

Sparse and low-rank feature selection for multi-label learning

Hyunki Lim*

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

[Abstract]

In this paper, we propose a feature selection technique for multi-label classification. Many existing feature selection techniques have selected features by calculating the relation between features and labels such as a mutual information scale. However, since the mutual information measure requires a joint probability, it is difficult to calculate the joint probability from an actual premise feature set. Therefore, it has the disadvantage that only a few features can be calculated and only local optimization is possible. Away from this regional optimization problem, we propose a feature selection technique that constructs a low-rank space in the entire given feature space and selects features with sparsity. To this end, we designed a regression-based objective function using Nuclear norm, and proposed an algorithm of gradient descent method to solve the optimization problem of this objective function. Based on the results of multi-label classification experiments on four data and three multi-label classification performance, the proposed methodology showed better performance than the existing feature selection technique. In addition, it was showed by experimental results that the performance change is insensitive even to the parameter value change of the proposed objective function.

▶ **Key words:** Multi-label learning, feature selection, low-rank approximation, sparsity, optimization method

[요 약]

본 논문에서는 다중 레이블 분류를 위한 특징 선별 기법을 제안한다. 기존 많은 특징 선별 기법들은 상호정보척도 등을 이용하여 특징과 레이블 사이의 연관성을 계산하여 특징을 선별하였다. 하지만 상호정보척도는 결합 확률을 요구하기 때문에 실제 전체 특징 집합에서 결합 확률을 계산하는 것은 어렵다. 따라서 소수의 특징만 계산이 가능하여 지역적 최적화만 가능하다는 단점을 가진다. 이런 지역적 최적화 문제를 피해, 주어진 특징 전체 공간에서 저랭크 공간을 구성하고, 희소성을 가진 특징들을 선별할 수 있는 특징 선별 기법을 제안한다. 이를 위해 뉴클리어 노름을 이용해 희귀 기반의 목적 함수를 설계하였고, 이 목적 함수의 최적화 문제를 풀기 위한 경사하강법 방식의 알고리즘을 제안하였다. 4가지의 데이터와 3가지 다중 레이블 분류 성능을 기준으로 다중 레이블 분류 실험 결과를 통해 제안하는 방법론이 기존 특징 선별 기법보다 좋은 성능을 나타내는 것을 보였다. 또한 제안하는 목적 함수의 파라미터 값 변화에도 성능 변화가 둔감한 것을 실험적인 결과로 확인하였다.

▶ **주제어:** 다중 레이블 학습, 특징 선별, 저랭크 근사, 희소성, 최적화 방법

-
- First Author: Hyunki Lim, Corresponding Author: Hyunki Lim
 - Hyunki Lim (hlim20@kyonggi.ac.kr), Div. of AI Computer Science and Engineering, Kyonggi University
 - Received: 2021. 04. 19, Revised: 2021. 05. 26, Accepted: 2021. 05. 26.

I. Introduction

데이터마이닝, 패턴인식, 기계학습 등의 연구 분야에서 고차원 데이터를 많이 다루게 된다[1]. 이 고차원 데이터는 높은 계산 복잡도와 많은 메모리를 요구하기 때문에 기계 학습 알고리즘을 활용하기 어렵다[2]. 이 문제는 학습 알고리즘의 성능을 악화시키고, 응용 가능성을 제한시킨다[3]. 이 문제를 극복하기 위해 특징 선별 기법들이 많이 제안되었다. 특징 선별은 주어진 고차원 데이터에서 불필요한 특징들을 제거하고 중요한 특징들만 남겨 저차원의 데이터로 변환하는 방법론이다. 정보력이 높은 소수의 특징을 가지는 되면 기계학습 방법론의 높은 성능과 해석력, 적은 계산 복잡도를 기대할 수 있다[4].

기존 많은 특징 선별 연구에서 상호정보척도(mutual information)를 이용하여 특징을 평가하였다. 상호정보척도는 특징과 레이블 사이의 종속성을 계산할 수 있어 전통적으로 특징 선별 연구에서 많이 사용되었다. 하지만 주어진 데이터 형태에 따라 결합확률(joint probability)의 계산이 어려운 경우가 많아 대부분 확률을 추정해야 한다는 문제가 있었다.

본 논문에서는 다중 레이블 분류를 위한 특징 선별 기법을 제안한다. 개별적인 특징을 평가하는 전통적인 방식이 아니라, 주어진 데이터의 랭크, 희소성을 계산하기 위하여 데이터 행렬 전체를 이용한다. 이를 위해 행렬 기반의 목적 함수가 설계되었고, 이 목적 함수 내에서 가중치 행렬을 이용하여 희소성과 저랭크 특징을 선별할 수 있도록 하였다. 행렬 기반의 목적 함수의 최적화를 위해 경사하강법(gradient descent) 기반 단순한 최적화 알고리즘을 제안하였다. 이 알고리즘은 하나의 행렬에 대해 지역 최적의 값(local optima)을 찾아준다. 실험 결과, 4개의 데이터에서 제안하는 방법이 좋은 성능을 나타내는 것을 확인하였다.

II. Related Works

다중 레이블 분류를 위한 특징 선별은 크게 내장형 방식(embedded approach), 필터 방식(filter approach), 래퍼 방식(wrapper approach)으로 나눌 수 있다[5]. 내장형 방식은 분류기(classifier) 내부적으로 포함되는 방식이다. 필터 방식은 분류기와 독립적으로 특징들을 평가하여 점수에 따라 특징을 선별하는 방식이다. 분류기와 독립적으로 동작하기 때문에 활용성이 좋지만 분류기마다 다른 성능을 보일 수 있다. 래퍼 방식은 분류기의 성능을 기반

으로 좋은 성능을 내는 특징 집합을 구하는 방식이다. 일반적으로 래퍼 방식은 특징 집합에 대해 클래스 분류 성능을 얻어야 하기 때문에 시간 복잡도가 높다. 상대적으로 필터 방식은 분류기를 사용하지 않기 때문에 시간적으로 빠른 장점이 있다. 본 논문에서는 분류기와 독립성, 빠른 알고리즘 속도의 장점을 가지는 필터 방식의 특징 선별 기법을 제안한다.

다중 레이블 특징 필터 방식은 레이블 변환(label transformation) 방식과 변환하지 않는 방식이 있다[5]. 변환하는 방식은 다중 레이블을 단일 레이블로 변환하여 기존 단일 레이블을 위한 특징 필터 방식을 이용하는 것이다[6]. 레이블 변환의 대표적인 방식인 레이블 멱집합(label powerset)의 경우, 레이블 집합이 나타낼 수 있는 모든 경우의 수를 클래스화한다. 이 방식은 직관적이지만, 각 클래스에 해당하는 패턴의 개수가 매우 적어진다는 문제를 야기한다. 이와 같이 이 방식은 기존 방식들을 바로 이용할 수 있다는 장점이 있지만 레이블 변환 과정에서 정보 손실이 발생할 수 있는 단점이 있다.

이런 단점을 해소하고자 레이블 변환없이 특징을 선별하는 방법(algorithm adaptation)들이 제안되었다[7-10]. 일반적으로 특징과 모든 개별 레이블과의 관계, 레이블과 레이블과의 관계를 고려하여 상호정보척도 등을 이용하여 계산하고 점수화한다. 변환 방식과 비교하여 정보 손실 발생이 적고, 좋은 성능을 나타내는 경우가 많다. 하지만 레이블이 많아짐에 따라 계산 양이 증가하는 단점이 있다. 다중 레이블 특징 선별 기법의 범주를 Fig. 1에서 확인할 수 있다.

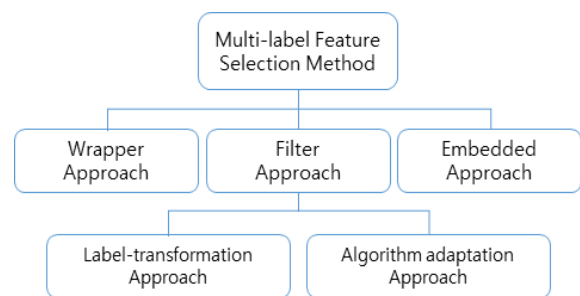


Fig. 1. Category of multi-label feature selection method

III. The Proposed Method

1. Objective function

들어가기 앞서, 몇 가지 표기법을 소개한다. 행렬은 대문자로 표기한다. 행렬 A 의 i 번째 행, j 번째 열에 속하는 성분을 a_{ij} 로 표기하고 a_i 는 i 번째 행벡터를 의미한다. $\|\cdot\|$

은 행렬 노름(matrix norm)을 의미하고, F 는 프로베니우스 노름(Frobenius norm)으로 다음과 같이 정의된다.

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \quad (1)$$

본 논문에서는 데이터를 $X \in \mathbb{R}^{n \times d}$ 로 표현한다. 이 때, d 는 특징의 개수, n 은 패턴의 개수로 정한다. 정규화 선형회귀(Regularized linear regression)의 일반적인 목적 함수는 다음과 같다[11].

$$\min_W \frac{1}{n} \|XW - Y\|_F^2 + \alpha \|W\|_F^2 \quad (2)$$

이 식에서 $W \in \mathbb{R}^{d \times c}$ 는 찾고자 하는 가중치, $Y \in \mathbb{R}^{n \times c}$ 는 다중 레이블이고, α 의 가중치로 W 가 희소성을 가질 수 있도록 정규화한다. 이 때, 저랭크성을 추가하기 위해 식 (2)를 다음과 같이 수정한다.

$$\min_W \frac{1}{n} \|XW - Y\|_F^2 + \alpha \|W\|_F^2 + \beta \|W\|_* \quad (3)$$

이 때 $\|W\|_*$ 는 뉴클리어 노름(nuclear norm)으로 새튼 p -노름(Schatten p -norm)의 한 형태이다. 새튼 1-노름은 다음과 같이 정의된다[12].

$$\|A\|_{S_p} = \left(\sum_i \sigma_i^p(A) \right)^{1/p} \quad (4)$$

이 때 $0 < p \leq 2$ 이며, $\sigma_i^p(A)$ 는 행렬 A 의 i 번째 큰 특이값(singular value)를 의미한다. 뉴클리어 노름은 식 (4)로부터 다음과 정의되고, A 의 랭크를 의미하게 된다.

$$\|A\|_* = \sum_i \sigma_i^0 = \text{Tr}(\sqrt{A^T A}) \quad (5)$$

여기에서 Tr 은 행렬의 대각합(trace)를 의미한다. 식 (5)로부터 최종적인 제안하는 목적 함수는 다음과 같다.

$$\min_W \frac{1}{n} \|XW - Y\|_F^2 + \alpha \|W\|_F^2 + \beta \text{Tr}(\sqrt{W^T W}) \quad (6)$$

2. Optimization

최적화 문제 식 (6)을 풀기 위해 다음과 같이 행렬 곱셈 형태로 변환한다.

$$\min_W \frac{1}{n} \text{Tr}((XW - Y)^T (XW - Y)) + \alpha \text{Tr}(W^T W) + \beta \text{Tr}(\sqrt{W^T W}) \quad (7)$$

이 식을 W 에 대한 함수 J 로 정의하고, J 의 도함수를 다음과 같이 정의할 수 있다.

$$\frac{\partial J}{\partial W} = -\frac{2}{n} (XW - Y)X^T + 2\alpha W + \beta U | \Sigma^T | \Sigma^{-1} V^T \quad (8)$$

식 (8)에서 U, Σ, V 는 W 특이값 분해(Singular value decomposition, SVD) $W = U \Sigma V^T$ 에서 얻어진 행렬들이다. 이 때 $\|W\|_*$ 에 대한 도함수는 다음과 같다[13].

$$\frac{\partial \|W\|_*}{\partial W} = U \Sigma^T \Sigma^{-1} V^T \quad (9)$$

식 (8)을 통해서 경사하강법 방식의 알고리즘을 다음과 같이 설계할 수 있다.

$$W^{t+1} = W^t - \eta \frac{\partial J}{\partial W} \quad (10)$$

이 때 η 는 학습률(learning rate)를 의미하며 라인 검색 알고리즘(line search algorithm) [14]등에서 결정할 수 있다. 최종적인 알고리즘은 다음과 같다.

Algorithm 1. Low-rank and sparsity based multi-label feature selection

```

1: Input: Input data  $X \in \mathbb{R}^{n \times d}$ , Label data  $Y \in \mathbb{R}^{n \times c}$ ,  $\alpha, \beta$ 
2: Initialization:  $t = 0$ ,  $W \in \mathbb{R}^{d \times c}$ ,  $m$ (Maximum iteration)
3: while  $t \leq m$  do
4:    $U, V, \Sigma$  ← from SVD of  $W$ 
5:    $W$  ←  $W - \eta \frac{\partial J}{\partial W}$  from Equation (8)
6:    $t$  ←  $t + 1$ 
7: end while
Output: Selected features corresponding to the largest values of  $\|w_i\|_2$ , which are sorted by descending order

```

IV. Experimental Results

이 장에서는 제안하는 방법의 성능을 비교하기 위해 실험 결과를 보인다. 다중 레이블 k -최근접 이웃 알고리즘 [15]과 다중 레이블 나이브 베이즈[16]를 분류 성능 비교를 위해 사용하였다. 홀드아웃(hold-out) 교차검증 방법을 사용하였다. 전체 패턴에서 80%를 훈련 데이터로, 20%를 테스트 데이터로 사용하였고, 10번 반복 실험하여 평균 값을 기입하였다. 실험은 Intel i5 2.9GHz, 8GB 메모리, HDD 하드웨어, Windows 10, MATLAB 2020b 소프트웨어 환경에서 수행되었다.

평가 방법으로 3가지 방법, 해밍 로스(Hamming loss, Hloss), 랭킹 로스(Ranking loss, Rloss), 다중 레이블 정확도(Multi-label accuracy, MlAcc)을 사용하였다[17]. 해밍 로스, 랭킹 로스는 낮을수록, 다중 레이블 정확도는 높을수록 좋은 분류 성능을 나타낸다.

4개의 데이터를 실험에 사용하였다. 각 데이터는 Emotions[18], Birds[19], Scene[20], Yeast[21]이며 자세한 내용은 Table 1에 기입하였다.

Table 1. Information about data sets

| Data | # Pattern | # Feature | # Label |
|----------|-----------|-----------|---------|
| Emotions | 593 | 72 | 19 |
| Birds | 645 | 260 | 6 |
| Scene | 2407 | 294 | 6 |
| Yeast | 2417 | 103 | 14 |

제안하는 방법과 비교할 5가지 기존 방법, AMI[7], MDMR[8], MLCFS[9], PPT+RF[6], QPFS[10]을 선정하였다. 각 방법은 상호정보척도 기반의 최신 방법들이다. 상호정보척도 계산을 위해 각 특징들은 이산화하였다[22]. 각 방법에 대해 선별된 특징 개수는 W 가 패턴 개수라고 할 때 \sqrt{W} 로 선정하였다[23]. 제안하는 방법의 파라미터 α 와 β 는 10^{-5} , 10^{-4} , ..., 10^3 까지 두고 비교 실험하여 가장 좋은 결과를 선정하여 기입하였다. Table 2-9는 4개의 데이터, 2개의 분류기에 따라 총 8가지의 결과를 보여준다. 각 테이블에서 첫 번째 행은 특징 선별 기법들, 두 번째 행부터 네 번째 행까지는 각각 Hloss, Rloss MIAcc 성능을 보여준다. 기법들 중에서 가장 좋은 성능을 보이는 기법 결과에 볼드체로 표기하였고, †는 해당 기법이 대응 표본 T 검정에서 다른 기법들보다 우수한 결과를 보일 때 표기하였다.

Table 2는 Emotions 데이터, MLkNN 분류기 실험 결과를 보여준다. 제안하는 방법이 모든 평가 방법에서 가장 좋은 성능을 보였다. 특히 모든 결과에서 통계적으로 좋은 성능을 보였다. Table 3는 Emotions 데이터, MLNB 분류기 실험 결과를 보여준다. 제안하는 방법이 모든 평가 방법에서 가장 좋은 성능을 보였다. Emotions 데이터는 대체로 MLkNN이 더 좋은 성능을 보여주었다. Table 3은 Birds 데이터, MLkNN 분류기 실험 결과를 보여준다. Rloss에서는 제안하는 방법이 가장 좋은 성능을 보였으나 Hloss, MIAcc에서는 MDMR이 가장 좋은 성능을 보였다. Table 4는 Birds 데이터, MLNB 분류기 실험 결과를 보여준다. Hloss, MIAcc에서는 제안하는 방법이 가장 좋은 성능을 보였고, Rloss에서는 MLCFS가 가장 좋은 성능을 보였다. Birds 데이터는 몇몇 특징이 이산 특징으로 상대적으로 상호정보척도를 이용하는 MDMR이나 MLCFS가 좋은 성능을 보인 것으로 판단된다. Table 6, 7은 Scene 데이터, Table 8, 9는 Yeast 데이터에서의 분류기 성능을 보여준다. 모든 평가 방법에서 제안하는 방법이 가장 좋은 성능을 보였다. Birds 데이터를 제외하고 모든 데이터에서

제안하는 방법이 좋은 성능을 보였고, Birds 데이터에서도 일부 평가 방법에서 좋은 성능을 보였다. 여러 데이터, 분류기, 평가 방법에 따라 수행된 다양한 실험 결과에서 제안하는 방법의 우수성을 볼 수 있다.

Table 2. Experimental result of MLkNN on Emotions data set

| | Hloss | Rloss | MIAcc |
|----------|----------------|----------------|----------------|
| AMI | 0.2171 | 0.2554 | 0.5178 |
| MDMR | 0.2172 | 0.2517 | 0.5131 |
| MLCFS | 0.2059 | 0.2438 | 0.5407 |
| PPT+RF | 0.2100 | 0.2471 | 0.5256 |
| QPFS | 0.2021 | 0.2406 | 0.5494 |
| Proposed | 0.1918† | 0.2179† | 0.5645† |

Table 3. Experimental result of MLNB on Emotions data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|----------------|----------------|
| AMI | 0.2654 | 0.1859 | 0.4894 |
| MDMR | 0.2602 | 0.1851 | 0.4893 |
| MLCFS | 0.2401 | 0.1757 | 0.5042 |
| PPT+RF | 0.2408 | 0.1888 | 0.4986 |
| QPFS | 0.2514 | 0.1840 | 0.4976 |
| Proposed | 0.2100 | 0.1583† | 0.5417† |

Table 4. Experimental result of MLkNN on Birds data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|---------------|---------------|
| AMI | 0.0481 | 0.2150 | 0.1801 |
| MDMR | 0.0468 | 0.2142 | 0.1802 |
| MLCFS | 0.0560 | 0.2270 | 0.1178 |
| PPT+RF | 0.0531 | 0.2452 | 0.1278 |
| QPFS | 0.0540 | 0.2393 | 0.1320 |
| Proposed | 0.0486 | 0.1868 | 0.1768 |

Table 5. Experimental result of MLNB on Birds data set

| | Hloss | Rloss | MIAcc |
|----------|----------------|---------------|----------------|
| AMI | 0.2409 | 0.1063 | 0.1017 |
| MDMR | 0.2302 | 0.1005 | 0.1062 |
| MLCFS | 0.0622 | 0.0853 | 0.1355 |
| PPT+RF | 0.2108 | 0.1354 | 0.0946 |
| QPFS | 0.1085 | 0.0858 | 0.1451 |
| Proposed | 0.0574† | 0.0930 | 0.1589† |

Table 6. Experimental result of MLkNN on Scene data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|----------------|----------------|
| AMI | 0.1448 | 0.2320 | 0.4867 |
| MDMR | 0.1271 | 0.2016 | 0.5610 |
| MLCFS | 0.1056 | 0.1674 | 0.6231 |
| PPT+RF | 0.1314 | 0.2266 | 0.5272 |
| QPFS | 0.1144 | 0.1824 | 0.5976 |
| Proposed | 0.0981 | 0.1493† | 0.6510† |

Table 7. Experimental result of MLNB on Scene data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|----------------|----------------|
| AMI | 0.2562 | 0.1651 | 0.4010 |
| MDMR | 0.2327 | 0.1244 | 0.4331 |
| MLCFS | 0.1764 | 0.0945 | 0.5285 |
| PPT+RF | 0.1985 | 0.1412 | 0.4694 |
| QPFS | 0.2171 | 0.1103 | 0.4580 |
| Proposed | 0.1510 | 0.0900† | 0.5680† |

Table 8. Experimental result of MLkNN on Yeast data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|----------------|----------------|
| AMI | 0.2121 | 0.2461 | 0.4924 |
| MDMR | 0.2106 | 0.2442 | 0.4956 |
| MLCFS | 0.2231 | 0.2722 | 0.4651 |
| PPT+RF | 0.2101 | 0.2518 | 0.4994 |
| QPFS | 0.2092 | 0.2427 | 0.5002 |
| Proposed | 0.2010 | 0.2310† | 0.5214† |

Table 9. Experimental result of MLNB on Yeast data set

| | Hloss | Rloss | MIAcc |
|----------|---------------|----------------|----------------|
| AMI | 0.2762 | 0.2491 | 0.4165 |
| MDMR | 0.2746 | 0.2483 | 0.4180 |
| MLCFS | 0.2194 | 0.1980 | 0.4301 |
| PPT+RF | 0.2357 | 0.2082 | 0.4397 |
| QPFS | 0.2473 | 0.2245 | 0.4420 |
| Proposed | 0.2126 | 0.1905† | 0.4509† |

제안하는 방법에 필요한 변수 α , β 에 따른 결과 변화를 실험적으로 확인하였다. Fig 2-5는 각각 Emotions, Birds, Scene, Yeast 데이터에 대한 제안하는 방법의 결과이다. 왼쪽 축이 α , 오른쪽 축이 β 를 나타내고 세로 축은 MIAcc 결과이다. Fig 2에서 Emotions 데이터는 α 와 β 값의 변함에 따라 성능의 변화는 적었고, 대체로 β 가 클 때 높은 성능을 보였다. β 는 저랭크에 주는 가중치로, Emotions 데이터는 저랭크 기반으로 특징을 선별할 때 좋은 결과임을 알 수 있다. Fig 3, Birds 데이터는 α 와 β 에 따라 상대적으로 변화의 폭이 컸다. 전체적으로 β 의 값이 클 때 좋은 성능을 보였다. 이는 Birds 데이터가 저랭크 가중치가 효과적이었다는 것을 보여준다. Fig 4, Scene 데이터의 경우 α 와 β 가 높을수록 좋은 성능을 보였다. 상대적으로 다른 데이터에 비해 가중치에 따른 성능의 편차가 큰 편이다. Fig 5, Yeast 데이터의 경우 α 와 β 에 따른 변화에 거의 성능 변화가 없어서 희소성과 저랭크에 가중치를 준 것으로 효과가 있었음을 보여준다.

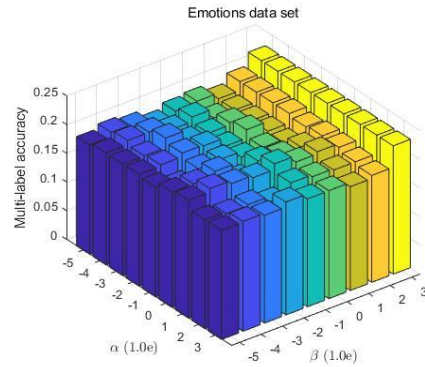


Fig. 2. Comparison of results according to changes in α and β on Emotions data set

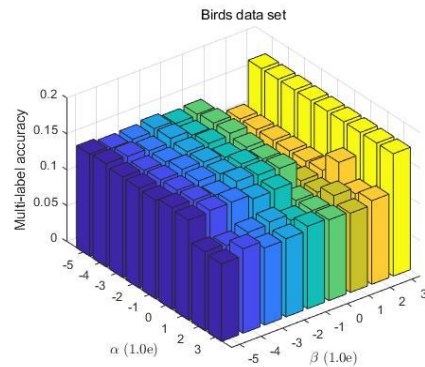


Fig. 3. Comparison of results according to changes in α and β on Birds data set

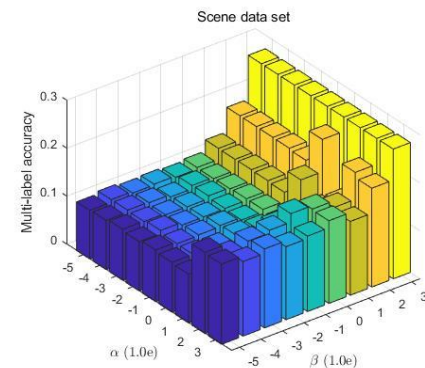


Fig. 4. Comparison of results according to changes in α and β on Scene data set

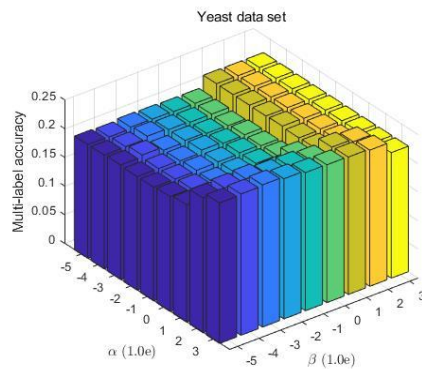


Fig. 5. Comparison of results according to changes in α and β on Yeast data set

V. Conclusions

본 논문에서는 다중 레이블 학습을 위한 특징 선별 기법을 제안하였다. 저랭크와 희귀성을 기반으로 불필요한 특징들을 제거하는 방법으로 기존 방법들에 비해 다중 레이블 학습에 더 효과적인 것을 확인하였다. 제안하는 방법의 빠르게 수렴하는 최적화 방법은 기존 상호정보척도를 계산해야하는 과정을 피할 수 있었다.

제안하는 방법의 장점들에도 불구하고, 추후 연구로서 몇가지 이슈들이 있다. 첫 번째, 희귀분석 형태의 목적 함수 때문에 이진 형태와 같은 범주형 데이터에 한계가 있다. 이를 개선하기 위해 상호정보척도 정보를 함께 고려할 수 있는 방법을 생각해볼 수 있다. 두 번째, 저랭크와 희귀성 사이의 가중치를 설정해야 한다. 가중치 설정을 위한 수학적 방법이나 실험적 방법을 통해 방향을 제시할 수 있는 방법이 필요할 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] R. Shang, Y. Meng, W. Wang, F. Shang, and L. Jiao, "Local discriminative based sparse subspace learning for feature selection," *Pattern Recognition*, Vol. 92, pp. 219-230, August 2019. DOI: 10.1016/j.patcog.2019.03.026
- [2] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 31, pp. 1532-1543, July 2018. DOI: 10.1109/TKDE.2018.2858782
- [3] Y. Zhang, Q. Wang, D.-W. Gong, and X.-F. Song, "Nonnegative laplacian embedding guided subspace learning for unsupervised feature selection," *Pattern Recognition*, Vol. 93, pp. 337-352, September 2019. DOI: 10.1016/j.patcog.2019.04.020
- [4] 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. DOI: 10.1016/j.patcog.2019.02.016
- [5] R. B. Pereira, A. Plastino, B. Zadrozny, and L. H. C. Merschmann, "Categorizing feature selection methods for multi-label classification," *Artificial Intelligence Review*, Vol. 49, pp. 57-78, 2018. DOI: 10.1007/s10462-016-9516-4
- [6] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the relieff algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, pp. 168-182, November 2015. DOI: 10.1016/j.neucom.2015.06.010
- [7] J. Lee, H. Lim, and D.-W. Kim, "Approximating mutual information for multi-label feature selection," *Electronics Letters*, vol. 54, pp. 76-77, July 2012. DOI: 10.1049/el.2012.1600
- [8] Y. Lin, Q. Hu, J. Liu, and J. Duan, "Multi-label feature selection based on max-dependency and min-redundancy," *Neurocomputing*, vol. 168 pp. 92-103, November 2015. DOI: 10.1016/j.neucom.2015.06.010
- [9] S. Jungjit, M. Michaelis, A. A. Freitas, J. Cinatl, "Two extensions to multi-label correlation-based feature selection: A case study in bioinformatics," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1519-1524, 2013.
- [10] H. Lim, J. Lee, and D.-W. Kim, "Optimization approach for feature selection in multi-label classification," *Pattern Recognition Letters*, vol. 89, pp. 25-30, April 2017. DOI: 10.1016/j.patrec.2017.02.004
- [11] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEE access*, Vol. 3, pp. 490-530, May 2015. DOI: 10.1109/ACCESS.2015.2430359
- [12] 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
- [13] 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, pp. 497-504, March 2017. DOI: 10.1109/TPAMI.2017.2688363
- [14] D. P. Bertsekas, "Constrained optimization and Lagrange

- multiplier methods," Academic press, 2014.
- [15] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, pp. 2038-2048, July 2007. DOI: 10.1016/j.patcog.2006.12.019
- [16] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Information Sciences*, vol. 179, pp. 3218-3229, September 2009. DOI: 10.1016/j.ins.2009.06.010
- [17] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 1819-1837, August 2014. DOI: 10.1109/TKDE.2013.39
- [18] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas, "Multi-label classification of music into emotions," *ISMIR*, pp. 325-330, 2008.
- [19] F. Briggs, H. Yonghong, R. Raich, and others, "New methods for acoustic classification of multiple simultaneous bird species in a noisy environment," *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1-8, 2013.
- [20] M. R. Boutell, J. Luo, X. Shen, and C. M. Xipeng, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, pp. 1757-1771, September 2004. DOI: 10.1016/j.patcog.2004.03.009
- [21] A. Elisseeff, J. Weston, and others, "A kernel method for multi-labelled classification," *NIPS*, pp. 681-687, 2001.
- [22] A. Cano, J. M. Luna, E. L. Gibaja, and S. Ventura, "LAIM discretization for multi-label data," *Information Sciences*, Vol. 330, pp. 370-384, 2016. DOI: 10.1016/j.ins.2015.10.032
- [23] J. Lee and D.-W. Kim, "Scfs: Multi-label feature selection based on scalable criterion for large label set," *Pattern Recognition*, vol. 66, pp. 342-352, June 2017. DOI: 10.1016/j.patcog.2017.01.014

Author



Hyunki Lim is currently an assistant professor in the Division of 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.