Pattern Recognition Letters 34 (2013) 349–357

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Feature selection for multi-label classification using multivariate mutual information

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 3 April 2012
Available online 2 November 2012

Communicated by S. Sarkar

Keywords:
Multi-label feature selection
Multivariate feature selection
Multivariate mutual information
Label dependency

ABSTRACT

Recently, classification tasks that naturally emerge in multi-label domains, such as text categorization, automatic scene annotation, and gene function prediction, have attracted great interest. As in traditional single-label classification, feature selection plays an important role in multi-label classification. However, recent feature selection methods require preprocessing steps that transform the label set into a single label, resulting in subsequent additional problems. In this paper, we propose a feature selection method for multi-label classification that naturally derives from mutual information between selected features and the label set. The proposed method was applied to several multi-label classification problems and compared with conventional methods. The experimental results demonstrate that the proposed method improves the classification performance to a great extent and has proved to be a useful method in selecting features for multi-label classification problems.

1. Introduction

Multi-label classification is a challenging problem that emerges in several modern applications such as text categorization, gene function classification, and semantic annotation of images (Schapire and Singer, 2000; Sebastiani, 2002; Lewis et al., 2004; Diplaris et al., 2005; Boutell et al., 2004). As in the traditional classification problem, the performance of multi-label classification is strongly influenced by the quality of input features. Theoretically, a pattern may lose its distinguishment owing to the irrelevant or redundant features since the similarity of each pair of patterns in same class can be decreased (Watanabe, 1969). These features could cause additional problems of confusing the learning algorithm and leading to poor classification performance (Guyon and Elisseeff, 2003; Saeys et al., 2007).

Consequently, most recent research concerned with multi-label classification naturally employed feature selection techniques (Yang and Pedersen, 1997; Chen et al., 2007; Doquire and Verleysen, 2011; Trohidis et al., 2008). The feature selection is a task of selecting relevant features directly to preserve the internal meaning of given features as it is. This is an important constraint in some applications; for example, the task of gene function classification considers the classification accuracy as well as the biological analysis of the selected features (Diplaris et al., 2005). In the present study, we focused on the feature selection approach to improve the performance of multi-label classification while preserving the inherent meaning of given features.

To select a set of relevant features from given data set, some multi-label feature selection algorithms optimize a set of parameters during feature selection process to tune the kernel function of multi-label classifier (Gu et al., 2011). However, it frequently encounters exhaustive calculations to find an appropriate hyperspace using pairwise comparisons of patterns. This process should be done in each iterative feature selection step, so it is impractical in the viewpoint of computational cost. There is another way of treating multi-label learning; this approach converts the multi-label problems into traditional single-label multi-class problem, and then each feature is evaluated in terms of dependency to transformed new single-label (Chen et al., 2007; Trohidis et al., 2008). This is the most simple approach and provides a connection between single-label learning researches and novel multi-label learning. However, it causes subsequent problems, since multiple labels are transformed to a single label, so that newly created label inherently contains too many classes, leading to difficulty of learning (Read, 2008).

In this paper, we propose a mutual information based multi-label feature selection criterion. The characteristic of our proposed method is that it does not involve any type of transformation method – it selects an effective feature subset by maximizing the dependency between selected features and labels. To the best of our knowledge, it is the first time of proposing a feature filter criterion that takes into account label interactions in evaluating the dependency of given features without resorting to problem transformation. This paper is organized as follows: Section 2 gives a

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

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.10.005