Information Sciences 293 (2015) 80–96

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

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

Article history:
Received 17 March 2014
Received in revised form 15 August 2014
Accepted 12 September 2014
Available online 20 September 2014

Keywords:
Multi-label feature selection
Memetic algorithm
Local refinement

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.

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).

http://dx.doi.org/10.1016/j.ins.2014.09.020
0020-0255/

converges to high-quality solutions faster than a GA for a single-label feature selection problem [21,38,39]. However, the extension of memetic single-label feature selection methods into a multi-label feature selection problem is non-trivial owing to a lack of consideration regarding the choice of meme for undergoing local refinement in multi-label classification. Therefore, in this paper, we propose a new memetic algorithm that unifies the specific issues related to the design of local refinement on multi-label feature selection problems. The proposed hybridization improves the multi-label classification accuracy and accelerates the search speed, refining the population of feature subsets generated by a GA by adding relevant features to, or removing redundant/irrelevant features from, multiple labels. To the best of our knowledge, this is the first study investigating multi-label feature selection on the basis of a memetic algorithm.

2. Related works

In the multi-label feature selection problem, a subset S composed of n selected features from F ($n \ll d$) that jointly have the largest dependency on labels Y is chosen. Popular conventional algorithms first transform label sets into a single label (a process called problem transformation), and then solve the resultant single-label feature selection problem.

Yang and Pederson [33] compared five multi-label feature selection methods based on five score functions for text categorization. However, the relations among the given labels are not captured because each label is treated independently. Chen et al. [2] proposed the use of an entropy-based label assignment (ELA), which assigns weights to a multi-label pattern based on the label entropy. Patterns that have too many labels are blurred out during the training phase because they are assigned low weights. Thus, it has been argued that a learning algorithm can avoid over-fitting from the learning process. A label powerset (LP) is applied to music information retrieval specifically for recognizing six emotions that are simultaneously evoked by a music clip [29]. The χ^2 statistics are used to select effective features with an LP to improve the recognition performance of multi-labeled music emotions. Although they used χ^2 statistics, a recent study reported that the use of ReliefF as a score function for assessing each feature in an LP yields successful classification results [26]. Read [23] proposed the use of a pruned problem transformation (PPT) to improve an LP. In a PPT, patterns with labels that occur very infrequently are merely removed from the training set by considering label sets with a predefined minimum occurrence. Doquire and Veleysen [8,9] proposed a PPT-based multi-label feature selection method to improve the classification performance of image annotation and gene function classification.

Although the problem transformation-based feature selection approach reduces the effort expended in designing a specific score function for multi-label problems, this process may cause subsequent problems. For example, problem transformation methods convert multiple labels into a single-label composed of multiple classes. If there are too many distinct label sets in the original labels, the resulting single-label will be composed of too many classes. Consequently, the performance of a learning algorithm may be degraded owing to a lack of training patterns for each class [27].

To tackle this, algorithm-adaptation approaches that directly handle multi-label problems can be considered [30]. Zhang et al. [34] proposed a multi-label feature selection method based on a GA that evaluates the benefits of a selected feature subset using the actual accuracy of a multi-label classifier. However, this method suffers from common drawbacks such as premature solutions and a slow convergence speed.

Ji and Ye [13] introduced an integrated learning framework in which the algorithm performs feature selection and multi-label classification simultaneously. This method assumes linear relations between the input features and multiple labels. Thus, the objective of this method is to find a combination of linear functions for each label. Because the coefficients of each linear function are trained using each label independently, it does not consider relations among labels. Nie et al. [20] designed an efficient feature selection method for genomic and proteomic biomarker selections that minimizes a joint $l_{2,1}$-norm based loss function and regularization. However, this method suffers from iterative matrix inversion calculations. Qian and Davidson [22] proposed a multi-label classification algorithm for a semi-supervised input dataset; here, patterns in an input dataset are partially assigned to a certain label subset, while the rest remain unlabeled. Similar to the work of Ji and Ye [13], Qian and Davidson assumed that the dependency between features and labels can be represented through linear combinations.

Gu et al. [12] proposed a multi-label feature selection method that minimizes errors in label ranking. Although this method may result in a higher generalized performance owing to the merit of support vector machines, it suffers from a high computational cost resulting from the exhaustive calculations required to find an appropriate hyperspace using pair-wise pattern comparisons. Kong and Yu [15] proposed a multi-label feature selection method for graph classification for use in drug activity predictions and toxicology analysis. This feature selection method optimizes the Hilbert–Schmidt independence score of selected feature subsets. Lee and Kim [17] introduced a multi-label feature selection method that maximizes the mutual information between a feature subset and multiple labels.

3. Proposed method

3.1. Motivation and approach

In this study, we enhance the performance of a population-based search, such as a GA, for multi-label feature selection by incorporating a multi-label-specific local refinement method. Fig. 1 shows a schematic of the cooperation process between a