Learning Discriminative Geodesic Flow Kernel for Unsupervised Domain Adaptation

Jianze Wei

Joint work with Jian Liang, Ran He, Jinfeng Yang
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
Domain adaptation

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

Domain adaptation

leveraging the prior knowledge from source domain on the similar task of target domain and alleviating the affect of manual labeling.
A large gap between machine (artificial) intelligence and human intelligence

Domain adaptation

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
Revisiting GFK

*geodesic flow*

Model domain shift

Middle space contain the information of source and target domains

$\Phi(0)$

$\Phi(t)$

$\Phi(1)$

Source space

Target space

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

Source space

Middle space

Target space

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

Source space

Middle space

Target space

Concentrating

Domain-invariant space

Gopalan et al. 2011
Gong et al. 2012
Cheng et al. 2011
Revisiting GFK

- **geodesic flow**
- Model domain shift

Source space \( \Phi(0) \)

Target space \( \Phi(1) \)

Middle space \( \Phi(t) \)

Source space

Middle space

Target space

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

Kernel method

Domain-invariant space

Concentrating
Motivation

Source space
Built with data & labels

Target space
Built with data

PLS
VS
PCA

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.

Source space
Built with data & labels

Target space
Built with data

Labeled source data
Unlabeled target data

PLS

VS

PCA

\[ \Phi(0) \rightarrow \Phi(1) \]
Motivation

Space built with labels

Space built without labels
The hidden script behind LP

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

Objective function:
\[
\min \ell(X, Y) + \lambda \sum_{i,j} H_{i,j} \| Y_i - Y_j \|^2_2
\]
\[s.t. \ H^T H = I\]

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.

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.

a) probabilistic transition matrix $H$ according to

$$h(x_j, x_i) = \frac{\exp\left\{ -\frac{(z_j - z_i)^2}{\sigma^2} \right\}}{\sum_{i=1}^{n_a+n_t} \exp\left\{ -\frac{(z_j - z_i)^2}{\sigma^2} \right\}} \frac{\exp\left\{ -\frac{(x_i - x_j)^T G(x_i - x_j)}{\sigma^2} \right\}}{\sum_{i=1}^{n_a+n_t} \exp\left\{ -\frac{(x_i - x_j)^T G(x_i - x_j)}{\sigma^2} \right\}}.$$ 

b) Compute the soft label $L$ using

$$L = \begin{bmatrix} \frac{\sum_{i=1}^{n_a+n_t} h(x_1, x_i) l_i}{\sum_{i=1}^{n_a+n_t} h(x_{n_a+n_t}, x_i) l_i} \\ \vdots \\ \frac{\sum_{i=1}^{n_a+n_t} h(x_{n_a+n_t}, x_i) l_i}{\sum_{i=1}^{n_a+n_t} h(x_{n_a+n_t}, x_i) l_i} \end{bmatrix} = \begin{bmatrix} h(x_1, x_1) & \cdots & h(x_1, x_{n_a+n_t}) \\ \vdots & \ddots & \vdots \\ h(x_{n_a+n_t}, x_1) & \cdots & h(x_{n_a+n_t}, x_{n_a+n_t}) \end{bmatrix} \begin{bmatrix} l_1 \\ \vdots \\ l_{n_a+n_t} \end{bmatrix} = H^T L.$$
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 $G$ according to

\[ G = \left[ \begin{array}{ccc} P_sU_1 & R_sU_2 \end{array} \right] \left[ \begin{array}{ccc} \Lambda_1 & \Lambda_2 & \Lambda_3 \end{array} \right] \left[ \begin{array}{ccc} U_1^T P_s^T & U_2^T P_s^T \end{array} \right], \]

and

\[ P_s^T P_s = U_1 \Gamma V^T, \quad R_s^T P_t = -U_2 \Sigma V^T \]

LP

a) Construct the probabilistic transition matrix $H$ according to

\[ h(x_j, x_i) = \frac{\exp\left\{-\frac{(x_j - x_i)^2}{\sigma^2}\right\}}{\sum_{i=1}^{n_s+n_t} \exp\left\{-\frac{(x_j - x_i)^2}{\sigma^2}\right\}} \]

b) Compute the soft label $L$ using

\[ L = \begin{bmatrix} \sum_{i=1}^{n_s+n_t} h(x_1, x_i) l_i \\ \vdots \\ \sum_{i=1}^{n_s+n_t} h(x_{n_s+n_t}, x_i) l_i \\ h(x_1, x_1) & \cdots & h(x_1, x_{n_s+n_t}) \\ \vdots & \vdots & \vdots \\ h(x_{n_s+n_t}, x_1) & \cdots & h(x_{n_s+n_t}, x_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} l_1 \\ \vdots \\ l_{n_s+n_t} \end{bmatrix} = H^T L. \]
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 $G$ according to

\[ G = [ P_s U_1 \quad R_s U_2 ] \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 \\ \end{bmatrix} \begin{bmatrix} U_1^T P_s^T \\ U_2^T R_s^T \end{bmatrix}, \]

and

\[ P_s^T P_t = U_1 \Gamma V^T, \quad R_s^T P_t = -U_2 \Sigma V^T. \]

**LP**

a) Construct the probabilistic transition matrix $H$ according to

\[ h(x_j, x_i) = \frac{\exp\left\{ -\frac{(x_i - x_j)^2}{\sigma^2} \right\}}{\sum_{i=1}^{n_s + n_t} \exp\left\{ -\frac{(x_i - x_i)^2}{\sigma^2} \right\}} \]

\[ = \frac{\exp\left\{ -\frac{(x_i - x_i)^2 G(x_i - x_i)}{\sigma^2} \right\}}{\sum_{i=1}^{n_s + n_t} \exp\left\{ -\frac{(x_i - x_i)^2 G(x_i - x_i)}{\sigma^2} \right\}}, \]

b) Compute the soft label $L$ using

\[ L = \begin{bmatrix} \sum_{i=1}^{n_s + n_t} h(x_1, x_i) l_i \\ \vdots \\ \sum_{i=1}^{n_s + n_t} h(x_{n_s+n_t}, x_i) l_i \\ h(x_1, x_1) & \cdots & h(x_1, x_{n_s+n_t}) \\ \vdots & \vdots & \vdots \\ h(x_{n_s+n_t}, x_1) & \cdots & h(x_{n_s+n_t}, x_{n_s+n_t}) \end{bmatrix} \begin{bmatrix} l_1 \\ \vdots \\ l_{n_s+n_t} \end{bmatrix} = H^T L. \]
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 \( G \)

LP
a) Construct the probabilistic transition matrix \( H \)
b) Compute the soft label \( L \)
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 $G$

LP
a) Construct the probabilistic transition matrix $H$
b) Compute the soft label $L$
Discriminative-GFK

**GFK**

a) Update the target basis $P_t$ using PLS

b) Compute the Geodesic Flow Kernel $G$

**LP**

a) Construct the probabilistic transition matrix $H$

b) Compute the soft label $L$
Discriminative-GFK

**GFK**

a) Update the target basis $P_t$ using PLS

b) Compute the Geodesic Flow Kernel $G$

**LP**

a) Construct the probabilistic transition matrix $H$

b) Compute the soft label $L$
Discriminative-GFK

GFK
a) Update the target basis $P_t$ using PLS
b) Compute the Geodesic Flow Kernel $G$

LP
a) Construct the probabilistic transition matrix $H$
b) Compute the soft label $L$
Experiment

- Object recognition

Office-Caltech dataset

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

- Object recognition

Gong et al. 2012
Zhang et al. 2017
Experiment

- Sentiment adaptation

Multi-domain sentiment dataset

- Four domains

- Features
  - Bag-of-SURF

- Classifier
  - 1-NN

Blitzer et al. 2007
Experiment

-Sentiment adaptation

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

- 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.


