Multi-View Non-negative Matrix Factorization Discriminant Learning via Cross Entropy Loss

Abstract:

Key Words:

1. INTRODUCTION

0 1.

6

1)

2)

3)

2 3

2. RELATED WORK

( ) 1

1

10, 1 - 21

, 22 - 2

, 11, 26

( ) 2

2 - 30

1

3. THE PROPOSED METHOD

3.1. Non-negative Matrix Factorization

$X \in \mathbb{R}_+^{m\times n}$,

n m

$W \in \mathbb{R}_+^{m\times k}$ $H \in \mathbb{R}_+^{n\times k}$

$$X \approx WH^T \quad (1)$$

$$_{W,H} \left\| X - WH^T \right\|_F^2 \quad (2)$$

$$W, H \geq 0 \quad (3)$$

F

W

H. , W H

$$_{W,H} \sum_{v=1}^{n_v} \left\| X^{(v)} - WH^T \right\|_F^2 + \Phi(W,H) \quad ( )$$

$$W, H \geq 0 \quad ( )$$

n_v , $X^{(v)}$

v- . $\Phi(\cdot)$

W H.

3.2. Multi-view Learning via DICS

31 - 3 ,

W_C

$W_S^{(v)}$ v-

$X^{(v)} = W_C H_C^T + W_S^{(v)} H_S^{(v)T}$

$$W = W_{CD}\ W_{CN}\ W_{SD}^{(v)}\ W_{SN}^{(v)} \qquad (6)$$
$$H = H_{CD}\ H_{CN}\ H_{SD}^{(v)}\ H_{SN}^{(v)}$$

W_{CD} H_{CD} , W_{CN} H_{CN} . , $W_{SD}^{(v)}$ $H_{SD}^{(v)}$, $W_{SN}^{(v)}$ $H_{SN}^{(v)}$. , $B = B_{CD}\ B_{SD}^{(v)}$, H_{CD} $H_{SD}^{(v)}$. ,

$$\min_{W,H,B} \sum_{v=1}^{n_v} \left\| X^{(v)} - WH^T \right\|_F^2 + \Phi(W,H) + \gamma \left\| Y - \left[B_{CD}\ B_{SD}^{(v)} \right] \begin{bmatrix} H_{CD}^T \\ H_{SD}^{(v)T} \end{bmatrix} \right\|_F^2 \qquad (\ )$$
$$\text{s.t. } W, H \geq 0 \qquad (\ )$$

$Y \in \mathbb{R}^{c \times n}$, , $y_{i,j} = 1$ j - i $y_{i,j} = 0$ . .

3.3. Improved Objective Function via Cross-Entropy Loss

. 10 . , , 0 1, , . - 0 1. - . - , -

$$J = -\sum_{i=1}^{c} y_i \ (p_i) \qquad (\ )$$

c , y_i , , y_i 1 0. p_i . ,

$$\min_{W,H,B} \sum_{v=1}^{n_v} \left\| X^{(v)} - WH^T \right\|_F^2 + \alpha \left\| W_D^T W_D \right\|_{1,1} + \beta \left\| H_D \right\|_{1,1} - \gamma \sum_{j=1}^{n} \sum_{i=1}^{c} y_{ij} \ p_{ij} \qquad (10)$$

α, β, γ - n_v , n , c .

$\| \cdot \|_{1,1}$ $l_{1,1}$ W_D , $\left\| W_D^T W_D \right\|_{1,1} = \sum_i w_{Di}^T w_{Di} + \sum_{i \neq j} w_{Di}^T w_{Dj}$. , , . , $l_{1,1}$ H_D . .

p_{ij} . ,

$$p_{ij} = \frac{e^{\sum_{k=1}^{k_1} b_{CD_{i,k}}\ h_{CD_{j,k}} + \sum_{k=1}^{k_3} b_{SD_{i,k}}^{(v)}\ h_{SD_{j,k}}^{(v)}}}{\sum_{t=1}^{n} e^{\sum_{k=1}^{k_1} b_{CD_{t,k}}\ h_{CD_{j,k}} + \sum_{k=1}^{k_3} b_{SD_{t,k}}^{(v)}\ h_{SD_{j,k}}^{(v)}}} \qquad (11)$$

$$f(W,H,B)$$
$$\sum_{v=1}^{n_v} \left\| X^{(v)} \ \sum_{i=1}^{k1} w_{CD_i} h_{CD_i}^T \ \sum_{i=1}^{k2} w_{CN_i} h_{CN_i}^T \ \sum_{i=1}^{k3} w_{SD_i}^{(v)} h_{SD_i}^{(v)T} \ \sum_{i=1}^{k} w_{SN_i}^{(v)} h_{SN_i}^{(v)T} \right\|_F^2$$
$$\alpha \quad \sum_{i\ 1}^{k1} \sum_{j\ 1}^{k1} w_{CD_i}^T w_{CD_j} \quad \sum_{i\ 1}^{k3} \sum_{j\ 1}^{k3} w_{SD_i}^{(v)T} w_{SD_j}^{(v)} \quad 2 \sum_{i\ 1}^{k1} \sum_{j\ 1}^{k3} w_{CD_i}^T w_{SD_j}^{(v)}$$
$$\beta 1_{1\ n} \quad \sum_{i} h_{CD_i}$$

$1_{1\times n}$ n

1.

$$\frac{\partial L}{\partial h_{CD_{ij}}} = -\sum_{j=1}^{n}\sum_{i,t=1}^{c} y_{ij}\cdot\frac{1}{p_{tj}}\cdot\frac{\partial p_{tj}}{\partial z_{ij}}\cdot b_{CD_{i,j}}$$

$$= \sum_{j=1}^{n}\sum_{i,t=1}^{c}\Big(-\frac{y_{ij}}{p_{ij}}\cdot p_{ij}\cdot(1-p_{ij}) + \sum_{t\neq i}\frac{y_{tj}}{p_{tj}}\cdot p_{tj}\cdot p_{ij}\Big)\cdot b_{CD_{i,j}}$$

$$= \sum_{j=1}^{n}\sum_{i,t=1}^{c}(-y_{ij} + p_{ij}\sum_{t} y_{tj})\cdot b_{CD_{i,j}} \quad (1\ )$$

Y 0 1 , $\sum_{t} y_{tj} = 1$.

$h_{CD_{i,j}}$

$$\frac{\partial L}{\partial h_{CD_{ij}}} = \sum_{j=1}^{n}\sum_{i=1}^{c}(p_{ij} - y_{ij})\cdot b_{CD_{i,j}}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{c}\Bigg(\frac{e^{\sum_{k=1}^{k_1} b_{CD_{i,\bar{k}}} h_{CD_{j,k}} + \sum_{k=1}^{k_3} b^{(v)}_{SD_{i,\bar{k}}} h^{(v)}_{SD_{j,k}}}}{\sum_{t=1}^{c} e^{\sum_{k=1}^{k_1} b_{CD_{t,\bar{k}}} h_{CD_{j,k}} + \sum_{k=1}^{k_3} b^{(v)}_{SD_{t,\bar{k}}} h^{(v)}_{SD_{j,k}}}} - y_{ij}\Bigg)\cdot b_{CD_{i,j}} \quad (1\ )$$

h_{CD_i}

$\frac{\partial L}{\partial h_{CD_i}} = \sum_{i=1}^{n}(p_i - y_i)^T \cdot b_{CD_i}$, p_i

.

$$L = -\gamma\sum_{j=1}^{n}\sum_{i=1}^{c} y_{ij} \quad p_{ij}$$

$$p_{ij} = \frac{e^{z_{ij}}}{\sum_{t=1}^{n} e^{z_{tj}}} \quad (13)$$

$$z_{ij} = \sum_{k=1}^{k_1} b_{CD_{i,\bar{k}}}\, h_{CD_{j,k}} + \sum_{k=1}^{k_3} b^{(v)}_{SD_{i,\bar{k}}}\, h^{(v)}_{CD_{j,k}}$$

$h_{CD_{i,j}}$

$$\frac{\partial L}{\partial h_{CD_{ij}}} = \frac{\partial L}{\partial p_{tj}}\cdot\frac{\partial p_{tj}}{\partial z_{ij}}\cdot\frac{\partial z_{ij}}{\partial h_{CD_{ij}}} \quad (1\ )$$

,

.

$$\frac{\partial L}{\partial p_{tj}} = -\sum_{j=1}^{n}\sum_{i,t=1}^{c}\frac{y_{ij}}{p_{tj}} \quad (1\ )$$

(1) $t = i$

$$\frac{\partial p_{tj}}{\partial z_{ij}} = \frac{\partial\left(\frac{e^{z_{ij}}}{\sum_{t=1}^{c} e^{z_{tj}}}\right)}{\partial z_{ij}} = \frac{e^{z_{ij}}\cdot\sum_{t=1}^{c} e^{z_{tj}} - e^{z_{ij}}\cdot e^{z_{ij}}}{(\sum_{t=1}^{c} e^{z_{tj}})^2} = p_{ij}\cdot(1-p_{ij}) \quad (16)$$

(2) $t \neq i$

$$\frac{\partial p_{tj}}{\partial z_{ij}} = \frac{\partial\left(\frac{e^{z_{tj}}}{\sum_{t=1}^{c} e^{z_{tj}}}\right)}{\partial z_{ij}} = -\frac{e^{z_{tj}}}{(\sum_{t=1}^{c} e^{z_{tj}})^2}\cdot e^{z_{ij}} = -p_{tj}\cdot p_{ij} \quad (1\ )$$

$b_{CD_{i,j}}$.

$max(0,\cdot)$,

$+$

η

$$w_{CD_i} = w_{CD_i} + \eta\Big[\sum_{v=1}^{n_v}\Big(R^{(v)}h_{CD_i} - \alpha(W_{CD}1_{k1\times1} + W^{(v)}_{SD}1_{k3\times1})\Big)\Big]_+ \quad (20)$$

$$w_{CN_i} = w_{CN_i} + \eta\Big[\sum_{v=1}^{n_v} R^{(v)}h_{CN_i}\Big]_+ \quad (21)$$

$$w^{(v)}_{SD_i} = w^{(v)}_{SD_i} + \eta\Big[R^{(v)}h^{(v)}_{SD_i} - \alpha(W_{CD}1_{k1\times1} + W^{(v)}_{SD}1_{k3\times1})\Big]_+ \quad (22)$$

$$w^{(v)}_{SN_i} = w^{(v)}_{SN_i} + \eta\Big[R^{(v)}h^{(v)}_{SN_i}\Big]_+ \quad (23)$$

$$h_{CD_i} = h_{CD_i} + \eta\left[\sum_{v=1}^{n_v}\begin{pmatrix} R^{(v)}w_{CD_i} - \frac{\beta}{2}1_{n\times1} \\ -\frac{\gamma}{2}Q^{(v)T}b_{CD_i}\end{pmatrix}\right]_+ \quad (2\ )$$

$$h_{CN_i} = h_{CN_i} + \eta\Big[\sum_{v}^{n_v} R^{(v)T}w_{CN_i}\Big]_+ \quad (2\ )$$

$$h^{(v)}_{SD_i} = h^{(v)}_{SD_i} + \eta\Big[R^{(v)T}w^{(v)}_{SD_i} - \frac{\beta}{2}1_{n\times1} - \frac{\gamma}{2}Q^{(v)T}b^{(v)}_{SD}\Big]_+ \quad (26)$$

$$h^{(v)}_{SN_i} = h^{(v)}_{SN_i} + \eta\Big[R^{(v)T}w^{(v)}_{SN_i}\Big]_+ \quad (2\ )$$

$R^{(v)}$ $Q^{(v)}$

$$R^{(v)} = X^{(v)} - W_{CD}H_{CD}^T - W_{CD}H_{CD}^T - W_{SD}^{(v)}H_{SD}^{(v)T} - W_{SN}^{(v)}H_{SN}^{(v)T} \quad (2\ )$$

$$Q^{(v)} = \ (B_{CD}H_{CD}^T + B_{SD}^{(v)}H_{SD}^{(v)T}) - Y \quad (2\ )$$

η , , B_{CD} $B_{SD}^{(v)}$, B_{CD} $B_{SD}^{(v)}$

$$B_{CD} = \frac{1}{n_v}\sum_{v=1}^{n_v}\left(Y - B_{SD}^{(v)}H_{SD}^{(v)T}\right)H_{CD}\left(H_{CD}^T H_{CD} + \lambda I\right)^{-1} \quad (30)$$

$$B_{SD}^{(v)} = \left(Y - B_{CD}H_{CD}^T\right)H_{SD}^{(v)}\left(H_{SD}^{(v)T}H_{SD}^{(v)} + \lambda I\right)^{-1} \quad (31)$$

I λ .

4. EXPERIMENT

, , .

4.1. Datasets

, , , , , , , , 1.

1.

	1200		6	2000
	26	2		2016
	6			6 6 33 66 6
	1	2		1 03
	1	2		1 03 6 1
	230	2		1 036 0
	26	2		1 03

, , , , , , , . ($k = 1$)

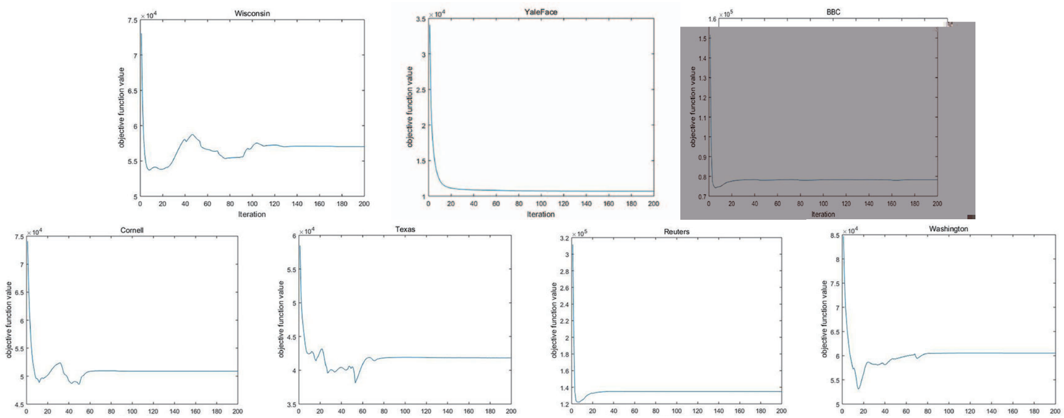
( ). (. .), . 3, λ 0. 0. 3 . 3, ( ) , λ $10^0, 10^1, 10^2, 10^3, 10$. 11, λ $10^0, 10^{-1}, 10^{-2}, 10^{-3}$. 26, α 100, β γ 0,100,200, 00,1000 . α 100 0 2 0, β 0.01, 0.02, 0.03 , γ 0.00 . , , , , α β $[0,1]$, γ 1.

4.2. Result

, 10 , . 2 . , , , , .

2.

	( )						
	0. 1.2	0.0 2.	3 .0 1.	1.0 1.	. 1.	6 6 2.2	2. 1.
	2. 0.2	6 .2 .2	3.1 0.2	. .	6 . 3.	.3 26	0.3 3.
	. 1.	33.3 6 .	. 26	6 0. .0	6 . .	6 2. 3.	6 .3 2.
	0.3 .0	.1 3.2	0.2 2.	2. 6 .1	16 .0	. 6 .0	.1 .
	0. 1.3	0.2 1.	1. 1.2	.3 2.2	3. 2.1	3. 2.2	36 1.



1.

4.3. Convergence Analysis

1,

100

0

4.4. Discriminant Matrix Visualization

( )

6

200 ,

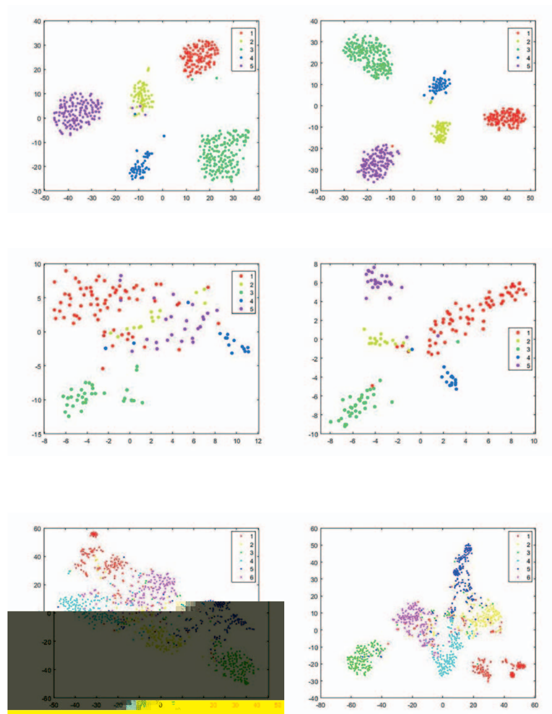
2 3

W_D

H_D ,

X .

2,



2.

5. Conclusion and Future Work

REFERENCES

[1] ... *Proceedings of the eleventh annual conference on Computational learning theory,* ... 1 ... 2-100.

[2] ... *Proceedings of the 24th international conference on Machine learning,* ... 200 ... 11 -166.

[3] ...

[1?] *Breakthroughs in statistics*, , 1 2, . 2-1 0.

[1?] ," *Advances in Neural Information Processing Systems*, , 200 , 2.

[1?] ," *Neural Computation*, 12, . 200 .

[20] ," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1, 1 1 . 201 .

[21] *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, . 21 0-21 .

[22] ," *ICML workshop on learning with multiple views*, 200 , . 20 2 .

[23] ," *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 201 , . 1 1-1 .

[2?] ," *IEEE Transactions on Systems, Man, and Cybernetics*, ( ), 0, . 6 , . 1 -1 6 , . 2010.

[2?] ," *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, 200 , . 11 1 6 .

[2?] ," *IEEE 16th International Conference on Data Mining (ICDM)*, 201 , . 12 12 0.

[2?] . Alpaydın, ," *Journal of machine learning research*, . 12, . 2211-22 , . 2011.

[2?] ," *Proceedings of the 28th International Conference on Machine Learning*, 2011, . 00.

[2?] ," *Proceedings of the 27th International Conference on Machine Learning*, 2010, . 11 11 2.

[30] ," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1, . 6 , . 12- 2 , . 2011.

[31] ," *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, . 1 11 .

[32] ," *Data Mining and Knowledge Discovery*, 1, . 201 .

[33]

," *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 201 , . 6 6 .

[3?] 12 12 6 (201)

[3?] ," *IEEE Signal Processing Letters*, 1, . 1, . 2010.

[3?] ," *Journal of machine learning research*, . 2 -2 0 , . 200 .