Online Transfer with Heterogeneous Source

Yan Yuguang

South China University of Technology

May 11, 2015

Problem

Online Multi-transfer Learning

The goal is to learn some prediction function $f(\mathbf{x}_t)$ on a target domain in an online fashion from a sequence of instances $\{(\mathbf{x}_t, y_t | t = 1, 2, \cdots, T\}$ in data space $\mathcal{X} \times \mathcal{Y}$.

- Homogeneous source domain:

$$\mathcal{X} = \mathcal{X}^k, \mathcal{Y} = \mathcal{Y}^k$$

- Heterogeneous source domain:

$$\mathcal{X} \cap \mathcal{X}^k = \emptyset, \mathcal{Y} = \mathcal{Y}^k$$

  co-occurrence information (?, ?)

OTL: A Framework of Online Transfer Learning (?, ?)

Online Transfer Learning (?, ?)

Related Work
OTL on homogeneous domain

$\mathcal{X}_s = \mathcal{X}$

Resemble learning strategy: HomOTL1.png

Online Transfer Learning (?, ?)

Related Work

OTL on homogeneous domain

HomOTL2.png

Online Transfer Learning (?, ?)

Related Work
OTL on heterogeneous domain

$\mathcal{X}_s \subset \mathcal{X}$

Multi-view approach: HetOTL.png

Methods

OHT

yanyg

$$\hat{y}_t = \operatorname{sign}\left(\sum_{k=1}^{N} \alpha_t^k \Pi\left(z_t^k\right) + \alpha_t \Pi\left(z_t\right) - \frac{1}{2} \right)$$

$$\hat{y}_t = \operatorname{sign}\left(\sum_{k=1}^{N} \alpha_t^k \operatorname{sign}\left(z_t^k\right) + \alpha_t \operatorname{sign}\left(z_t\right) \right)$$

target domain & homogeneous source domain:

$$z_t = \mathbf{w}_t^\top \mathbf{x}_t, z_t^k = \mathbf{v}^{k\top} \mathbf{x}_t$$

heterogeneous source domain:

$$z_t^k = \sum_{\mathbf{x}^k \in D^K} sim(\mathbf{x}^k, \mathbf{x}_t) y_i^k$$

where $sim(\mathbf{x}^k, \mathbf{x}_t)$ is calculated by co-occurrence information (?, ?), and D^K is the set of K neast neighbors.

Mistake bound

a3.jpg

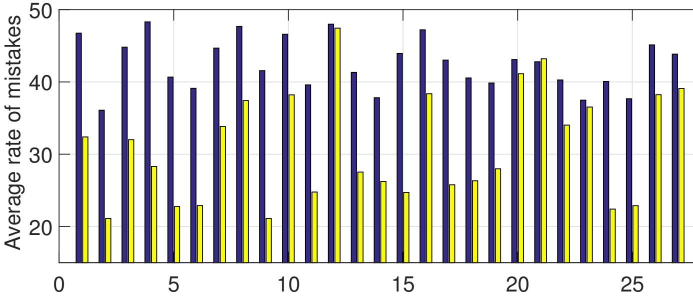
a1.jpg

NUS-WIDE dataset

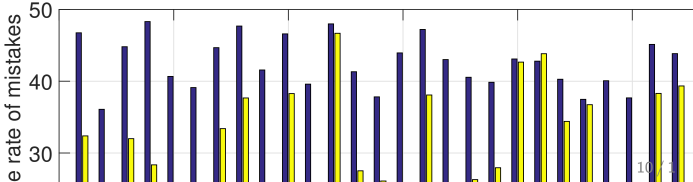
- Target domain: Image
- Heterogeneous source: Text
- Co-occurrence data: co-occurred image-tag pairs

- Passive-Aggressive algorithms
  Do not exploit knowledge from the source domain
- Kernel function
  Gaussian Kernel
- Number of nearest neighbors
  $K = 100$

[!htb] [PA-II vs. OHT1-II]

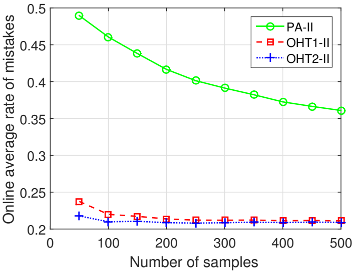


[PA-II vs. OHT2-II]



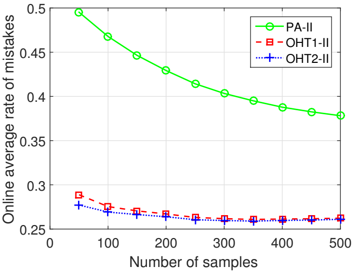
Experiments

Results



[!htb] [Task 2]

[Task 14]



[Task 36]

Paired t-test ($\alpha = 0.01$)

- OHT1 vs. PA: 44/0/1
- OHT2 vs. PA: 42/2/1

Cohen's d value ($d > 0.8$: large promotion, $0.2 < d < 0.8$: middle promotion)

- OHT1: 41/3
- OHT2: 40/3

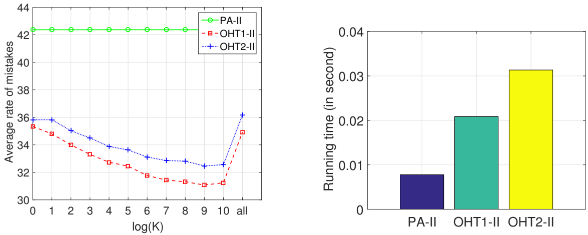


Figure : (a) The average rate of mistakes under varying values of K. (b) The average running time of different algorithms when all instances in heterogeneous source are considered.

THANK YOU
FOR
YOUR ATTENTION!