

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

Abstract:

Multi-view data has been widely used in many applications. However, the existing methods for multi-view data analysis are often based on the assumption that the data is linearly separable. In this paper, we propose a novel method for multi-view data analysis based on non-negative matrix factorization (NMF) and discriminant learning via cross entropy loss. The proposed method can handle non-linearly separable data and is able to learn discriminative features from the data. The experimental results show that the proposed method outperforms the existing methods in terms of classification accuracy.

Key Words:

1. INTRODUCTION

In recent years, multi-view data has become a popular topic in machine learning. Multi-view data refers to data that is collected from multiple different sources or perspectives. For example, in a medical diagnosis system, a patient's data might be collected from multiple different sensors or tests. In a recommendation system, a user's preferences might be collected from multiple different sources, such as their browsing history, purchase history, and social media activity. Multi-view data is often more informative than single-view data, but it is also more challenging to analyze. One of the main challenges is that the data is often high-dimensional and sparse. Another challenge is that the data is often noisy and contains outliers. In this paper, we propose a novel method for multi-view data analysis based on non-negative matrix factorization (NMF) and discriminant learning via cross entropy loss. The proposed method can handle non-linearly separable data and is able to learn discriminative features from the data. The experimental results show that the proposed method outperforms the existing methods in terms of classification accuracy.

The proposed method is based on the idea of learning discriminative features from the data. In NMF, the data matrix is decomposed into two non-negative matrices, W and H , such that $V \approx WH$. The matrix W represents the basis features, and the matrix H represents the coefficients of these features. In discriminant learning, the goal is to learn a set of features that are discriminative for the different classes. This is done by minimizing the cross entropy loss, which is a measure of the difference between the predicted and actual class probabilities. The proposed method combines NMF and discriminant learning to learn discriminative features from the data. The experimental results show that the proposed method outperforms the existing methods in terms of classification accuracy.

The proposed method is based on the idea of learning discriminative features from the data. In NMF, the data matrix is decomposed into two non-negative matrices, W and H , such that $V \approx WH$. The matrix W represents the basis features, and the matrix H represents the coefficients of these features. In discriminant learning, the goal is to learn a set of features that are discriminative for the different classes. This is done by minimizing the cross entropy loss, which is a measure of the difference between the predicted and actual class probabilities. The proposed method combines NMF and discriminant learning to learn discriminative features from the data. The experimental results show that the proposed method outperforms the existing methods in terms of classification accuracy.

The proposed method is based on the idea of learning discriminative features from the data. In NMF, the data matrix is decomposed into two non-negative matrices, W and H , such that $V \approx WH$. The matrix W represents the basis features, and the matrix H represents the coefficients of these features. In discriminant learning, the goal is to learn a set of features that are discriminative for the different classes. This is done by minimizing the cross entropy loss, which is a measure of the difference between the predicted and actual class probabilities. The proposed method combines NMF and discriminant learning to learn discriminative features from the data. The experimental results show that the proposed method outperforms the existing methods in terms of classification accuracy.

2010-2012

0 1.

2 - 0

1

1

1

2

3. THE PROPOSED METHOD

3.1. Non-negative Matrix Factorization

$$\begin{aligned}
 & X \in \mathbb{R}_+^{m \times n}, \\
 & W \in \mathbb{R}_+^{m \times k} \quad H \in \mathbb{R}_+^{n \times k} \\
 & X \approx WH^T \tag{1}
 \end{aligned}$$

$$\begin{aligned}
 & \min_{W,H} \|X - WH^T\|_F^2 \\
 & \dots W, H \geq 0 \tag{2}
 \end{aligned}$$

$$H \cdot W, W \cdot H$$

2. RELATED WORK

1.

$$\begin{aligned}
 & \sum_{v=1}^{n_v} \|X^v - WH^T\|_F^2 + \Phi(W, H) \\
 & \dots W, H \geq 0
 \end{aligned}$$

$$\begin{aligned}
 & n_v \quad X^v \\
 & \dots \Phi \\
 & W \quad H
 \end{aligned}$$

3.2. Multi-view Learning via DICS

1 - ,

$$\begin{aligned}
 & W_C \\
 & W_S^v \\
 & X^v = W_C H_C^T + W_S^v H_S^{vT}
 \end{aligned}$$

$$\begin{aligned}
W &= W_{CD} W_{CN} W_{SD}^v W_{SN}^v \\
H &= H_{CD} H_{CN} H_{SD}^v H_{SN}^v \\
B &= \begin{bmatrix} W_{CD} & H_{CD} \\ W_{CN} & H_{CN} \end{bmatrix} \begin{bmatrix} W_{SD}^v & H_{SD}^v \\ W_{SN}^v & H_{SN}^v \end{bmatrix}
\end{aligned}$$

$$B = \begin{bmatrix} B_{CD} & B_{SD}^v \\ B_{CN} & B_{SN}^v \end{bmatrix} \begin{bmatrix} H_{CD} & H_{SD}^v \\ H_{CN} & H_{SN}^v \end{bmatrix}$$

$$\begin{aligned}
& \sum_{v=1}^{n_v} \|X^v - WH^T\|_F^2 + \Phi(W, H) \\
& + \gamma \left\| Y - \begin{bmatrix} B_{CD} & B_{SD}^v \end{bmatrix} \begin{bmatrix} H_{CD}^T \\ H_{SD}^{vT} \end{bmatrix} \right\|_F^2 \\
& \dots W, H \geq 0
\end{aligned}$$

$$Y \in \mathbb{R}^{c \times n}$$

$$\begin{aligned}
y_{i,j} &= 1 \quad j = i \\
y_{i,j} &= 0 \quad j \neq i
\end{aligned}$$

3.3. Improved Objective Function via Cross-Entropy Loss

$$f(W, H, B) = \sum_{v=1}^{n_v} \|X^v - WH^T\|_F^2 + \Phi(W, H) \quad (10)$$

$$W_{CD_i}^T W_{CD_j} = \delta_{ij} \quad 0 \leq i, j \leq c-1 \quad (11)$$

$$W_{SD_i}^v W_{SD_j}^v = \delta_{ij} \quad 0 \leq i, j \leq c-1 \quad (12)$$

$$J = -\sum_{i=1}^c y_i \log p_i$$

$$\begin{aligned}
& \dots y_i \\
& \dots y_i \\
& 0. \quad p_i
\end{aligned}$$

$$\begin{aligned}
& \sum_{v=1}^{n_v} \|X^v - WH^T\|_F^2 + \alpha \|W_D^T W_D\|_{1,1} \\
& + \beta \|H_D\|_{1,1} - \gamma \sum_{j=1}^n \sum_{i=1}^c y_{ij} \log p_{ij}
\end{aligned} \quad (10)$$

$$\alpha, \beta, \gamma \geq 0, \quad n_v, c, n$$

$$\begin{aligned}
& \|\cdot\|_{1,1} = \sum_i \|w_{Di}\|_{1,1} \\
& \|W_D^T W_D\|_{1,1} = \sum_i w_{Di}^T w_{Di} + \sum_{i \neq j} w_{Di}^T w_{Dj}
\end{aligned}$$

$$p_{ij} = \frac{e^{\sum_{k=1}^{k_1} b_{CD_i,k} h_{CD_j,k} + \sum_{k=1}^{k_2} b_{SD_i,k}^v h_{SD_j,k}^v}}{\sum_{t=1}^n e^{\sum_{k=1}^{k_1} b_{CD_t,k} h_{CD_j,k} + \sum_{k=1}^{k_2} b_{SD_t,k}^v h_{SD_j,k}^v}} \quad (11)$$

$$f(W, H, B)$$

$$\begin{aligned}
& \sum_{v=1}^{n_v} \left\| \begin{bmatrix} X^v \\ W_{SD_i}^v h_{SD_i}^{vT} \\ W_{CD_i}^T W_{CD_j} \end{bmatrix} - \begin{bmatrix} W_{CN_i} h_{CN_i}^T \\ W_{SN_i}^v h_{SN_i}^{vT} \\ W_{SD_i}^v W_{SD_j}^v \end{bmatrix} \right\|_F^2 \\
& + \alpha \sum_{i=1}^{c-1} \sum_{j=1}^{c-1} \left\| \begin{bmatrix} W_{CD_i}^T W_{CD_j} \\ W_{SD_i}^v W_{SD_j}^v \end{bmatrix} \right\|_F^2 \\
& + 2 \sum_{i=1}^{c-1} \sum_{j=1}^{c-1} w_{CD_i}^T w_{SD_j}^v
\end{aligned}$$

$$\beta \mathbf{1}_n \cdot h_{CD_i}$$

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

$$p_{ij} = \frac{e^{z_{ij}}}{\sum_{t=1}^c e^{z_{ij}}}$$

$$z_{ij} = \sum_{k=1}^{k_1} b_{CD_i,k} \cdot h_{CD_j,k} + \sum_{k=1}^k b_{SD_i,k}^v \cdot h_{CD_j,k}^v$$

$$\frac{\partial L}{\partial h_{CD_j}} = \frac{\partial L}{\partial p_{ij}} \cdot \frac{\partial p_{ij}}{\partial z_{ij}} \cdot \frac{\partial z_{ij}}{\partial h_{CD_j}}$$

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

1 $t = i$

$$\begin{aligned} & \frac{\partial p_{ij}}{\partial z_{ij}} \\ &= \frac{\partial \left(\frac{e^{z_{ij}}}{\sum_{t=1}^c e^{z_{ij}}} \right)}{\partial z_{ij}} \\ &= \frac{e^{z_{ij}} \cdot \sum_{t=1}^c e^{z_{ij}} - e^{z_{ij}} \cdot e^{z_{ij}}}{\left(\sum_{t=1}^c e^{z_{ij}} \right)^2} \\ &= p_{ij} \cdot (1 - p_{ij}) \end{aligned}$$

2 $t \neq i$

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

$$b_{CD_i,j}$$

$$\begin{aligned} \frac{\partial L}{\partial h_{CD_j}} &= -\sum_{j=1}^n \sum_{i=1}^c y_{ij} \cdot \frac{1}{p_{ij}} \cdot \frac{\partial p_{ij}}{\partial z_{ij}} \cdot b_{CD_i,j} \\ &= \sum_{j=1}^n \sum_{i=1}^c -\frac{y_{ij}}{p_{ij}} \cdot p_{ij} \cdot (1 - p_{ij}) + \sum_{i \neq j} \frac{y_{ij}}{p_{ij}} \cdot p_{ij} \cdot p_{ij} \cdot b_{CD_i,j} \\ &= \sum_{j=1}^n \sum_{i=1}^c -y_{ij} + p_{ij} \sum_i y_{ij} \cdot b_{CD_i,j} \end{aligned}$$

1

$$\sum_t y_{ij} = 1$$

$$\begin{aligned} \frac{\partial L}{\partial h_{CD_j}} &= \sum_{j=1}^n \sum_{i=1}^c p_{ij} - y_{ij} \cdot b_{CD_i,j} \\ &= \sum_{j=1}^n \sum_{i=1}^c \frac{e^{\sum_{k=1}^{k_1} b_{CD_i,k} \cdot h_{CD_j,k} + \sum_{k=1}^k b_{SD_i,k}^v \cdot h_{SD_j,k}^v}}{\sum_{t=1}^c e^{\sum_{k=1}^{k_1} b_{CD_i,k} \cdot h_{CD_j,k} + \sum_{k=1}^k b_{SD_i,k}^v \cdot h_{SD_j,k}^v}} \\ &\quad - y_{ij} \cdot b_{CD_i,j} \end{aligned}$$

1

$$\begin{aligned} \frac{\partial L}{\partial h_{CD_i}} &= \sum_{i=1}^n p_i - y_i^T \cdot b_{CD_i} \\ &= \max(0, \cdot) \end{aligned}$$

η

$$w_{CD_i} = w_{CD_i} + \eta \left[\sum_{v=1}^{n_v} \left(R^v h_{CD_i} - \alpha W_{CD} 1_{k \times 1} + W_{SD}^v 1_{k \times 1} \right) \right]_+ \quad 20$$

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

$$w_{SD_i}^v = w_{SD_i}^v + \eta \left[R^v h_{SD_i}^v - \alpha W_{CD} 1_{k \times 1} + W_{SD}^v 1_{k \times 1} \right]_+ \quad 22$$

$$w_{SN_i}^v = w_{SN_i}^v + \eta \left[R^v h_{SN_i}^v \right]_+ \quad 2$$

$$h_{CD_i} = h_{CD_i} + \eta \left[\sum_{v=1}^{n_v} \left(R^v w_{CD_i} - \frac{\beta}{2} 1_{n \times 1} \right) \right]_+ \quad 2$$

$$h_{CN_i} = h_{CN_i} + \eta \left[\sum_v R^v w_{CN_i} \right]_+ \quad 2$$

$$h_{SD_i}^v = h_{SD_i}^v + \eta \left[R^v w_{SD_i}^v - \frac{\beta}{2} 1_{n \times 1} - \frac{\gamma}{2} Q^v w_{SD_i}^v \right]_+ \quad 2$$

$$h_{SN_i}^v = h_{SN_i}^v + \eta \left[R^v w_{SN_i}^v \right]_+ \quad 2$$

$$R^v \quad Q^v$$

$$\begin{aligned}
 R^v - X^v - W_{CD}H_{CD}^T - W_{CD}H_{CD}^T & \quad 2 \\
 -W_{SD}^v H_{SD}^{vT} - W_{SN}^v H_{SN}^{vT} & \\
 Q^v = & \quad B_{CD}H_{CD}^T + B_{SD}^v H_{SD}^{vT} - Y \quad 2 \\
 \eta &
 \end{aligned}$$

$$\begin{aligned}
 & \quad B_{CD} \quad B_{SD}^v \\
 & \quad B_{CD} \quad B_{SD}^v
 \end{aligned}$$

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

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

4. EXPERIMENT

4.1. Datasets

	1200		2000
	2	2	201
	1	2	1 0
	1	2	1 0 1
	2 0	2	1 0 0
	2	2	1 0

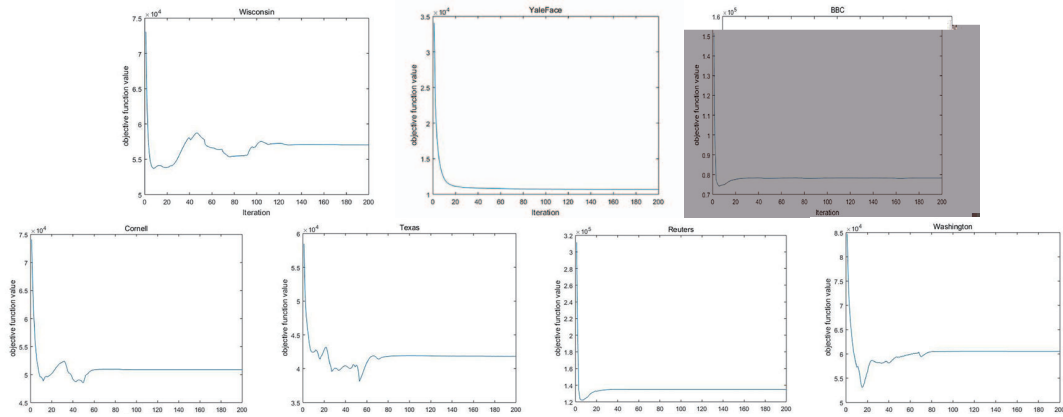
$k = 1$

λ
 $0.1, 0.01, \dots$
 λ
 $10^0, 10^1, 10^2, 10^{-1}, 10^{-2}, \dots$
 $11, \dots$
 λ
 $10^0, 10^{-1}, 10^{-2}, 10^{-3}, \dots$
 $2, \dots$
 $\alpha = 100, \dots, \beta$
 $\gamma = 0, 100, 200, 500, 1000, \dots$
 α
 $100, 0.2, 0, \dots, \beta$
 $0.01, 0.02, 0.0, \dots, \gamma = 0.00$

4.2. Result

α
 β
 $[0,1], \dots, \gamma = 1.$
 $10,$
 2

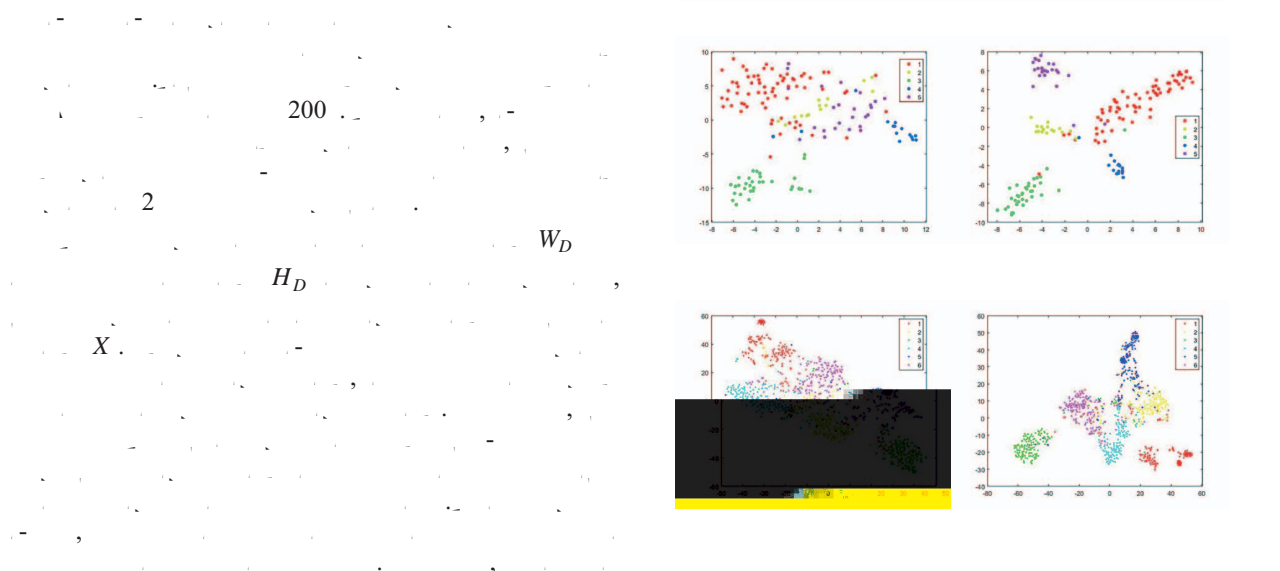
	0. 1.2	0.0 2.	.0 1.	1.0 1.	. 1.	. 2.2	2. 1.
	2. 0.2	.2 .2	.1 0.2 2.	0. .
	. 1.	. .	. 2.	0. .0	. .	2. .	. 2.
	0. .0	.1 .2	0.2 2.	2. .1	1. .0	. .0	.1 .
	0. 1.	0.2 1.	1. 1.2	. 2.2	. 2.1	. 2.2	. 1.



4.3. Convergence Analysis



4.4. Discriminant Matrix Visualization



2. *Journal of Machine Learning Research*, 2003, 4:1151-1190.

3. *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 11-11.

4. *Journal of Machine Learning Research*, 2008, 9:11-11.

2. *Journal of Machine Learning Research*, 2003, 4:1151-1190.

3. *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 11-11.

4. *Journal of Machine Learning Research*, 2008, 9:11-11.

5. Conclusion and Future Work

In this paper, we have presented a new method for learning linear models. The method is based on the idea of using a linear model to approximate the non-linear relationship between the input and output variables. The method is simple and efficient, and it can be applied to a wide range of problems. The results of the experiments show that the method is effective and can be used as a practical tool for learning linear models.

REFERENCES

1. *Journal of Machine Learning Research*, 2003, 4:1151-1190.
2. *Proceedings of the eleventh annual conference on Computational learning theory*, 2000, pp. 2-100.

1 *Breakthroughs in statistics*, 1, 2, 1 2-1 0.

1 *Advances in Neural Information Processing Systems*, 200, 2.

1 *Neural Computation*, 1, 12, 2 2, 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.

2 *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 201, 1 1-1.

2 *IEEE Transactions on Systems, Man, and Cybernetics*, 0, 1 -1, 2010.

2 *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, 200, 11 11.

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

2 Alpaydin, *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.

0 *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1, 1 12-1 2, 2011.

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

2 *Data Mining and Knowledge Discovery*, 2, 1, 201.

Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 201, 2, 12 12 201.

IEEE Signal Processing Letters, 1, 1, 2010.

Journal of machine learning research, 2, 2 0, 200.