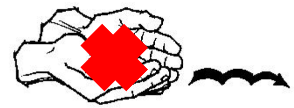
Exploiting Privileged Information from Web Data for Visual Recognition

Li Niu presents

Learning from Web is increasingly popular due to freely available web data. However, this problem is challenging due to following main issues.

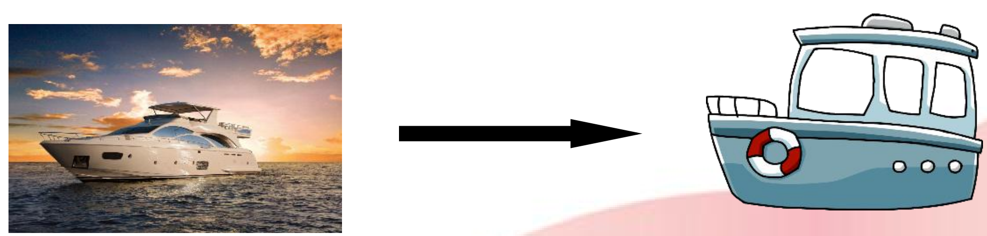
- Label noise: Query "boat"



- Privileged information

Azimut 95 Luxury Yacht at the Miami International Boat Show 2012 Azimut-Benetti Yachts sees 20 per cent gain in new luxury yacht sales

- Domain distribution mismatch



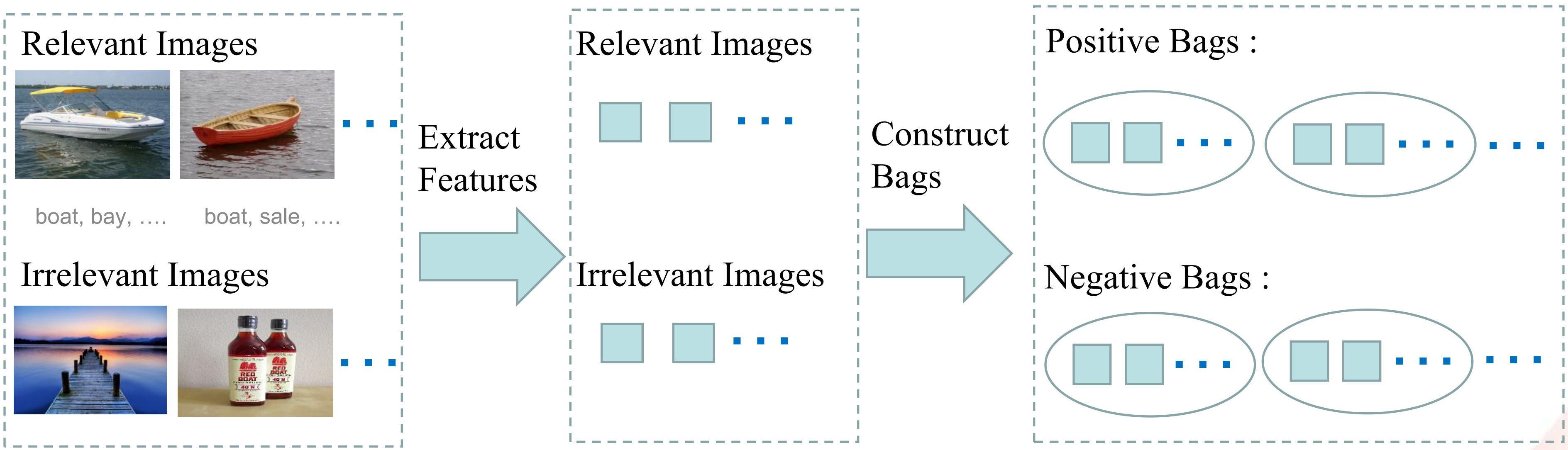
source domain

target domain

NANYANG TECHNOLOGICAL UNIVERSITY

Background: Multi-instance Learning

Multi-instance learning (MIL) method treat each cluster as a “bag” and the images in each bag as “instances”

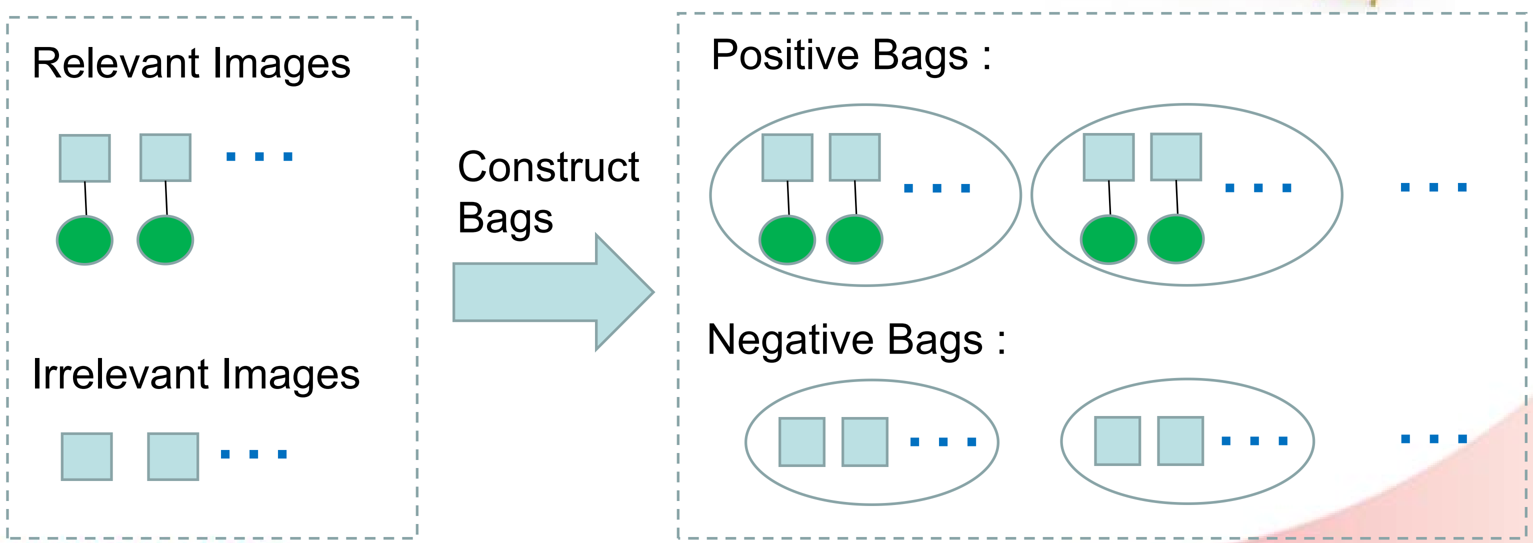


NANYANG TECHNOLOGICAL UNIVERSITY

Background: Learning Using Privileged Information [1]

Online images are generally associated with textual descriptions which are not available for consumer photos.

Training: attending classes in the classroom
Testing: taking an exam
Privileged Information: teacher's instruction



NANYANG TECHNOLOGICAL UNIVERSITY

Background: Learning Using Privileged Information [1]

SVM+
(Primal Form)

$$\min_{\tilde{\mathbf{w}}, \tilde{b}, \mathbf{w}, b} \quad \frac{1}{2}\left(\|\mathbf{w}\|^2 + \gamma\|\tilde{\mathbf{w}}\|^2\right) + C\sum_{i=1}^{n} \xi(\tilde{\mathbf{x}}_i),$$

$$\text{s.t.} \quad y_i(\mathbf{w}'\phi(\mathbf{x}_i) + b) \geq 1 - \xi(\tilde{\mathbf{x}}_i), \quad \xi(\tilde{\mathbf{x}}_i) \geq 0, \quad \forall i,$$

$$\xi(\tilde{\mathbf{x}}_i) = \tilde{\mathbf{w}}'\tilde{\phi}(\tilde{\mathbf{x}}_i) + \tilde{b}$$

oracle function

$$\min_{\mathbf{w}, b} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \xi_i$$

$$\text{s.t.} \quad y_i(\mathbf{w}'\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

primal form of SVM

x_i : visual feature

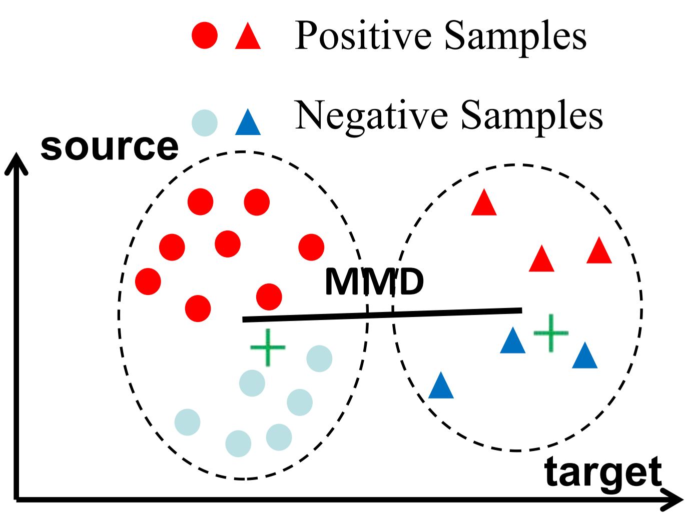
$\tilde{x}_i$: textual feature

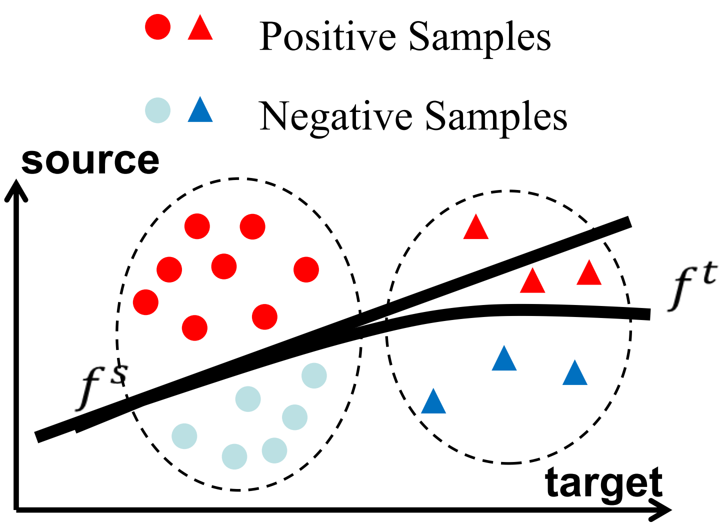
NANYANG TECHNOLOGICAL UNIVERSITY

[1] Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged infromatin. Neural Networks **22** (2009) 544–557

Background: Domain Adaptation

minimize the Maximum Mean Discrepancy (MMD) [2] between source domain and target domain by reweighting training samples



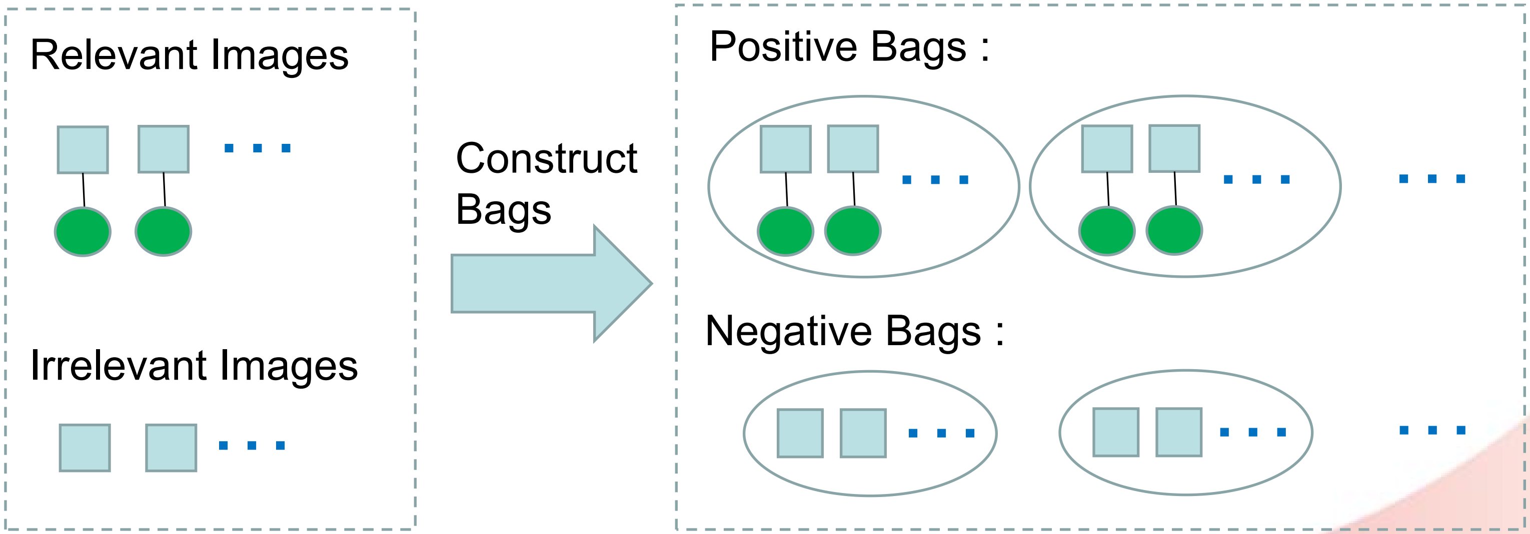


NANYANG TECHNOLOGICAL UNIVERSITY

[2] Huang, J., Smola, A., Gretton, A., Borgwardt, K., Scholkopf, B.: Correcting sample selection bias by unlabeled data. In: NIPS. (2007)

We unified MIL, LUPI and DA into one formulation, which can handle label noise, utilize privileged information and tackle with domain distribution mismatch at the same time.

- Label Noise → Multi-instance Learning
- Privileged Information → Learning using Privileged Information
- Domain Distribution Mismatch → Domain Adaptation



NANYANG TECHNOLOGICAL UNIVERSITY

Bag-level MIL Method: sMIL-PI (Primal Form)

averaged bag feature

$$\min_{\mathbf{w},b,\tilde{\mathbf{w}},\tilde{b},\boldsymbol{\eta}} \quad \frac{1}{2}\left(\|\mathbf{w}\|^2+\gamma\|\tilde{\mathbf{w}}\|^2\right)+C_1\sum_{j=1}^{L^+}\underline{\xi(\tilde{\mathbf{z}}_j)}+C_2\sum_{j=L^++1}^{m}\eta_j,$$

$$\text{s.t.} \quad \mathbf{w}'\underline{\mathbf{z}_j}+b\geq p_j-\underline{\xi(\tilde{\mathbf{z}}_j)}, \quad \xi(\tilde{\mathbf{z}}_j)\geq 0, \quad \forall j=1,\ldots,L^+,$$

$$\mathbf{w}'\underline{\mathbf{z}_j}+b\leq -1+\eta_j, \quad \eta_j\geq 0, \quad \forall j=L^++1,\ldots,m,$$

$$\underline{\xi(\tilde{\mathbf{z}}_j)}=\tilde{\mathbf{w}}'\underline{\tilde{\mathbf{z}}_j}+\tilde{b}$$

$$p_j=\sigma-(1-\sigma)=2\sigma-1$$

margin for sMIL

positive ratio

$$\mathbf{z}_j=\frac{1}{|\mathcal{B}_j|}\sum_{i\in\mathcal{I}_j}\phi(\mathbf{x}_i)$$

bag size

$$\tilde{\mathbf{z}}_j=\frac{1}{|\mathcal{B}_j|}\sum_{i\in\mathcal{I}_j}\tilde{\phi}(\tilde{\mathbf{x}}_i)$$

NANYANG TECHNOLOGICAL UNIVERSITY

Bag-level MIL Method: sMIL-PI (Dual Form)

Kernel based on visual feature

Kernel based on textual feature

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad -\mathbf{p}'\boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}'(\mathbf{K}\circ\mathbf{y}\mathbf{y}')\boldsymbol{\alpha} + \frac{1}{2\gamma}(\hat{\boldsymbol{\alpha}} + \boldsymbol{\beta} - C_1\mathbf{1})'\tilde{\mathbf{K}}(\hat{\boldsymbol{\alpha}} + \boldsymbol{\beta} - C_1\mathbf{1}),$$

$$\text{s.t.} \quad \boldsymbol{\alpha}'\mathbf{y} = 0, \quad \mathbf{1}'(\hat{\boldsymbol{\alpha}} + \boldsymbol{\beta} - C_1\mathbf{1}) = 0, \quad \bar{\boldsymbol{\alpha}} \leq C_2\mathbf{1}, \quad \boldsymbol{\alpha} \geq \mathbf{0}, \quad \boldsymbol{\beta} \geq \mathbf{0},$$

$$\boldsymbol{\alpha} = [\hat{\boldsymbol{\alpha}}', \bar{\boldsymbol{\alpha}}']'$$

$$\mathbf{p} = [p_1, \ldots, p_{L^+}, \mathbf{1}'_{m-L^+}]'$$

$$\mathbf{y} = [\mathbf{1}'_{L^+}, -\mathbf{1}'_{m-L^+}]'$$

Positive bags

Negative bags

NANYANG TECHNOLOGICAL UNIVERSITY

Domain Adaptation Method: sMIL-PI-DA (Dual Form)

MMD

Dual form of sMIL-PI

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta}} \quad \boxed{H(\boldsymbol{\alpha},\boldsymbol{\beta})} + \frac{\mu}{2}\boxed{\left\|\frac{1}{m}\sum_{i=1}^{m}\theta_i\mathbf{z}_i^s - \frac{1}{n_t}\sum_{i=1}^{n_t}\mathbf{z}_i^t\right\|^2}$$

$$\text{s.t.} \quad \boldsymbol{\alpha}'\mathbf{y} = 0, \quad \mathbf{1}'(\hat{\boldsymbol{\alpha}} + \boldsymbol{\beta} - C_1\mathbf{1}) = 0, \quad \bar{\boldsymbol{\alpha}} \leq C_2\mathbf{1}, \quad \boldsymbol{\beta} \geq \mathbf{0}$$

$$\underline{\mathbf{0} \leq \boldsymbol{\alpha} \leq C_3\boldsymbol{\theta}}, \quad \mathbf{1}'\boldsymbol{\theta} = m$$

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Retrieval

- Dataset
  NUS-WIDE: 269,648 images, 81 categories
  WebQuery: 71,478 images, 353 queries

- Experimental setting

  - NUS-WIDE:
    1) entire dataset is split into 60% training set and 40% test set
    2) construct 25 positive bags and 25 negative bags with bagsize 15

  - WebQuery:
    1) entire dataset is split into 60% training set and 40% test set
    2) discard queries with fewer than 100 training images
    3) remaining 19,665 training images, 13,114 test images, 163 queries
    4) set bagsize as 5, construct positive bags as many as possible, construct equal number of negative bags

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Retrieval

- **Features**
  - Visual feature: 4096-dim DeCAF[6] features
  - Textual feature: 200-dim term-frequency (TF) feature

- **Baselines**
  - SVM
  - MIL methods:
    1) sMIL [Bunescu et al. ICML 2007]
    2) mi-SVM [Andrews et al. NIPS 2003]
    3) MIL-CPB [Li et al. ICCV 2011]
  - LUPI methods:
    1) SVM+ [Vapnik et al. T-NN 2009]
    2) Rank Transfer [Sharmanska et al. ICCV 2013]
  - Multi-view methods
    1) KCCA [Hardoon et al. Neural Computation 2004]
    2) SVM-2K [Farquhar et al. NIPS 2005]
  - Classeme [Torresani et al. ECCV 2010]

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Retrieval

- Results

MAPs (%) of different methods for image retrieval.

Dataset	NUS-WIDE	WebQuery
SVM	54.41	48.51
pSVM+	57.92	50.35
RT	42.63	31.92
Classeme	54.14	48.48
KCCA	54.62	47.86
SVM-2K	54.43	49.04
sMIL(PI)	56.72 (**60.88**)	51.42 (**52.63**)
mi-SVM(PI)	57.46 (**58.97**)	48.90 (**51.83**)
MIL-CPB(PI)	57.40 (**59.96**)	50.69 (**53.02**)

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Categorization

- Source domain
  NUS-WIDE: 269,648 images, 81 categories
  Flickr: we crawl 142,081 Flickr images using the class names in Caltech-256 as queries.

- Target domain
  Caltech-256: 29,780 images

- Experimental setting
  256 overlapped concepts between Flickr and Caltech-256
  17 overlapped concepts between NUS-WIDE and Caltech-256

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Categorization

- Baselines

- include the baselines for image retrieval

- Domain adaptation baselines
  1) SA [Fernando et al. ICCV 2013]
  2) TCA [Pan et al. T-NN 2011]
  3) DIP [Baktashmotlagh et al. ICCV 2013]
  4) KMM [Huang et al. NIPS 2007]
  5) GFK [Gong et al. CVPR 2012]
  6) SGF [Gopalan et al. ICCV 2011]
  7) DASVM [Bruzzone et al. T-PAMI 2010]
  8) STM [Chu et al. CVPR 2013]

- (1)~(6) combined with our classifier sMIL-PI

NANYANG TECHNOLOGICAL UNIVERSITY

Experiments: Image Categorization

MAPs (%) of different methods without domain adaptation

Training Set	NUS-WIDE	Flickr
SVM	65.33	31.41
pSVM+	66.61	35.84
RT	55.53	19.09
Classeme	66.58	34.57
KCCA	65.94	35.69
SVM-2K	66.61	35.09
sMIL	67.73	35.26
sMIL-PI	**68.55**	**39.49**

MAPs (%) of different methods with domain adaptation

Training Set	NUS-WIDE	Flickr
SVM	65.33	31.41
sMIL-PI	68.55	39.49
sMIL-PI-DA	**70.56**	**41.35**
DASVM	67.96	33.52
STM	65.73	28.52
SA	56.13(68.73)	30.15(39.61)
TCA	61.28(66.64)	27.91(37.57)
DIP	61.08(65.32)	26.49(35.16)
KMM	60.32(68.78)	32.08(37.85)
GFK	62.98(64.60)	23.90(29.24)
SGF	66.29(68.57)	30.08(37.46)

NANYANG TECHNOLOGICAL UNIVERSITY

Thanks for your attention!