Learning Discriminative Geodesic Flow Kernel for Unsupervised Domain Adaptation

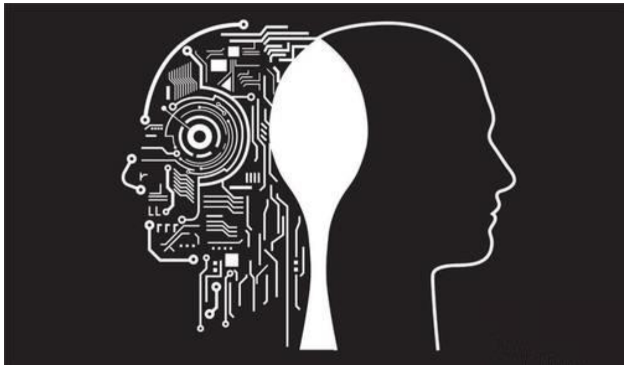
Jianze Wei

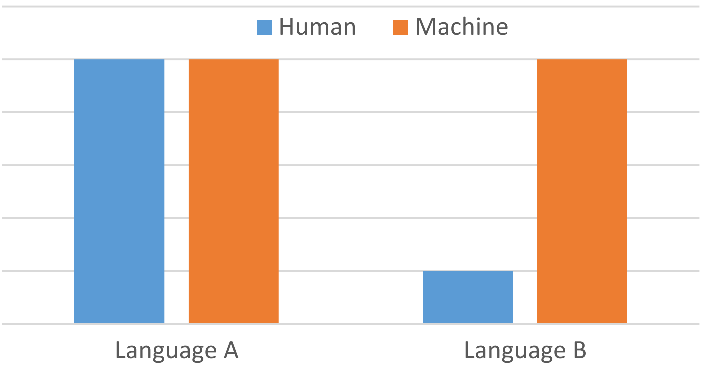
Joint work with Jian Liang, Ran He, Jinfeng Yang

Long *et al.* 2013
Long *et al.* 2014
Li *et al.* 2014
Sun *et al.* 2016

Domain adaptation

A large gap between machine (artificial) intelligence and human intelligence

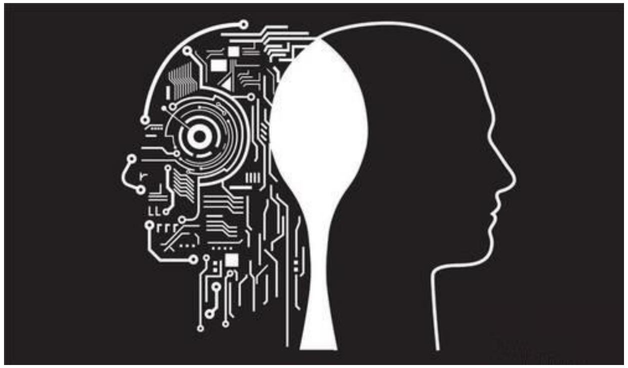


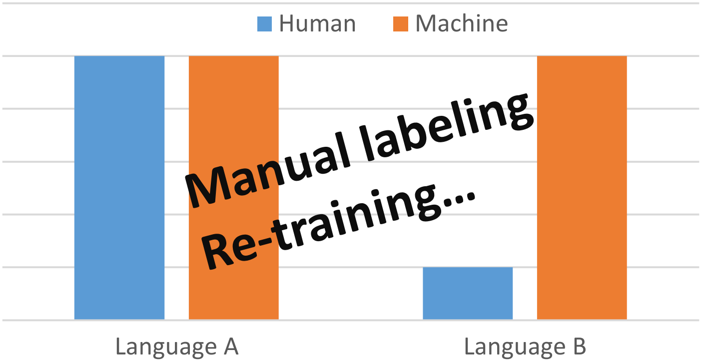


Long *et al.* 2013
Long *et al.* 2014
Li *et al.* 2014
Sun *et al.* 2016

Domain adaptation

A large gap between machine (artificial) intelligence and human intelligence

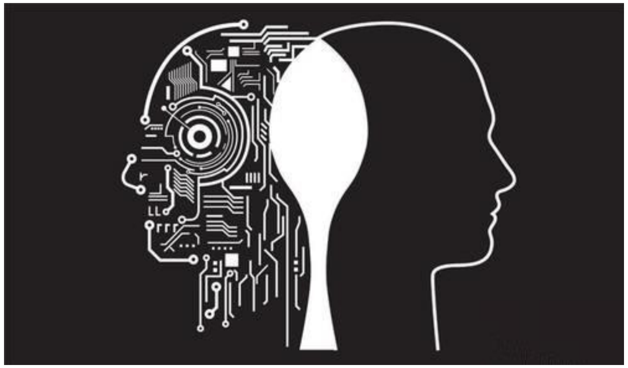


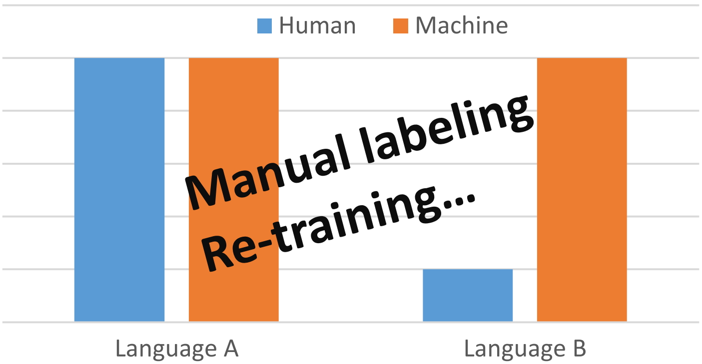


Long *et al.* 2013
Long *et al.* 2014
Li *et al.* 2014
Sun *et al.* 2016

Domain adaptation

A large gap between machine (artificial) intelligence and human intelligence





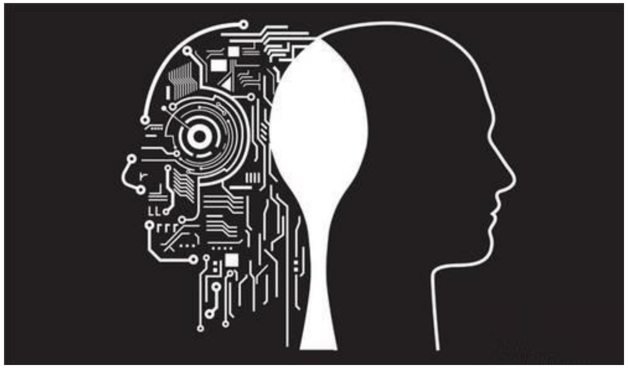
<u>Domain adaptation</u>

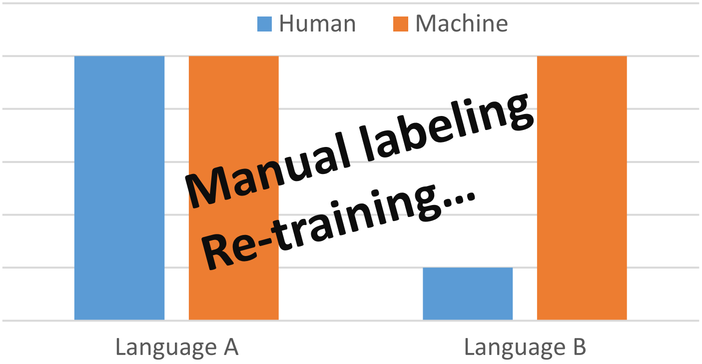
leveraging the prior knowledge from source domain on the similar task of target domain and alleviating the affect of manual labeling.

Long *et al.* 2013
Long *et al.* 2014
Li *et al.* 2014
Sun *et al.* 2016

Domain adaptation

A large gap between machine (artificial) intelligence and human intelligence





<u>Domain adaptation</u>

leveraging the prior knowledge from source domain on the similar task of target domain and alleviating the affect of manual labeling.

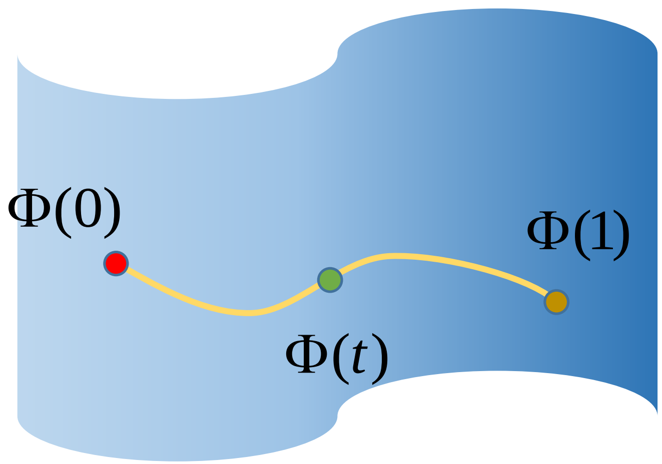
the label of the target data is unavailable

Unsupervised domain adaptation

Gopalan *et al.* 2011
Gong *et al.* 2012
Cheng *et al.* 2011

Revisiting GFK

geodesic flow
Model domain shift



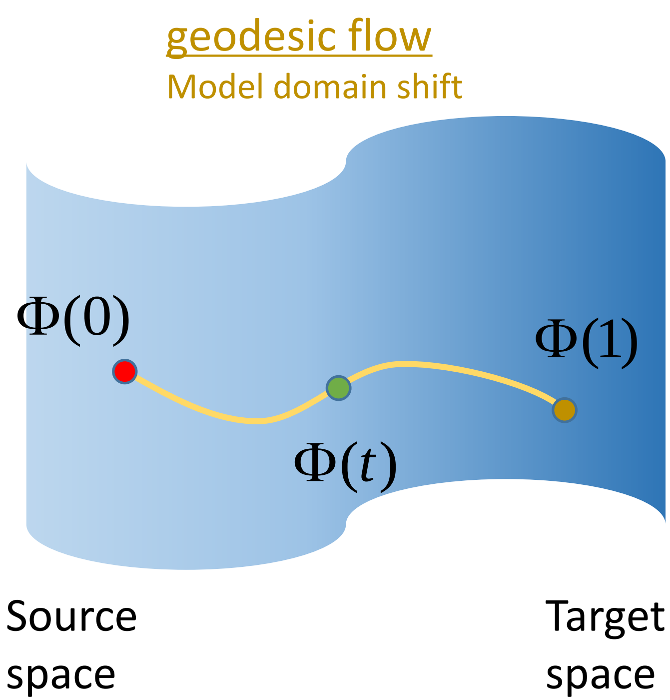
Source space

Target space

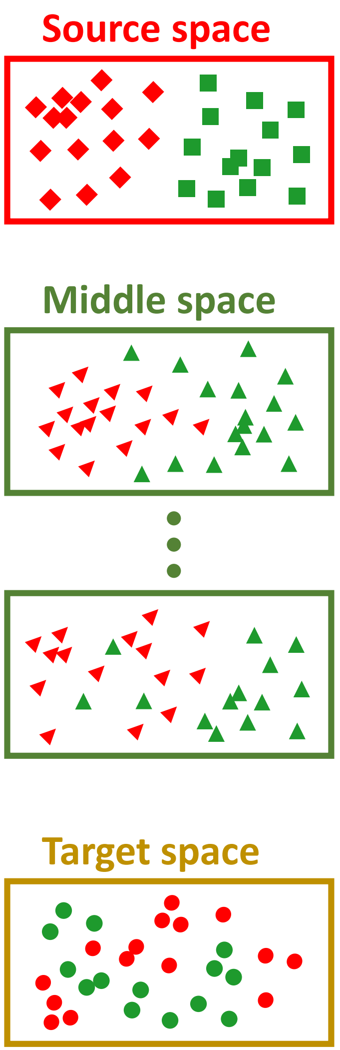
Middle space
contain the information of source and target domains

Gopalan *et al.* 2011
Gong *et al.* 2012
Cheng *et al.* 2011

Revisiting GFK

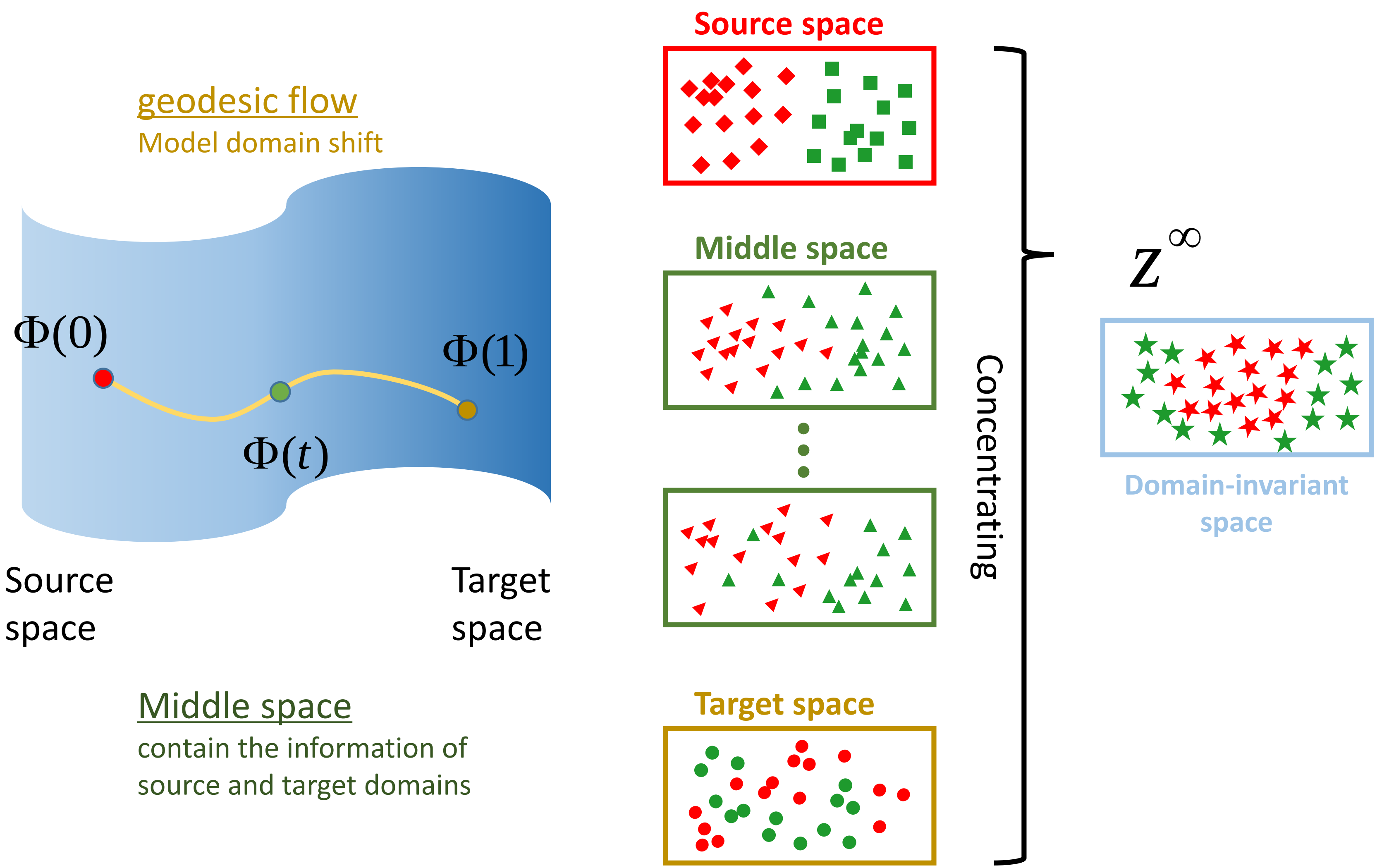


Middle space
contain the information of source and target domains



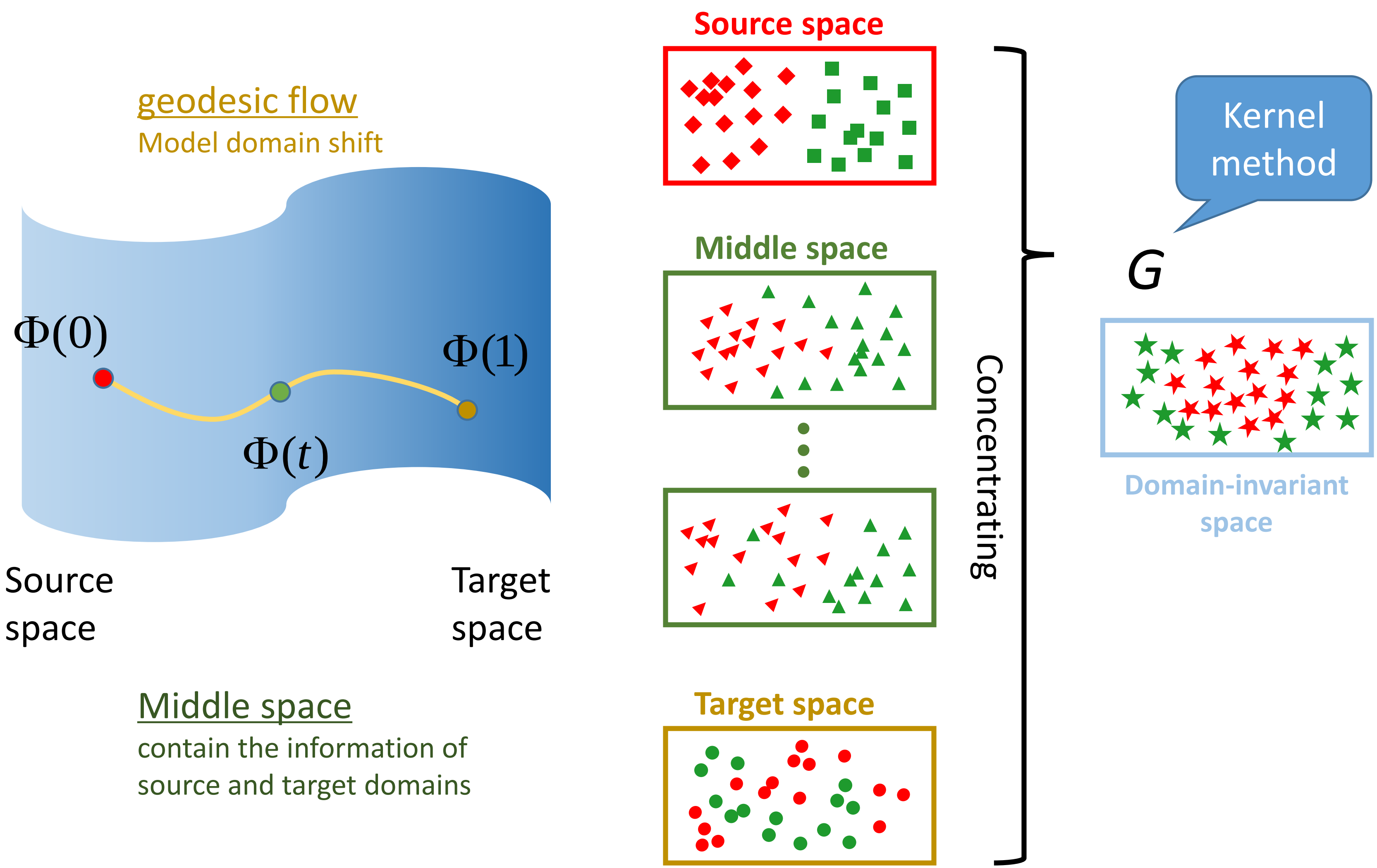
Gopalan *et al.* 2011
Gong *et al.* 2012
Cheng *et al.* 2011

Revisiting GFK



Gopalan *et al.* 2011
Gong *et al.* 2012
Cheng *et al.* 2011

Revisiting GFK

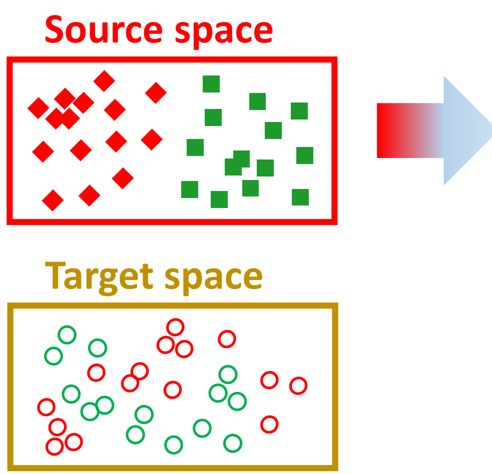


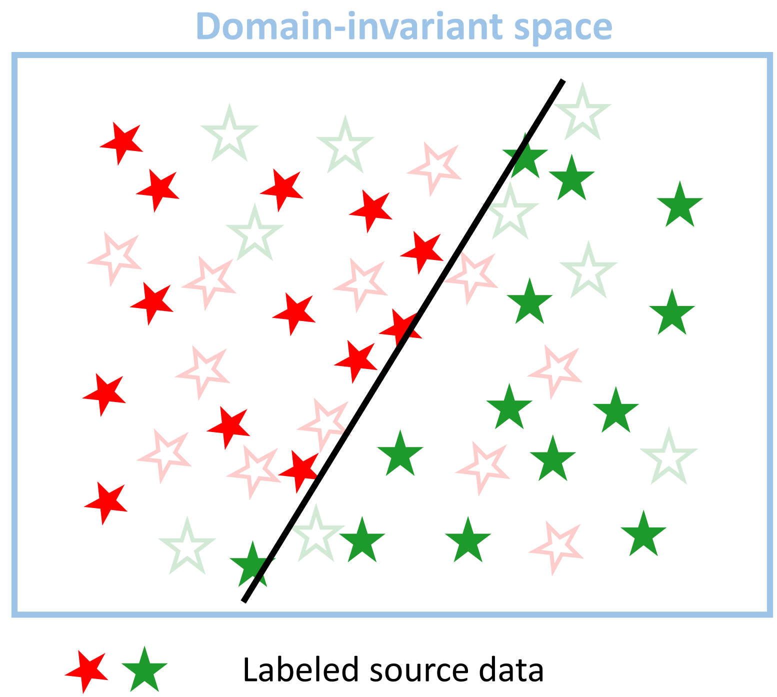
Motivation

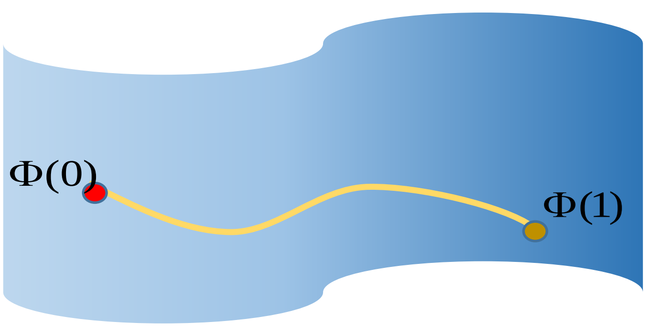
PLS

vs

PCA





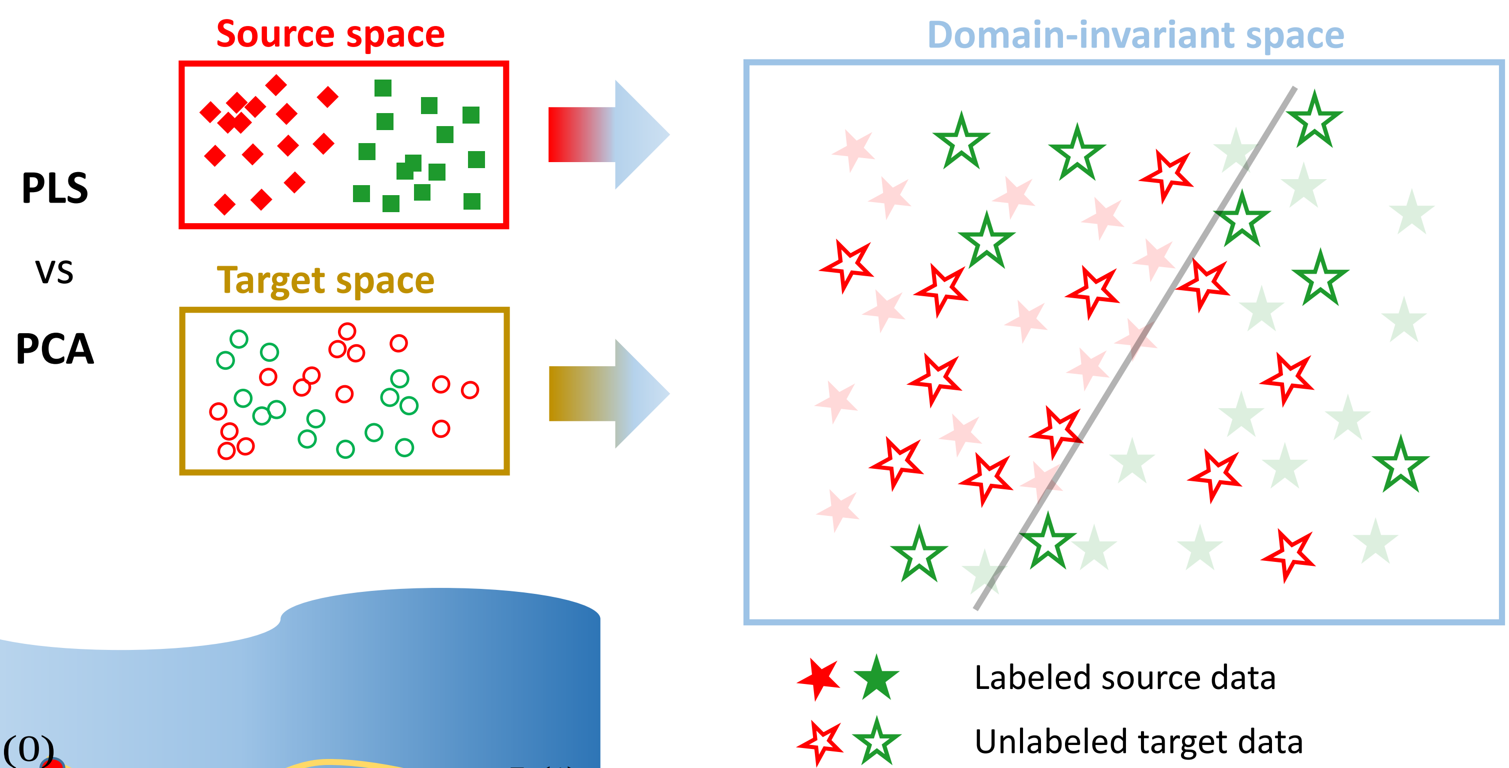


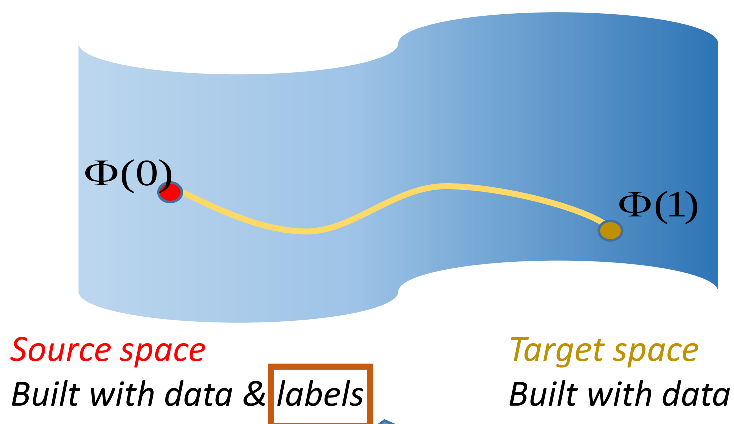
Source space
Built with data & labels

Target space
Built with data

Domain-invariant space has a good property of domain-invariance

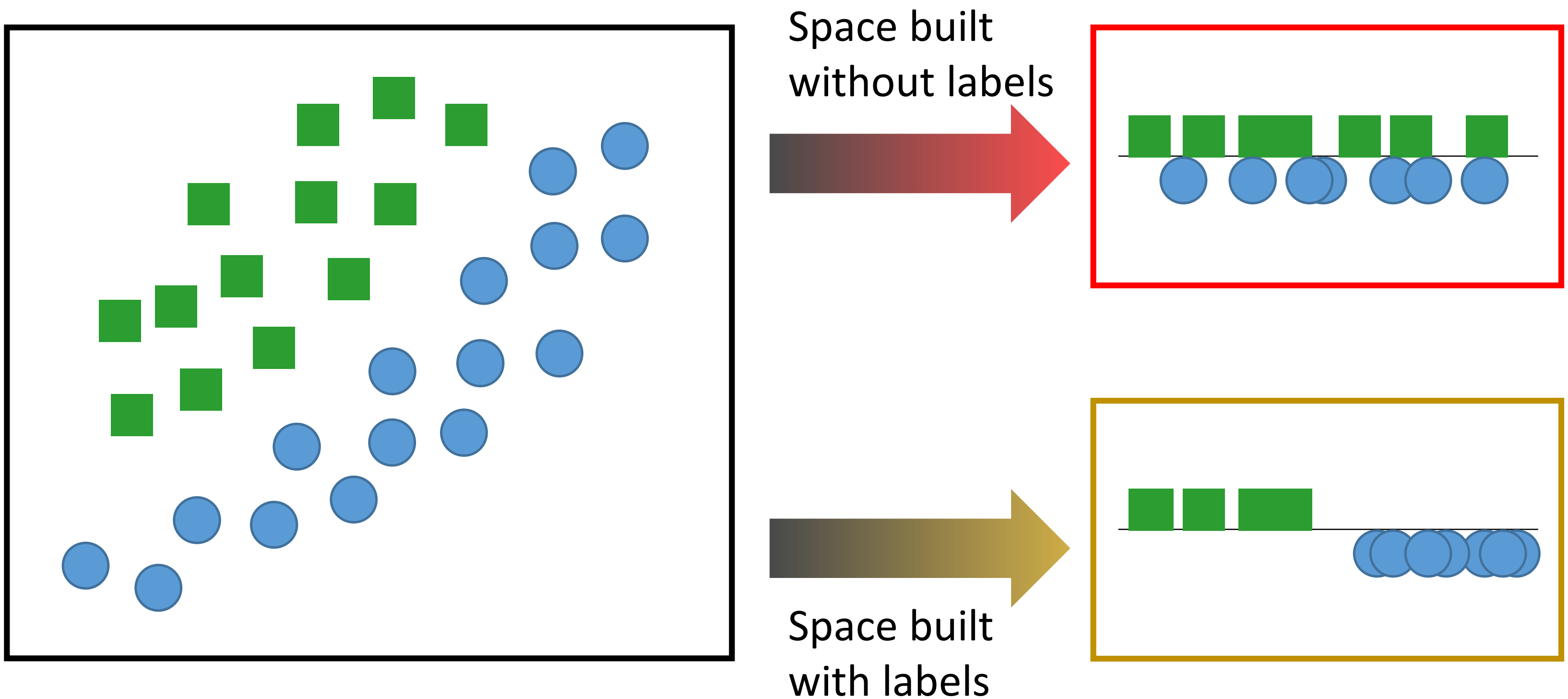
Motivation





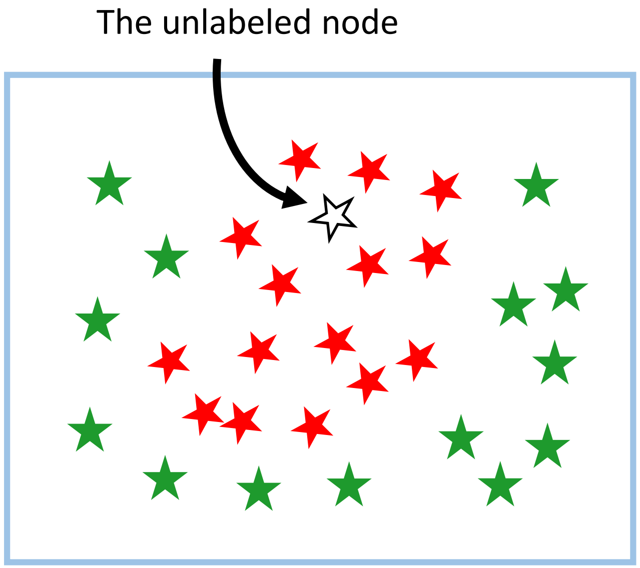
Domain-invariant space has a good property of domain-invariance, but it is **not** class-discriminative

Motivation



Zhu *et al*. 2002
Fujiwara *et al*. 2014
Long *et al*. 2014

Label propagation

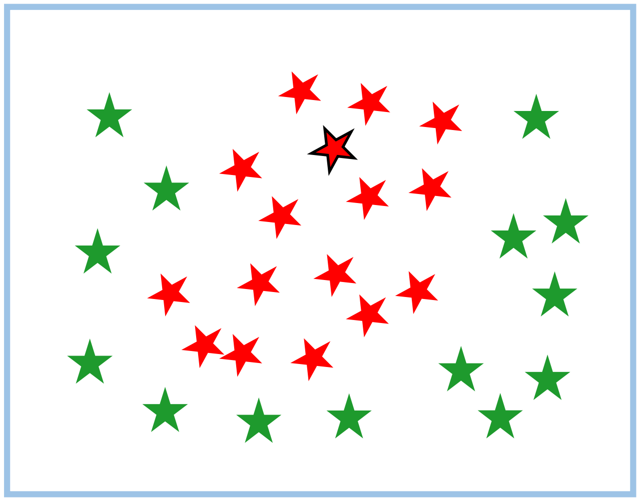


The hidden script behind LP

The sample should be more likely to have the same label as the nearest samples.

Zhu *et al.* 2002
Fujiwara *et al.* 2014
Long *et al.* 2014

Label propagation



The hidden script behind LP

~~*The sample should be more likely to have the same label as the nearest samples.*~~

Maximizing the consistency between pseudo label structure and data structure.

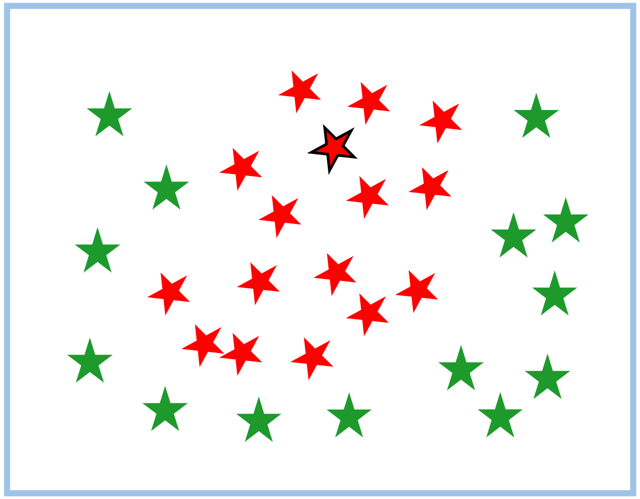
Objective function:

$$\mathrm{Min}\ \ell(X,Y)+\lambda\boxed{\sum_{i,j} H_{i,j}\left\|Y_i-Y_j\right\|_2^2}$$

$$s.t.\ \ H^{\mathrm{T}}H=I$$

Zhu *et al.* 2002
Fujiwara *et al.* 2014
Long *et al.* 2014

Label propagation



<u>The hidden script behind LP</u>

~~*The sample should be more likely to have the same label as the nearest samples.*~~

Maximizing the consistency between pseudo label structure and data structure.

a) probabilistic transition matrix *H* according to

$$h(\mathbf{x}_j, \mathbf{x}_i) = \frac{\exp\{-\frac{(\mathbf{z}_i^\infty - \mathbf{z}_j^\infty)^2}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{z}_i^\infty - \mathbf{z}_j^\infty)^2}{\sigma^2}\}} = \frac{\exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{G} (\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{G} (\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}},$$

b) Compute the soft label *L* using

$$\mathbf{L} = \begin{bmatrix} \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_1, \mathbf{x}_i)\mathbf{l}_i \\ \vdots \\ \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_i)\mathbf{l}_i \end{bmatrix} = \begin{bmatrix} h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & h(\mathbf{x}_1, \mathbf{x}_{n_s+n_t}) \\ \vdots & & \vdots \\ h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_1) & \cdots & h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_{n_s+n_t} \end{bmatrix} = \mathbf{H}^{\mathrm{T}}\mathbf{L},$$

Discriminative-GFK

GFK

a) Initializ the source basis P_s and the target basis P_t using PLS and PCA respectively

b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$ according to

$$\mathbf{G} = \begin{bmatrix} \mathbf{P}_s\mathbf{U}_1 & \mathbf{R}_s\mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_2 \\ \boldsymbol{\Lambda}_2 & \boldsymbol{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^{\mathrm{T}}\mathbf{P}_s^{\mathrm{T}} \\ \mathbf{U}_2^{\mathrm{T}}\mathbf{R}_s^{\mathrm{T}} \end{bmatrix},$$

and

$$\mathbf{P}_S^{\mathrm{T}}\mathbf{P}_T = \mathbf{U}_1\boldsymbol{\Gamma}\mathbf{V}^{\mathrm{T}}, \mathbf{R}_S^{\mathrm{T}}\mathbf{P}_T = -\mathbf{U}_2\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$$

LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$ according to

$$\begin{aligned} h(\mathbf{x}_j, \mathbf{x}_i) &= \frac{\exp\{-\frac{(\mathbf{z}_i^{\infty} - \mathbf{z}_j^{\infty})^2}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{z}_i^{\infty} - \mathbf{z}_j^{\infty})^2}{\sigma^2}\}} \\ &= \frac{\exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}}, \end{aligned}$$

b) Compute the soft label $\boldsymbol{L}$ using

$$\begin{aligned} \mathbf{L} &= \begin{bmatrix} \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_1, \mathbf{x}_i)\mathbf{l}_i \\ \vdots \\ \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_i)\mathbf{l}_i \end{bmatrix} \\ &= \begin{bmatrix} h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & h(\mathbf{x}_1, \mathbf{x}_{n_s+n_t}) \\ \vdots & & \vdots \\ h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_1) & \cdots & h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_{n_s+n_t} \end{bmatrix} \\ &= \mathbf{H}^{\mathrm{T}}\mathbf{L}, \end{aligned}$$

Discriminative-GFK

GFK

a) Initializ the source basis P_s and the target basis P_t using PLS and PCA respectively

b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$ according to

$$\mathbf{G} = \begin{bmatrix} \mathbf{P}_s\mathbf{U}_1 & \mathbf{R}_s\mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{\Lambda}_2 \\ \mathbf{\Lambda}_2 & \mathbf{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^{\mathrm{T}}\mathbf{P}_s^{\mathrm{T}} \\ \mathbf{U}_2^{\mathrm{T}}\mathbf{R}_s^{\mathrm{T}} \end{bmatrix},$$

and

$$\mathbf{P}_S^{\mathrm{T}}\mathbf{P}_T = \mathbf{U}_1\mathbf{\Gamma}\mathbf{V}^{\mathrm{T}}, \mathbf{R}_S^{\mathrm{T}}\mathbf{P}_T = -\mathbf{U}_2\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$$

LP

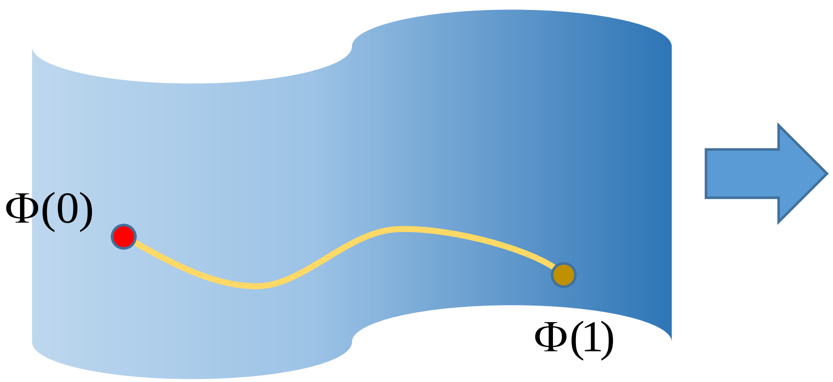
a) Construct the probabilistic transition matrix $\boldsymbol{H}$ according to

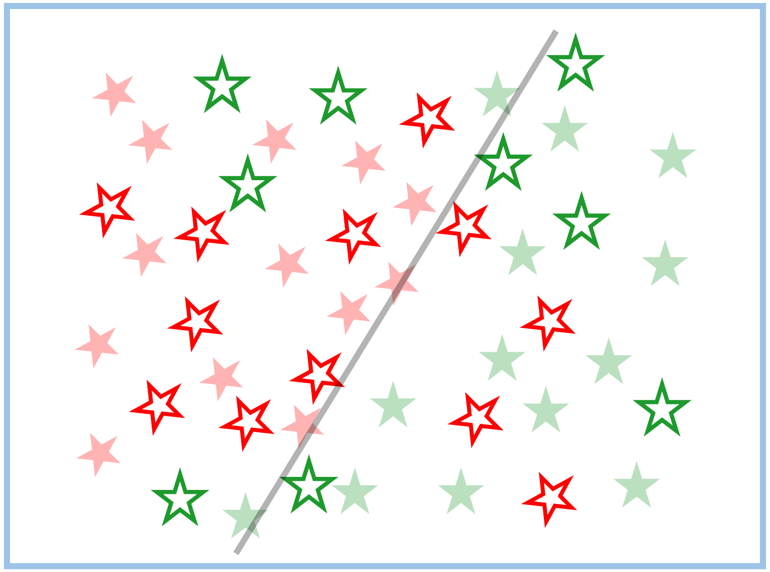
$$\begin{aligned} h(\mathbf{x}_j, \mathbf{x}_i) &= \frac{\exp\{-\frac{(\mathbf{z}_i^{\infty} - \mathbf{z}_j^{\infty})^2}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{z}_i^{\infty} - \mathbf{z}_j^{\infty})^2}{\sigma^2}\}} \\ &= \frac{\exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}}{\sum_{i=1}^{n_s+n_t} \exp\{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{G}(\mathbf{x}_i - \mathbf{x}_j)}{\sigma^2}\}}, \end{aligned}$$

b) Compute the soft label $\boldsymbol{L}$ using

$$\begin{aligned} \mathbf{L} &= \begin{bmatrix} \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_1, \mathbf{x}_i)\mathbf{l}_i \\ \vdots \\ \sum_{i=1}^{n_s+n_t} h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_i)\mathbf{l}_i \end{bmatrix} \\ &= \begin{bmatrix} h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & h(\mathbf{x}_1, \mathbf{x}_{n_s+n_t}) \\ \vdots & & \vdots \\ h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_1) & \cdots & h(\mathbf{x}_{n_s+n_t}, \mathbf{x}_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_{n_s+n_t} \end{bmatrix} \\ &= \mathbf{H}^{\mathrm{T}}\mathbf{L}, \end{aligned}$$

Discriminative-GFK





GFK

a) Initializ the source basis P_s and the target basis P_t using PLS and PCA respectively

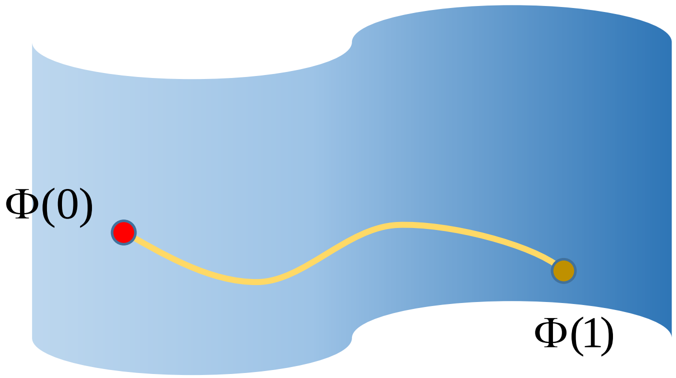
b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$

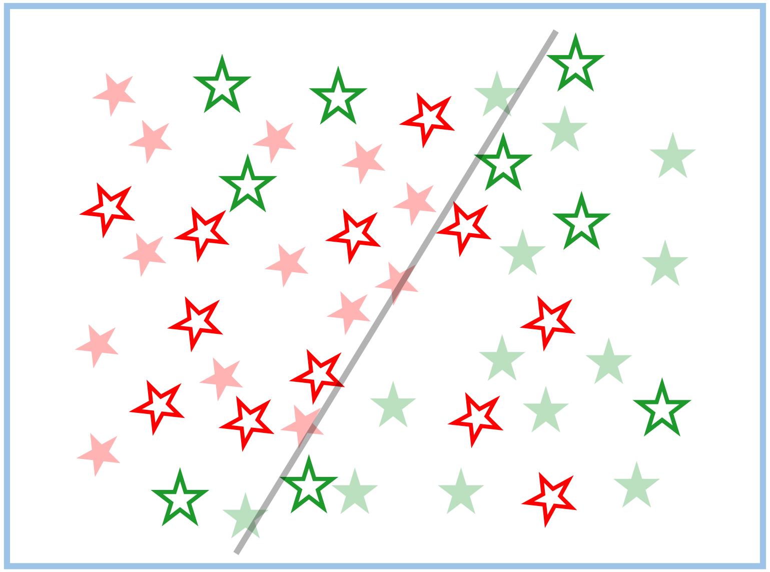
LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$

b) Compute the soft label $\boldsymbol{L}$

Discriminative-GFK





GFK

a) Initialize the source basis P_s and the target basis P_t using PLS and PCA respectively

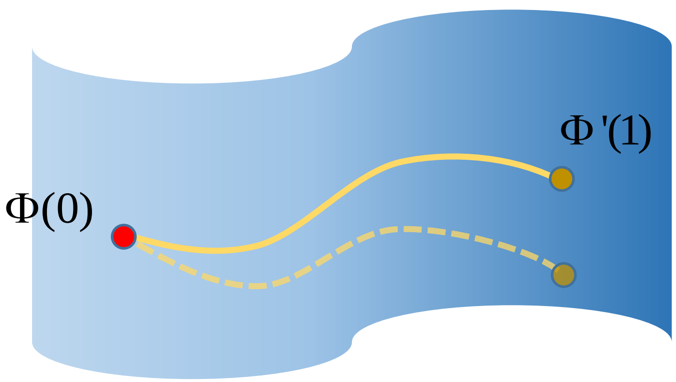
b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$

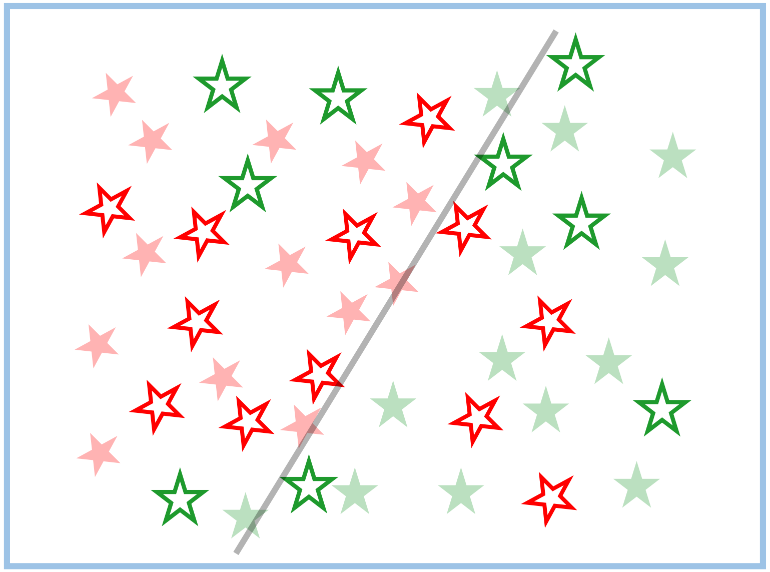
LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$

b) Compute the soft label $\boldsymbol{L}$

Discriminative-GFK





GFK

a) Update the target basis P_t using PLS

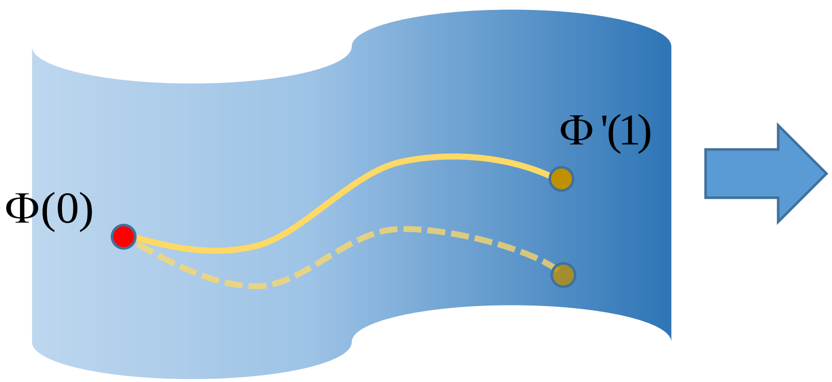
b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$

LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$

b) Compute the soft label $\boldsymbol{L}$

Discriminative-GFK



GFK

a) Update the target basis P_t using PLS

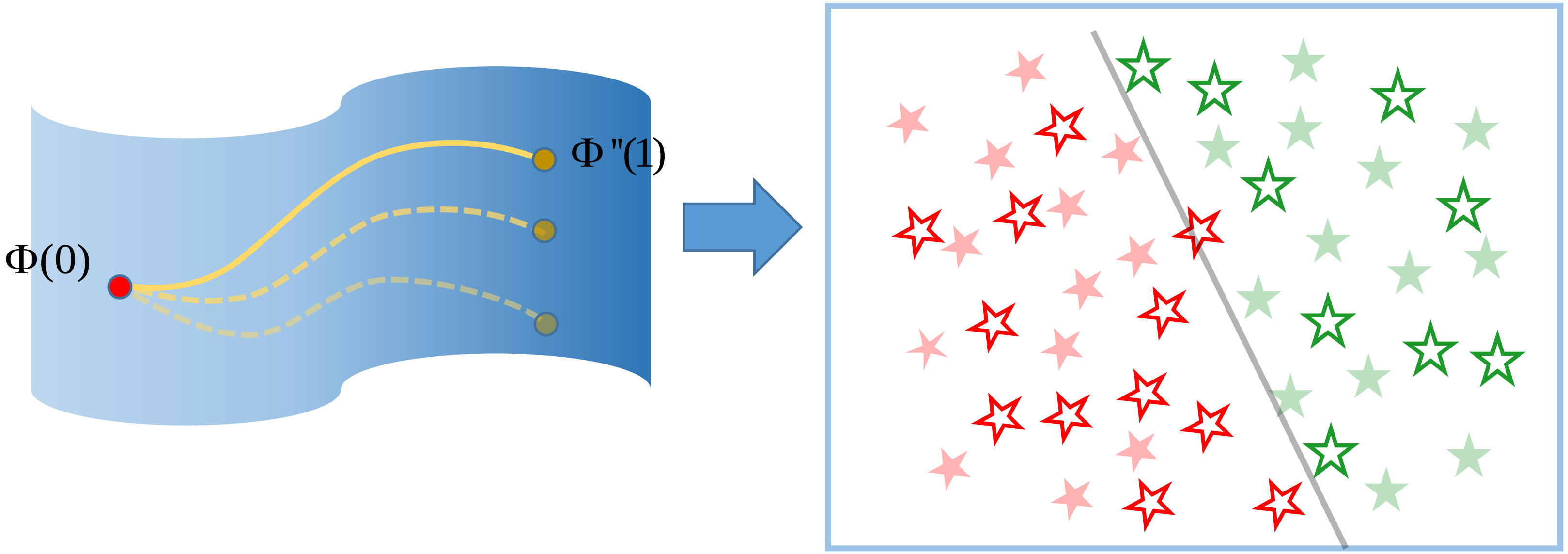
b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$

LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$

b) Compute the soft label $\boldsymbol{L}$

Discriminative-GFK



GFK

a) Update the target basis P_t using PLS

b) Compute the Geodesic Flow Kernel $\boldsymbol{G}$

LP

a) Construct the probabilistic transition matrix $\boldsymbol{H}$

b) Compute the soft label $\boldsymbol{L}$

Gong *et al.* 2012

Experiment

- Object recognition

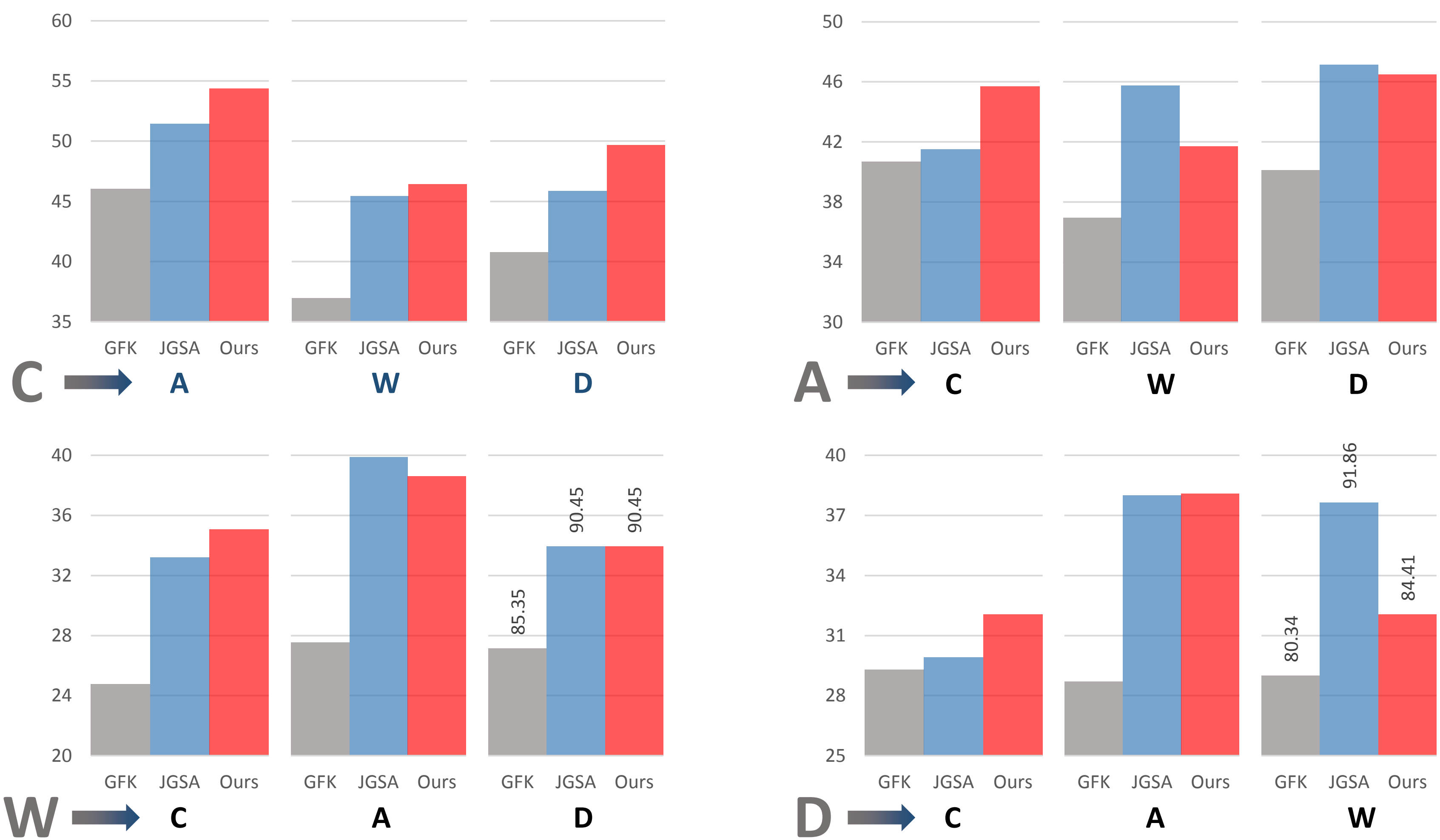
<u>Office-Caltech dataset</u>

- Four domains
- Features
  - Bag-of-SURF
- Classifier
  - 1-NN

Gong *et al.* 2012
Zhang *et al.* 2017

Experiment

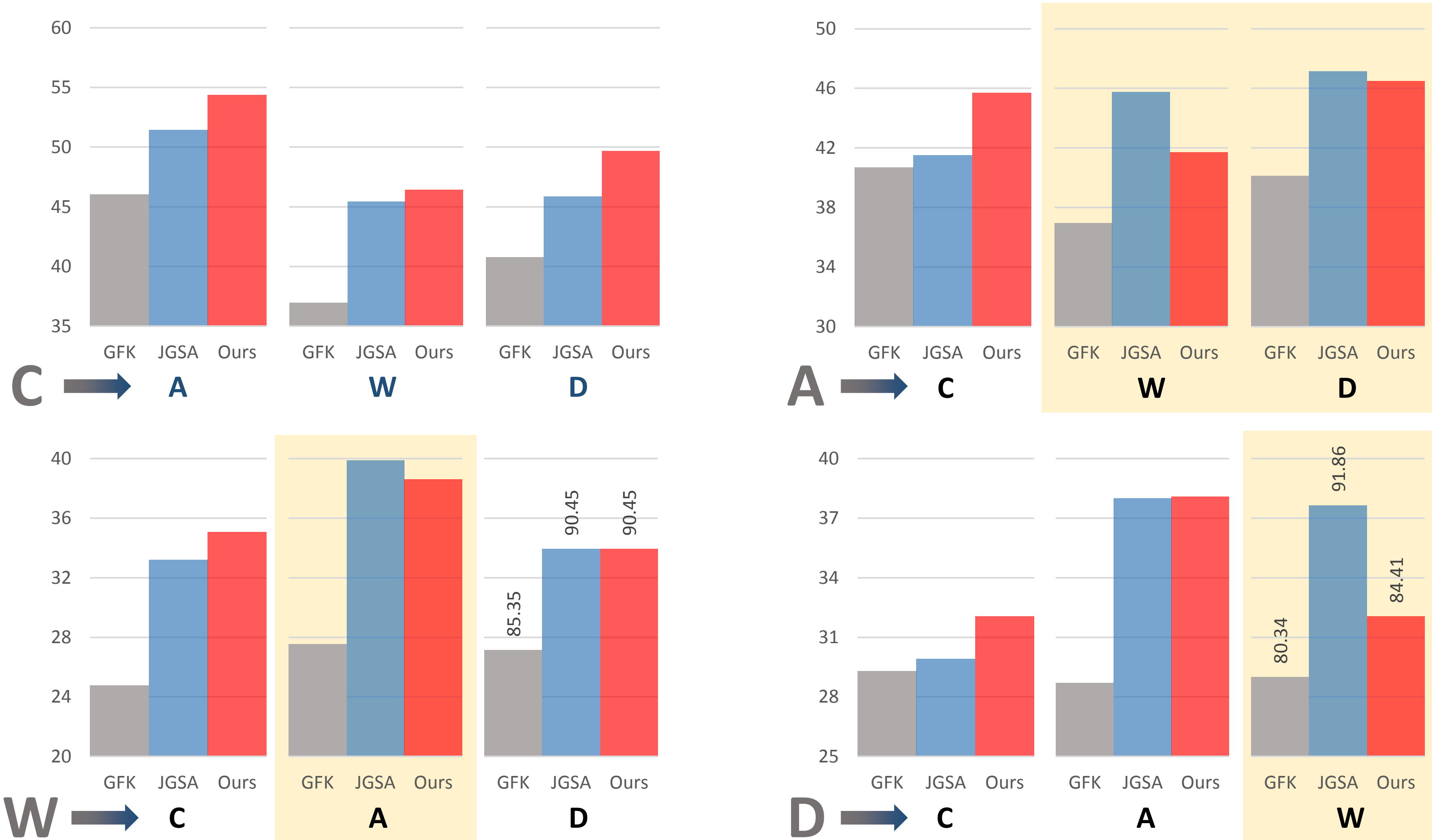
- Object recognition



Gong *et al.* 2012
Zhang *et al.* 2017

Experiment

- Object recognition



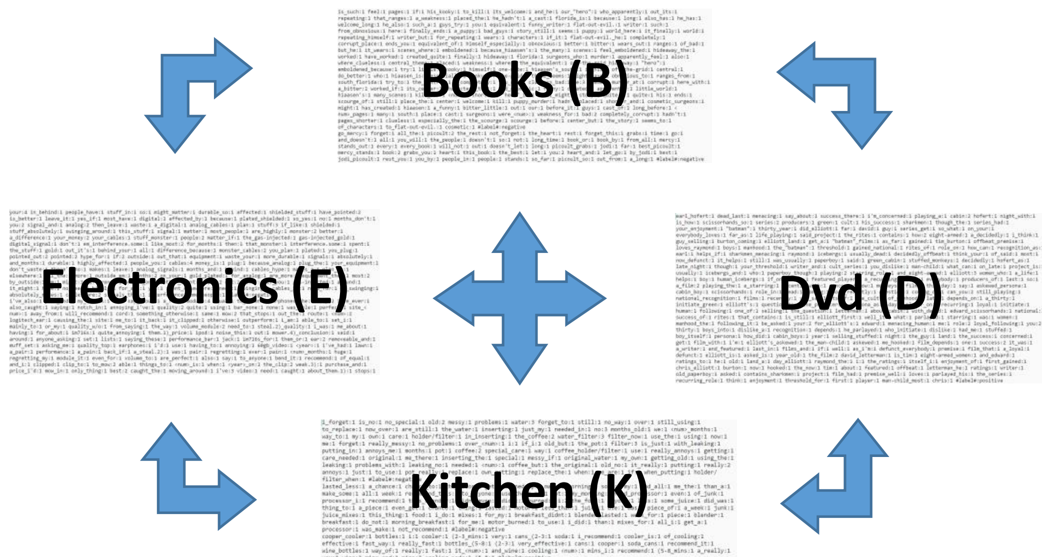
Blitzer *et al.* 2007

Experiment

- Sentiment adaptation

<u>Multi-domain sentiment dataset</u>

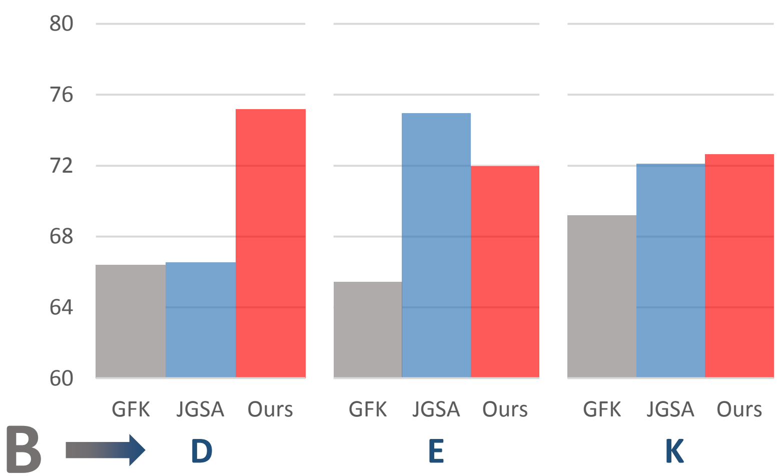
- Four domains
- Features
  - Bag-of-SURF
- Classifier
  - 1-NN

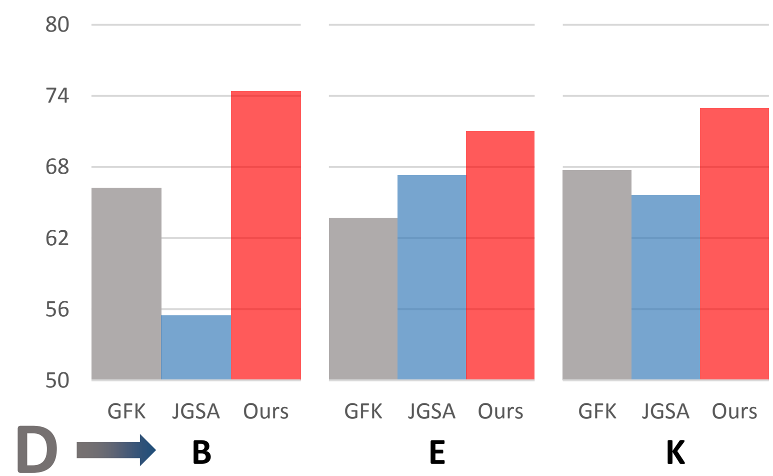


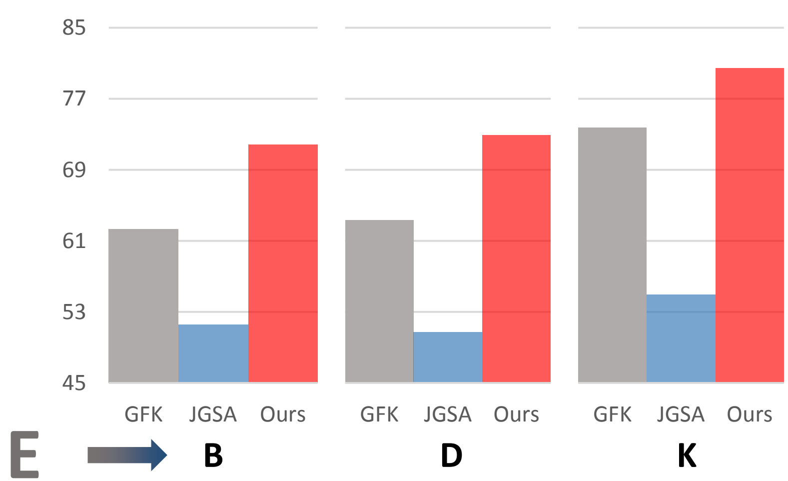
Blitzer *et al.* 2007
Gong *et al.* 2012
Zhang *et al.* 2017

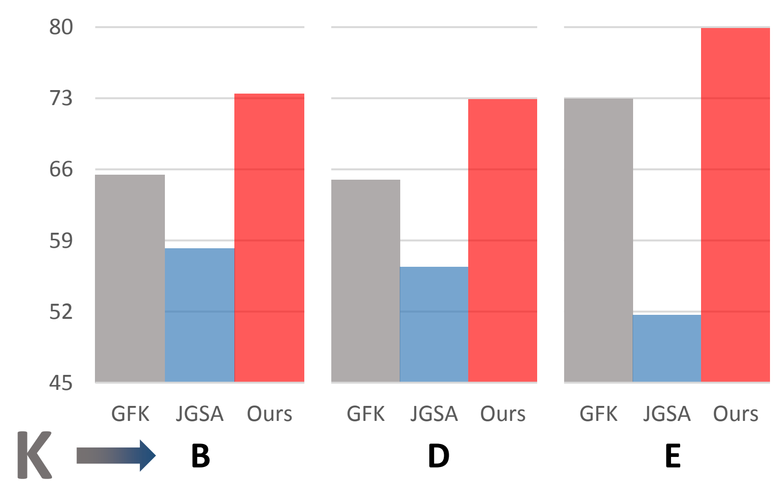
Experiment

- Sentiment adaptation





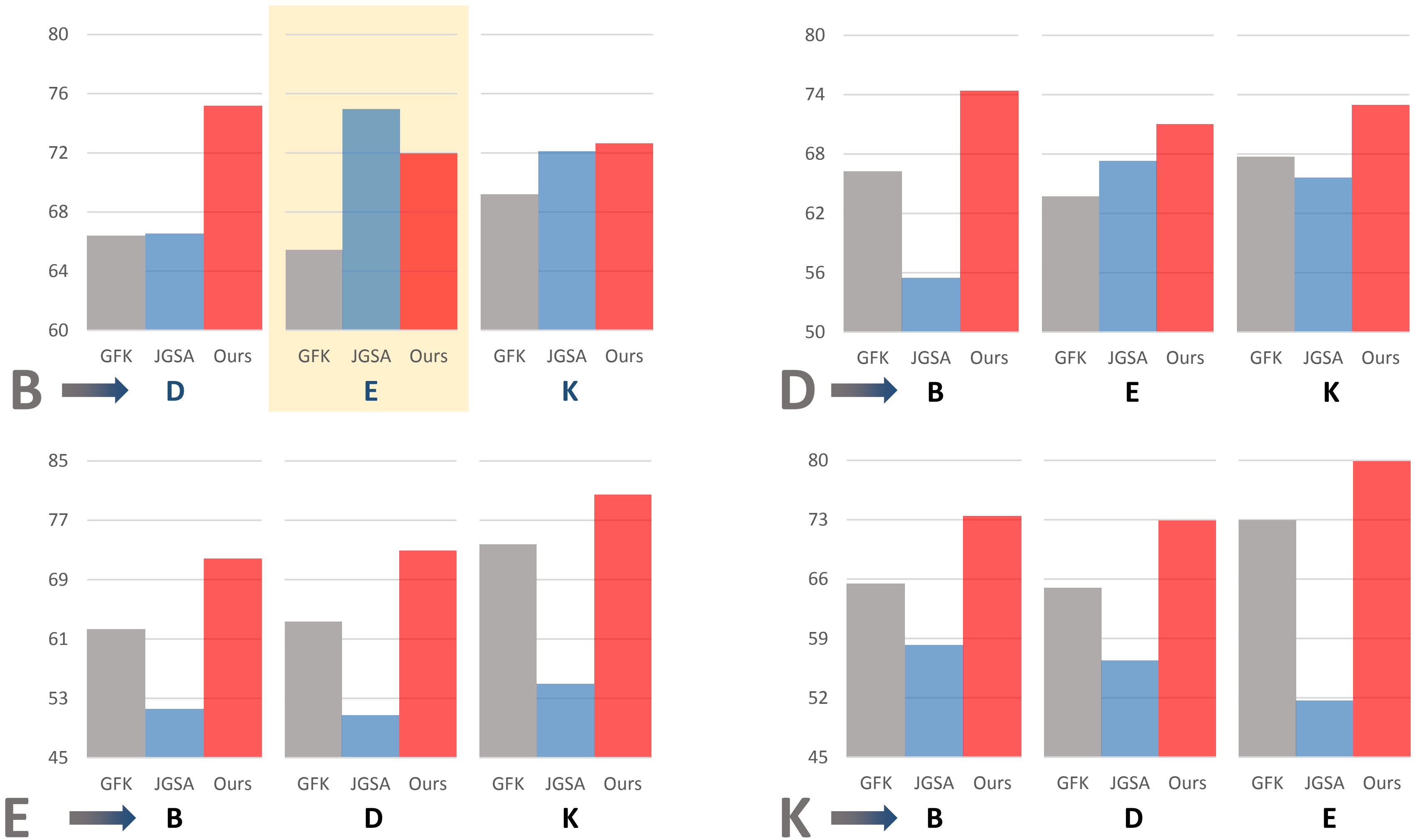




Blitzer *et al.* 2007
Gong *et al.* 2012
Zhang *et al.* 2017

Experiment

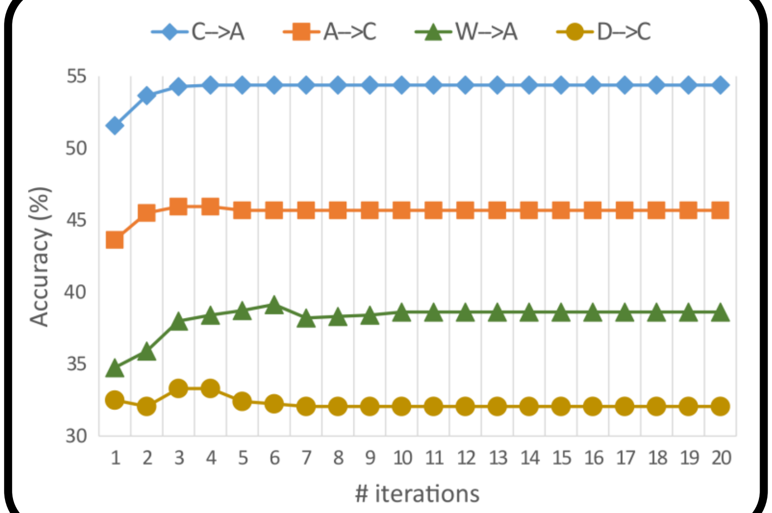
- Sentiment adaptation



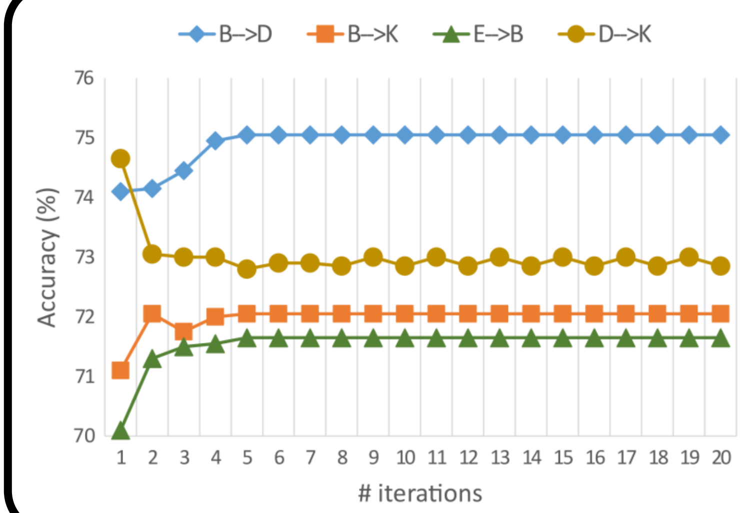
Experiment

- Convergence analysis

Office-Caltech dataset



Multi-domain sentiment dataset



Degradation cases are marked as the lines with dots.

Ours model can quickly converge **within 10 iterations**.

Thank you!

[Blitzer *et al.* 2007] J. Blitzer, M. Dredze, and F. Pereira, “Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification,” in ACL, 2007, pp. 440–447.

[Cheng *et al.* 2011] V. Cheng and C. Li, “Classification probabilistic pca with application in domain adaptation,” in PAKDD, 2011, pp. 75–86.

[Fujiwara *et al.* 2014] Y. Fujiwara and G. Irie, “Efficient label propagation,” in ICML, 2014, pp. 784–792.

[Gopalan *et al.* 2011] R. Gopalan, R. Li, and R. Chellappa, “Domain adaptation for object recognition: An unsupervised approach,” in ICCV, 2011, pp. 999–1006.

[Gong *et al.* 2012]B. Gong, Y. Shi, F. Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” in CVPR, 2012, pp. 2066–2073.

[Li et al. 2014] W. Li, L. Duan, D. Xu, and I. Tsang, “Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation,” TPAMI, vol. 36, no. 6, pp. 1134–1148, 2014.

[Long et al. 2014] M. Long, J. Wang, G. Ding, Jiaguang Sun, and Philip S Yu, “Transfer joint matching for unsupervised domain adaptation,” in CVPR, 2014, pp. 1410–1417.

[Long et al. 2013] M. Long, J. Wang, G. Ding, Jiaguang Sun, and Philip S Yu, “Transfer feature learning with joint distribution adaptation,” in ICCV, 2013, pp. 2200–2207.

[Long et al. 2014]M. Long, J. Wang, G. Ding, Sinno Jialin Pan, and S Yu Philip, “Adaptation regularization: A general framework for transfer learning,” TKDE, vol. 26, no. 5, pp. 1076–1089, 2014.

[Sun et al. 2016] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” in AAAI, 2016, pp. 2058–2065.

[Zhang *et al.* 2017] J. Zhang, W. Li, and P. Ogunbona, “Joint geometrical and statistical alignment for visual domain adaptation,” in CVPR, 2017, pp. 1859–1867.

[Zhu *et al*. 2002] X. Zhu and Z. Ghahramani, “Learning from labeled and unlabeled data with label propagation,” Technical Report CMUCALD-02-107, 2002.