Dissimilarity-based Sparse Subset Selection

paper repot by

Mohsen Nabian

- in this paper, goal is given <u>pairwise dissimilarities</u> d_{ij} between the elements of "<u>Source set</u>" and "<u>target set</u>". we consider the problem of finding a <u>subset of source set</u>, called "<u>representative</u>" that can efficiently <u>describe the target set</u>.

Formulations:

$\mathbb{X} = \{x_1, \dots x_M\}$ → source set

$Y = \{y_1, \dots, y_N\}$ → target set

or $X = \begin{bmatrix} x_1 & x_2 & \dots & x_M \end{bmatrix}$

$Y = \begin{bmatrix} y_1 & y_2 & \dots & y_N \end{bmatrix}$

we are given pairwise dissimilarities between each x to each y : $\{d_{ij}\}_{i=1,\dots,M}^{j=1,\dots,N}$

- d_{ij} shows how well x_i represents y_j.
- Smaller d_{ij}, the better x_i represent y_j

$$\underset{\text{is given}}{D} \triangleq \begin{bmatrix} d_1^T \\ \vdots \\ d_M^T \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ \vdots & \vdots & & \vdots \\ d_{M1} & d_{M2} & \dots & d_{MN} \end{bmatrix} \in \mathbb{R}^{M \times N}$$

Scanned by CamScanner

we have to find $Z \in \mathbb{R}^{M \times N}$ matrix that meets the following objectives : $Z_{M\times N} = \begin{bmatrix} & & \\ & z_{ij} & \\ . & & \end{bmatrix}$ related to x_i

related to y_j

- $z_{ij} \in \{0,1\}$

- $\sum_{i=1}^{M} z_{ij} = 1 \quad \forall j$ means every element in y, should be presented by only 1 x_i

$\Rightarrow$ $d_{ij} z_{ij}$ is the cost encoding y_j by x_i

so if z_{ij} is not 0, d_{ij} should be small.

So we can formulate the problem as follows:

$$\begin{cases} \min \quad \lambda \underbrace{\sum_{i=1}^{M} I(\|z_i\|_p)}_{\text{number of representative}} + \underbrace{\sum_{j=1}^{N}\sum_{i=1}^{M} d_{ij} z_{ij}}_{\text{total cost of encoding } Y} \\ \text{s.t.} \quad \sum_{i=1}^{M} z_{ij} = 1, \forall j; \quad z_{ij} \in \{0,1\} \quad \forall i,j \end{cases} \qquad (*)$$

we want $\sum_{i=1}^{M} I(\|z_i\|_p)$ to be minimized so that we have less points involved on representing target set.

in source set

We want more all zero rows in Z.

Problem (*) is NP-hard and non Convex :

Convex form :

(**)

$$\min_{\{z_{ij}\}} \ \lambda \sum \|z_i\|_p + \sum_{j=1}^{N} \sum_{i=1}^{M} d_{ij} z_{ij}$$

$$\text{s.t.} \ \sum_{i=1}^{M} z_{ij} = 1 , \ \forall j \ ; \quad z_{ij} \geq 0 \ \forall i,j$$

equivalent Matrix form

$$\min_{Z} \ \lambda \|Z\|_{1,p} + tr(D^T Z)$$

$$\text{s.t.} \ 1^T Z = 1^T , \quad Z \geq 0$$

where

$$\|Z\|_{1,p} \triangleq \sum_{i=1}^{M} \|z_i\|_p$$

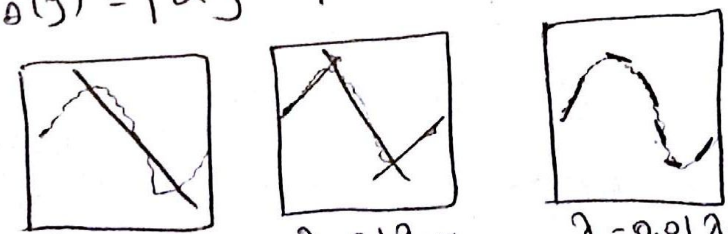
λ : if $\lambda \to 0$ $\Rightarrow$ each y_j chooses the nearest x_i

if $\lambda \to \infty$ $\Rightarrow$ only one representative will be selected.

as example :

we have a noisy data $\{y_j\}_{j=1}^{N}$ and $X = \{\theta_i\}_{i=1}^{N}$ where each $\theta_i = (a_i, b_i)$ represent a model. θ_i are obtained as follows : for each y_j and its $K=4$ nearest neighbors, we find θ_i (a_i, b_i) by minimizing $l_\theta(y) = |a^T y - b|$ for the $k+1$ points.

by changing λ we see :

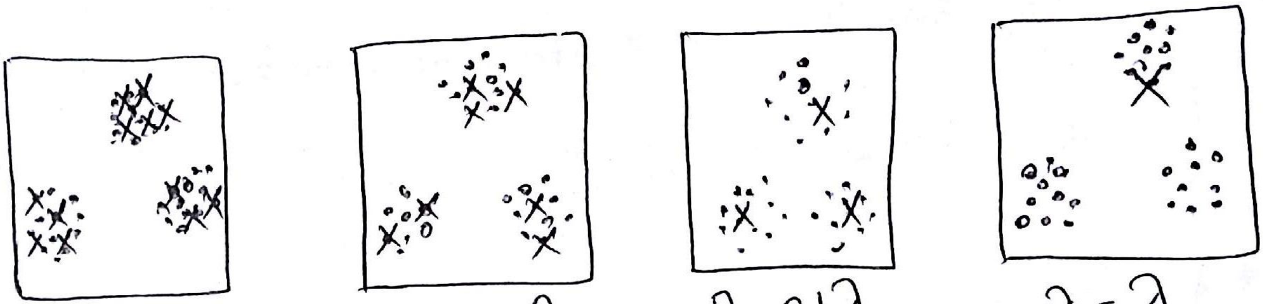


$\lambda = \lambda_{max}$ $\quad$ $\lambda = 0.1 \lambda_{max}$ $\quad$ $\lambda = 0.01 \lambda_{max}$

in another experiment: $X = Y = \underline{3}$ gaussian representative

• Y

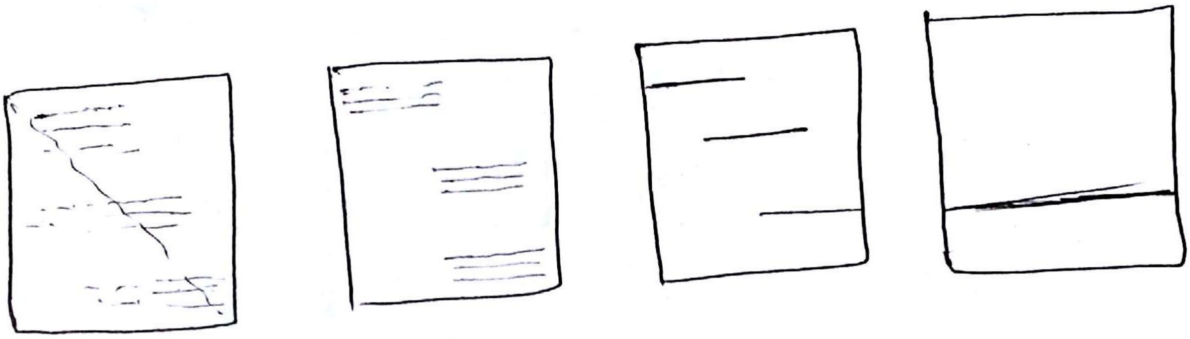
X represent ative



$\lambda = 0.002\lambda_{max}$ $\lambda = 0.01\lambda_{max}$ $\lambda = 0.1\lambda_{max}$ $\lambda = \lambda_{max}$

Z matrix

— ↳ non zero values



Dealing with outliers:

outlier in

- Source set → if X_i is outlier, it can not represent any Y_j, therefore outamatically ~~removed~~ not selected.
- Target set
  ↳ we define e_j Corresponding to each target y_j and set: $\sum z_{ij} + e_j = 1$

optimization robust to outlier:

(***)

$$\min_{\{z_{ij}\},\{e_j\}} \lambda \sum_{i=1}^{M} \|z_i\|_p + \sum_{j=1}^{N}\sum_{i=1}^{M} d_{ij} z_{ij} + \sum_{j=1}^{N} w_j e_j$$

$$\text{s.t.} \left(\sum_{i=1}^{M} z_{ij}\right) + e_j = 1, \quad \forall j; \quad z_{ij} \geq 0$$

$$e_j \geq 0 \quad \forall j$$

Since $\left(\sum_{i=1}^{M} z_{ij}\right) + e_j = 1$,

if y_j is outlier, minimization will put $e_j = 1$ and $\left(\sum_{i=1}^{M} z_{ij}\right) = 0$ therefor, y_j is not represented at all.

opt (***) in Matrix form: (equivalent to ***)

$$\min_{Z,e} \lambda \|Z\|_{1,p} + tr\left(\begin{bmatrix} D \\ w^T \end{bmatrix}^T \begin{bmatrix} Z \\ e^T \end{bmatrix}\right)$$

$$\text{s.t.} \quad 1^T \begin{bmatrix} Z \\ e^T \end{bmatrix} = 1^T, \quad \begin{bmatrix} Z \\ e^T \end{bmatrix} \geq 0$$

also w_j in optimization (***) can be chosen as

$$\underline{w_j = \beta e^{-\frac{\min_i d_{ij}}{\tau}}} \quad \leftarrow \text{suggestion}$$

another choice: $w_j = w$ for all js.

Clustering via Representative :

① Optimal solution Z^* indicates elements of x that are representative to Y.

② They also provides membership information.

~~in other~~ i.e. for $y_j \rightarrow [z^*_{1j}, \ldots, z^*_{Mj}]$ is probability vector for presenting y_j by elements of X. ⇒ This is like soft assignment for clustering.

We can also do ~~to~~ hard assignment by taking minimum d_{ij} for those selected $\{x_{\ell 1}, \ldots, x_{\ell k}\}$ ~~to~~ for presenting y_j. i.e. we take the closest among selected x_is to y_j (hard assignment)

DS3 Implementation

author has introduced an algorithm to efficiently solve minimization (**) using ADMM.

This algorithm is provided in page 7 of the paper.

The computational complexity is of $O(MN)$ very fast in table 7, the average computational time of the proposed ADMM algorithm is compared to the CVX (Sedumi Solver) which showed that proposed ADMM outperforms significantly in all trials with $\lambda = 0.01 \cdot \lambda_{max}$.

data size $N \times N$, experiment on: $\begin{cases} N = \{30, 50, 100, 200, 500, 1000, 2000\} \\ p = \{2, \infty\} \end{cases}$

Regularization parameter effect

regularization puts a trade of between the number of representative and encoding cost.

Theorem 7: if $\begin{cases} \lambda_{max,2} \triangleq \max\limits_{i \neq l} \dfrac{\sqrt{N}}{2} \dfrac{\|d_i - d_{l^*}\|_2^2}{1^T(d_i - d_l^*)} & l^* = \arg\min_i 1^T d_i \\ \lambda_{max,\infty} \triangleq \max\limits_{i \neq l} \dfrac{\|d_i - d_{l^*}\|_1}{2} \end{cases}$

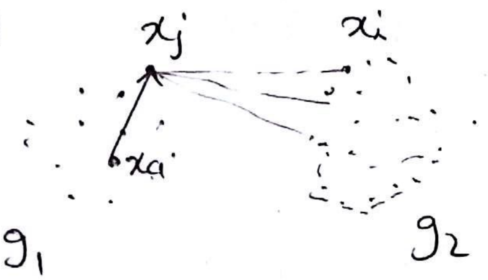
Then entire Y will be presented only by one single x^*

Clustering Gaurantee

under 2 conditions, clustering is guaranteed

figure 7 shows the conditions.

①

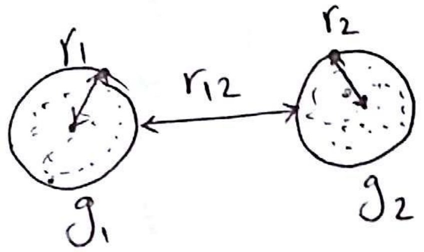


in g_1

distance x_j to x_{ci} is lower than distance x_j to any point in g_2'

Same condition for g_2 against g_1

②



$\lambda < r_{12} - \max\{r_1, r_2\}$

Experiments :

1) classification using representative:

$K = 15$ class 4485 images 210 or 410 images per class

80% training 20% test

we find representative of the training data in each class and use them as reduced data to perform NN classification on the test data.

after selecting η fraction of training samples in each class using each algorithm.

$$err(\eta) = accuracy(1) - accuracy(\eta)$$

Tabel 2 showed the classification performance using different methods:

Algorithm	Rand	Kmedioids	AP	DS3
$\eta = 0.05$			*	
$\eta = 0.10$				*
$\eta = 0.20$				*
$\eta = 0.35$				*

(*) → best performance

also

Confusion matrix	Confusion matrix	Confusion matrix
$\eta = 0.05$	$\eta = 0.35$	$\eta = 1$

② Modeling segmentation of dynamic data:

- modeling and segmentation of human activities in motion capture data
- data set = time series of different subjects each performing several activities.

9

- used 14 most informative joints

Seq number	7	2	...	14
# frames	865	2115	...	1204
#activities	4	8	...	4

SC error%

SBiC error%

Kmedoid error%

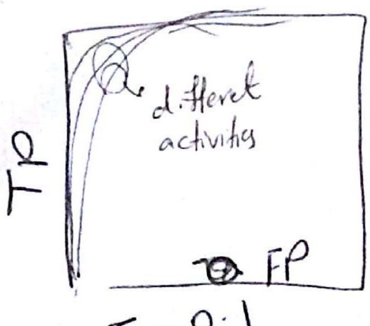
AP error

DS3 error%

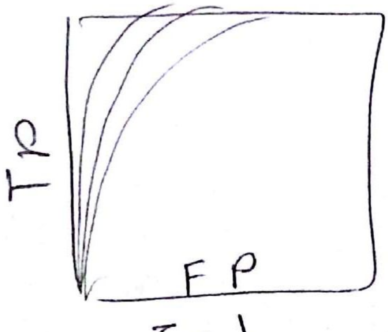
↳ outperformed others in almost all sequences.

Outlier

The author also showed algorithms robustness to outlier by ploting True positive rate vs false positive for some outliers in prev. experiment for $\tau = 0.1$ & $\tau = 1.0$



$\tau = 0.1$



$\tau = 1$

⇒ Algorithm is very robust to outlier

10