1 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.

1.1 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 mean and standard deviation of mistake rates.

1.2 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 in 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 OHT-II algorithm. In addition, we set the number of nearest neighbors to be considered $K = 10$. Sensitivity of parameters will be examined in subsequent sections.

1.3 Results and Discussion

Table 1: Results of all 45 tasks

Task	PA	OHT1	OHT2
1	46.9680 ± 2.0210	33.5300 ± 0.3860	**33.4120 ± 0.2709**
2	**36.3640 ± 1.7050**	37.6100 ± 1.5417	37.3700 ± 1.8823
3	44.7760 ± 1.8058	**37.0000 ± 0.4989**	37.5260 ± 0.4165
4	48.6120 ± 2.1255	30.0240 ± 0.3207	**29.9780 ± 0.2665**
5	41.0060 ± 1.8923	24.5420 ± 0.2818	**24.4820 ± 0.1452**
6	39.5800 ± 1.8349	**24.2720 ± 0.2336**	24.3000 ± 0.1407
7	**45.1820 ± 1.9846**	45.2080 ± 1.9012	46.3180 ± 2.3319
8	47.3560 ± 2.0675	31.8680 ± 0.3763	**31.5920 ± 0.2770**
9	41.8000 ± 1.9664	20.8040 ± 0.2445	**20.6780 ± 0.1299**
10	47.2940 ± 2.0270	43.1020 ± 1.7316	**41.9280 ± 0.5276**
11	40.0120 ± 1.7352	26.7260 ± 0.2766	**26.7120 ± 0.1591**
12	47.7680 ± 2.2787	**47.5620 ± 2.2548**	48.8500 ± 2.2433
13	41.9080 ± 1.6537	**27.1000 ± 0.2470**	27.1340 ± 0.1730
14	38.4440 ± 1.8358	26.7880 ± 0.2571	**26.7540 ± 0.2027**
15	44.1560 ± 1.9865	**25.3060 ± 0.2264**	25.3200 ± 0.1729
16	47.3700 ± 2.3904	41.3460 ± 1.5358	**40.5680 ± 0.4087**
17	43.4260 ± 1.7323	29.0940 ± 0.3272	**28.9240 ± 0.1747**
18	40.9240 ± 1.5931	25.1700 ± 0.2385	**25.1140 ± 0.1954**
19	40.5360 ± 1.6404	30.1420 ± 0.4425	**29.9840 ± 0.2881**
20	43.5760 ± 1.9861	**43.2420 ± 1.7934**	44.7280 ± 2.2167
21	43.3680 ± 1.8529	**42.7500 ± 1.7351**	44.3600 ± 1.7702
22	40.8480 ± 1.6315	36.9640 ± 2.0682	**36.6520 ± 0.3070**
23	38.1040 ± 2.0535	**36.8040 ± 1.3680**	38.7380 ± 1.6810
24	40.0740 ± 1.7843	**27.0860 ± 0.2640**	27.1220 ± 0.1528
25	37.7980 ± 1.7306	**23.8620 ± 0.2044**	23.8840 ± 0.1308
26	**45.3960 ± 1.8828**	48.0940 ± 0.6795	46.6660 ± 2.0584
27	44.0160 ± 1.7706	**31.7160 ± 0.3401**	31.8080 ± 0.3203
28	34.2800 ± 1.6704	17.9340 ± 0.2413	**17.8720 ± 0.1379**
29	47.6480 ± 2.3180	**45.4580 ± 0.5772**	46.6200 ± 1.2092
30	35.2360 ± 1.6471	19.9200 ± 0.2108	**19.8540 ± 0.1359**
31	48.8520 ± 2.3983	**40.5660 ± 0.5416**	41.0820 ± 0.5235
32	42.9840 ± 2.1348	24.9140 ± 0.2400	**24.8640 ± 0.1202**
33	42.8700 ± 1.7923	**41.4680 ± 1.6964**	43.3180 ± 1.1824
34	43.9460 ± 2.0912	**34.3320 ± 0.5644**	34.4200 ± 0.2719
35	38.5280 ± 1.6079	**32.7740 ± 1.6567**	32.7980 ± 0.2060
36	38.7800 ± 1.8058	**26.4900 ± 0.3183**	26.5300 ± 0.1667
37	44.5600 ± 1.8790	32.1100 ± 0.3380	**32.0020 ± 0.2486**
38	39.8800 ± 1.6143	**33.7560 ± 1.1056**	34.4460 ± 0.2966
39	40.5700 ± 1.5981	35.5400 ± 1.5259	**35.1480 ± 0.4279**
40	47.8280 ± 1.9353	**46.3060 ± 0.5261**	47.2580 ± 0.9894
41	41.8340 ± 2.0317	37.3080 ± 2.3227	**36.3240 ± 0.3491**
42	36.7060 ± 1.3907	**32.5580 ± 0.2189**	32.6100 ± 0.2946
43	42.0780 ± 1.6208	34.3080 ± 1.3717	**33.6300 ± 0.2890**
44	40.6600 ± 1.5188	**33.7740 ± 0.7395**	34.2240 ± 0.2731
45	42.1760 ± 2.0556	29.8760 ± 0.3514	**29.7740 ± 0.2721**

Table summarizes the mistake rates of all 45 tasks. We see that in most tasks, PA has the very high mistake rate, which prove the dificulty of image classification task without any auxiliary source information. The observation that our proposed OHT algorithms generally outperform PA validates the effectivity of heterogeneous transfer.

Figure illustrates the dynamic process of several online learning tasks, respectively. We observe that in some tasks(e.g., 35, 39 and 41), the mistake rates of all three algorithms decrease during the period, and OHT algorithms always achieve better performance than PA. Furthermore, in some tasks(e.g., 8, 11 and 24), 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.

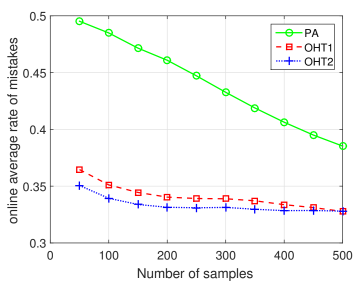
We also analyze the performance difference between PA and two OHT algorithms. Statistical significance against PA was assessed by paired t-test at 0.05 level. For each task, a win (or loss) is counted when OHT algorithm is significatly better (or worse) than PA algorithm over 100 trials. Otherwise, a tie is recorded. The win/tie/loss results is 30/1/1 for competition between OHT1 and PA, and 30/1/1 for competition between OHT2 and PA. This result validates that our OHT algorithms is statistically better than PA algorithm.

Besides, we make use of Cohen's d value to measure the improvement of our algorithms. $d >= 0.8$ generally indicates a large promotion. OHT1 algorithm achieves large improvement in 30 tasks and middle improvement in 3 tasks. For OHT2 algorithms, the numbers are 32 and 2.

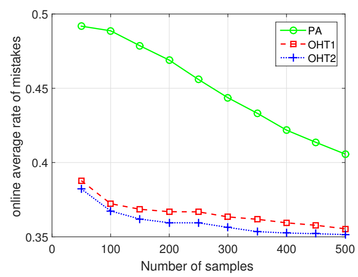
1.4 Parameters and Running time

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. Therefor, we only investigate how different values of parameter K affect the classification accuracy of the algorithms. We select a number of tasks randomly to evaluate the parameter sensitivity. Table shows the performance of the proposed algorithms with varied values of parameter K in task 1. We observe that the performance of the algorithms is stable. Similar obvervation are showed in other tasks. In consideration of that we employ the weighted K nearest neighbors strategy, it is easy to understand. The larger distance a instance in the heterogeneous source domain has, the less impact the instance makes.

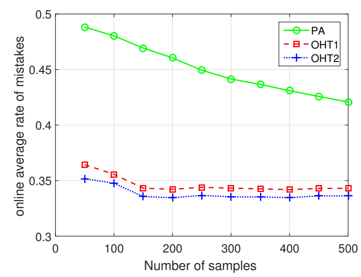
Running time Table shows the mean and standard deviation of running time of different algorithms in several randomly selected tasks. All of the algorithms were implemented in Matlab, and all experiments were run in a Linux machine with 3.2 GHz CPU and 3.8 GB memory. From the table, we can see that PA without exploiting knowledge from the source domain is the most efficient. Two OHT algorithms are less efficient. The main reason of more running time is the searching process for the nearest neighbors. Because of the insensitivity of parameter K, we can simply make use of all instances in the heterogeneous source domain to decrease the running time and obtain comparable performance.



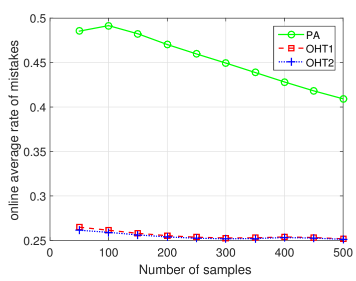
(a) Task 35



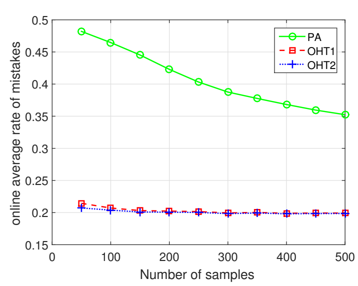
(b) Task 39



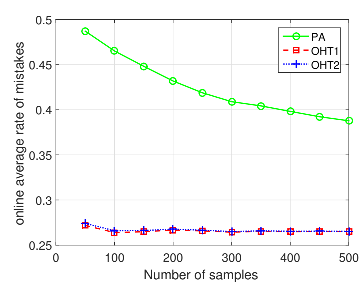
(c) Task 43



(d) Task 18



(e) Task 30



(f) Task 36

Figure 1: Online mistake rates