

# Effective Multi-label Feature Selection based on Large Offspring Set created by Enhanced Evolutionary Search Process

Hyunki Lim\*, Wangduk Seo\*\*, Jaesung Lee\*\*

## Abstract

Recent advancement in data gathering technique improves the capability of information collecting, thus allowing the learning process between gathered data patterns and application sub-tasks. A pattern can be associated with multiple labels, demanding multi-label learning capability, resulting in significant attention to multi-label feature selection since it can improve multi-label learning accuracy. However, existing evolutionary multi-label feature selection methods suffer from ineffective search process. In this study, we propose a evolutionary search process for the task of multi-label feature selection problem. The proposed method creates large set of offspring or new feature subsets and then retains the most promising feature subset. Experimental results demonstrate that the proposed method can identify feature subsets giving good multi-label classification accuracy much faster than conventional methods.

▶ Keyword: Multi-label Learning, Multi-label Feature Selection, Evolutionary Search, Memetic Offspring Creation

## 1. Introduction

최근 데이터 수집 기술의 발달로 인하여, 많은 데이터들이 다중 레이블 형태로 수집되어 행동 인식 [1, 2], 데이터 흐름 분류 [3, 4], 사람 인지 [5] 와 같은 응용에서 활용되고 있다. 더 나아가 사람 본연의 정보를 활용하여 인공지능 인문학과의 접목 연구 또한 시도 되고 있다 [6, 7, 8]. 일반적으로 다중 레이블 데이터에서 하나의 패턴은  $F$ 라는 특징 집합으로서 표현할 수 있고, 이진 벡터 형태의 레이블  $L$ 에 할당된다. 이 때 각 패턴  $w_i \in W$  ( $1 \leq i \leq |W|$ )는 하나의 레이블 부분집합  $\lambda_i \subseteq L$ 에 할당되며 여기서  $L = l_1, \dots, l_{|L|}$ 은 레이블의 유한한 부분집합으로 나타낼 수 있다.

일반적으로 학습 정확도, 에너지 소모량, 소모 시간 등과 같은 척도를 통해 어떤 응용의 성능을 측정한다 [9]. 성능 향상을 위한 방법으로 많은 연구들이 특징 선별을 제안하고 있다 [9-11]. 그러나 대부분의 연구들이 단일 레이블을 위한 방법론

들이고, 다중 레이블을 위한 연구는 적은 편이다. 다중 레이블 특징 선별에서는, 적절한 특징을 선택하고 중복적인 특징들은 버려서 특징 부분 집합  $S \subset F$ 를 생성한다. 특징 선별을 수행하기 위한 방법론 중에 진화 알고리즘(Evolutionary Algorithm, EA) 기반의 특징 선별 기법이 촉망받고 있다 [13].

진화 알고리즘은 진화 과정 중에 특징 집합들이 후보로서 생성되고, 현재의 후보들보다 더 좋은 후보를 찾기 위해 동작한다. 따라서 더 좋은 후보를 가질 확률이 높은 후보들을 우선적으로 선정된다. 그러나 후보 특징의 적합도를 평가할 수 있는 횟수 또는 적합도 평가 횟수(Fitness Function Calls, FFC)가 제한되어 있으므로 모든 후보 특징 집합을 평가할 수 없다. 그렇기 때문에 유망하지 않은 후보 특징들이 FFC를 증가시키는 것은 매우 불필요하다 [14-17]. 본 논문에서는 메메틱(memetic) 탐색을 기반으로 하는 효과적인 다중 레이블 특징 선별 기법을 제안한다. 제안하는 방법은 많은 수의 후보 특징

• First Author: Hyunki Lim, Corresponding Author: Jaesung Lee

\*Hyunki Lim (hyunkilim@kist.re.kr), Korea Institute of Science and Technology

\*\*Wangduk Seo (seowangduk@gmail.com), School of Computer Science and Engineering, Chung-Ang University

\*\*Jaesung Lee (curseor@cau.ac.kr), School of Computer Science and Engineering, Chung-Ang University

• Received: 2018. 06. 28, Revised: 2018. 08. 02, Accepted: 2018. 08. 17.

• This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A6A3A01078538)

집합을 만들고, 정보 이론(information theory) 기반의 평가 함수를 통해 소수 개의 유망한 특징 집합 유지시킨다. 따라서 불필요한 후보 특징의 FFC를 줄인다. 실험 결과, 제안하는 방법이 유망한 특징을 잘 찾을 수 있었고, 같은 FFC 안에서 기존 방법론들보다 더 빠르게 수렴하는 것을 확인할 수 있었다.

## II. Related Work

다중 레이블 특징 선별은 학습 시간을 줄이면서 클래스 분류의 성능은 향상시킬 수 있는 장점이 있다. 특징 선별 방법은 크게 필터(filter)와 래퍼(wrapper)로 나뉜다. 필터 방법은 특징의 중요도를 평가하는 척도에 의해 특징의 순위를 매긴다. 한편, 래퍼 방법은 특징 부분 집합을 생성하고 평가하고 향상시킨다.

다중 레이블 특징 선별을 위한 가장 단순한 방법은 레이블 멱집합(Label Powerset) 방법과 같이 레이블을 단일 레이블로 변환하는 방법이다 [18]. 이 방법은 기존 단일 레이블 데이터를 위한 특징 선별 방법을 바로 이용할 수 있다는 장점이 있다. 하지만 이러한 필터 방법은 레이블 간의 연관성을 고려하지 않고, 변환하는 과정에서 클래스 간 불균형 문제로 클래스 분류 성능에 한계를 보였다 [19].

다양한 EA 기반 특징 선별 방법들은 래퍼 방법으로서 활용되었다. 그 중 유전자 알고리즘(Genetic Algorithm, GA)은 EA 기반 특징 선별 방법에서 자주 사용되는 알고리즘이다 [14]. 그러나 GA는 FFC가 낮을 때 불안정한 상태로 수렴하는 큰 단점이 있다 [20]. 최근 EA 기반 특징 선별 방법은 선택되는 특징의 개수가 두 번째 목적 변수로 여겨지고 있다. 하나 이상의 목적 변수를 가질 때 다목적 최적화 문제를 위한 설계된 비지배적 정렬(non-dominated sort)을 수행한다. 최근 GA 기반 비지배적 정렬(NSGA-II) 방법 [15]과 입자 군집 최적화(MPSOFS) [16]이 다중 레이블 특징 선별을 위해 제안되었다. 하지만 이 방법들은 불필요한 정답 후보들이 자주 생성되고 평가될 수 있어 비효율적이다.

최근 EA 기반 다중 레이블 특징 래퍼 방법은 유망한 특징 후보와 유망하지 않은 후보를 모두 도출해내는 경향이 있다. 왜냐하면 이 방법들이 새로운 특징 후보들을 생성할 때 특징의 중요성을 무시하기 때문이다. 이를 개선하기 위해 메메틱 특징 선별 기법을 제안했다 [20]. 메메틱 탐색은 특징들의 중요도를 평가하는 국소 탐색(local search)을 포함하고 있다. 하지만 이 방법은 국소탐색 과정에서 걸러지지 않은 무작위적인 후보 특징들이 있을 수 있어 효과적이지 못하다.

## III. Proposed Method

EA 기반 탐색 방법 중, EDA는 다양한 문제를 해결하기에 효과적이라 증명되었다 [21, 22]. 새로운 특징 집합을 생성할 때 일반적인 EA 기반 탐색 방법과 달리 EDA는 유전자 연산(genetic operator)를 사용하지 않는다. 대신에 EDA는 확률 모델을 이용하여 새로운 후보들을 만들고, 후보들로부터 추정하여 확률 모델을 갱신한다. 다양한 EDA 방법들 중에서, population-based incremental learning(PBIL)이 많은 응용에서 좋은 성능을 나타내었다[24]. PBIL은 단순함과 높은 효과 때문에 다양한 문제에서 많이 활용되었다 [19]. PBIL은 확률 모델에 다음과 같이 결정한다.

$$p^{t+1}(i) = p^t \times (1 - LR) + F^t(i) \times LR \quad (1)$$

여기에서  $p^t(i)$ 는  $t$ 번째 세대에  $i$ 번째 특징의 선택 확률이므로,  $F^t(i)$ 는 평가 값을 기준으로 상위 50% 특징의  $t$ 번째 세대에 유망한  $i$ 번째 특징과 연관된 확률이고,  $LR$ 은 학습 계수(learning rate)이다. (1)에 따라  $i + 1$ 번째 세대에서의 특징이 선택될 확률  $P^{t+1}$ 이 계산되고,  $t + 1$ 번째 세대의 특징 집합이 결정된다.

특징 선별 문제에서 PBIL은 거대한 공간을 탐색해야한다. 따라서 유망한 특징 집합을 찾기 위해서는 막대한 계산량이 필요하다. 그러나 PBIL과 같은 단순한 모델은 구현하기 쉽지만 유망한 특징 집합 찾기와 같은 복잡한 문제를 풀기에 적당하지 않다 [25]. 이를 위해 유망한 후보 생성을 위한 과정을 개선한다. PBIL은 막대한 양의 특징 집합을 생성하고 유망한 특징 집합을 결정하는 필터 방법을 이용하여 특징들을 선택한다.

좋은 특징 집합이 필터 방법에 의해 생성되더라도 유망하지 않은 특징 또한 생성될 수 있다. 왜냐하면 생성 과정이 확률에 따라 결정되기 때문이다. 유망하지 않은 특징 집합은 FFC를 소모하고 탐색 효율에 부정적인 영향을 준다. 이를 극복하기 위해 FFC를 소모하지 않는 특징 집합 평가 방법을 제안한다. 제안하는 방법은 각 특징 집합에 대해 정보 이론 기반으로 적합도(relevance)와 중복도(redundancy)를 계산한다. 따라서 제안하는 방법은 최대 적합도와 최소 중복도를 가지는 특징 집합을 선택하게 된다. 제안하는 특징 집합이 국소적으로 유일하게 유망한 확률을 가지게 되므로, 제안하는 방법은 선택 알고리즘으로 룰렛 선택 알고리즘을 사용한다 [26]. 따라서 유망하지 않은 특징 부분 집합은 FFC 사용 없이 걸러지게 된다.

FFC 소모없이 유망한 특징 집합을 찾기 위해, 다중 레이블 특징 필터 과정에서 교배를 이용한다. 본 논문에서는 FFC 소모 없이 빠르게 좋은 특징 집합으로 수렴할 수 있도록 PBIL 기반 메메틱 탐색 방법을 제안한다.

## 알고리즘 1: 제안하는 방법

```

1: Input: population size  $m$ , max FFC  $v$ 
2: Output: best feature subset  $S_g$ 
3: initializing  $O(t)$  and probabilistic model  $p$ 
4: evaluating  $O(t)$ 
5:  $u \leftarrow 0$ 
6: store the global best feature subset to  $S_g$ 
7: while  $u \leq v$  do
8:   update  $P$  by Equation (1)
9:    $E(t) \leftarrow$  create explosively reproduced set
10:   $O(t+1) \leftarrow$  select feature subsets in  $E(t)$ 
11:  evaluate  $O(t+1)$ 
12:   $u \leftarrow u + m$ 
13:  update global best feature subset  $S_g$ 
14: end while
    
```

알고리즘 1은 제안하는 방법을 보여준다. 전체 인구 (population) 크기  $m$ , 최대 FFC  $v$ 를 입력 값으로 하여 특징 집합  $O(t)$ 와 확률 모델  $P$ 를 초기화 한다(line 3). 제안하는 방법은 최대  $|F|$  크기의 무작위 이진 비트를 통해  $m$ 개의 특징 집합을 생성한다. 확률 모델  $P$ 는  $|F|$  크기의 벡터이고 벡터의 각 요소는 대응하는 특징의 선택 확률을 나타낸다. 각 요소는  $O(t)$ 의 특징 분포를 위해 초기화되고 생성된  $O(t)$ 는 평가된다. 그리고 FFC 소모량  $u$ 는 0으로 초기화되고 가장 좋은 특징 집합을  $S_g$ 에 저장한다(line 6). 확률모델  $P$ 는 식 (1)에 의해 갱신되고 특징 집합  $E$ 가 거대하게 재생성된다. 이 때  $E$ 는 생성 함수의 필터 방식으로 생성된다(line 9). 그리고  $m$ 개의 특징 집합이 생성 함수의 가중치 기반 룰렛 방식으로 선택이 되고 새로운 세대  $O(t+1)$ 이 도출된다(line 10). 특징 집합  $O(t+1)$ 은 평가되고(line 11)  $u$ 는 증가한다(line 12). 가장 좋은 특징 집합  $S_g$ 은 갱신된다(line 13). 지정된 최대 FFC  $v$ 가 다 소모되면 알고리즘은  $S_g$ 를 출력한다.

## 알고리즘 2: 생성 함수

```

1: Input: explosive reproduced population size  $e$ , probabilistic model  $P$ 
2: Output: explosive reproduced set  $E(t)$ 
3:  $E(t) \leftarrow \emptyset$ 
4: for  $k=1$  to  $e$  do
5:    $n \leftarrow$  random integer value in  $[0, |F|]$ 
6:    $S_k \leftarrow \emptyset$ 
7:   for  $i=1$  to  $n$  do
8:      $R \leftarrow \{f_1, f_2, f_3, \dots\}$ 
9:     for  $j=1$  to  $|F|$  do
10:      if  $P^i(f_j) >$  random value in  $[0, 1]$  then
11:         $S_k \leftarrow S_k \cup j$ 
12:      break
13:    end if
14:  end for
15: end for
16:  $E(t) \leftarrow E(t) \cup S_k$ 
17: end for
    
```

알고리즘 2는 거대한 특징 집합  $E$ 를 생성하는 생성 함수의 과정을 보여준다. 각 특징 집합은 0에서  $|F|$  사이의 무작위 정수만큼의 특징을 선택한다(line 5). 유망한 특징 집합을 생성하기 위해 각 특징은 필터 방식인 식 (5)에 의해서 순위가 결정된다. 식 (5)는 선택되지 않은 특징과 현재의 특징 집합  $S_k$ 의 중복도, 그리고 선택되지 않은 특징과 레이블 사이의 적합성을 고려한다. 그러면 이 함수는  $P$ 에 의해 결정된 가장 중요한 특징 집합으로부터 하나의 특징을 선택 여부를 결정한다(line 9-13). 만약에 한 특징이 선택되면, 이 특징은 특징 집합  $S_k$ 에 포함된다(line 11). 이 특징 집합이 생성되면 이후  $E$ 에 추가된다.

## 알고리즘 3: 선택 함수

```

1: Input: explosively reproduced set  $E$ , population size  $m$ 
2: Output: filtered set  $G(t)$ 
3: rank feature subsets in set  $E$  by Equation (4)
4: select  $m$  feature subsets by roulette wheel selection
5:  $G(t) \leftarrow$  selected feature subsets
    
```

알고리즘 3은 제안하는 특징 집합 평가 식 (4)를 이용하여 거대하게 재생성된 집합에서의 유망한 특징 집합을 선택하는 함수이다. 이 함수는  $E$ 안에서 특징 집합의 순위를 결정하고 룰렛 선택 방법으로  $m$ 개의 특징 집합  $G(t)$ 를 선택한다. 이 때 룰렛 선택 방법은 식 (4)에 의해 가중치가 결정된다 (line 4).

FFC 소모없이 특징 집합을 평가하기 위해, 우리는 전체 특징 집합  $F$ 에서 선택된 각 특징의 중요성을 적합성 척도[24]를 이용하여 평가한다.

$$I(f_i) = Rel(f_i) - Red(f_i) \quad (2)$$

이 식에서  $Rel(f_i)$ 와  $Red(f_i)$ 는  $i$ 번째 특징의 적합도와 중복도를 의미하고  $I(f_i)$ 는  $i$ 번째 특징의 중요성을 의미한다. 연구 [25]에서, 특징 집합  $S_{i-1}$ 에서  $i$ 번째 특징의 중요성을 적합도와 중복도를 이용하여 다음과 같이 정의했다.

$$I(f_i) = \sum_{l \in L} M(f_i; l) - \sum_{f \in S_{i-1}} M(f_i; f) \quad (3)$$

이 식에서  $M(x, y) = H(x) - H(x, y) + H(y)$ 은 변수  $x$ 와  $y$  사이의 상호정보척도(mutual information)을 나타내고,  $H(x) = -\sum P(x) \log P(x)$ 는 확률밀도함수  $P(x)$ 로부터 계산된 엔트로피(joint entropy)이다 [9]. 식 (3)을 통해서 특징 집합  $S$ 의 평가 함수는 다음과 같이 정의된다.

$$E(S) = \sum_{f_i \in S} \sum_{l \in L} M(f_i; l) - \sum_{f_i \in S} \sum_{f_j \in S} M(f_i; f_j) \quad (4)$$

우리는 특징의 중요성을 평가하기 위한 최신의 필터 방법을 이용한다. 필터 방법은 평가 함수를 통해 특징의 순위를 결정한다. 후보 집합 재생성 단계에서 각 특징 집합에 대해 우리는 필터 방법을 통해 중요성을 정하고 특징 집합  $S$ 를 고려하는 확률  $P'(i)$ 에 기반하여  $i$ 번째 특징을 선택한다. 만약에  $i$ 번째 특징이 선택되지 않는다면 특징이 선택될 때까지 두 번째 가장 중요한 특징을 선택한다. 연구 [12]에서 큰 레이블 데이터를 위해 확장 가능한 필터 방법(scalable criterion for large label sets, SCLS)을 제안하였고 기존 필터 방법을 압도하는 결과를 보여주었다. 우리는 이 방법을 사용하여 특징의 중요성을 측정한다. SCLS는  $i$ 번째 특징의 중요성을 다음과 같이 계산한다.

$$I(f_i) = \sum_{l \in L} M(f_i; l) - \sum_{f \in S} \sum_{l \in L} \frac{M(f; l)}{H(f_i)} M(f; f) \quad (5)$$

#### IV. Experimental Results

여러 응용분야에서 14개의 데이터를 실험에 활용하였다. Bird 데이터는 새들의 소리를 담고 있는 오디오 데이터이다. Enron과 Language Log(LLOG) 데이터는 텍스트 마이닝 분야에서 생성되었고, 단어의 유무와 특정 주제에 문서와 연관되어 있는 레이블로 이루어져 있다. Mediamill은 자동 감지 시스템으로부터 얻어진 영상 데이터이다. Medical 데이터는 임상 관련 텍스트로부터 자연어처리 과정을 거쳐서 얻어졌다. TMC2007는 복잡 공간 시스템(complex space system)의 안전 보고서 내용을 담고 있다. 나머지 8개의 데이터는 Yahoo 데이터로부터 얻어졌다. 우리는 비지도 차원 감소(unsupervised dimensional reduction) 방법을 이용해, 10,000개 이상의 특징을 TMC2007과 Yahoo데이터의 차원을 줄였다. 제안하는 방법론이 정보이론을 사용하기 때문에, 숫자형(numerical) 특징인 데이터를 [29]의 방법을 통해 이산화(discretization)하였다. Table 1은 실험에 사용된 데이터의 명세를 보여준다.

Table 1. Characteristics of employed data sets

	$ W $	$ F $	$ L $
Birds	645	260	19
Enron	1,702	1,001	53
LLOG	1,460	1,004	75
Mediamill	43,907	120	101
Medical	978	1,449	45
TMC2007	28,596	49,060	22
Business	11,214	1,096	30
Education	12,030	1,377	33
Entertain	12,730	1,600	21
Health	9,205	1,530	32
Reference	8,027	1,984	33
Science	6,428	1,859	40
Social	12,111	2,618	39
Society	14,512	1,590	27

우리는 제안하는 방법론을 GA [14], NSGA-II [15], MPSOFS [16]와 비교하였다. 특징 선별의 결과로 클래스 분류 성능을 비교하기 위해 다중 레이블 나이브 베이브(multi-label naive Bayes, MLNB) 분류기를 사용하였다 [14]. 각 데이터에 대해서 교차검증을 통해 결과를 얻었다. 데이터 전체의 80%를 훈련 데이터, 20%를 검증 데이터로 사용하였다. 전체 인구는 20, 최대 FFC는 100으로 설정하였다. 제안하는 방법은 확률 모델을 통해 최대 500개의 특징을 생성했고, 학습 계수는 0.4로 두었다. GA와 NSGA-II는 한 세대에서 두 개의 자손(offspring)과 돌연변이(mutation) 과정을 통한 한 개의 특징 집합을 만들었다. MPSOFS는 가장 좋은 특징 집합을 유지하고 각 입자(particle)로 개별적으로 가장 좋은 집합을 유지한다. 그 후에 MPSOFS는 속도(velocity)를 갱신한다. 모든 실험은 10번을 반복하였고, 평균 값을 통해 비교하였다.

특징 선별 방법의 성능을 비교하기 위해, 4개의 클래스 분류 평가 방법, 해밍 로스(hamming loss), 랭킹 로스(ranking loss), 다중 레이블 정확도(multi-label accuracy), 정규화 커버리지(normalized coverage)를 사용하였다. 다중 레이블 정확도는 다음과 같이 정의된다.

$$mlacc(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{|\lambda_i \cap Y_i|}{|\lambda_i \cup Y_i|}$$

여기서  $T$ 는 주어진 검증 데이터이고  $\lambda$ 는 정답 레이블 집합을 의미한다. 해밍 로스는 다음과 같이 정의된다.

$$hloss(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{1}{|L|} |\lambda_i \Delta Y_i|$$

여기서  $\Delta$ 는 두 집합 사이의 대칭 차이(symmetric difference)를 의미한다. 랭킹 로스는 다음과 같다.

$$rloss(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \frac{|\{(a,b) | a \in \lambda_i, b \in \bar{\lambda}_i, \Psi_{i,a} \leq \Psi_{i,b}\}|}{|\lambda_i| |\bar{\lambda}_i|}$$

여기서  $\bar{\lambda}_i$ 은  $\lambda_i$ 의 보수(complement) 집합을 의미한다. 랭킹 로스는 연관성 있는 레이블과 연관성 없는 레이블 쌍 사이의  $\Psi_{i,a} \leq \Psi_{i,b}$ 이 가능한  $(a,b)$  쌍이 평균 값을 측정한다. 정규화 커버리지는 다음과 같이 정의된다.

$$ncov(T) = \frac{1}{|L|} \left( \frac{1}{|T|} \sum_{i=1}^{|T|} \max_{l \in \lambda_i} rank(l) - 1 \right)$$

여기서  $rank(\cdot)$ 는  $\Psi_{i,l}$ 에 따라 대응되는 연관된  $l \in \lambda_i$ 의 순위를 출력한다. 다중 레이블 정확도는 높을수록, 해밍 로스, 랭킹 로스, 정규화 커버리지는 낮을수록 좋은 클래스 분류 성능을 의미한다.

Table 2-5는 4개의 클래스 분류 성능을 나타내고 있다. 가장 좋은 성능을 보인 방법은 진하게 표시하였다. 각 Table에 마지막 열은 14개 데이터에 대해 평균 순위를 표기하였다. 다중 레이블 정확도와 랭킹 로스 성능에서, 제안하는 방법이 모든 데이터에 대해서 GA, NSGA-II, MPSOFS를 압도하였다. 해밍 로스 성능에서는 TMC2007 데이터를 제외하고 제안하는 방법이 가장 좋은 성능을 보였다. 정규화 커버리지 성능에서는 LLOG 데이터를 제외하고 제안하는 방법이 가장 좋은 성능을 보였다.

Table 2. Comparison results in terms of multi-label accuracy

Method	Birds	Enron	LLOG	Mediamill
Proposed	<b>0.527</b>	<b>0.383</b>	<b>0.249</b>	<b>0.362</b>
GA	0.491	0.284	0.209	0.336
NSGA-II	0.480	0.282	0.208	0.347
MPSOFS	0.453	0.206	0.042	0.163
Method	Medical	TMC2007	Business	Education
Proposed	<b>0.427</b>	<b>0.441</b>	<b>0.672</b>	<b>0.318</b>
GA	0.303	0.435	0.657	0.316
NSGA-II	0.297	0.434	0.662	0.318
MPSOFS	0.286	0.420	0.634	0.283
Method	Entertain	Health	Reference	Science
Proposed	<b>0.396</b>	<b>0.537</b>	<b>0.436</b>	<b>0.288</b>
GA	0.361	0.499	0.422	0.231
NSGA-II	0.362	0.495	0.429	0.237
MPSOFS	0.365	0.496	0.414	0.234
Method	Social	Society	Avg.Rank	
Proposed	<b>0.546</b>	<b>0.371</b>	<b>1.00</b>	
GA	0.517	0.258	2.79	
NSGA-II	0.526	0.267	2.64	
MPSOFS	0.527	0.239	3.57	

Table 3. Comparison results in terms of Hamming loss

Method	Birds	Enron	LLOG	Mediamill
Proposed	<b>0.061</b>	<b>0.060</b>	<b>0.016</b>	<b>0.034</b>
GA	0.072	0.100	0.075	0.048
NSGA-II	0.064	0.104	0.072	0.054
MPSOFS	0.135	0.198	0.292	0.174
Method	Medical	TMC2007	Business	Education
Proposed	<b>0.020</b>	0.088	<b>0.029</b>	<b>0.042</b>
GA	0.023	0.088	0.035	0.046
NSGA-II	0.022	<b>0.086</b>	0.037	0.048
MPSOFS	0.023	0.117	0.079	0.061
Method	Entertain	Health	Reference	Science
Proposed	<b>0.055</b>	<b>0.039</b>	<b>0.034</b>	<b>0.035</b>
GA	0.070	0.051	0.055	0.051
NSGA-II	0.068	0.054	0.047	0.045
MPSOFS	0.105	0.067	0.0886	0.110
Method	Social	Society	Avg.Rank	
Proposed	<b>0.025</b>	<b>0.054</b>	<b>1.07</b>	
GA	0.042	0.062	2.71	
NSGA-II	0.040	0.060	2.29	
MPSOFS	0.070	0.144	3.93	

Table 4. Comparison results in terms of ranking loss

Method	Birds	Enron	LLOG	Mediamill
Proposed	<b>0.115</b>	<b>0.100</b>	<b>0.155</b>	<b>0.060</b>
GA	0.129	0.133	0.164	0.066
NSGA-II	0.125	0.149	0.163	0.067
MPSOFS	0.132	0.194	0.180	0.159
Method	Medical	TMC2007	Business	Education
Proposed	<b>0.115</b>	<b>0.073</b>	<b>0.062</b>	<b>0.089</b>
GA	0.145	0.075	0.070	0.100
NSGA-II	0.139	0.076	0.067	0.100
MPSOFS	0.140	0.078	0.096	0.101
Method	Entertain	Health	Reference	Science
Proposed	<b>0.111</b>	<b>0.085</b>	<b>0.111</b>	<b>0.118</b>
GA	0.140	0.098	0.130	0.154
NSGA-II	0.137	0.097	0.128	0.150
MPSOFS	0.153	0.098	0.140	0.157
Method	Social	Society	Avg.Rank	
Proposed	<b>0.075</b>	<b>0.135</b>	<b>1.00</b>	
GA	0.088	0.151	2.86	
NSGA-II	0.085	0.153	2.29	
MPSOFS	0.097	0.121	3.86	

Table 5. Comparison results in terms of normalized coverage

Method	Birds	Enron	LLOG	Mediamill
Proposed	<b>0.194</b>	<b>0.277</b>	0.202	<b>0.196</b>
GA	0.223	0.337	0.201	0.205
NSGA-II	0.212	0.336	<b>0.199</b>	0.208
MPSOFS	0.222	0.413	0.201	0.330
Method	Medical	TMC2007	Business	Education
Proposed	<b>0.155</b>	<b>0.203</b>	<b>0.132</b>	<b>0.148</b>
GA	0.180	0.208	0.139	0.159
NSGA-II	0.178	0.209	0.141	0.159
MPSOFS	0.179	0.211	0.168	0.159
Method	Entertain	Health	Reference	Science
Proposed	<b>0.197</b>	<b>0.157</b>	<b>0.157</b>	<b>0.182</b>
GA	0.221	0.167	0.177	0.215
NSGA-II	0.219	0.167	0.171	0.213
MPSOFS	0.233	0.166	0.179	0.215
Method	Social	Society	Avg.Rank	
Proposed	<b>0.127</b>	<b>0.242</b>	<b>1.21</b>	
GA	0.140	0.257	2.86	
NSGA-II	0.136	0.261	2.43	
MPSOFS	0.145	0.317	3.50	

모든 데이터에 대해서 각 특징 선별 기법들의 성능을 측정 한 후 우리는 통계적 검증을 하였다. 우리는 통계적 검증에서 다양 하게 쓰이는 프리드먼 검정(Friedman test)을 여러 개의 방법 들을 비교하기 위해서 사용하였다 [30].  $k$ 개의 방법과  $N$ 개의 데이터가 있을 때,  $R_j$ 를 귀무 가설(null hypothesis)에 대한  $j$  번째 방법의 평균 순위라고 하자. 그러면 다음 프리드먼 통계  $F_F$ 는 다음과 같은  $F$ -분포를 가지게 된다.

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2}$$

여기서  $\chi_F^2$ 은 다음과 같이 정의된다.

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[ \sum_{j=1}^k R_j^2 - \frac{k(k+1)^2}{4} \right]$$

Table 6. Summary of the Friedman statistics  $F_f$  ( $k=4$ ,  $N=14$ ) and critical value in terms of each evaluation measure

Evaluation measure	$F_F$	Critical value ( $\alpha=0.05$ )
Multi-label accuracy	30.333	2.845
Hamming loss	65.624	
Ranking loss	75.472	
Normalized coverage	16.355	

만약  $F_F$ 가 유의 수준(significance level)  $\alpha$ 에서의 임계 값(critical value)보다 크다면, 귀무 가설은 기각되고 비교하는 방법들은 서로 다른 성능을 가지는 것을 의미한다. 귀무 가설이 기각되면 제안하는 방법이 다른 방법보다 통계적으로 유의미하게 좋은 성능을 보였다고 분석할 수 있다. 프리드먼 검정을 위해 유의 수준  $\alpha$ 를 0.05로 두었다. Table 6은 프리드먼 검정을 정리한 내용을 담고 있다. 프리드먼 검정 통계  $F_F$ 는 모든 클래스 분류 성능 측정에서 임계 값보다 컸다. 그러므로 귀무 가설은 모두 기각되었다.

## V. Conclusions

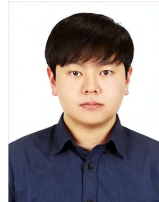
다중 레이블 데이터를 다루기 위해 우리는 거대한 재생성 기반 메테릭 탐색 방법을 제안하였다. 우리는 특징 집합을 거대하게 생성하고 과도한 컴퓨터 연산량의 소모 없이 유망한 특징 집합을 선택했다. 실험 결과 제안하는 방법이 기존의 방법들보다 빠르게 수렴하는 것을 볼 수 있었다. 향후 연구로 우리는 다양한 응용 분야에 이 방법을 활용하고 좀 더 다양한 특징 선별 기법과 성능 비교를 진행할 것이다.

## REFERENCES

- [1] Kumar, R., Qamar, I., Viridi, J.S., Krishnan, N.C., "Multi-label Learning for Activity Recognition," Intelligent Environments (IE), 2015 International Conference on. IEEE, pp. 152-155, 2015.
- [2] Alhamoud, A., Muradi, V., Böhnstedt, D., Steinmetz, R., "Activity recognition in multi-user environments using techniques of multi-label classification," Proceedings of the 6th International Conference on the Internet of Things. ACM, pp. 15-23, 2016.
- [3] Sousa, R., Gama, J., "Multi-label classification from high-speed data streams with adaptive model rules and random rules," Progress in Artificial Intelligence, pp. 1-11, 2018.
- [4] Read, J., Bifet, A., Holmes, G., Pfahringer, B., "Scalable and efficient multi-label classification for evolving data streams," Machine Learning, Vol. 88, pp. 243-272.
- [5] Read, J., Martino, L., Olmos, P.M., Luengo, D., "Scalable multi-output label prediction: From classifier chains to classifier trellises," Pattern Recognition, Vol. 48, pp. 2096-2109, 2012.
- [6] Wang, Tinghuai, Jean-Yves Guillemaut, and John Collomosse, "Multi-label propagation for coherent video segmentation and artistic stylization," Image Processing (ICIP), 2010 17th IEEE International Conference on, pp. 3005-3008, 2010.
- [7] Persing, Isaac, and Vincent Ng, "Modeling thesis clarity in student essays," Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol. 1, pp. 260-269, 2013.
- [8] Park, Jong-Beom, "Development of the Poker Game Achievement Engine for Artificial Intelligence," Journal of the Korea Society of Computer and Information, Vol. 14, No. 11, pp. 41-52, 2009.
- [9] Ghasemzadeh, H., Amini, N., Saeedi, R., Sarrafzadeh, M., "Power-aware computing in wearable sensor networks: An optimal feature selection," IEEE Trans. Mob. Comput., Vol. 14, pp. 800-812, 2015.
- [10] Yan, K., Zhang, D., "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," Sensors and Actuators B: Chemical, Vol. 212, pp. 353-363, 2015.
- [11] Li, D., Zhou, Y., Hu, G., Spanos, C.J., "Optimal Sensor Configuration and Feature Selection for AHU Fault Detection and Diagnosis," IEEE Transactions on Industrial Informatics, Vol. 13, pp. 1369-1380, 2017.
- [12] Lee, J., Kim, D.W., "SCLS: Multi-label feature selection based on scalable criterion for large label set," Pattern Recognit. Vol. 66, pp. 342-352, 2017.
- [13] Lee, J., Seo, W., Kim, D.W., "Effective Evolutionary Multilabel Feature Selection under a Budget Constraint," Complexity, pp. 1-14, 2018.
- [14] Zhang, M.L., Peña, J.M., Robles, V., "Feature selection for multi-label naïve Bayes classification," Inf. Sci., Vol. 179, pp. 3218-3229, 2009.
- [15] Yin, J., Tao, T., Xu, J., "A multi-label feature selection

- algorithm based on multi-objective optimization,” *Int. Joint Conf. Neural Networks. IEEE*, pp. 1–7, 2015.
- [16] Zhang, Y., Gong, D.W., Sun, X.Y., Guo, Y.N., “A PSO-based multi-objective multi-label feature selection method in classification.” *Sci. Rep. Vol. 7*, pp. 1–12, 2017.
- [17] Gupta, P., Dallas, T., “Feature selection and activity recognition system using a single triaxial accelerometer,” *IEEE Transactions on Biomedical Engineering*, Vol. 61, pp. 1780–1786, 2014.
- [18] Read, J., “A pruned problem transformation method for multi-label classification,” *Proc. New Zealand Computer Science Research Student Conf.*, pp. 143–150, 2008.
- [19] Spolaôr, N., Cherman, E.A., Monard, M.C., Lee, H.D., “A Comparison of Multi-label Feature Selection Methods using the Problem Transformation Approach,” *Electron. Notes Theor. Comput. Sci.*, pp. 135–151, 2013.
- [20] Lee, J., Kim, D.W., “Memetic feature selection algorithm for multi-label classification,” *Inf. Sci.*, pp. 80–96, 2015.
- [21] Xue, B., Zhang, M., Browne, W.N., Yao, X., “A survey on evolutionary computation approaches to feature selection,” *IEEE Trans. Evol. Comput.*, Vol. 20, pp. 606–626, 2016.
- [22] Perez, M., Rubin, D.M., Marwala, T., Scott, L.E., Stevens, W., “A population-based incremental learning approach to microarray gene expression feature selection,” In *Proc. IEEE 26th Conv. Electr. Electron. Eng. Israel.*, pp. 000010–000014, 2010.
- [23] Zhou, A., Sun, J., Zhang, Q., “An estimation of distribution algorithm with cheap and expensive local search methods,” *IEEE Trans. Evol. Comput.*, Vol. 19, pp. 807–822, 2015.
- [24] Baluja, S., “Population-based incremental learning. a method for integrating genetic search based function optimization and competitive learning,” *Technical report, Carnegie-Mellon Univ Pittsburgh Pa Dept Of Computer Science*, 1994.
- [25] Pelikan, M., Goldberg, D.E., Lobo, F.G., “A survey of optimization by building and using probabilistic models,” *Comput. Optim. Appl.*, Vol. 21, pp. 5–20, 2002.
- [26] Lipowski, A., Lipowska, D., “Roulette-wheel selection via stochastic acceptance,” *Physica A: Statistical Mechanics and its Applications*, Vol. 391, pp. 2193–2196, 2012.
- [27] Lee, J., Kim, D.W., “Feature selection for multi-label classification using multivariate mutual information,” *Pattern Recognit. Lett.*, Vol. 34, pp. 349–357, 2013.
- [28] Lee, J., Lim, H., Kim, D.W., “Approximating mutual information for multi-label feature selection,” *Electron. Lett. Vol. 48, No. 15*, pp. 929–930, 2012.
- [29] Cano, A., Luna, J.M., Gibaja, E.L., Ventura, S., “LAIM discretization for multi-label data,” *Inf. Sci.*, pp. 370–384, 2016.
- [30] Demsar, J., “Statistical comparisons of classifier over multiple data sets,” *J. Mach. Learn. Res.*, Vol. 7, pp. 1–30, 2006.

### Authors



Hyunki Lim is a postdoctoral researcher in the Image and Media Research Center, KIST (Korea Institute of Science and Technology) in Seoul, Korea. He received Ph.D., M.S. and B.S. at Chung-Ang University. His interests include feature selection and

time series analysis in machine learning.



Wangduk Seo is currently a Master student at Chung-Ang Univ. in Seoul, Korea, in the school of computer science and engineering, which he joined in 2017.



Jaesung Lee received the B.S., M.S., and Ph.D. degrees from Chung-Ang University, Seoul, South Korea, in 2007, 2009, and 2013, all in computer science. He is currently an assistant professor. His research interest includes the data mining

with applications to affective computing and ambient intelligence. In theoretical domain, he also studies classification, feature selection, and especially multilabel learning with information theory.