

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 achieve better performance than the existing methods. The experimental results show that the proposed method can effectively learn the discriminative information from the multi-view data and achieve better performance than the existing methods.

Key Words:

1. INTRODUCTION

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 achieve better performance than the existing methods. The experimental results show that the proposed method can effectively learn the discriminative information from the multi-view data and achieve better performance than the existing methods.

(26 201 02)

2. RELATED WORK

(1) [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100].

3. THE PROPOSED METHOD

3.1. Non-negative Matrix Factorization

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

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

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

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

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

$$\min_{W, H} \sum_{v=1}^{n_v} \|X^{(v)} - WH^T\|_F^2 + \Phi(W, H) \quad (4)$$

$$W, H \geq 0 \quad (5)$$

$$\min_{W, H} \sum_{v=1}^{n_v} \|X^{(v)} - WH^T\|_F^2 + \Phi(W, H) \quad (6)$$

$$W, H \geq 0 \quad (7)$$

$$\min_{W, H} \sum_{v=1}^{n_v} \|X^{(v)} - WH^T\|_F^2 + \Phi(W, H) \quad (8)$$

$$W, H \geq 0 \quad (9)$$

3.2. Multi-view Learning via DICS

$$X^{(v)} = W_C H_C^T + W_S^{(v)} H_S^{(v)T} \quad (10)$$

$$\begin{aligned}
W &= W_{CD} W_{CN} W_{SD}^{(v)} W_{SN}^{(v)} \\
H &= H_{CD} H_{CN} H_{SD}^{(v)} H_{SN}^{(v)} \\
B &= B_{CD} B_{SD}^{(v)}
\end{aligned}
\quad (6)$$

$$\begin{aligned}
&\sum_{v=1}^{n_v} \|X^{(v)} - WH^T\|_F^2 + \Phi(W, H) \\
&+ \gamma \left\| Y - \begin{bmatrix} B_{CD} & B_{SD}^{(v)} \end{bmatrix} \begin{bmatrix} H_{CD}^T \\ H_{SD}^{(v)T} \end{bmatrix} \right\|_F^2 \\
&\dots W, H \geq 0
\end{aligned}
\quad (7)$$

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

3.3. Improved Objective Function via Cross-Entropy Loss

The objective function is reformulated using cross-entropy loss. The matrix Y is defined as $Y_{ij} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. The objective function is then expressed as:

$$J = -\sum_{i=1}^c y_i \ln(p_i) \quad (8)$$

where p_i is the probability of the i -th class, and y_i is the corresponding label. The constraint $\sum_{i=1}^c p_i = 1$ is imposed.

$$\begin{aligned}
&\sum_{v=1}^{n_v} \|X^{(v)} - WH^T\|_F^2 + \alpha \|W_D^T W_D\|_{l_1} \\
&+ \beta \|H_D\|_{l_1} - \gamma \sum_{j=1}^n \sum_{i=1}^c y_{ij} p_{ij}
\end{aligned}
\quad (10)$$

where α, β, γ are regularization parameters. The l_1 norm is used for sparsity. The matrix W_D is defined as $W_D = [W_{CD} \quad W_{CN} \quad W_{SD}^{(v)} \quad W_{SN}^{(v)}]$.

$$\|W_D^T W_D\|_{l_1} = \sum_i w_{Di}^T w_{Di} + \sum_{i \neq j} w_{Di}^T w_{Dj}$$

The matrix H_D is defined as $H_D = [H_{CD} \quad H_{CN} \quad H_{SD}^{(v)} \quad H_{SN}^{(v)}]$. The matrix p_{ij} is the joint probability distribution.

$$p_{ij} = \frac{e^{\sum_{k=1}^{k_1} b_{CDi,k} h_{CDj,k} + \sum_{k=1}^{k_3} b_{SDi,k}^{(v)} h_{SDj,k}^{(v)}}}{\sum_{t=1}^n e^{\sum_{k=1}^{k_1} b_{CDt,k} h_{CDj,k} + \sum_{k=1}^{k_3} b_{SDt,k}^{(v)} h_{SDj,k}^{(v)}}} \quad (11)$$

The objective function $f(W, H, B)$ is then defined as:

$$\begin{aligned}
&\sum_{v=1}^{n_v} \left\| X^{(v)} - \begin{bmatrix} \sum_{i=1}^{k_1} w_{CDi} h_{CDi}^T & \sum_{i=1}^{k_2} w_{CNi} h_{CNi}^T \\ \sum_{i=1}^{k_3} w_{SDi}^{(v)} h_{SDi}^{(v)T} & \sum_{i=1}^k w_{SNi}^{(v)} h_{SNi}^{(v)T} \end{bmatrix} \right\|_F^2 \\
&+ \alpha \sum_{i=1}^{k_1} \sum_{j=1}^{k_1} w_{CDi}^T w_{CDj} + \sum_{i=1}^{k_3} \sum_{j=1}^{k_3} w_{SDi}^{(v)T} w_{SDj}^{(v)} \\
&+ 2 \sum_{i=1}^{k_1} \sum_{j=1}^{k_3} w_{CDi}^T w_{SDj}^{(v)} \\
&+ \beta \sum_{i=1}^n h_{CDi}
\end{aligned}$$

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

$$L = -\gamma \sum_{j=1}^n \sum_{i=1}^c y_{ij} \cdot p_{ij} \quad (13)$$

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

$$z_{ij} = \sum_{k=1}^{k_1} b_{CD_i, \bar{k}} h_{CD_j, k} + \sum_{k=1}^{k_2} b_{SD_i, \bar{k}}^{(v)} h_{CD_j, k}^{(v)}$$

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

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

$$(1) \quad t = i$$

$$\frac{\partial p_{ij}}{\partial z_{ij}} = \frac{\partial \left(\frac{e^{z_{ij}}}{\sum_{t=1}^c e^{z_{ij}}} \right)}{\partial z_{ij}} \quad (6)$$

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

$$(2) \quad t \neq i$$

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

$$b_{CD_i, j}$$

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

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

$$\begin{aligned} \frac{\partial L}{\partial h_{CD_j}} &= \sum_{j=1}^n \sum_{i=1}^c (p_{ij} - y_{ij}) \cdot b_{CD_i, j} \\ &= \sum_{j=1}^n \sum_{i=1}^c \left(\frac{e^{\sum_{k=1}^{k_1} b_{CD_i, \bar{k}} h_{CD_j, k} + \sum_{k=1}^{k_2} b_{SD_i, \bar{k}}^{(v)} h_{SD_j, k}^{(v)}}}{\sum_{t=1}^c e^{\sum_{k=1}^{k_1} b_{CD_i, \bar{k}} h_{CD_j, k} + \sum_{k=1}^{k_2} b_{SD_i, \bar{k}}^{(v)} h_{SD_j, k}^{(v)}}} - y_{ij} \right) \cdot b_{CD_i, j} \end{aligned} \quad (1)$$

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

$$\max(0, \cdot) \cdot \eta$$

$$w_{CD_i} = w_{CD_i} + \eta \left[\sum_{v=1}^{n_v} \left(R^{(v)} h_{CD_i} - \alpha (W_{CD} 1_{k_1 \times 1} + W_{SD}^{(v)} 1_{k_2 \times 1}) \right) \right]_+ \quad (20)$$

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

$$w_{SD_i}^{(v)} = w_{SD_i}^{(v)} + \eta \left[R^{(v)} h_{SD_i}^{(v)} - \alpha (W_{CD} 1_{k_1 \times 1} + W_{SD}^{(v)} 1_{k_2 \times 1}) \right]_+ \quad (22)$$

$$w_{SN_i}^{(v)} = w_{SN_i}^{(v)} + \eta \left[R^{(v)} h_{SN_i}^{(v)} \right]_+ \quad (23)$$

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

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

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

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

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

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

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

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

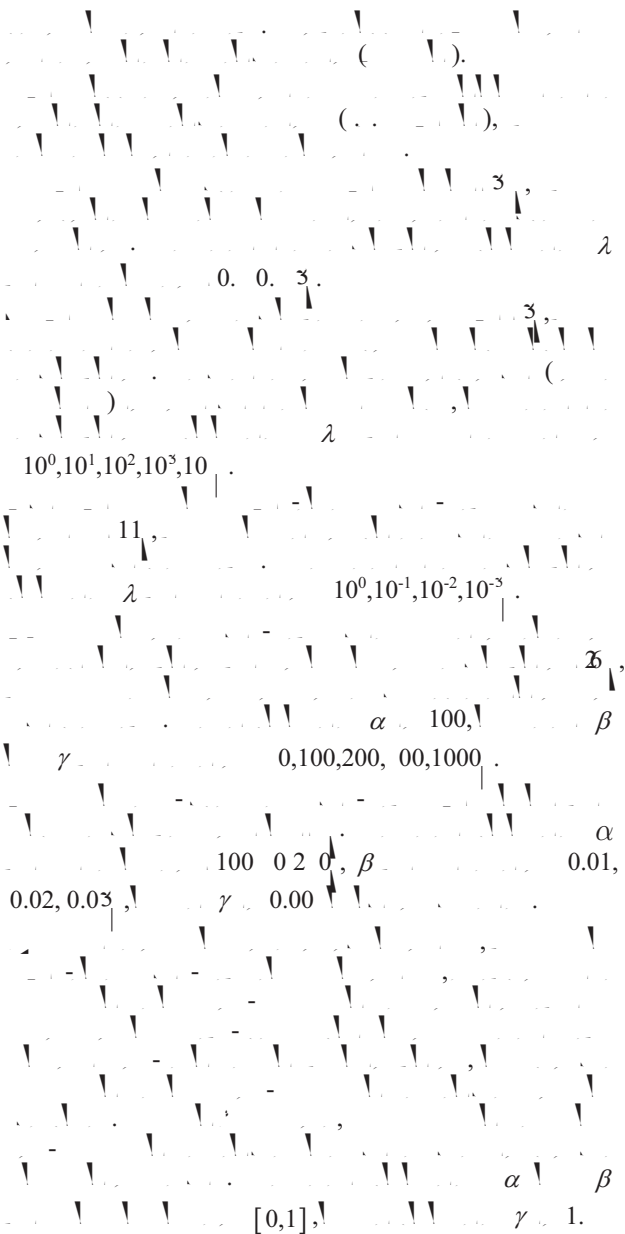
$$B_{SD}^{(v)} = (Y - B_{CD}H_{CD}^T) H_{SD}^{(v)} (H_{SD}^{(v)T} H_{SD}^{(v)} + \lambda I)^{-1} \quad (31)$$

4. EXPERIMENT

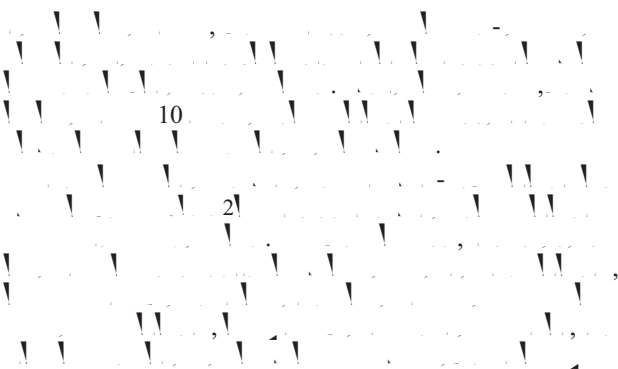
4.1. Datasets

Dataset	Number of samples	Number of features	Number of classes
1	1200	6	2000
2	26	2	206
3	6		6 6 33 66 6
4	1	2	1 03
5	1	2	1 03 6 1
6	230	2	1 036 0
7	3	2	1 03

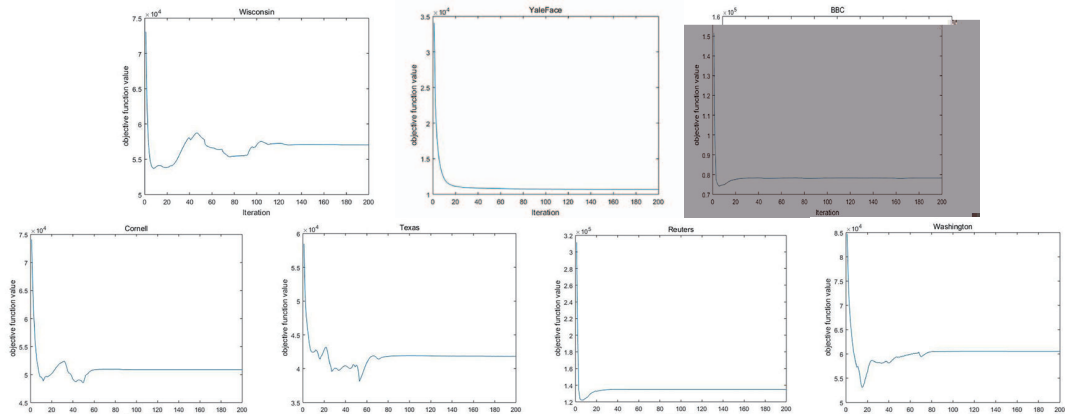
(k=1)



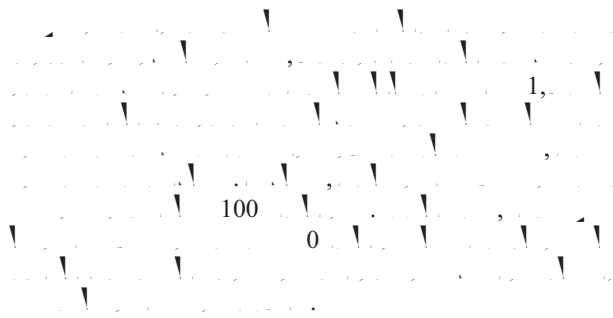
4.2. Result



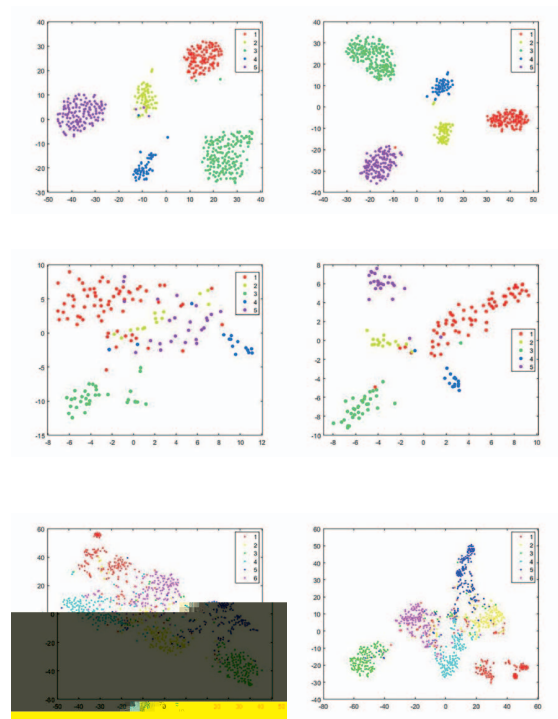
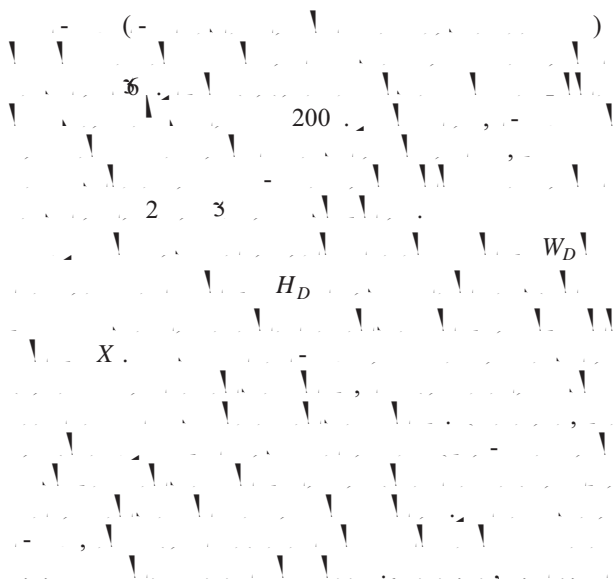
	()						
	0. 1.2	0.0 2.	3.0 1.	1.0 1.	. 1.	6 6 2.2	2. 1.
	2. 0.2	6 .2 .2	3.1 0.2	. .	6 . 3.	.3 26	0.3 3.
	. 1.	33.3 6 .	. 26	6 0. .0	6 . .	6 2. 3.	6 .3 2.
	0.3 .0	.1 3.2	0.2 2.	2. 6.1	16 .0	. 6.0	.1 .
	0. 1.3	0.2 1.	1. 1.2	.3 2.2	3. 2.1	3. 2.2	36 1.



4.3. Convergence Analysis



4.4. Discriminant Matrix Visualization



2. ...
 3. ...

2. ...

5. Conclusion and Future Work

...

REFERENCES

1. ...
 2. ...

1 *Breakthroughs in statistics*, ... 1 2, ... 6 2-1 0.

1 *Advances in Neural Information Processing Systems*, ... 200, ... 3 6 2.

1 *Neural Computation*, ... 6, ... 12, ... 3 3 6, ... 200.

20 *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ... 3, ... 1, ... 1 1, ... 2016.

21 *2012 IEEE Conference on Computer Vision and Pattern Recognition*, ... 2012, ... 26 0-26.

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

23 *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, ... 2016, ... 1 1-1.

2 *IEEE Transactions on Systems, Man, and Cybernetics*, ... (...), ... 0, ... 6, ... 1 3 -1 6, ... 2010.

2 *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, ... 200, ... 11 166.

28 *IEEE 16th International Conference on Data Mining (ICDM)*, ... 2016, ... 12 12 0.

2 Alpaydin, ... *Journal of machine learning research*, ... 12, ... 2211-226, ... 2011.

2 *Proceedings of the 28th International Conference on Machine Learning*, ... 2011, ... 3 3 00.

2 *Proceedings of the 27th International Conference on Machine Learning*, ... 2010, ... 113 11 2.

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

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

32 *Data Mining and Knowledge Discovery*, ... 6, ... 1, ... 2013.

33 *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ... 2011, ... 6 6.

3 *IEEE Signal Processing Letters*, ... 1, ... 1, ... 2010.

6 *Journal of machine learning research*, ... 2 -6 0, ... 200.