Chapter 1.21
Multi-Label Classification: An Overview

Grigorios Tsoumakas
Aristotle University of Thessaloniki, Greece

Ioannis Katakis
Aristotle University of Thessaloniki, Greece

ABSTRACT

Multi-label classification methods are increasingly required by modern applications, such as protein function classification, music categorization, and semantic scene classification. This article introduces the task of multi-label classification, organizes the sparse related literature into a structured presentation and performs comparative experimental results of certain multi-label classification methods. It also contributes the definition of concepts for the quantification of the multi-label nature of a data set.

INTRODUCTION

Traditional *single-label* classification is concerned with learning from a set of examples that are associated with a single label l from a set of disjoint labels L, $|L| > 1$. If $|L| = 2$, then the learning problem is called a *binary* classification problem (or *filtering,* in the case of textual and Web data), while if $|L| > 2$, then it is called a *multiclass* classification problem.

In *multi-label* classification, the examples are associated with a set of labels $Y \subseteq L$. In the past, multi-label classification was mainly motivated by the tasks of text categorization and medical diagnosis. Text documents usually belong to more than one conceptual class. For example, a newspaper article concerning the reactions of the Christian church to the release of the *Da Vinci Code* film can be classified into both of the categories *society\religion* and *arts\movies*. Similarly, in medical diagnosis, a patient may be suffering, for example, from diabetes and prostate cancer at the same time.

Nowadays, we notice that multi-label classification methods are increasingly required by modern applications, such as protein function classification (Elisseeff & Weston, 2002), music categorization (Li & Ogihara, 2003), and semantic scene classification (Boutell, Luo, Shen, & Brown, 2004). In semantic scene classification, a photograph can belong to more than one conceptual class at the same time, such as *sunsets* and *beaches*. Similarly, in music categorization, a song may belong to more than one genre. For example, several hit songs of the popular rock band *Scorpions* can be characterized as both *rock* and *ballad*.

This article aims to serve as a starting point and reference for researchers interested in multi-label classification. The main contributions are: a) a structured presentation of the sparse literature on multi-label classification methods with comments on their relative strengths and weaknesses, and, when possible, the abstraction of specific methods to more general and thus more useful schemata; b) the introduction of an undocumented multi-label method; c) the definition of a concept for the quantification of the multi-label nature of a data set; and d) preliminary comparative experimental results about the performance of certain multi-label methods.

The remainder of the article is organized as follows. The next section discusses tasks that are related to multi-label classification. Subsequently follows an organized presentation of multi-label classification methods. The next section introduces the concept of label density, and presents the metrics that have been proposed in the past for the evaluation of multi-label classifiers. The following section presents the comparative experiments and discusses the results, while the concluding section summarizes this work and points to future research directions.

RELATED TASKS

A task that also belongs to the general family of supervised learning and is very relevant to multi-label classification is that of *ranking*. In ranking, the task is to order a set of labels *L*, so that the topmost labels are more related to the new instance. There exist a number of multi-label classification methods that learn a ranking function from multi-label data. However, a ranking of labels requires postprocessing in order to give a set of labels, which is the proper output of a multi-label classifier.

In certain classification problems, the labels belong to a *hierarchical structure*. The *dmoz* open directory, for example (http://www.dmoz.org/), maintains a hierarchy of conceptual classes for the categorization of Web pages. A Web page may be labelled using one or more of those classes, which can belong to different levels of the hierarchy. The top level of the Munich Information Centre for Protein Sequences (MIPS) hierarchy (http://mips.gsf.de/) consists of classes, such as *metabolism*, *energy*, *transcription,* and *protein synthesis*. Each of these classes is then subdivided into more specific classes, and these are in turn subdivided, and then again subdivided, so the hierarchy is up to four levels deep (Clare & King, 2001). When the labels in a data set belong to a hierarchical structure, then we call the task *hierarchical classification*. If each example is labelled with more than one node of the hierarchical structure, then the task is called *hierarchical multi-label classification*. In this article, we focus on flat (nonhierarchical) multi-label classification methods.

Jin and Ghahramani (2002) call *multiple-label problems*, the semisupervised classification problems, where each example is associated with more than one class, but only one of those classes is the true class of the example. This task is not that common in real-world applications as the one we are studying.

Multiple-instance learning is a variation of supervised learning, where the task is to learn a concept given positive and negative bags of instances (Maron & Lozano-Perez, 1997). Each bag may contain many instances, but a bag is labelled positive even if only one of the instances in it falls within the concept. A bag is labelled negative only if all the instances in it are negative.