Online Transfer Learning with Heterogeneous Source

Author1
Address line

Author2
Address line

Author3
Address line

Abstract

AAAI creates proceedings, working notes, and technical reports directly from electronic source furnished by the authors. To ensure that all papers in the publication have a uniform appearance, authors must adhere to the following instructions.

Introduction

Related Work

Online Heterogeneous Transfer

Experimental Results

In this section, we empirically evaluate the performance of proposed online heterogeneous transfer learning algorithms and classic online Passive-Aggressive algorithms (PA). Encouraging results demonstrate that the proposed algorithms outperform these algorithms.

Dataset

Our experiments are conducted for image classification by leveraging information from text data. We use NUS-WIDE dataset to generate learning tasks. The NUS-WIDE dataset is extracted from Flickr. It includes 269,648 images and the associated tags from Flickr, with a total number of 5,018 unique tags. An image instance is represented by a feature vector based on SIFT descriptions, and a text instance is represented by a feature vector based on tags. There are 81 ground-truth class labels in the dataset. We randomly selected 10 classes (bird, boat, car, flower, food, rock, sun, toy, tree) and built $C_{10}^2 = 45$ binary classification tasks.

We refer the images as data in the target domain, and the tags as the text data in the heterogeneous source domain. Each binary classification task has 500 image instances in the target domain, 1,200 text instances in the heterogeneous source domain, and 1,500 co-occurred image-text pairs. In order to obtain stable results, we draw 100 times of random permutation of the image instances in the target domain and evaluate the performance of learning algorithms based on average rate of mistakes.

Baseline Methods

We compare the proposed methods with Passive-Aggressive (PA) online learning algorithms. PA algorithm proposed by Crammer et al. does not exploit knowledge from the source domain. It deals with the traditional online learning problem in the target domain.

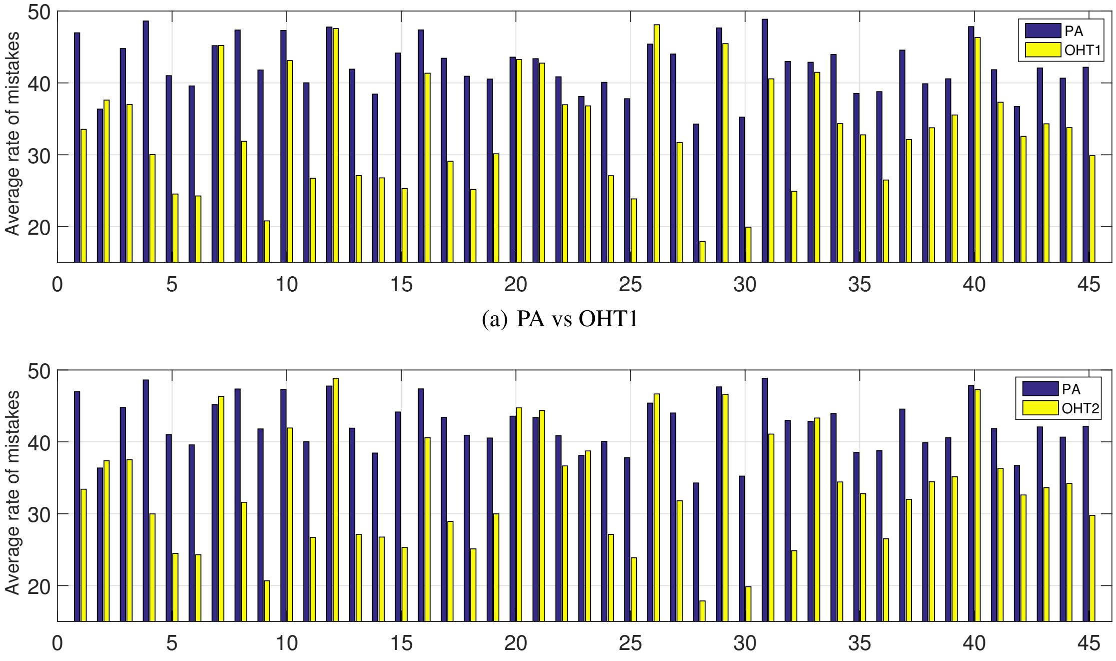
For fair comparison and simplicity, we adopt Gaussian kernel function on all the algorithms and tasks. The kernel parameter $\sigma = 8$ for the target domain. The regularization parameter $C = 5$, $\beta = \frac{\sqrt{T}}{\sqrt{T}+\sqrt{2\ln 4}}$ for OHT2 algorithm. In addition, we set the number of nearest neighbors to be considered $K = 10$. Sensitivity of parameters will be examined in subsequent sections.

Results and Discussion

Figure 1 summarizes the mistake rates of all 45 binary classification tasks. The x-axis of the figure refers to the 45 tasks. We see that on most tasks, PA has the very high mistake rate, which prove the dificulty of image classification task without any auxiliary source information and the necessity of knowledge transfer. The observation that our proposed OHT algorithms in general outperform PA validates the effectivity of heterogeneous transfer learning.

Figure 2 illustrates the dynamic process of several sample online learning tasks, respectively. We observe that on some tasks(e.g., 35, 39 and 43), the online mistake rates of all three algorithms decrease during the period, and OHT algorithms always achieve better performance than PA. Furthermore, on some tasks(e.g., 18, 30 and 36), OHT algorithms are able to obtain a good performance at the beginning stage and remain stable in the future. These observations verifies that the OHT algorithms indeed transfer useful knowledge from the heterogeneous source domain to the target domain.

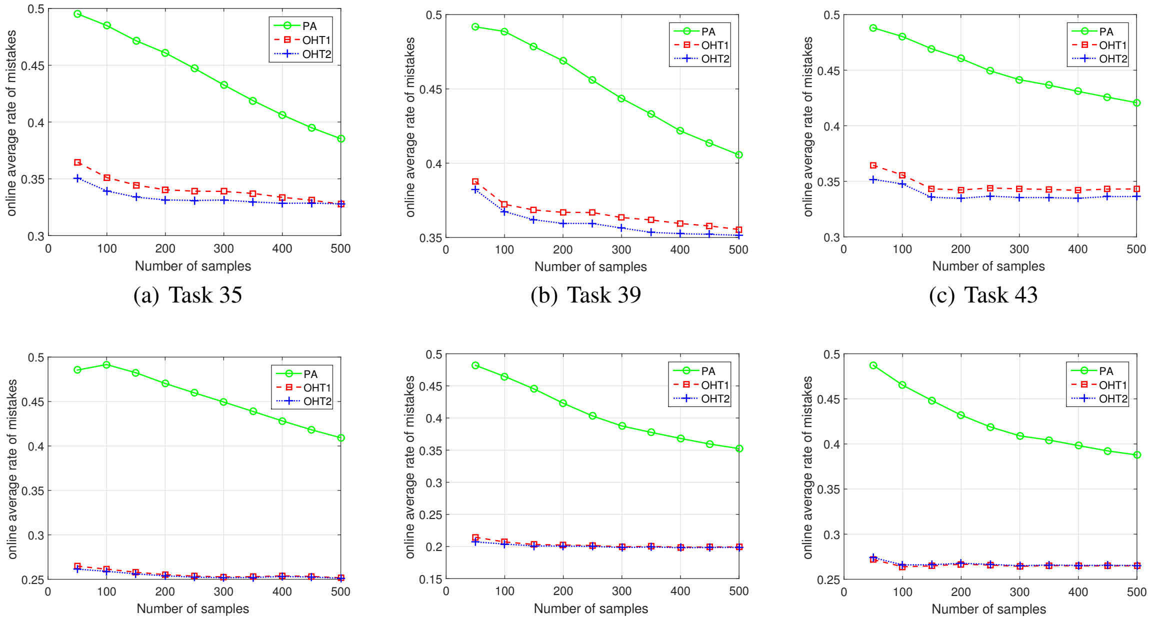
We also analyze the performance difference between PA and two OHT algorithms separately. Statistical significance against PA was assessed by paired t-test at 0.05 level. For each task, a win (or loss) is counted when OHT algorithm is significatly better (or worse) than PA algorithm over 100 trials. Otherwise, a tie is recorded. The win/tie/loss results is 42/1/2 for competition between OHT1 and PA, and 37/0/8 for competition between OHT2 and PA. This result validates that our OHT algorithms is statistically better than PA algorithm, and OHT1 is more stable than OHT2.



(a) PA vs OHT1

(b) PA vs OHT2

Figure 1: Average rate of mistakes on all 45 tasks



(a) Task 35

(b) Task 39

(c) Task 43

(d) Task 18

(e) Task 30

(f) Task 36

Figure 2: Online average rate of mistakes on example tasks

Besides, we utilize Cohen's d value to measure the improvement of our algorithms. Generally, $d > 0.8$ indicates a large promotion, and $0.2 < d < 0.8$ indicates a middle promotion. In out experiments, OHT1 algorithm achieves large improvement on 39 tasks and middle improvement on 3 tasks. For OHT2 algorithms, the numbers are 35 and 2. Actually, d is much larger than 0.8 on most of tasks.

Parameters and Running time

Parameters Experiments in paper about online transfer learning illustrated that the performance of online transfer learning algorithms is generally insensitive to the parameter C and β. Consequently, we only investigate how different values of parameter K affect the classification accuracy of the algorithms.

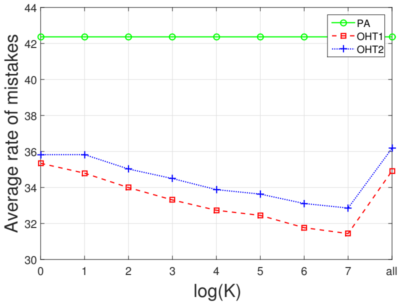


Figure 3: Results under varying values of K

Average mistake rates over all 45 tasks under varying values of K. Figure 3 shows the average mistake rates with varied values of parameter K over all 45 tasks. We observe that the performance of the proposed methods consistently outperform PA algorithm, which indicates that K nearest neighbors do provide valuable advice for the classification task.

Running time All of the algorithms were implemented in Matlab, and all experiments were run in a Linux machine with 3.2 GHz CPU and 3.8 GB memory. The main reason of more running time for OHT algorithms is probably the searching process for the nearest neighbors. We can simply make use of all instances in the heterogeneous source domain to obtain generally comparable running time to PA, and at the same time, achieve better performance than PA.

Conclusion