

## Multi-label Feature Selection Using Redundancy and Relevancy based on Regression Optimization

Hyunki Lim\*

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

### [Abstract]

High-dimensional data causes difficulties in machine learning due to high time consumption and large memory requirements. In particular, in a multi-label environment, higher complexity is required as much as the number of labels. This paper proposes a feature selection method to improve classification performance in multi-label settings. The method considers three types of relationships: between features, between features and labels, and between labels themselves. To achieve this, a regression-based objective function is designed. This objective function calculates the linear relationships between features and labels and uses mutual information to compute relationships between features and between labels. By minimizing this objective function, the optimal weights for feature selection are found. To optimize the objective function, a gradient descent method is applied to develop a fast-converging algorithm. The experimental results on six multi-label datasets show that the proposed method outperforms existing multi-label feature selection techniques. The classification performance of the proposed method, averaged over six datasets, showed a Hamming loss of 0.1285, a ranking loss of 0.1811, and a multi-label accuracy of 0.6416. Compared to the AMI(Approximating Mutual Information) algorithm, the performance was better by 0.0148, 0.0435, and 0.0852, respectively.

▶ **Key words:** Feature Selection, Multi-label Classification, Regression Model Optimization, Mutual Information, Machine Learning

### [요 약]

고차원 데이터는 기계학습에서 높은 시간 소모와 큰 메모리 요구로 학습에 어려움을 유발한다. 특히 다중 레이블 환경에서는 레이블의 개수만큼 더 높은 복잡도를 요구한다. 본 논문에서는 다중 레이블 환경에서 분류 성능 향상을 위한 특징 선별 기법을 제안한다. 중요한 특징을 선별하기 위해 특징과 특징, 특징과 레이블, 레이블과 레이블, 세 가지 관계를 고려했으며 이를 위해 회귀 기반 목적 함수를 설계하였다. 이 목적 함수는 특징과 레이블 사이의 선형적인 관계를 계산하고 상호 정보 기반의 특징과 특징, 레이블과 레이블 사이의 관계를 계산한다. 이 목적 함수를 최소화하는 가중치를 찾아 특징을 선별할 수 있다. 이 목적 함수 최적화를 위해 경사하강법 방식을 제시하여 빠르게 수렴할 수 있는 알고리즘을 제안하였다. 여섯 개의 다중 레이블 데이터에 대한 실험 결과 제안된 방법이 기존 다중 레이블 특징 선별 기법보다 높은 성능을 보여주었다. 여섯 개의 데이터 평균으로 본 제안 방법의 분류 성능은 해밍 로스 0.1285, 랭킹 로스 0.1811, 다중 레이블 정확도 0.6416으로, 비교 대상 알고리즘인 AMI(Approximating Mutual Information)에 비해 각각 0.0148, 0.0435, 0.0852 더 우수한 성능을 보였다.

▶ **주제어:** 특징 선별, 다중 레이블 분류, 회귀 모델 최적화, 상호 정보, 기계학습

- First Author: Hyunki Lim, Corresponding Author: Hyunki Lim
- \*Hyunki Lim (hlim20@kyonggi.ac.kr), Div. of AI Computer Science and Engineering, Kyonggi University
- Received: 2024. 08. 20, Revised: 2024. 11. 14, Accepted: 2024. 11. 14.

## I. Introduction

기계학습, 패턴인식 분야에서 고차원 데이터를 다루어야 할 경우가 있다 [1]. 이 고차원 데이터는 중복되거나 학습에 방해되는 정보를 포함할 수 있을 뿐만 아니라 학습 알고리즘의 시간이 오래 걸리고 많은 메모리를 요구하기 때문에 해당 분야에서 처리하기 어렵다 [2]. 또한 고차원 데이터는 학습 알고리즘의 성능을 저하시키고 응용을 어렵게 한다. 더 나아가 주어진 데이터가 다중 레이블(Label)인 경우 다루어야 할 복잡도는 더욱 커지게 된다 [3]. 다중 레이블 환경에서 고차원 데이터의 문제를 극복하기 위해 많은 연구들에서 특징 선별 기법을 도입하였다. 다중 레이블 특징 선별 기법은 특징 평가 척도에 따라 중요한 특징을 선택하고 주어진 특징 집합에서 불필요한 특징을 제거하는 알고리즘이다. 특징 선별의 결과로, 선택된 특징 집합은 기계학습의 정확도를 향상시킬 수 있고, 학습하는 시간을 줄이며, 데이터의 해석력을 증가시킬 수 있다. 또한 차원의 저주, 과적합 등의 위험 요소를 줄일 수 있다 [4].

특징 선별 방법은 클래스 레이블 정보의 주어지는 정도에 따라 세 가지 범주로 나눌 수 있다. 레이블 정보가 모두 주어진 경우 지도 학습, 레이블 정보가 전혀 주어지지 않은 경우 비지도 학습, 레이블 정보가 일부 주어진 경우 준지도 학습으로 나눈다 [5]. 지도 학습은 레이블 정보를 사용하여 클래스 분류 성능을 높일 수 있는 주요한 특징을 선별한다. 준지도 학습은, 적은 수의 레이블 데이터와 많은 수의 레이블이 없는 데이터를 이용하여 특징을 선별한다. 비지도 학습은 레이블 정보 없이 분산과 같은 정보를 이용하여 주어진 데이터의 형태를 유지할 수 있는 특징을 선별한다. 본 논문에서는 다중 레이블 분류 성능을 향상시킬 수 있는 지도 학습을 목적으로 한다.

다중 레이블 특징 선별 기법들은 상호 정보 (Mutual Information)를 이용하여 여러 레이블들과 특징들 사이의 연관성을 계산하여 주요한 특징을 선별하고자 한다. 하지만 다중 레이블 특성상 상호 정보 계산에 높은 복잡도가 요구되고, 부분 최적(Sub-optimal)에 빠지기 쉽다 [6].

본 논문에서는 회귀 방식의 목적 함수와 일차 상호 정보 기반의 목적 함수를 병합하여 새로운 목적 함수를 제안한다. 이 목적 함수는 단순하지만 특징 선별을 위한 주요 방식을 반영하였다. 또한 경사하강법 방식을 이용하여 빠르게 수렴할 수 있는 알고리즘을 제안하였다. 실험 결과, 제안하는 방법이 기존 다중 레이블 특징 선별 기법들보다 좋은 분류 성능을 보여주었다.

본 논문의 주요 기여 요소는 다음과 같다.

새로운 목적 함수 제안: 다중 레이블 특징 선별 기법에서 기존의 상호 정보 기반 방법의 한계를 극복하기 위해, 회귀 기반 목적 함수와 상호 정보 기반 목적 함수를 병합한 새로운 목적 함수를 제안하였다. 이 목적 함수는 단순하면서도 특징 선별에 필요한 주요 정보를 반영할 수 있도록 설계되었다.

효율적인 알고리즘 개발: 제안된 목적 함수의 최적화를 위해 경사하강법을 기반으로 한 효율적인 알고리즘을 개발하였다. 이 알고리즘은 빠르게 수렴할 수 있으며, 기존 방법들보다 계산 복잡도를 줄이는 데 초점을 맞추었다.

다중 레이블 환경에서의 성능 향상: 여섯 개의 다중 레이블 데이터에 대한 실험을 통해 제안된 방법이 기존의 다중 레이블 특징 선별 기법보다 우수한 성능을 보였음을 확인하였다.

이와 같은 기여를 통해, 본 연구는 다중 레이블 분류 문제에서 특징 선별을 위한 새로운 접근 방식을 제시하며, 다양한 응용 분야에서의 활용 가능성을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 다중 레이블 분류 및 특징 선별 기법에 대해 설명한다. 3장에서는 제안된 방법론에 대해 상세히 설명하며, 회귀 기반 목적 함수와 상호 정보 기반 목적 함수의 병합에 대해 기술한다. 4장에서는 여섯 개의 다중 레이블 데이터 세트(Dataset)를 사용한 실험 결과를 제시하고, 제안된 기법의 성능을 기존 기법들과 비교하여 분석한다. 마지막으로 5장에서는 본 연구의 결론을 요약하고, 향후 연구 방향을 제시한다.

## II. Related Works

다중 레이블을 다루기 위해 레이블을 단일 레이블로 변환하는 방식과 다중 레이블을 그대로 활용하는 방식으로 나눌 수 있다 [6]. 단일 레이블로 변환하게 되면 기존 단일 레이블을 위한 특징 선별 기법을 그대로 활용할 수 있다는 장점이 있다. 하지만 변환 과정에서 정보 손실이 발생할 수 있다는 단점이 있다. 레이블 멱집합(Label Powerset, LP) 방식은 다중 레이블을 가능한 모든 경우를 고려하여 단일 레이블로 변환하는 방식이다 [7]. 이 방식은 단순하지만 레이블 개수가  $L$ 개일 때 최대  $2^L$ 개의 클래스 레이블을 생성하게 되어 불균형이 심해지는 단점이 있다. LP 방식에서 자주 나타나지 않는 레이블을 삭제하여 불균형을 완화시키는 PPT(Pruned Problem Transformation)이 제안되었다 [8]. 하지만 이 방식은 레이블 정보의 손실을 받

생시키게 된다.

레이블을 변환하지 않고 특징 선별하는 방식은 크게 필터(Filter) 방법, 래퍼(Wrapper) 방법, 내장형(Embedded) 방법으로 나눌 수 있다. 래퍼 방식으로 클래스 분류기의 성능을 직접적으로 이용하여 최대화할 수 있는 특징을 선별한다. 최적화를 위해 부스팅(Boosting), 유전 알고리즘(Genetic Algorithm) [9] 등이 사용된다. 이 방식은 분류기의 성능을 계산해야 하는 단점이 있다. 내장형 방법은 클래스 분류기가 내부적으로 특징을 선별하는 과정을 포함한다. 의사결정트리(Decision Tree) 방식이 분류기 내부에서 특징 선별 과정을 포함하고 있다 [10].

필터 방법은 클래스 분류기와 독립적인 평가 방법을 이용하여 특징을 선별하는 방식이다. 이때 상호 정보를 이용한 방식을 이용한 평가 방식이 많이 제안되었다. PMU(Pairwise Multi-label Utility) 방식은 레이블 변환 없이 레이블 간 연관성을 상호 정보를 이용하여 계산하는 필터 방식이다 [11]. 하지만 이 방식은 지역 최적해만을 찾게 된다. QPFS(Quadratic Programming Feature Selection) 방식은 특징과 레이블 간의 상호 정보를 수치 최적화 형태로 변환하여 지역 최적해 문제를 해소하고자 하였다 [3]. 하지만 많은 상호 정보 계산을 요구하게 된다.

회귀 분석 방식의 내장형 방법들이 제안되었다. LFFS(MLFS Method based on Label Correlations and Feature Redundancy) 방식은 릿지 회귀(Ridge Regression)를 이용하여 행렬을 설계하고  $l_{2,1}$ -norm 방식을 활용하여 다중 레이블 특징 선별 기법을 제안하였다 [12]. RFSR(Robust Flexible Sparse Regularization Norm) 방식은  $l_1$ -norm 방식과  $l_{2,1}$ -norm 방식 사이에서 유연하게 특징을 선별할 수 있는 기법을 제안하였다 [13]. LCIFS(Learning Correlation Information for Multi-label Feature Selection) 방식은 스펙트럼 그래프(Spectral Graph)를 이용하여 레이블간 연관성을 고려하고 동시에 중복되는 특징을 고려할 수 있는 방식을 제안하였다 [14]. Y. Li 외는 2차 이상의 레이블간 연관성을 고려하기 위한 행렬을 설계하였고,  $l_{2,1}$ -norm 방식을 이용하여 중복되는 특징을 고려하는 다중 레이블 특징 선별 기법을 제안하였다 [15]. RWFS(Relevance based on Weight Feature Selection) 방식은 특징과 레이블 사이의 상호 정보를 가중치 행렬로 표현하여 목적 함수를 설계하였고 이를 기반으로 하는 다중 레이블 특징 선별 기법을 제안하였다 [16]. J. Dai 외는 퍼지 상호 정보(Fuzzy Mutual Information)를 이용하여 중요하게 연관된 레이블들에 가

중치를 주고 해당 레이블들을 조건부로 계산하는 조건부 상호 정보(Conditional Mutual Information) 기반으로 특징이 선별할 수 있는 알고리즘을 제안하였다 [17]. M. Faraji 외는 레이블 간 상관관계와 불균형 분포를 고려하여 다중 특징 선별 기법을 제안하였다 [18]. 이를 위해 특징 행렬과 레이블 행렬 사이의 공유되는 패턴을 찾을 수 있도록 목적 함수를 설계하였다. CLE-MLFS(Multi-label Feature Selection based on Correlation Label Enhancement) 방식은 레이블 간 연관성의 차이를 고려하여 특징 선별을 하도록 설계되었다 [19]. 이를 위해 데이터 샘플(Sample) 사이의 연관성을 고려하여 레이블에 이 정보를 투여하였고, 희소 선형 회귀 모델에 적용하였다. MFSJMI(Multi-label Feature Selection Considering Join Mutual Information and Interaction Weight) 방식은 레이블 간의 연관성을 하나의 레이블 쌍만 고려하도록 하여 하나의 특징이 여러 레이블을 고려할 수 있는 상호 정보를 계산할 수 있도록 설계되었다 [20]. Y. Yong 외는 레이블 간 상관관계와 레이블 고유의 특징을 모두 고려한 새로운 방법을 제안하였다 [21]. 레이블 상관관계의 탐색 과정을 제한하고 레이블 고유 특징을 학습하며, 노이즈(Noise)와 이상값의 영향을 줄이기 위해 탐색 방법을 사용하였다. 제안하는 방식은 내장형 방법이다.

### III. Proposed Method

#### 1. Objective Function

$n$ 개의 패턴과  $d$ 개의 특징으로 이루어진 데이터  $X \in \mathbb{R}^{n \times d}$ 와  $c$ 개의 레이블로 이루어진 다중 레이블 집합  $Y = \{y_1, \dots, y_c\} \subseteq \{0, 1\} \in \mathbb{R}^{n \times d}$ 라고 하자. 이때  $Y_{ij}$ 가 1인 것은  $i$ 번째 패턴이  $j$ 번째 레이블에 속하는 것을 의미하고 그렇지 않으면  $Y_{ij}$ 은 0으로 표현된다. 이때 회귀 기반 특징 선별 기법의 기본적인 목적 함수는 다음과 같다.

$$\min_w \|XW - Y\|_F^2, \quad (1)$$

$$\text{subject to } W \geq 0$$

여기서  $W \in \mathbb{R}^{d \times c}$ 는 가중치 행렬이며 비음수 제한을 가진다. 특징의 선택을 양수 값으로, 특징이 선택되지 않음을 0으로 표현하기 위해 가중치 행렬에 비음수 제한을 둔다. 절댓값이 클수록 해당 신뢰성 높은 특징으로 여길 수 있다.  $w_{(i)}$ 를  $W$ 에서  $i$ 번째 행이라고 할 때, 최적화를 통해 얻게 된  $W$ 에서 이 행의 크기가 큰 값이 특징으로 선

별된다. 이 목적 함수는 데이터를  $W$ 에 의해 사영한  $XW$ 와 다중 레이블  $Y$ 사이의 차이를 최소화하는 특징을 찾고자 설계되었다. 목적 함수 설계를 위해 프로베니우스(Frobenius)-norm을 사용하였다. 이 목적 함수는  $W$ 에 대해 볼록(Convex) 함수가 되어 효율적으로 최적화 문제를 풀 수 있다. 이 목적 함수의 최적화 과정을 통해 결정된  $W$ 의 각 행의 크기가 큰 상위  $k$ 개 특징이 선별된다. 하지만 이 방식은 데이터와 레이블 사이의 선형적인 관계만 고려한다는 한계를 가진다.

많은 연구에서 상호 정보를 활용한 특징 선별 방식이 제안되었다. 상호 정보를 통해 특징과 레이블 사이의 연관성을 의미하는 관련도(Relevancy)와 특징과 특징 사이의 연관성을 의미하는 중복도(Redundancy)를 고려한다 [12,17,21-23]. 임의의 두 확률변수  $x$ 와  $y$  사이의 상호 정보는 다음과 같이 정의된다.

$$MI(A,B) = \sum_{a \in A} \sum_{b \in B} P(a,b) \log \left( \frac{P(a,b)}{P(a)P(b)} \right). \quad (2)$$

여기서  $P(a,b)$ 는  $A$ 와  $B$ 의 결합 확률 분포 함수를 나타내고,  $P(a)$ 는  $A$ 의 주변 분포 함수를 나타낸다.  $i$ 번째 특징 벡터  $x_i$ ,  $j$ 번째 레이블 벡터를  $y_j$ 라고 할 때, 상호 정보를 이용한 관련도와 중복도는 각각  $MI(x_i, y_j)$ ,  $MI(x_i, x_j)$ 로 표현된다. 다중 레이블 환경에서는 추가적으로 레이블과 레이블 사이의 연관성을  $MI(y_i, y_j)$ 로 고려한다 [24]. 이 연관성을 계산하여 세 개의 행렬  $Q \in \mathbb{R}^{d \times d}$ ,  $R \in \mathbb{R}^{c \times c}$ ,  $S \in \mathbb{R}^{d \times c}$ 로 표현될 수 있고, 행렬의 각 성분은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} q_{ij} &= MI(x_i, x_j), \\ r_{ij} &= 1 - MI(y_i, y_j), \\ s_{ij} &= 1 - MI(x_i, y_j). \end{aligned} \quad (3)$$

행렬들의 각 성분은 특징과 레이블 사이의 연관성을 나타내고, 상호 정보를 계산하기 위해 데이터를 이진화 한다면 모든 성분은 0보다 큰 값을 가지게 된다.  $Q$ 는 중복도를 나타내고 특징 선별에서 중복되는 특징은 최소화해야 한다.  $R$ 은 레이블 사이의 연관성으로 구성되어 있고 음수화를 통해 전체 값은 최소화되어야 레이블 간 연관성이 높은 특징이 선별될 수 있다.  $S$ 는 특징과 레이블 사이의 연관성으로 음수화를 통해 전체 값은 최소화되어야 레이블과 연관성이 높은 특징이 선별될 수 있다. 이 값들을 회귀식 (1)과 병합하여 다음 (4) 식과 같은 목적 함수  $J$ 를 설계할 수 있다.

$$\begin{aligned} J = \min_W & \|XW - Y\|_F^2 + \alpha \text{Tr}(W^T Q W) + \\ & \beta \text{Tr}(R W^T W) + \gamma \text{Tr}(S^T W), \\ & \text{subject to } W \geq 0. \end{aligned} \quad (4)$$

여기서  $\alpha$ ,  $\beta$ ,  $\gamma$ 는 각각 중복도, 레이블간 연관성, 관련도에 대한 가중치를 의미하고 모두 양수이다.  $\text{Tr}$ 은 정사각 행렬의 대각합(Trace)를 의미한다.  $\text{Tr}(W^T Q W)$ 에서  $i$ 번째 값  $w_{(i)}^T Q w_{(i)}$ 은  $i$ 번째 특징이 다른 특징들 간의 중복도를 모두 합한 값을 의미하게 된다. 비슷한 방식으로  $R w_{(i)}^T w_{(i)}$ 은  $i$ 번째 특징과 레이블 사이의 연관성의 정도인  $S^T w_{(i)}$ 은  $i$ 번째 특징의 모든 레이블 간의 관련도의 합을 나타낸다. Fig 1은 제안하는 방식이 고려하는 요소들과 이를 바탕으로 제안하는 최종 목적 함수를 보여준다. 제안하는 방법은 특징 선별을 위한 네 가지 요소를 고려하였고, 이를 하나의 목적 함수로 설계하였다.

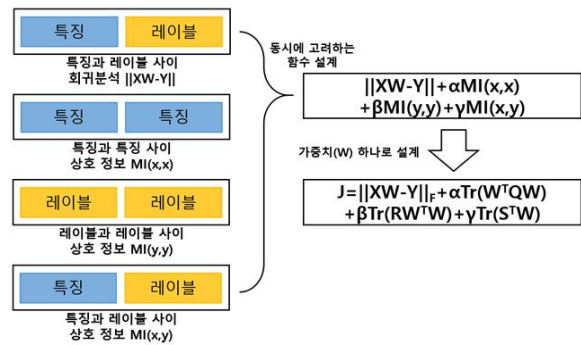


Fig. 1. Factors considered for the proposed feature selection approach and the proposed objective function

## 2. Optimization

제약 조건을 가지는 목적 함수의 최적화 문제를 풀기 위하여 라그랑주 승수법(Lagrange Multiplier Method)을 사용한다. 이에 따라  $W$ 에 대한 비음수 제약 조건을 제거하여 제약 조건이 없는 다음과 같은 목적 함수로 최적화 문제가 변경된다.

$$\begin{aligned} L(W, \phi) = \min_W & \|XW - Y\|_F^2 + \\ & \alpha \text{Tr}(W^T Q W) + \beta \text{Tr}(R W^T W) + \\ & \gamma \text{Tr}(S^T W) - \phi \odot W. \end{aligned} \quad (5)$$

여기서  $\phi \in \mathbb{R}^{d \times c}$ 는 라그랑주 승수(Lagrange Multiplier)이며 최적화 변수의 개수만큼 필요하다.  $\odot$ 는 아다마르 곱(Hadamard Product)으로 크기가 같은 행렬에서 상응하는 원소끼리 곱하는 연산이다. 라그랑주 승수는 이때 이 최적화 문제의 KKT(Karush-Kuhn-Tucker Conditions) 조건은 다음과 같다.

- 1)  $\nabla J(W^*) - \phi = 0$  (Stationarity)
- 2)  $\phi \odot W^* = 0$  (Complementary Slackness)
- 3)  $\phi \geq 0$  (Dual Feasibility)
- 4)  $W \geq 0$  (Primal Feasibility)

여기서  $W^*$ 는 주어진 최적화 문제에서의 최적의 값을 의미한다. 이 목적 함수는 이차 볼록 함수의 형태로 다음과 같이 기울기를 계산할 수 있다.

$$\frac{\partial L}{\partial W} = 2X^T XW - 2X^T Y + 2\alpha QW + 2\beta WR + \gamma S - \phi. \quad (6)$$

이 기울기가 0인 식은  $W$ 에 대한 실베스터 등식 (Sylvester Equation) 형태로 일반해가 존재하므로 KKT 조건을 만족하는  $W^*$ 와  $\phi$ 가 존재할 수 있음을 알 수 있고, 이 기울기를 이용하여  $W$ 에 대한 최적화 알고리즘을 쉽게 설계할 수 있다. 이 기울기를 이용하여 다음과 같이  $W$ 가 최소화되는 경사하강법 알고리즘을 다음 Algorithm 1과 같이 설계할 수 있다. 이 알고리즘은 하이퍼파라미터 (Hyperparameter)  $\alpha$ ,  $\beta$ ,  $\gamma$ 를 설정한 후 주어진 데이터에 대해 특징 간의 상호 정보 및 레이블 간의 상호 정보를 이용하여  $Q$ ,  $R$ ,  $S$ 를 계산한다. 이후 가중치 행렬  $W$ 를 반복적으로 업데이트하는 방식으로 특징의 중요도를 계산한다. 최종적으로 각 특징의 중요도를  $W$  기반으로 정렬하여, 중요한 특징들을 선별한다.

제안하는 알고리즘의 복잡도는 목적 함수 내의 상호 정보 값들을 나타내는  $Q$ ,  $R$ ,  $S$ 의 계산과 경사하강법의 계산으로 이루어진다.  $Q$ 는 특징 사이의 상호 정보로  $O(d^2)$ ,  $R$ 은 레이블 사이의 상호 정보로  $O(c^2)$ ,  $S$ 는 특징과 레이블 사이의 상호 정보로  $O(dc)$ 만큼의 계산이 요구된다. 여기서  $O(\cdot)$ 는 빅오 표기법(Big-O Notation)을 의미한다. 경사하강법에서 경사를 계산하는 데에는  $O(d^2c)$ 만큼의 시간이 요구되고 수렴하기 위한 반복 횟수는 상수만큼 요구되어 최종적인 복잡도는  $O(d^2c + c^2)$ 이다.

#### Algorithm 1. Proposed method

- 1 입력:  $X \in \mathbb{R}^{d \times n}, \alpha, \beta, \gamma$
- 2 초기화:  $t = 0, W \geq 0, \maxiter$
- 3 while  $t \leq \maxiter$  do
- 4  $V \leftarrow 2X^T XW + 2\alpha QW + 2\beta WR + \gamma S$
- 5  $W \leftarrow W \odot \frac{V}{2X^T Y}$
- 6 end while
- 7 출력:  $W$ 의 각 행의 크기  $w_{(i)}$ 를 내림차순으로 정렬한 특징 번호

## IV. Experiments

이 장에서는 제안하는 방법의 클래스 분류 성능 향상의 정도를 비교하기 위해 실험 결과를 보인다. 다중 레이블  $k$ -근접 이웃 (Multi-label  $k$ -nearest Neighbors, ML $k$  NN) 분류기[25]와 선형 SVM(Linear Support Vector Machine, LinSVM) 분류기[13,16,20]를 이용하여 클래스 분류 성능을 비교하였다. ML $k$ NN에서  $k$ 값은 5로 설정하였다. ML $k$ NN은 많은 연구에서 다중 레이블 특징 선별의 성능을 비교하기 위해 사용되었다 [3-4,12-21]. 훈련 데이터와 테스트 데이터를 무작위로 8:2 나누는 교차 검증 방식으로 10번씩 결과를 내어 평균값을 기입하였다.

여섯 개의 데이터를 실험에 이용하였다 [26-30]. emotions 데이터 세트는 593개의 노래로 구성된 음악 데이터로, 72개의 음악 특징을 포함하며, 각 곡은 6개의 주요 감정 레이블로 분류된다. enron 데이터 세트는 어떤 회사의 이메일 데이터를 기반으로 한 텍스트 데이터 세트이다. genbase 데이터 세트는 단백질 기능 분류를 위한 생물학 데이터로, 각 인스턴스는 단백질을, 각 레이블은 단백질의 기능을 나타낸다. medical 데이터 세트는 978개의 임상 자유 텍스트 보고서로 구성되어 있으며, 각 보고서는 최대 45개의 질병 코드(Code)로 레이블링(Labeling)되어 있다. scene 데이터 세트는 2407개의 풍경 이미지를 포함하며, 294개의 시각적 특징과 6개의 클래스 레이블로 주석이 달려 있다. yeast 데이터 세트는 2417개의 효모 유전자 데이터를 포함한 생물학적 데이터 세트로, 각 효모 유전자는 최대 14개의 레이블(예: 에너지, 전사 등)로 주석이 달려 있다. 각 데이터에 대한 세부 사항은 Table 1에 정리하였다.

Table 1. Information about datasets

Data	# Sample	# Feature	# Label
emotions	593	72	6
enron	1702	1001	53
genbase	662	1185	27
medical	978	1494	45
scene	2407	294	6
yeast	2417	103	14

평가 방법으로 해밍 로스(Hamming loss, hloss), 랭킹 로스(Ranking loss, rloss), 다중 레이블 정확도 (Multi-label accuracy, mlacc)를 사용하였다 [4]. 해밍 로스와 랭킹 로스는 낮을수록 좋은 분류 성능을 나타내고, 다중 레이블 정확도는 높을수록 좋은 성능을 나타낸다.

선별한 특징의 개수는 데이터의 패턴 개수  $N$ 개 기준  $\sqrt{N}$ 를 결정하여 비교하였다[31]. 제안하는 방법과 비교할 5가지 기존 방법, AMI[32], MDMR[33], FIMF[4], QPFS[3]를 선정하였다. AMI는 특징과 레이블 사이의 1차 상호 정보를 고려하여 특징을 선별한다. MDMR은 특징과 레이블, 특징과 특징 사이의 상호 정보를 고려하여 새로운 특징 평가 함수를 설계하여 특징을 선별한다. FIMF는 빠른 특징 선별을 위해 고려할 레이블의 개수를 제한하는 상호 정보 기반 평가 함수로 특징을 선별한다. 비교 대상과 제안하는 방법에 필요한 매개변수  $\alpha, \beta, \gamma$ 는  $10^{-3}, 10^{-2}, \dots, 10^3$ 까지 두고 비교 실험하여 가장 좋은 결과를 선정하여 기입하였다.

Table 2. MLkNN performance results on emotions dataset (24 selected features)

	hloss	rloss	mlacc
AMI	0.2456	0.2674	0.4682
MDMR	0.2380	0.2657	0.4849
FIMF	0.2470	0.2715	0.4730
QPFS	0.2470	0.2734	0.4631
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.2237</b> ( $10^{-3}, 10^{-1}, 10^{-3}$ )	<b>0.2471</b> ( $10^{-3}, 1, 10$ )	<b>0.5133</b> ( $10^2, 10^{-3}, 10^{-3}$ )

Table 3. MLkNN performance results on enron dataset (41 selected features)

	hloss	rloss	mlacc
AMI	0.0620	0.3113	0.2631
MDMR	0.0570	0.2812	0.2992
FIMF	0.0531	0.2831	0.3445
QPFS	0.0610	0.2985	0.2793
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0506</b> ( $10^2, 1, 1$ )	<b>0.2591</b> ( $10^2, 10^{-3}, 10$ )	<b>0.3931</b> ( $10^{-3}, 10^2, 10$ )

Table 4. MLkNN performance results on genbase dataset (26 selected features)

	hloss	rloss	mlacc
AMI	0.0040	0.0445	0.9568
MDMR	0.0040	0.0445	0.9564
FIMF	0.0043	0.0445	0.9527
QPFS	0.0039	0.0445	0.9577
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0019</b> ( $10^{-2}, 10, 10^{-1}$ )	<b>0.0386</b> ( $10^2, 10^{-1}, 10^{-2}$ )	<b>0.9791</b> ( $10^{-3}, 1, 10^{-1}$ )

Table 5. MLkNN performance results on medical dataset (31 selected features)

	hloss	rloss	mlacc
AMI	0.0077	0.1189	0.7746
MDMR	0.0076	0.1127	0.7772
FIMF	0.0074	0.1111	0.7816
QPFS	0.0085	0.1251	0.7467
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0023</b> ( $10^2, 1, 1$ )	<b>0.0478</b> ( $10^2, 10^{-1}, 1$ )	<b>0.9486</b> ( $10^{-2}, 1, 1$ )

Table 6. MLkNN performance results on scene dataset (49 selected features)

	hloss	rloss	mlacc
AMI	0.1723	0.3213	0.4463
MDMR	0.1569	0.3156	0.4902
FIMF	0.1266	0.2210	0.5546
QPFS	0.1558	0.3150	0.4924
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.1256</b> ( $10^3, 10^{-1}, 10^2$ )	<b>0.2066</b> ( $10^3, 10^{-1}, 10^{-3}$ )	<b>0.5812</b> ( $10^3, 10^{-1}, 10^{-3}$ )

Table 7. MLkNN performance results on yeast dataset (49 selected features)

	hloss	rloss	mlacc
AMI	0.2399	0.2842	0.4296
MDMR	0.2398	0.2876	0.4287
FIMF	0.2392	0.2850	0.4296
QPFS	0.2404	<b>0.2838</b>	0.4281
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.2383</b> ( $10^{-3}, 10^2, 10^{-3}$ )	0.2876 ( $10^2, 1, 10$ )	<b>0.4344</b> ( $10^{-3}, 10^2, 10^{-3}$ )

Table 2-7는 6개의 데이터에 대해 각각의 MLkNN 분류 성능을 보여준다. 각 테이블에서 첫 번째 행은 특징 선별 기법들, 두 번째부터 네 번째 행은 해밍 로스 랭킹 로스, 다중 레이블 정확도 성능을 나타낸다. 기법들 중에서 가장 좋은 성능을 보이는 기법 결과에 볼드체로 표기하였다. 제안하는 방법의 경우 해당 결과를 나타낸 매개변수  $\alpha, \beta, \gamma$  값을 기입하였다.

실험 결과 yeast 데이터를 제외하고, 모든 데이터에서 제안하는 방법이 가장 좋은 성능을 보였다. 특히 enron, medical 데이터에서 제안하는 방식이 기존 방법에 비해 큰 차이를 내며 좋은 성능을 보였다. genbase는 특징 값이 희소하여 성능의 차이가 많이 나타나지 않았고, yeast 데이터에서는 랭킹 로스에서 QPFS의 성능이 가장 좋았다.

Table 8. LinSVM performance results on emotions dataset (24 selected features)

	hloss	rloss	mlacc
AMI	0.2290	0.1954	0.4622
MDMR	0.2270	0.1946	0.4736
FIMF	0.2239	0.1949	0.4713
QPFS	0.2304	0.1940	0.4551
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.2168</b> ( $10^3, 10, 10^2$ )	<b>0.1780</b> ( $10, 10^3, 10^2$ )	<b>0.4887</b> ( $10^3, 10, 10^2$ )

Table 9. LinSVM performance results on enron dataset (41 selected features)

	hloss	rloss	mlacc
AMI	0.0571	0.1551	0.2270
MDMR	0.0535	0.1348	0.2844
FIMF	0.0511	0.1377	0.3411
QPFS	0.0562	0.1465	0.2378
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0481</b> ( $10^2, 10^{-3}, 10$ )	<b>0.1351</b> ( $10^2, 10^{-2}, 10$ )	<b>0.3896</b> ( $10^2, 10^{-3}, 10$ )

Table 10. LinSVM performance results on genbase dataset (26 selected features)

	hloss	rloss	mlacc
AMI	0.0031	0.0070	0.9650
MDMR	0.0031	0.0067	0.9650
FIMF	0.0033	0.0063	0.9624
QPFS	0.0030	0.0072	0.9662
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0010</b> ( $10^{-3}, 10^{-2}, 1$ )	<b>0.0027</b> ( $10^3, 1, 10^{-1}$ )	<b>0.9889</b> ( $10^{-3}, 10^{-2}, 1$ )

Table 11. LinSVM performance results on medical dataset (31 selected features)

	hloss	rloss	mlacc
AMI	0.0054	0.0393	0.8103
MDMR	0.0053	0.0370	0.8156
FIMF	0.0050	0.0365	0.8304
QPFS	0.0063	0.0466	0.7912
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.0001</b> (10,10,1)	<b>0.0079</b> ( $10^{-1}, 10, 1$ )	<b>0.9737</b> ( $10^{-2}, 1, 1$ )

Table 12. LinSVM performance results on scene dataset (49 selected features)

	hloss	rloss	mlacc
AMI	0.1469	0.1891	0.2842
MDMR	0.1339	0.1524	0.3726
FIMF	0.1198	0.1277	0.5031
QPFS	0.1328	0.1520	0.3773
Proposed ( $\alpha, \beta, \gamma$ )	<b>0.1168</b> ( $10^3, 10^{-1}, 10^2$ )	<b>0.1112</b> ( $10^3, 10, 10^2$ )	<b>0.5453</b> ( $10^3, 10^{-2}, 10^{-3}$ )

Table 13. LinSVM performance results on yeast dataset (49 selected features)

	hloss	rloss	mlacc
AMI	<b>0.2094</b>	<b>0.2209</b>	<b>0.4446</b>
MDMR	0.2100	0.2231	0.4440
FIMF	0.2104	0.2248	0.4444
QPFS	0.2104	0.2228	0.4381
Proposed ( $\alpha, \beta, \gamma$ )	0.2126 ( $10^{-3}, 1, 10^2$ )	0.2328 ( $10^{-3}, 10^2, 10^2$ )	0.4274 ( $10^{-2}, 10^2, 10^{-2}$ )

Table 8-13은 6개의 데이터에 대해 각각의 LinSVM 분류 성능을 보여준다. 각 테이블에서 첫 번째 행은 특징 선별 기법들, 두 번째부터 네 번째 행은 해밍 로스 랭킹 로스, 다중 레이블 정확도 성능을 나타낸다. 기법들 중에서 가장 좋은 성능을 보이는 기법 결과에 볼드체로 표기하였다. 제안하는 방법의 경우 해당 결과를 나타낸 매개변수  $\alpha, \beta, \gamma$  값을 기입하였다.

실험 결과 yeast 데이터를 제외하고, 모든 데이터에서 제안하는 방법이 가장 좋은 성능을 보였다. MLkNN의 결과와 유사하게 enron, medical 데이터에서 제안하는 방식이 기존 방법에 비해 큰 차이를 내며 좋은 성능을 보였다. genbase는 MLkNN의 결과와 다르게 제안하는 방법

이 기존 방법에 비해 유의미한 성능 차이를 보였다. 하지만 yeast 데이터에서는 모든 성능 평가에서 AMI가 가장 우수한 성능을 보였다.

Table 14. Comparison Hamming loss results based on parameters change in the emotions dataset ( $\alpha$  is fixed to 0.1)

$\beta \backslash \gamma$	0.01	0.1	1	10	100
0.01	0.2459	0.2395	0.2333	0.2449	0.2366
0.1	0.2381	0.2367	0.2395	0.2424	0.2374
1	0.2383	0.2381	0.2352	0.2322	0.2319
10	0.2408	0.2377	0.2342	0.2388	0.2356
100	0.2288	0.2362	0.2319	0.2381	0.2357

Table 15. Comparison Hamming loss results based on parameters change in the emotions dataset ( $\beta$  is fixed to 0.1)

$\alpha \backslash \gamma$	0.01	0.1	1	10	100
0.01	0.2415	0.2308	0.2347	0.2356	0.2419
0.1	0.2381	0.2367	0.2395	0.2424	0.2374
1	0.2404	0.2350	0.2314	0.2275	0.2346
10	0.2322	0.2332	0.2328	0.2411	0.2352
100	0.2333	0.2374	0.2374	0.2342	0.2394

Table 16. Comparison Hamming loss results based on parameters change in the emotions dataset ( $\gamma$  is fixed to 0.1)

$\alpha \backslash \beta$	0.01	0.1	1	10	100
0.01	0.2388	0.2308	0.2331	0.2417	0.2411
0.1	0.2395	0.2367	0.2381	0.2377	0.2362
1	0.2414	0.2350	0.2352	0.2316	0.2398
10	0.2288	0.2332	0.2346	0.2424	0.2403
100	0.2410	0.2374	0.2305	0.2431	0.2376

Table 17. Comparison ranking loss results based on parameters change in the scene dataset ( $\alpha$  is fixed to 0.1)

$\beta \backslash \gamma$	0.01	0.1	1	10	100
0.01	0.2254	0.2189	0.2128	0.2195	0.2243
0.1	0.2162	0.2275	0.2260	0.2174	0.2197
1	0.2178	0.2223	0.2280	0.2178	0.2222
10	0.2207	0.2216	0.2166	0.2264	0.2157
100	0.2162	0.2178	0.2191	0.2151	0.2234

Table 18. Comparison ranking loss results based on parameters change in the scene dataset ( $\beta$  is fixed to 0.1)

$\alpha \backslash \gamma$	0.01	0.1	1	10	100
0.01	0.2182	0.2243	0.2149	0.2235	0.2195
0.1	0.2162	0.2275	0.2260	0.2174	0.2197
1	0.2226	0.2230	0.2164	0.2176	0.2203
10	0.2193	0.2148	0.2218	0.2195	0.2220
100	0.2158	0.2220	0.2173	0.2240	0.2157

Table 19. Comparison ranking loss results based on parameters change in the scene dataset ( $\gamma$  is fixed to 0.1)

$\alpha \backslash \beta$	0.01	0.1	1	10	100
0.01	0.2210	0.2243	0.2176	0.2298	0.2216
0.1	0.2189	0.2275	0.2223	0.2216	0.2178
1	0.2236	0.2230	0.2106	0.2162	0.2215
10	0.2132	0.2148	0.2199	0.2237	0.2193
100	0.2208	0.2220	0.2267	0.2208	0.2237

Table 14는 emotions 데이터에 대해 제안하는 방법의 매개변수 중  $\alpha$ 를 0.1로 고정하고  $\beta$ 와  $\gamma$ 를 변경함에 따라 해밍 로스의 변화를 보여준다. Table 15는  $\beta$ 를 0.1로 고정, Table 16은  $\gamma$ 는 0.1로 고정하였고 다른 매개변수의 변화에 따른 해밍 로스 결과를 나타낸다. 매개변수의 변화에 따른 해밍 로스의 결과가 크게 변하지 않고 안정적이며 대체로 0.2 정도의 결과를 나타내어 제안하는 방식이 매개변수에 안정적임을 확인할 수 있다. Table 17은 scene 데이터에 대해 제안하는 방법의 매개변수 중  $\alpha$ 를 0.1로 고정하고  $\beta$ 와  $\gamma$ 를 변경함에 따라 랭킹 로스의 변화를 보여준다. Table 18은  $\beta$ 를 0.1로 고정, Table 19는  $\gamma$ 는 0.1로 고정하였고 다른 매개변수의 변화에 따른 랭킹 로스 결과를 나타낸다. emotions 데이터의 결과와 유사하게 랭킹 로스가 0.2 정도에서 안정적인 결과를 보여주고 있다. scene 데이터에 대한 기존 방식들이 0.3 정도의 결과를 많이 보여주는 것에 비해 제안하는 방식은 매개변수 변화에도 불구하고 0.3을 넘기지 않고, 0.2 초반으로 제안하는 방법이 안정적인 결과를 나타냄을 보여주고 있다.

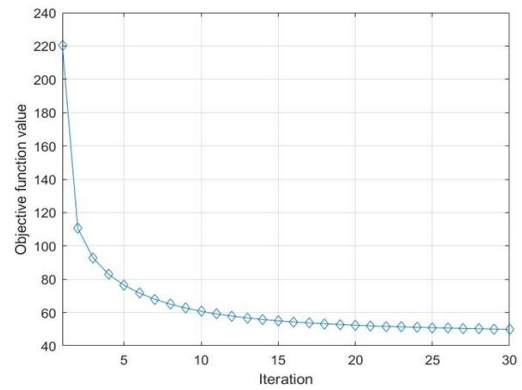


Fig. 2. Convergence rate of the proposed method in the genbase dataset

Fig. 2는 genbase 데이터에 대해 제안하는 방법의 수렴도를 보여준다. 가로축은 제안하는 알고리즘의 반복 횟수, 세로축은 제안하는 방법의 목적 함수값이다. 반복 횟수 처음부터 3번째까지 목적 함수값이 급격하게 감소하는 것을 볼 수 있고, 10번째 이전에 수렴한 것으로 볼 수 있다. 이를 통해 제안하는 방법의 알고리즘이 효율적으로 동작하는 것을 알 수 있다.

### V. Conclusions

본 연구에서는 특징과 특징, 특징과 레이블, 레이블과 레이블 사이의 상호 정보 값을 하나의 회귀분석 기반 목적 함수를 설계하여, 다중 레이블 분류 성능 향상을 위한 특징 선별 기법을 제안하였다. 두 변수 사이의 상호 정보만을 계산하여 기존 방식들보다 낮은 시간 복잡도를 가지며 경사하강법 기반의 단순한 최적화 방식을 제안하여 주어진 목적 함수의 최솟값을 빠르게 찾을 수 있었다.

제안된 방법의 우수한 성능에도 불구하고, 추가 연구가 필요한 몇 가지 문제가 있다. 첫째, 연속형 데이터에 대한 상호 정보 계산이 어렵기 때문에 직접적인 계산을 활용하기 어렵다. 둘째, 매개변수 설정에 대한 방향성이 부족하다. 데이터마다 우선적으로 고려해야 할 상호 정보가 다를 수 있지만 직접적으로 판단할 수 있는 방법을 찾기 어렵다. 향후의 연구 주제로서 이 문제들을 고려하여 좀 더 실용적인 다중 레이블 분류를 위한 특징 선별 기법을 제시할 수 있을 것이다.



## ACKNOWLEDGEMENT

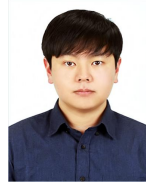
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number 2020R1A6A1A03040583).

## 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] 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
- [4] J. Lee, and D.-W. Kim, "Fast multi-label feature selection based on information-theoretic feature ranking," *Pattern Recognition*, Vol. 48, pp. 2761-2771, September 2015. DOI: 10.1016/j.patcog.2015.04.009
- [5] 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
- [6] R. B. Pereira, A. Plastino, B. Zadrozny, L.H.C. Merschmann, "Categorizing feature selection methods for multi-label classification," *Artificial Intelligence Review*, Vol. 49, pp. 57-78, September 2018. DOI: 10.1007/s10462-016-9516-4
- [7] G. Tsoumakas, and I. Vlahavas, "Random k-labelsets: an ensemble method for multilabel classification," In *Proceedings of the European Conference on Machine Learning*, pp 406-417, 2007.
- [8] J. Read, "A pruned problem transformation method for multi-label classification," In *Proceedings of the New Zealand Computer Science Research Student Conference*, pp 143-150, 2008.
- [9] M.L. Zhang, J.M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," Vol. 179, pp. 3218-3229, *Information Sciences*, September 2009. DOI: 10.1016/j.ins.2009.06.010
- [10] A. Clare and R.D. King, "Knowledge discovery in multi-label phenotype data," In *Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 42-53, 2001.
- [11] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters*, Vol. 34, pp. 349-357, February 2013. DOI: 10.1016/j.patrec.2012.10.005
- [12] Y. Fan, B. Chen, W. Huang, J. Liu, and W. Weng, "Multi-label feature selection based on label correlations and feature redundancy," Vol. 241, pp. 108256, *Knowledge-Based Systems*, April 2022. DOI: 10.1016/j.knosys.2022.108256
- [13] Y. Li, L. Hu, and W. Gao, "Multi-label feature selection via robust flexible sparse regularization," Vol. 134, pp. 109074, *Pattern Recognition*, February 2023. DOI: 10.1016/j.patcog.2022.109074
- [14] Y. Fan, J. Liu, J. Tang, P. Liu, Y. Lin, and Y. Du, "Learning correlation information for multi-label feature selection," Vol. 145, pp. 109899, *Pattern Recognition*, January 2024. DOI: 10.1016/j.patcog.2023.109899
- [15] Y. Li, L. Hu, and W. Gao, "Label correlations variation for robust multi-label feature selection," Vol. 609, pp. 1075-1097, *Information Sciences*, September 2022. DOI: 10.1016/j.ins.2022.07.154
- [16] L. Hu, L. Gao, Y. Li, P. Zhang, and W. Gao, "Feature-specific mutual information variation for multi-label feature selection," Vol. 593, pp. 449-471, *Information Sciences*, May 2022. DOI: 10.1016/j.ins.2022.02.024
- [17] J. Dai, W. Huang, C. Zhang, and J. Liu, "Multi-label feature selection by strongly relevant label gain and label mutual aid," Vol. 145, pp. 109945, *Pattern Recognition*, January 2024. DOI: 10.1016/j.patcog.2023.109945
- [18] M. Faraji, S.A. Seyedi, F.A. Tab, and R. Mahmoodi, "Multi-label feature selection with global and local label correlation," Vol. 246, pp. 123198, *Expert Systems with Applications*, July 2024. DOI: 10.1016/j.eswa.2024.123198
- [19] Z. He, Y. Lin, C. Wang, L. Guo, and W. Ding, "Multi-label feature selection based on correlation label enhancement," Vol. 647, pp. 119526, *Information Sciences*, November 2023. DOI: 10.1016/j.ins.2023.119526
- [20] P. Zhang, G. Liu, and J. Song, "MFSJMI: Multi-label feature selection considering join mutual information and interaction weight," Vol. 138, pp. 109378, *Pattern Recognition*, June 2023. DOI: 10.1016/j.patcog.2023.109378
- [21] Y. Yang, H. Chen, Y. Mi, C. Luo, S.-J. Horng, and Tianrui Li, "Multi-label feature selection based on stable label relevance and label-specific features," Vol. 648, pp. 119525, *Information Sciences*, November 2023. DOI: 10.1016/j.ins.2023.119525
- [22] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, pp. 1226-1238, June 2005. DOI: 10.1109/TPAMI.2005.159

- [23] T. Klonecki, P. Teisseyre, and J. Lee, "Cost-constrained feature selection in multilabel classification using an information-theoretic approach," *Pattern Recognition*, Vol. 141, pp. 1-18, September 2023. DOI: 10.1016/j.patcog.2023.109605
- [24] J. Lee and D.-W. Kim, "Mutual information-based multi-label feature selection using interaction information," *Expert Systems with Applications*, Vol. 42, pp. 2013-2025, March 2015. DOI: 10.1016/j.eswa.2014.09.063
- [25] 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
- [26] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multilabel classification of music into emotions," *International Conference on Music Information Retrieval*, pp. 325-330, 2008.
- [27] S. Diplaris, G. Tsoumakas, P. Mitkas and I. Vlahavas, "Protein Classification with Multiple Algorithms," *Panhellenic Conference on Informatics*, pp. 448-456, 2005.
- [28] J.P. Pestian, C. Brew, P. Matykiewicz, D.J. Hovermale, N. Johnson, K.B. Cohen, and W. Duch, "A shared task involving multi-label classification of clinical free text," *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, pp. 97-104, 2007.
- [29] M.R. Boutell, J. Luo, X. Shen, and C.M. Brown, "Learning multi-labelscene classification," *Pattern Recognition*, Vol. 37, pp. 1757-1771, September 2004. DOI: 10.1016/j.patcog.2004.03.009
- [30] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," *Advances in Neural Information Processing Systems*, 2001.
- [31] J. Lee and D.-W. Kim, "SCLS: Multi-label feature selection based on scalable criterion for large label set," *Pattern Recognition*, Vol. 66, pp. 342-352, June 2017. DOI:
- [32] J. Lee, H. Lim, and D.-W. Kim, "Approximating mutual information for multi-label feature selection," *Electronics Letters*, Vol. 48, pp. 1-2, July 2012. DOI: 10.1049/el.2012.1600
- [33] 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

## Authors



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 and related optimization methods with innovative applications such as music emotion recognition and smart factory.