

## Unsupervised feature selection using orthogonal decomposition and low-rank approximation

Hyunki Lim\*

\*Assistant Professor, Div. of AI Computer Science and Engineering, Kyonggi University, Suwon, Korea

### [Abstract]

In this paper, we propose a novel unsupervised feature selection method. Conventional unsupervised feature selection method defines virtual label and uses a regression analysis that projects the given data to this label. However, since virtual labels are generated from data, they can be formed similarly in the space. Thus, in the conventional method, the features can be selected in only restricted space. To solve this problem, in this paper, features are selected using orthogonal projections and low-rank approximations. To solve this problem, in this paper, a virtual label is projected to orthogonal space and the given data set is also projected to this space. Through this process, effective features can be selected. In addition, projection matrix is restricted low-rank to allow more effective features to be selected in low-dimensional space. To achieve these objectives, a cost function is designed and an efficient optimization method is proposed. Experimental results for six data sets demonstrate that the proposed method outperforms existing conventional unsupervised feature selection methods in most cases.

▶ **Key words:** Feature selection, Unsupervised learning, Low-rank approximation, Orthogonal projection, Regularization

### [요 약]

본 논문에서는 새로운 비지도 특징 선별 기법을 제안한다. 기존 비지도 방식의 특징 선별 기법들은 특징을 선별하기 위해 가상의 레이블 데이터를 정하고 주어진 데이터를 이 레이블 데이터에 사영하는 회귀 분석 방식으로 특징을 선별하였다. 하지만 가상의 레이블은 데이터로부터 생성되기 때문에 사영된 공간이 비슷하게 형성될 수 있다. 따라서 기존의 방법들에서는 제한된 공간에서만 특징이 선택될 수 있었다. 이를 해소하기 위해 본 논문에서는 직교 사영과 저랭크 근사를 이용하여 특징을 선별한다. 이 문제를 해소하기 위해 가상의 레이블을 직교 사영하고 이 공간에 데이터를 사영할 수 있도록 한다. 이를 통해 더 주요한 특징 선별을 기대할 수 있다. 그리고 사영을 위한 변환 행렬에 저랭크 제한을 두어 더 효과적으로 저차원 공간의 특징을 선별할 수 있도록 한다. 이 목표를 달성하기 위해 본 논문에서는 비용 함수를 설계하고 효율적인 최적화 방법을 제안한다. 여섯 개의 데이터에 대한 실험 결과는 제안된 방법이 대부분의 경우 기존의 비지도 특징 선별 기법보다 좋은 성능을 보여주었다.

▶ **주제어:** 특징 선별, 비지도 학습, 저랭크 근사, 직교 사영, 정규화

- 
- First Author: Hyunki Lim, Corresponding Author: Hyunki Lim
  - Hyunki Lim (hlim20@kyonggi.ac.kr), Div. of AI Computer Science and Engineering, Kyonggi University
  - Received: 2022. 04. 25, Revised: 2022. 05. 10, Accepted: 2022. 05. 11.

## I. Introduction

데이터 마이닝, 패턴 인식 및 컴퓨터 비전과 같은 인공지능 연구 주제에서 고차원 데이터를 자주 접하게 된다 [1]. 이러한 데이터는 중복되는 정보를 포함할 뿐만 아니라 알고리즘의 실행 시간이 오래 걸리고 메모리 요구 사항이 크기 때문에 다양한 기계학습 과정에서 처리하기 어렵다 [2]. 또한 이러한 문제는 학습 알고리즘의 성능을 저하시키고 응용을 어렵게 한다. 이 문제를 극복하기 위해 많은 연구들에서 특징 선별 기법을 도입하였다. 특징 선별 기법은 특정 평가 방법에 따라 중요한 특징을 선택하고 주어진 특징 집합에서 불필요한 특징을 제거하는 알고리즘이다. 특징 선별의 결과로, 선택된 특징 집합은 기계 학습 과정에 성능적, 시간적인 도움이 되고, 데이터의 해석력을 증가시키며, 차원의 저주, 과적합 위험 등의 비용을 줄일 수 있다 [3].

특징 선별 방법은 클래스 정보의 유무에 따라 지도 학습, 비지도 학습, 준지도 학습, 세 가지 범주로 나눌 수 있다 [4]. 지도 학습의 경우, 레이블 정보를 사용하여 주요한 특징을 선별한다. 준지도 학습의 경우, 적은 수의 레이블링 된 데이터와 많은 수의 레이블링되지 않은 데이터를 이용하여 특징을 선별한다. 비지도 학습의 경우, 특징 선별 기법은 분산과 같은 정보를 이용하여 주어진 데이터의 형태를 유지하도록 하는 특징을 선별한다. 많은 경우 레이블링 비용 때문에 레이블 정보를 얻는 것이 어려운 경우가 많아 비지도 특징 선별 기법이 활용되는 경우가 많다. 본 논문에서는 비지도 특징 선별 기법에 중점을 둔다.

많은 비지도 특징 선별 기법들은 패턴 사이의 유사도 정보를 담은 인접 행렬을 이용하여 지역적 구조를 유지할 수 있도록 하는 특징을 선별한다 [5]. 그러나 패턴 사이의 유사도는 주어진 모든 특징을 기반으로 계산되기 때문에, 특징 선별 과정 중에 유사도가 수정될 수 없다. 따라서 주어진 데이터의 구조가 변하지 않도록 하는 특징이 선별되어 소수의 특징을 선택하는 데에 한계를 가질 수 있다.

본 논문에서는 주어진 데이터를 직교하는 공간으로 변환하는 사영(projection) 방식과 저랭크(low-rank) 근사를 통해 핵심적인 특징을 선택할 수 있는 방법을 제안한다. 이를 위해 인접 행렬 정보를 사용하지 않고, 사영 방식과 저랭크 근사를 할 수 있는 목적 함수를 설계하고, 이를 풀기 위한 최적화 방식을 제안한다. 실험 결과, 제안하는 방법이 기존 비지도 특징 선별 기법들보다 좋은 군집화 성능을 보여주었다.

## II. Related Works

특징 선택을 위한 기본 접근 방식은 가중치 행렬과 레이블에 의해 사영된 특징 간의 차이를 최소화하는 기능을 찾는 것이다 [6]. 비지도 특징 선택에서 레이블 정보는  $k$ -means와 같은 군집화 알고리즘에서 파생된 잠재 레이블(latent label)로 대체된다. 이 목표를 달성하기 위해 잠재 레이블을 결정하거나 지역적 구조(local structure)를 보존하는 다양한 방법이 존재한다.

많은 비지도 특징 선별 기법은 패턴 간의 유사성 정보를 사용하여 지역적 구조 정보를 고려했다. X. He는 비지도 특징 선별을 위해 라플라시안 점수를 제안하였다 [7]. 라플라시안 점수는 각 특징에 대해 가장 가까운 이웃 정보를 사용하여 지역적인 정보를 활용할 수 있도록 계산된다. D. Cai는 데이터의 다중 군집의 구조를 보존하는 특징을 선택하는 특징 선별 기법을 제안하였다 [8]. 해당 최적화 문제는 희소 문제와  $l_1$ -정규화된 최소 제곱 문제만 포함하므로 효율적으로 해결할 수 있었다. Z. Li는 스펙트럼 군집화(spectral clustering)와 특징 선택을 동시에 수행하는 차별적 특징 선택(NDFS)을 제안하였다 [9]. 보다 정확한 군집화 결과를 얻기 위해 가중치의 양수 제약 조건이 추가되었다. S. Wang은 행렬 변환 없이 가능한 희소 학습을 통해 특징 선별 과정을 군집화 알고리즘에 내장하는 내장형 비지도 특징 선별 기법(EUFS)을 제안하였다 [10].

최근에는 모든 특징을 가진 로컬 구조 정보 외에 다른 잠재 정보를 사용하는 많은 비지도 특징 선택 방법이 도입되었다. P. Zhu는 부분공간(subspace) 군집화 내에서 저랭크에 기반한 제한적 자기표현(RSR) 모델을 제안하였다 [11]. RSR에서 각 특징은 연관되어 있는 특징들의 선형 조합(linear combination)으로 표현된다. D. Han은 새로운 유형의 레이블 행렬을 가진 정규화된 회귀 기반 공식에 기초한 직교 기반 군집화 특징 선별 기법(SOCFS)을 제안하였다. 레이블 행렬은 직교 기반 군집화를 수행하여 사영된 데이터의 잠재 레이블을 찾고자 한다. S. Du는 행렬 분해(matrix factorization)을 통해 부분적 기하 구조를 고려하고, 인접 행렬을 이용하는 새로운 비지도 특징 선별 기법을 제안하였다[13]. J. Lee는 정보이론을 기반으로 하는 비지도 특징 선별 기법을 제안하였다 [13]. 이 방법의 목적 함수는 선택된 특징들 사이의 상호작용정보(interaction information)이 최대화되는 특징들을 선택하도록 설계되었다 [14].

### III. Proposed Method

#### 1. Objective function

$n$ 개의 패턴과  $d$ 개의 특징으로 이루어진 데이터  $X \in \mathbb{R}^{d \times n}$ 가 있을 때, 정규화(regularization)된 회귀 기반 비지도 특징 선별 기법의 대표적인 목적 함수는 다음과 같다.

$$\min_W \| | W^T X - Y | \|_F^2 + \alpha \| | W | \|_{2,1} \quad (1)$$

여기서  $W \in \mathbb{R}^{d \times c}$ 는 가중치 행렬이며,  $c$ 는 임시적인 군집의 개수이며,  $Y = [y_1, \dots, y_n] \in \{0, 1\}^{c \times n}$ 는 군집화 결과를 저장하는 벡터이다. 여기서  $y_i \in \{0, 1\}^c$ 는  $x_i$ 에 대한 특정 군집에 속함을 0, 1로 나타낸다.  $\alpha \geq 0$  정규화를 위한 매개변수이다. 이 목적 함수는 데이터를  $W$ 에 의해 사영한  $W^T X$ 와 군집 레이블  $Y$ 사이의 차이를 최소화하는 특징을 찾고자 설계되었다. 이 때 임시 군집 결과  $Y$  대신에 직교하는 공간으로 사영하고자 하면 다음과 같이 목적 함수를 나타낼 수 있다.

$$\min_{W,B,E} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1}, \quad (2)$$

$$s.t. B^T B = I, E^T E = I, E \geq 0$$

여기서  $I$ 는 단위행렬이다. 이 목적 함수는 SOCFs의 목적 함수로부터 전개되었다 [12]. 직교성 제한을 위한  $B^T B = I$ 는  $B$ 의 열들이 독립적인 관계를 가질 수 있도록 한다.  $E$ 와 관련되어 있는 직교성과 양수 제한은  $E$ 의 각 행이 오직 하나의 0이 아닌 값들을 가지게 해준다 [15]. 이 제한성들을 유지하며 인수분해하여, 사영된 데이터  $W^T X$ 가 직교성을 가지는 군집화를 수행할 수 있도록 해준다. 이 목적 함수를 최소화하는 문제는  $W$ 에 의해 선택된 특징들이 직교성을 가지는 군집화를 유도하게 된다.

특징들 사이의 상관관계를  $W$ 의 랭크 최소화를 통해 고려해볼 수 있다.  $W$ 에 대한 저랭크 제한성을 식 (2)에 반영하여 수정하면 다음과 같이 정리할 수 있다.

$$\min_{W,B,E} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1} + \beta \text{Rank}(W), \quad (3)$$

$$s.t. B^T B = I, E^T E = I, E \geq 0$$

여기서  $\text{Rank}(W)$ 는  $W$ 의 랭크를 의미하고  $\beta$ 가 높을수록  $W$ 가 랭크가 작아지도록 유도한다. 가중치 행렬  $W$ 의 최소화는 특징들 사이의 고유한 상관관계를 찾기 위해 낮은 랭크 안에서 탐색을 한다. 그러나 식 (3)은 랭크 함수  $\text{Rank}(W)$ 의 연속하지 않고 볼록(convex)하지 않은 특징

때문에 풀기 어렵다 [16]. 뉴클리어 노름(nuclear norm)은 랭크 함수를 볼록 함수로 전환하여 랭크를 근사하여 계산할 수 있도록 해준다. B. Recht의 연구에서 랭크 함수의 하한 볼록 함수를 제시하였다. 이를 반영하여 식 (3)을 다시 정리할 수 있다 [17].

$$\min_{W,B,E} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1} + \beta \| | W | \|_*, \quad (4)$$

$$s.t. B^T B = I, E^T E = I, E \geq 0$$

여기서  $\| | W | \|_*$ 은 뉴클리어 노름을 나타내고 Schatten  $p$ -노름의 한 종류이다. 행렬  $A$ 의 Schatten  $p$ -노름은 다음과 같이 정의된다 [18].

$$\| | A | \|_{s_p} = \left( \sum_i \sigma_i^p(A) \right)^{1/p} \quad (5)$$

여기서  $0 < p \leq 2$ 이고,  $\sigma_i(A)$ 는 행렬  $A$ 의  $i$ 번째 큰 특이값(singular value)이다.  $p$ 를 1로 설정한 후 정의에 따라 뉴클리어 노름은 다음과 같이 정리할 수 있다.

$$\| | A | \|_* = \sum_i \sigma_i = \text{Tr}(\sqrt{A^T A}) \quad (6)$$

여기서  $\text{Tr}$ 은 행렬의 대각합(trace)를 의미하고, 이 식은 행렬  $A$ 의 랭크 근사가 된다. 이 정의를 이용하여 식 (4)를 다시 정의하면 다음과 같다.

$$\min_{W,B,E} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1} + \beta \text{Tr}(\sqrt{W^T W}), \quad (5)$$

$$s.t. B^T B = I, E^T E = I, E \geq 0$$

이 식은 제한된  $W$ 의 랭크 내에서 특징을 선택하게 된다.

#### 2. Optimization

$E$ 에 대한 직교성 제한을 보장하기 위해, 보조 변수  $G$ 를 이용하여 식 (5)를 다음과 같이 재정의한다.

$$\min_{W,B,E,G} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1} + \beta \text{Tr}(\sqrt{W^T W}) + \gamma \| | G - E | \|_F^2, \quad (6)$$

$$s.t. B^T B = I, E^T E = I, E \geq 0$$

여기서  $\gamma > 0$ 은 가중치 변수로 다른 항들과 범위를 조절한다. 이 식은 다른 변수들이 고정돼 있을 때 하나의 변수에 대해 볼록하다. 따라서 이 최적화 문제를 한 번에 하나의 변수에 대해 풀 수 있다.

첫 번째,  $W$ 에 대해 식을 표현하면 다음과 같다.

$$\min_{W,B,E,G} \| | W^T X - BE^T | \|_F^2 + \alpha \| | W | \|_{2,1} + \beta \text{Tr}(\sqrt{W^T W}) \quad (7)$$

이 식을  $\mathcal{J}$ 라 하고,  $W$ 에 대한 도함수를 정리하면 다음

과 같다.

$$\frac{\partial J}{\partial W} = 2XX^T W - 2XEB^T + 2\alpha DW + \beta UI_{dc} V^T = 0 \quad (8)$$

도함수를 0으로 만드는  $W$ 는 다음과 같다.

$$W = (2XX^T + 2\alpha D)^{-1} (2XEB^T - \beta UI_{dc} V^T) \quad (9)$$

여기서  $D$ 는 대각 성분  $D_{ii} = 1/2 \| | W^i | \|_2$ 인 대각 행렬이다. 이 형태는  $\| | W | \|_{2,1}$ 의  $W$ 에 대한 도함수를 실용적으로 표현되는 방식이다 [12].  $U$ 와  $V$ 는  $W = U\Sigma V^T$ 의 SVD(singular value decomposition)로 유도된 행렬들이다. 그리고  $Tr(\sqrt{W^T W})$ 의 도함수는 다음 정의로부터 전개되었다 [19].

$$\frac{\partial \| | W | \|_*}{\partial W} = UI_{dc} V^T \quad (10)$$

여기서  $I_{dc} \in \mathbb{R}^{d \times c}$ 은 행렬로  $i$ 와  $j$ 가 같을 때  $I_{i,j}$ 가 1이고 나머지는 0인 행렬이다.

두 번째  $B$ 에 대해 식을 표현하면 다음과 같다.

$$\min_B \| | W^T X - BE^T | \|_F^2, s.t. B^T B = I \quad (11)$$

여기에 직교 행렬  $E$ 를 이용하여 다음과 같이 재정의하면 다음과 같다.

$$\min_B \| | (W^T X - BE^T)E | \|_F^2, s.t. B^T B = I \quad (12)$$

여기서  $W^T X E$ 를  $A$ 라 하면 다음과 같이 표현할 수 있다.

$$\min_B \| | B - A | \|_F^2, s.t. B^T B = I \quad (13)$$

이 식을 라그랑주 승수법을 이용하여 다음과 같이 재정의할 수 있다.

$$L(B, \Lambda) = \| | B - A | \|_F^2 + Tr(\Lambda(B^T B - I)) = Tr(BB^T - 2BA^T + AA^T) + Tr(\Lambda(B^T B - I)) \quad (14)$$

식 (14)의 도함수는 다음과 같이 정의된다.

$$\frac{\partial L}{\partial B} = 2B - 2A + 2\Lambda B = 0$$

$$\rightarrow A = B(I + \Lambda)$$

$$\rightarrow B = A(A^T A)^{-1/2} (\because A^T A = (I + \Lambda)^2) \quad (15)$$

$A$ 의 SVD를  $U\Sigma V^T$ 라 하면  $A^T A = V\Sigma^2 V^T$ 가 된다. 따라서  $B = U\Sigma V^T (V\Sigma V^T)^{-1/2} = UV^T$ 가 된다.

세 번째,  $E$ 에 대해 식을  $B$ 의 방식과 유사하게 정리하면 다음과 같다.

$$\min_E \| | W^T X - BE^T | \|_F^2 + \gamma \| | E - G | \|_F^2$$

$$s.t. E^T E = I$$

$$= \min_E \| | E - X^T W B | \|_F^2 + \gamma \| | E - G | \|_F^2$$

$$s.t. E^T E = I \quad (16)$$

$X^T W B$ 를  $A$ 라 하고 라그랑주 승수법을 이용하여 다음과 같이 재정의할 수 있다.

$$L(E, \Lambda) = \| | E - A | \|_F^2 + \gamma \| | E - G | \|_F^2$$

$$+ Tr(\Lambda(E^T E - I))$$

$$= Tr(EE^T - 2EA^T + AA^T) +$$

$$\gamma Tr(EE^T - 2EG^T + GG^T) + Tr(\Lambda(E^T E - I)) \quad (17)$$

식 (17)의 도함수는 다음과 같이 정의된다.

$$\frac{\partial L}{\partial E} = 2E - 2A + 2\gamma B + 2\Lambda E = 0$$

$$\rightarrow A + \gamma G = E((1 + \gamma)I + \Lambda) \quad (18)$$

이 방정식의 해는 식 (15)과 유사하게  $E = UV^T$ 이고,

$U$ 와  $V$ 는  $X^T W B + \gamma G$ 의 SVD로부터 정의된다.

네 번째,  $G$ 에 대해서 식을 정리하면 다음과 같다.

$$\min_G \| | G - E | \|_F^2, s.t. G \geq 0 \quad (19)$$

이 최적화 문제의 해는  $G = 1/2(E + | E |)$ 로 결정된다. 최종적인 알고리즘 Algorithm 1과 같다.

#### Algorithm 1. Proposed method

---

```

1 입력:  $X \in \mathbb{R}^{d \times n}, \alpha, \beta, \gamma$ 
2 초기화:  $t = 0, D = I, B, E, G, \maxIter$ 
3 while  $t \leq \maxIter$  do
4    $E \leftarrow UV^T$ ,  $U, V$ 는  $X^T W B + \gamma G$ 의 SVD
5    $G \leftarrow 1/2(E + | E |)$ 
6    $B \leftarrow UV^T$ ,  $U, V$ 는  $W^T X E$ 의 SVD
7    $W \leftarrow (2XX^T + 2\alpha D)^{-1} (2XEB^T - \beta UI_{dc} V^T)$ 
8 end while
9 출력:  $\| | w^i | \|_2$ 의 값을 내림차순으로 정렬한 특징 번호
```

---

## IV. Experiments

이 장에서는 제안하는 방법의 성능을 비교하기 위해 실험 결과를 보인다.  $k$ -means 군집화 방법을 이용하여 비지도 학습 성능을 비교하였다. 전체 패턴에 대해서 비지도 특징 선별을 진행한 후  $k$ -means를 30번 반복 실험하여 평균 값을 기입하였다.

평가 방법으로 군집화 정확도(Clustering Accuracy, CLACC), 정규화 상호정보척도(Normalized Mutual Information, NMI)를 사용하였다 [12]. 두 척도는 군집화 결과와 레이블 사이의 정확도, 상관관계를 비교하여 높을 수록 좋은 군집화 성능을 나타낸다.

Table 1. Information about data sets

Data	# Pattern	# Feature	# Class
lymphoma	96	4026	13
madelon	2600	500	6
orl	400	1024	12
tox171	171	5748	7
umist	575	644	14
yaleb	2414	1024	2

6개의 실세계 데이터를 실험에 사용하였다. 각 데이터는 lymphoma, madelon, orl, tox171, umist, yaleb이다. lymphoma 데이터는 림프종에 대한 정보를 담고 있는 생물학 데이터이며, tox171 데이터는 독성학을 이용한 프로파일을 담고 있는 생물학 데이터이다. madelon 데이터는 NeurIPS 2003 특징 선별 경쟁을 위해 제작된 인공 데이터이다. orl, umist, yaleb 데이터는 사람의 얼굴 사진을 담고 있는 이미지 데이터이다. 데이터의 크기 등 자세한 내용은 Table 1에 기입하였다 [12, 20, 21]. 특징 개수가 많은 lymphoma, tox171 데이터에서는 50, 100, ..., 300개의 특징을 선별하였고, 나머지 데이터에서는 10, 20, 30, ..., 100개의 특징을 선별하였다. 제안하는 방법과 비교할 5가지 기존 방법, MaxVar[22], LapScore[8], NDFS[9], EUFS[10], RSR[11]을 선정하였다. 비교 대상과 제안하는 방법에 필요한 매개변수  $\alpha$ ,  $\beta$ ,  $\gamma$ 는  $10^{-3}$ ,  $10^{-2}$ , ...,  $10^3$ 까지 두고 비교 실험하여 가장 좋은 결과를 선정하여 기입하였다. 비교 방법들의 초기화는 해당 연구의 방식을 따라 설정하였고, 제안하는 방법 초기화를 위해  $W$ ,  $G$ 는 0에서 1사이의 무작위 값으로 초기화하였다. 직교 제한이 있는  $B$ 는 0과 1사이의 무작위 값으로 행렬을 생성한 후에 QR 분해로 초기화 하였다. 최대 반복횟수  $\maxIter$ 는 30으로 설정하였다.

Table 2. Experimental result on lymphoma data set

	CLACC	NMI
MaxVar	0.5542	0.7345
LapScore	0.5177	0.7300
NDFS	0.5122	0.6784
EUFS	0.5535	0.7272
RSR	0.5326	0.6991
Proposed	<b>0.5865</b>	<b>0.7610</b>

Table 3. Experimental result on madelon data set

	CLACC	NMI
MaxVar	0.5238	0.0052
LapScore	0.5208	0.0045
NDFS	0.5700	0.0142
EUFS	0.5796	0.0184
RSR	0.6154	0.0388
Proposed	<b>0.6191</b>	<b>0.0413</b>

Table 4. Experimental result on orl data set

	CLACC	NMI
MaxVar	0.5084	0.8700
LapScore	0.3591	0.8101
NDFS	0.3399	0.7978
EUFS	0.4476	0.8538
RSR	0.4886	0.8640
Proposed	<b>0.5186</b>	<b>0.8713</b>

Table 5. Experimental result on tox171 data set

	CLACC	NMI
MaxVar	0.4246	0.4664
LapScore	0.4107	0.4292
NDFS	0.4191	0.4227
EUFS	0.4236	0.4569
RSR	0.4164	0.3978
Proposed	<b>0.4967</b>	<b>0.5327</b>

Table 6. Experimental result on umist data set

	CLACC	NMI
MaxVar	0.4210	0.8201
LapScore	0.3345	0.7815
NDFS	0.4364	0.8060
EUFS	0.4913	0.8350
RSR	0.4587	0.8348
Proposed	<b>0.5392</b>	<b>0.8590</b>

Table 7. Experimental result on yaleb data set

	CLACC	NMI
MaxVar	0.0953	0.7286
LapScore	0.0909	0.7243
NDFS	0.1182	0.6996
EUFS	0.1189	0.6992
RSR	0.0912	0.7308
Proposed	<b>0.1746</b>	<b>0.7507</b>

Table 2-7는 6개의 데이터에 대해 각각의 군집화 성능을 보여준다. 각 테이블에서 첫 번째 행은 특징 선별 기법들, 두 번째 행은 CLACC, 세 번째 행은 NMI 성능을 보여준다. 기법들 중에서 가장 좋은 성능을 보이는 기법 결과에 볼드체로 표기하였다.

실험 결과 모든 데이터에서 제안하는 방법이 가장 좋은 성능을 보였다. lymphoma 데이터에서 제안하는 방법의 CLACC가 0.5865, NMI가 0.7610으로 가장 좋은 성능을

보였고, MaxVar가 두번째로 좋은 성능을 보였다. madelon 데이터에서 RSR과 제안하는 방법이 좋은 성능을 보였고, 미세하게 제안하는 방법이 더 좋은 성능을 나타냈다. ori 데이터에서는 MaxVar가 두번째로 좋은 성능을 보였고, 제안하는 방법이 가장 좋은 성능을 보였다. tox171 데이터에서는 제안하는 방법이 다른 방법들에 비해 압도적으로 좋은 성능을 보여주었다. umist 데이터에서는 EUFS가 두번째로 좋은 성능을 보였고, 제안하는 방법이 가장 좋은 성능을 보였다. yaleb 데이터에서 CLACC 성능이 제안하는 방법이 다른 방법들보다 압도적으로 좋은 성능을 보였다.

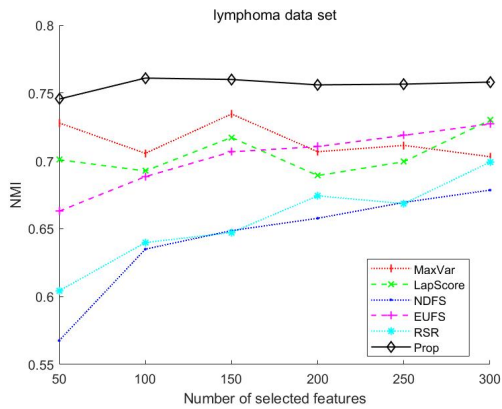


Fig. 1. NMI comparison based on number of features selected in the lymphoma data set

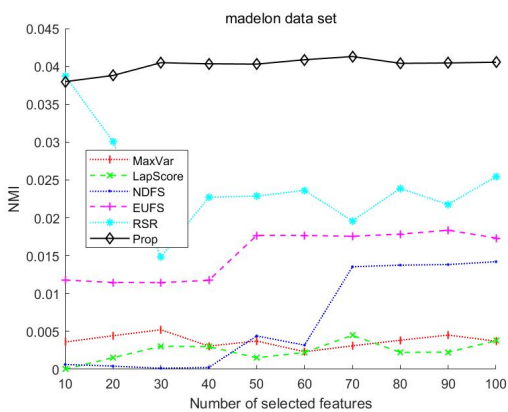


Fig. 2. NMI comparison based on number of features selected in the madelon data set

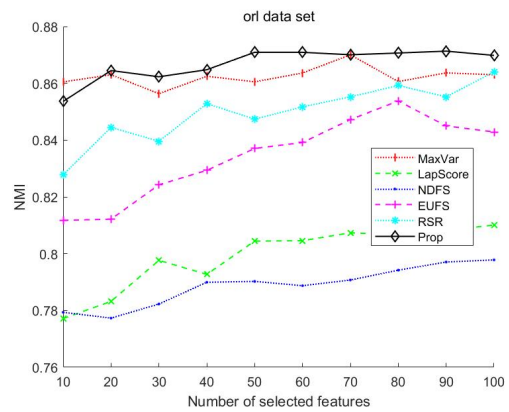


Fig. 3. NMI comparison based on number of features selected in the ori data set

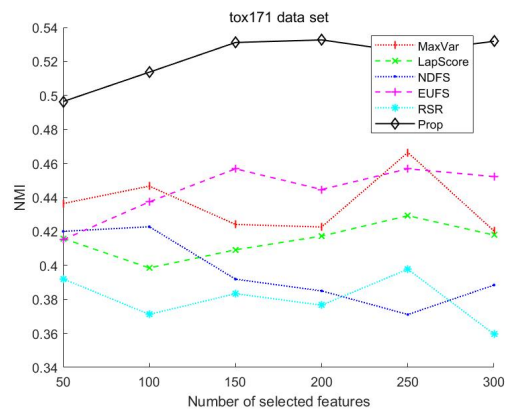


Fig. 4. NMI comparison based on number of features selected in the tox171 data set

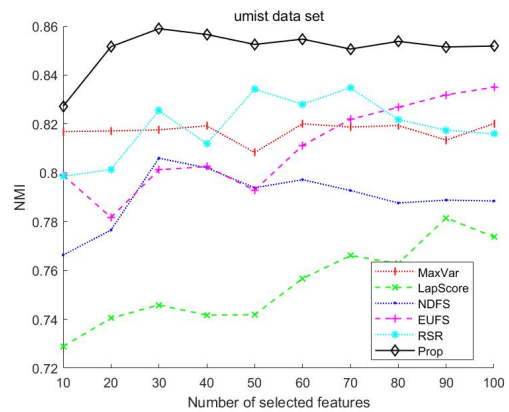


Fig. 5. NMI comparison based on number of features selected in the umist data set

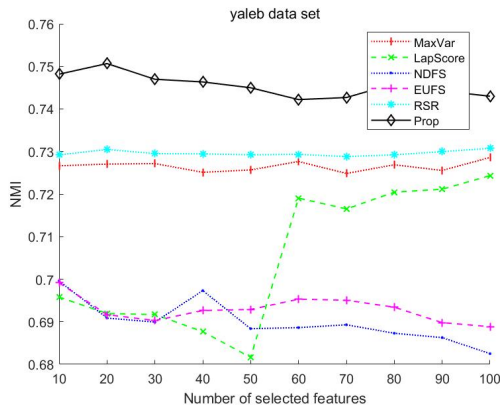


Fig. 6. NMI comparison based on number of features selected in the yaleb data set

Fig 1-6는 6개의 데이터에 대해 선택된 특징 개수에 따른 각 방법의 군집화 성능을 보여준다. 가로 축은 선택된 특징의 개수, 세로 축은 NMI 성능을 나타낸다. 기존의 방법론들은 특징 개수가 늘어남에 따라 성능의 증감이 있었다. 대체적으로 제안하는 방법은 특징 개수에 따라 성능의 변화가 적은 편이었고, 개수에 관계없이 기존 방법들보다 좋은 성능을 보였다.

본 실험 중 Fig 2의 madelon 데이터에서 RSR은 특징 개수가 10개에서 30개까지 성능이 감소하였고, NDFS나 EDFS는 40~50개부터 성능이 증가하였다. 제안하는 방법은 특징 개수가 늘어남에도 성능 차이가 별로 없었고, 적은 특징 개수에서 나타났던 가장 좋은 성능이 유지되었다. 제안하는 방법은 저랭크한 공간에서의 특징을 선별하기 때문에 비교적 적은 특징 개수에서 좋은 성능을 나타냈고, 그 이후에 추가되는 특징이 성능을 저해하지 않았던 것으로 판단된다.

## V. Conclusions

본 연구에서는 저랭크를 기반으로 하는 비지도 특징 선별 기법을 제안하였다. 제안된 방법은 낮은 랭크와 희소성을 기반으로 주요한 특징을 선별하여 기존 방법들에 비해 비지도 특징 선별에서 더 효과적이라는 것을 보여주었다.

제안된 방법의 우수한 성능에도 불구하고, 추가 연구가 필요한 몇 가지 문제가 있다. 첫째, 범주형 데이터에 대해 목적 함수가 잘 풀 수 없는 한계를 가진다. 둘째, 저랭크와 희소성 사이의 가중치 설정을 해야 한다는 문제를 가진다. 향후의 연구 주제로서 이 문제들을 고려하여 더욱 효과적인 비지도 특징 선별 기법을 제시할 수 있을 것이다.

## ACKNOWLEDGEMENT

This work was supported by the GRRC program of Gyeonggi province. [GRRC KGU 2020-B01, Research on Intelligent Industrial Data Analytics]

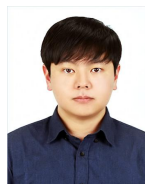
## REFERENCES

- [1] E. Smirni, and G. Ciardo, "Workload-Aware Load Balancing for Cluster Web Servers," *IEEE Trans. on Parallel and Distributed Systems*, Vol. 16, No. 3, pp. 219-232, March 2005. DOI: 10.1016/j.patcog.2019.03.026
- [2] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 11, pp. 2765-2781, March 2013. DOI: 10.1109/TPAMI.2013.57
- [3] S. Sharmin, M. Shoyaib, A. A. Ali, M. A. H. Khan, and O. Chae, "Simultaneous feature selection and discretization based on mutual information", *Pattern Recognition*, Vol. 91, pp. 162-174, July 2019. 10.1016/j.patcog.2019.02.016
- [4] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A survey on semi-supervised feature selection methods", *Pattern Recognition*, Vol. 64, April 2017. 10.1016/j.patcog.2016.11.003
- [5] S. Du, Y. Ma, S. Li, and Y. Ma, "Robust unsupervised feature selection via matrix factorization", *Neurocomputing*, Vol. 241, pp. 115-127, June 2017. 10.1016/j.neucom.2017.02.034
- [6] F. Nie, H. Huang, X. Cai, C. H. Ding, "Efficient and robust feature selection via joint  $l_2, 1$ -norms minimization", *Advances in Neural Information Processing Systems*, pp. 1813-1821, 2010.
- [7] X. He and P. Niyogi, "Locality preserving projections", *Advances in neural information processing systems*, Vol. 16, 2003.
- [8] D. Cai, Deng, C. Zhang, and He, Xiaofei, "Unsupervised feature selection for multi-cluster data", *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 333-342, 2010.
- [9] Z. Li, Y. Yang, J. Liu, X. Zhou, and H. Lu, "Unsupervised feature selection using nonnegative spectral analysis", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 2, pp. 1026-1032, 2012.
- [10] S. Wang, J. Tang, and H. Liu, "Embedded unsupervised feature selection", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29, pp. 470-476, 2015.
- [11] P. Zhu, W. Zuo, L. Zhang, Q. Hu, and S.C. Shiu, "Unsupervised feature selection by regularized self-representation", *Pattern Recognition*, Vol. 48, No. 2, pp. 438-446, February 2015. DOI:

10.1016/j.patcog.2014.08.006

- [12] D. Han and J. Kim, "Unsupervised simultaneous orthogonal basis clustering feature selection", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5016-5023, June 2015.
- [13] S. Du, Y. Ma, S. Li, and Y. Ma, "Robust unsupervised feature selection via matrix factorization", Neurocomputing, Vol. 241, pp. 115-127, June 2017. DOI: 10.1016/j.neucom.2017.02.034
- [14] J. Lee, W. Seo, D.-W. Kim, "Efficient information-theoretic unsupervised feature selection", Electronics Letters, Vol. 54, No. 2, pp. 76-77, January 2017.
- [15] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering", Proceedings of the International Conference on Data Mining, pp. 606-610, November 2005.
- [16] X. Zhong, L. Xu, Y. Li, Z. Liu, and E. Chen, "A nonconvex relaxation approach for rank minimization problems", Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 29, No. 1, January 2015.
- [17] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization", SIAM Review, Vol. 52, No. 3, pp. 471-501, 2010. DOI: 10.1137/070697835
- [18] H. Zhang, J. Qian, B. Zhang, J. Yang, C. Gong, and Y. Wei, "Low-rank matrix recovery via modified Schatten- $p$  norm minimization with convergence guarantees", IEEE Transactions on Image Processing, Vol. 29, pp. 3132-3142, December 2019. DOI: 10.1109/TIP.2019.2957925
- [19] X. Zhen, M. Yu, X. He, and S. Li, "Multi-target regression via robust low-rank learning", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, No. 2, pp. 497-504, February 2017. DOI: 10.1109/TPAMI.2017.2688363
- [20] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 8, pp. 1548-1560, August 2010. DOI: 10.1109/TPAMI.2010.231
- [21] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 5, pp. 1178-1192, September 2012. DOI: 10.1109/TPAMI.2012.197
- [22] J. G. Dy and C. E. Brodley, "Feature selection for unsupervised learning", Journal of Machine Learning Research, Vol. 5, pp. 845-889, August 2004.

## Author



Hyunki Lim is currently an assistant professor in the Division of AI Computer Science and Engineering, Kyonggi Univ. (KGU) in Suwon, Korea. Prior to coming to KGU, he did his postdoc at KIST, and

Ph.D., M.S. and B.S. at Chung-Ang Univ., Korea. His research interest includes advanced machine algorithms with innovative applications such as music emotion recognition and smart factory.