb{I}[\hat{r}_t = 1]$ and $\beta_{I_t} + \mathbb{I}[\hat{r}_t = 0]$. At time t, the choice of arm $a_t \in \{1, 2, \cdots, n\}$ is obtained by sampling θ_i from posteriors and computing the maximal expected rewards, i.e., $a_t = \arg\max_{i=1}^{n} \hat{\pi}_i$ where $\hat{\pi}_i$ is sampled from $Beta(\alpha_i, \beta_i)$.

3.1.3 Sample Selection Using Thompson Sampling

Suppose we have source dataset $\mathcal{S} = \{x_j, y_j, z_j\}_{j=1}^{N_S}$, target training data $\mathcal{T}_1 = \{x_j, y_j, z_j\}_{j=1}^{N_{T1}}$, and target validation data $\mathcal{T}_2 = \{x_j, y_j z_j\}_{j=1}^{N_{T2}}$, where $x,\ y$, and z indicate predictors, labels, and features or auxiliary information, respectively, and y_j's in $\mathcal{T}_2$ are suppposed to be unknown in real problems. Suppose $N_{T2} \gg N_{T1}$ and $N_S \gg N_{T1}$, then the lack of labelled target data requires the model to be pre-trained on the source data, and the computational budget requires selecting a subset of source. To formulate this sample selection problem in the multi-armed bandit setting, we first cluster $\mathcal{S}$ into n arms $\mathcal{S}_1, \cdots, \mathcal{S}_n$ according to z. Starting with an empty training set and a random assigned model, we add samples from cluster a_t to the training set at time t. (In the simulation experiments, I started with a model trained with some randomly selected data selected from source. Need to change either the words here, or the experiment.) After updating the model and evaluating the reward

$$r_t = \mathbb{I}[\text{the model accuracy improves on the target validation set}].$$

we can update posteriors according to Thompson Sampling in the Beta-Bernoulli bandit theory. Overtime, the model improves incrementally. Table 1 shows the algorithm where convergence is obtained when the change in rewards is less than 0.001. After that, the model is further trained on the labelled target set $\mathcal{T}_1$ before making predictions for unlabelled target set $\mathcal{T}_2$.

Sample selection: Thompson Sampling with Beta-Bernoulli Bandit
given $\alpha_i = \alpha, \beta_i = \beta, i \in \{1, \cdots, n\}$
Initialize model m randomly.
evaluate m on the target set and obtain accuracy a.
Let $a_1, \cdots, a_n = a$.
while $\lvert a_t - a_{t-1} \rvert > 0.001$ **do**
$a_t = \arg\max_i \hat{\pi}_i$, where $\hat{\pi}_i \sim Beta(\alpha_i, \beta_i)$
randomly select a batch of samples h_t from C_{a_t}.
update the pre-trained model m with h_t.
compute the prediction accuracy a_t of updated m on target.
if $a_t > a_{t-1}$:
$\alpha_{a_t} = \alpha_{a_t} + 1$
else
$\beta_{a_t} = \beta_{a_t} + 1$

Table 1: Sample selection: Thompson Sampling with Beta-Bernoulli Bandit

3.2 Dimension Reduction

4 Data

4.1 Simulation

We simulate two datasets with 2-dimensional 2,000 observations in two classes each (Figure 1) (the size can change if needed). Data are generated in the forms

$$x_2 = sin(x_1) + \log(|x| + 1), x_1 \in [-10, 10] \tag{3}$$

and

$$x_2 = 0.001 \times x_1^3 - 0.003 \times x_1^2 + 0.002 \times x_1 + 1, x_1 \in [-100, 100], \tag{4}$$

and then rescaled according to the mean and standard error. Normally distributed errors are added to all points, with mean zero and variance equal to the sample variance divided by 10. Points are assigned "wrong" labels with probability 0.01.

(The second simulated dataset fails this algorithm, and we may not include such failure.)

Each of the simulated dataset is clustered by x_1 into 4 non-overlapping groups, where the clusters are randomly assigned with probability 0.05 (Figure 2). One of the cluster is set as the target, while the rest are 3 clusters of the source (1555 observations). The target dataset is then randomly divided into two sets $\mathcal{T}_1$ and $\mathcal{T}_2$ with the same size (222 observations each). The former represents the target training dataset, where the training set is evaluated and the Thompson Sampling is updated accordingly; the latter represents the

target validation dataset, which is unseen in the sample selection procedure, and evaluates the final result. In both simulation datasets, cluster 1 and 2 are closer to the target compared with cluster 3, so we expect the algorithm to choose cluster 1 and 2 more frequently.

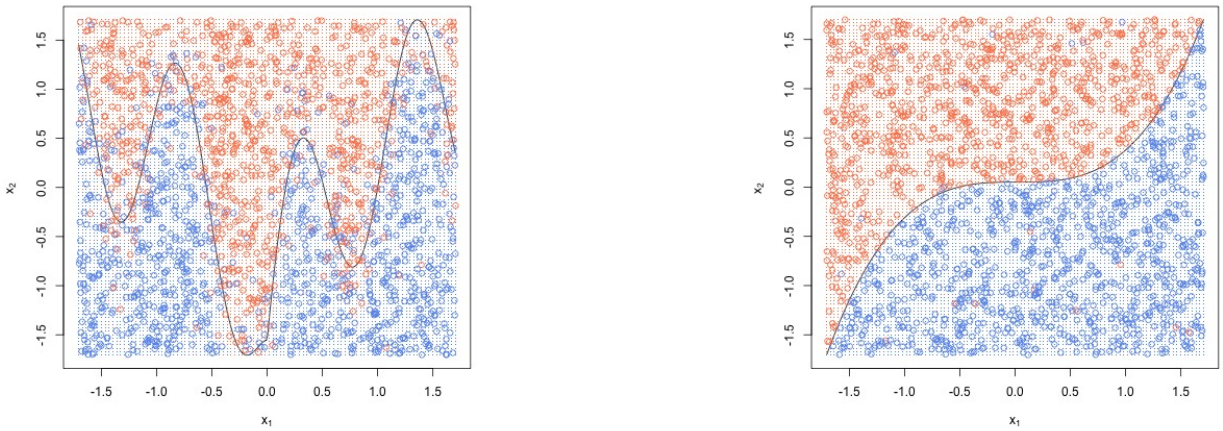

Figure 1: Simulated Dataset with the ideal decision boundary. In each panel, the coral and blue points represent samples from two classes, and the black line is the indicates the ideal decision boundary that separates classes regardless of errors.

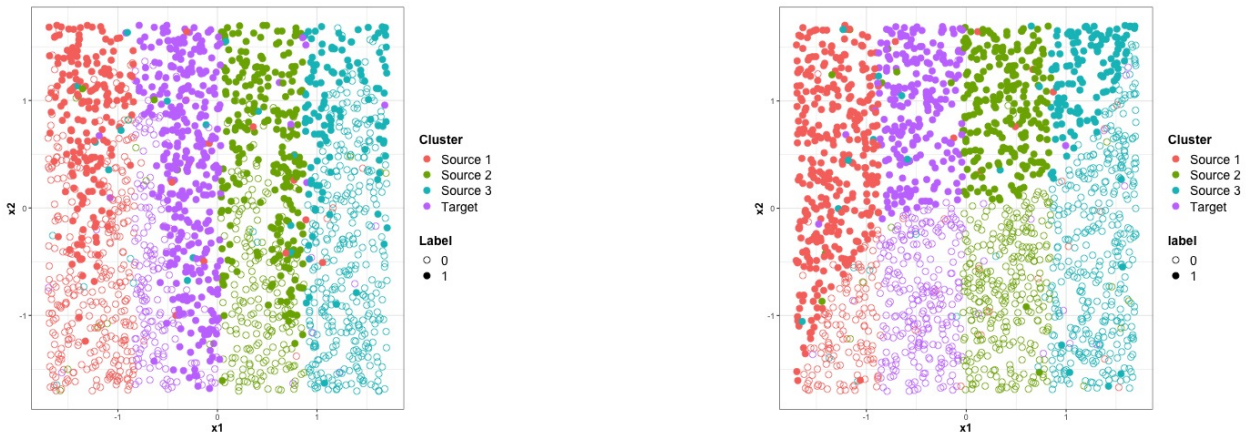

Figure 2: Simulated Dataset clustered according to x_1.

4.2 Deep Globe

5 Experiments

Some variations in the experiment design include:

1. Characteristics of the dataset, including data size, data diversity, the shape of decision boundaries, relationship between source and target.

2. The choice of model and its hyper parameters.

3. The choice of training dataset: whether or not use source; how to choose source.

4. Parameters of the incremental sample selection, including number of clusters, number of samples added in one round.

5.1 Simulation

We compare our sample selection algorithm with baseline models trained in the following ways under the same computational budget.

1. The model is trained on the labelled target data $\mathcal{T}_1$ only, without using the source data $\mathcal{S}$.

2. The model is trained sequentially on the randomly-selected samples from the source data $\mathcal{S}$ as well as the labelled target data $\mathcal{T}_1$.

3. The model is trained on the source data that are "nearest" to target as well as the labelled target data $\mathcal{T}_1$.

5.2 Performance Evaluation

1. The model accuracy on the target validation set under a certain computational budget.

2. The time until convergence.

3. The model robustness.

6 Results

6.1 Simulation

Figure 1 show the accuracy of the model trained on sequentially selected source dataset and evaluated on the target validation set. For both datasets, the algorithm tends to select cluster 1 &2, which is as expected. The model of the first simulated dataset improves overtime, while that of the second never performs better than the initial one after adding source samples. The failure is probably because the decision boundary for the second simulated dataset is also flat. From Figure 4, the decision boundary learned in the beginning separated the classes almost perfectly, which became worse after adding samples from the source.

Table 2 and 3 show the final model performance for two simulated datasets. Different from Figure 3, the models here are further trained on the target training set $\mathcal{T}_1$. (I haven't performed the baseline algorithm and randomly add samples from the source. But I expect this baseline will perform greater or equal to the proposed algorithm for the second simulated dataset, especially with respect to the computation budget.)

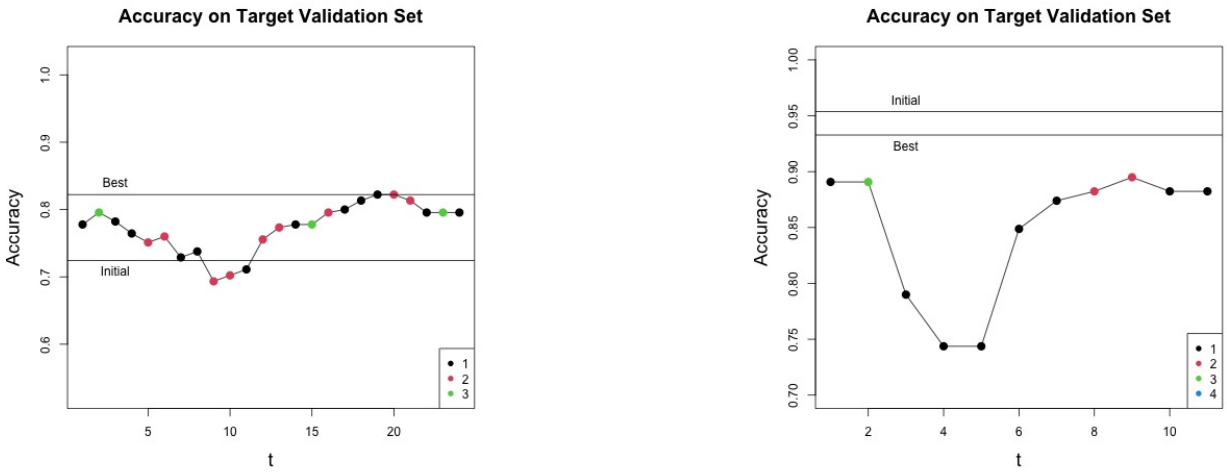

Figure 3: Accuracy on the target validation set for the model trained sequentially selected source dataset overtime for both simulated datasets. In each panel, the two horizontal lines annotated as "best" and "initial" represent the performances of the model trained using $\mathcal{T}_1 \bigcup \mathcal{S}$ and the initial model trained on 20 randomly selected points from $\mathcal{S}$ before the sample selection.

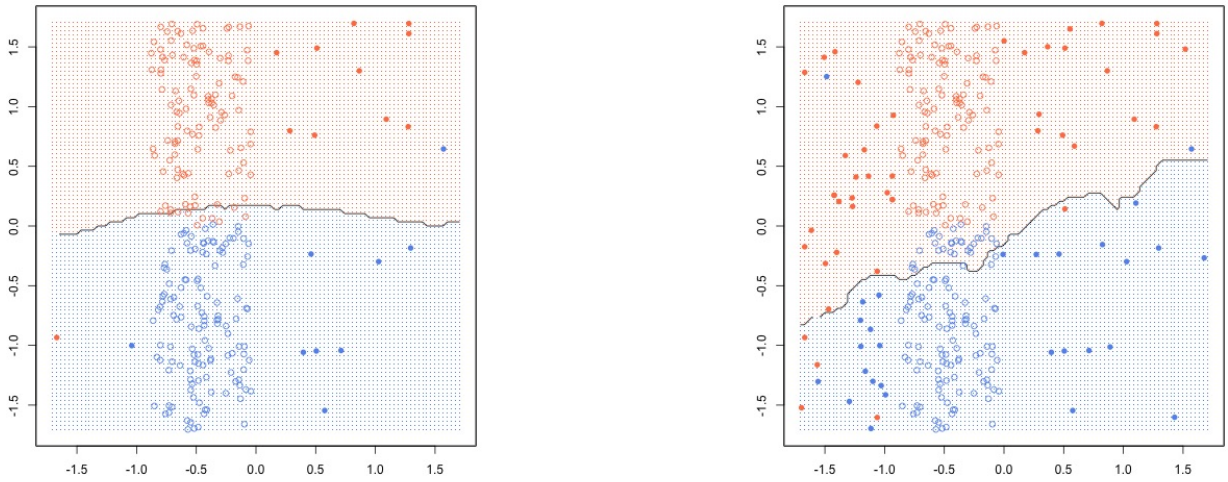

Figure 4: The decision boundaries of the second simulated dataset at the initial (left) and final (right) time points.

Training data	10-NN	another model
Target ($\mathcal{T}_1$) only	0.805	
Target ($\mathcal{T}_1$) & Source (S) selected by TS	0.866	
Target ($\mathcal{T}_1$) & Source (S) selected randomly		

Table 2: Model performance on simulated dataset 1 with different selection rules of training data.

Training data	10-NN	another model
Target ($\mathcal{T}_1$) only	0.983	
Target ($\mathcal{T}_1$) & Source (S) selected by TS	0.987	
Target ($\mathcal{T}_1$) & Source (S) selected randomly		

Table 3: Model performance on simulated dataset 2 with different selection rules of training data.

References

[1] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[2] Benjamín Gutiérrez, Loïc Peter, Tassilo Klein, and Christian Wachinger. A multi-armed bandit to smartly select a training set from big medical data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 38–45. Springer, 2017.

[3] Djallel Bouneffouf, Romain Laroche, Tanguy Urvoy, Raphael Féraud, and Robin Allesiardo. Contextual bandit for active learning: Active thompson sampling. In *International Conference on Neural Information Processing*, pages 405–412. Springer, 2014.