Memetic feature selection algorithm for multi-label classification

Jaesung Lee, Dae-Won Kim *

School of Computer Science and Engineering, Chung-Ang University, 221, Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Republic of Korea

ARTICLE INFO

ABSTRACT

The use of multi-label classification, i.e., assigning unseen patterns to multiple categories, has emerged in modern applications. A genetic-algorithm based multi-label feature selection method has been considered useful because it successfully improves the accuracy of multi-label classification. However, genetic algorithms are limited to identify fine-tuned feature subsets that are close to the global optimum, which results in a long runtime. In this paper, we present a memetic feature selection algorithm for multi-label classification that prevents premature convergence and improves the efficiency. The proposed method employs memetic procedures to refine the feature subsets found through a genetic search, resulting in an improvement in multi-label classification. Empirical studies using various tests show that the proposed method outperforms conventional multi-label feature selection methods.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Multi-label classification is a challenging problem that has emerged in several modern areas of application, such as text categorization [14], gene function classification [11], and the semantic annotation of images [1]. Let $W \subset \mathbb{R}^d$ denote an input space constructed from a set of features F, where $|F| = d$ and patterns drawn from W are assigned to a certain label subset $\lambda \subseteq Y$, where $Y = \{y_1, \ldots, y_\psi\}$ is a finite set of labels with $|Y| = \psi$. Thus, multi-label classification is the task of assigning unseen patterns to multiple labels. However, this is a difficult task because its efficacy can be varied according to the number of labels, features, patterns, and evaluation measures used to assess the quality of the predicted labels from different aspects [4,5,18,23,28,29,36,37].

Based on exhaustive experiments, researchers have reported that feature selection can improve the performance of multi-label classification [2,12,13,15,16,20,26,33–35]. Researchers have considered various approaches to performing feature selection for multi-label learning. Among them, the genetic-algorithm (GA) based multi-label feature selection method has shown strength in terms of classification performance [34] because it evaluates the fitness of the feature subsets using a multi-label classifier directly. However, owing to its inherent characteristics, a GA consumes enormous time to find a feature subset and sometimes may not find the optimum subset with sufficient precision; thus, it often converges to prematured solutions.

Recent studies on memetic algorithms (MAs) have demonstrated that they converge to high-quality solutions more efficiently than GAs for complex problems. Zhu et al. and Oh et al. presented a memetic algorithm-based feature selection that

---

* Corresponding author. Tel.: +82 2 820 5304; fax: +82 2 820 5301.
E-mail address: dwkim@cau.ac.kr (D.-W. Kim).