Pattern Recognition 44 (2011) 2436–2449

Contents lists available at ScienceDirect

Pattern Recognition

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

# Hierarchical annotation of medical images

Ivica Dimitrovski [a,b,*], Dragi Kocev [a], Suzana Loskovska [b], Sašo Džeroski [a]

[a] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b] Department of Computer Science and Computer Engineering, Faculty of Electrical Engineering and Information Technologies, Rugjer Boshkovikj bb, 1000 Skopje, Macedonia

ARTICLE INFO

*Article history:*
Received 16 July 2010
Received in revised form
28 February 2011
Accepted 19 March 2011
Available online 29 March 2011

*Keywords:*
Automatic image annotation
Hierarchical multi-label classification
Predictive clustering trees
Feature extraction from images

ABSTRACT

We present a hierarchical multi-label classification (HMC) system for medical image annotation. HMC is a variant of classification where an instance may belong to multiple classes at the same time and these classes/labels are organized in a hierarchy. Our approach to HMC exploits the annotation hierarchy by building a single predictive clustering tree (PCT) that can simultaneously predict all annotations of an image. Hence, PCTs are very efficient: a single classifier is valid for the hierarchical semantics as a whole, as compared to other approaches that produce many classifiers, each valid just for one given class. To improve performance, we construct ensembles of PCTs. We evaluate our system on the IRMA database that consists of X-ray images. We investigate its performance under a variety of conditions. To begin with, we consider two ensemble approaches, bagging and random forests. Next, we use several state-of-the-art feature extraction approaches and combinations thereof. Finally, we employ two types of feature fusion, i.e., low and high level fusion. The experiments show that our system outperforms the best-performing approach from the literature (a collection of SVMs, each predicting one label at the lowest level of the hierarchy), both in terms of error and efficiency. This holds across a range of descriptors and descriptor combinations, regardless of the type of feature fusion used. To stress the generality of the proposed approach, we have also applied it for automatic annotation of a large number of consumer photos with multiple annotations organized in semantic hierarchy. The obtained results show that this approach is general and easily applicable in different domains, offering state-of-the-art performance.



## 1. Introduction

Digital imaging in medicine is in constant growth due to the increasing availability of imaging equipment in hospitals. Average-sized radiology departments now produce several tera-bytes of data annually. This prompts for efficient systems for image annotation, storage, retrieval and mining. Typically, medical image databases are accessed via textual information through the standard picture archiving and communication system (PACS) [1,2]. PACS integrates imaging modalities and interfaces with hospital and departmental information systems to manage storage and distribution of images to medical personnel, researchers, clinics, and imaging centers. An important requirement of PACS is the provision of an efficient search function to access the required images.

An universal format for PACS image storage and retrieval is the digital imaging and communications in medicine (DICOM) standard [3]. DICOM is a well-known standard for handling, storing, printing, and transmitting information in medical imaging. The DICOM header contains tags to decode the body part examined, the patient position and the acquisition modality. Some of the tags are automatically set by the digital system according to the imaging protocol used to capture the pixel data. Other part of the tags are set manually by the physicians or radiologists during the routine documentation. This procedure cannot always be considered very reliable, since frequently some entries are either missing, false, or do not describe the anatomic region precisely [4]. Furthermore, manual annotation of images is an expensive and time-consuming procedure, especially given the large and constantly growing databases of medical images. Thus, completely automated categorization in terms of DICOM tags is currently not possible, but is highly desirable.

Automatic image annotation or image classification is an important step in image retrieval. In the medical domain, using information directly extracted from images to annotate/categorize them will improve the quality of image annotation in particular, and more generally the quality of patient care. Properly classified medical image

* Corresponding author at: Department of Computer Science and Computer Engineering, Faculty of Electrical Engineering and Information Technologies, Rugjer Boshkovikj bb, 1000 Skopje, Macedonia. Tel.: +389 2 3099 159.
*E-mail addresses:* ivicad@feit.ukim.edu.mk (I. Dimitrovski), Dragi.Kocev@ijs.si (D. Kocev), suze@feit.ukim.edu.mk (S. Loskovska), Saso.Dzeroski@ijs.si (S. Džeroski).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.03.026

data can help medical professionals in fast and effective access to data in their teaching, research, training, and diagnostic problems. The results of the classification step can also be used for multilingual image annotation as well as for DICOM header correction [5].

Automatic image annotation can be used for retrospective annotation (pre DICOM). It can also be used as help for human annotators (i.e., radiologists), where the annotations that are suggested by the system are corrected/verified/confirmed by the human annotator. The limits of performance of an automated annotation system that learns from example images annotated by humans, is the rate/probability of operator error/agreement of annotators.

Automatic image annotation uses a computer system which automatically assigns metadata in the form of captions or keywords to a digital image. Typically, image analysis first extracts feature vectors. Then, together with the training annotations, they are used by a machine learning algorithm to learn to automatically assign annotations. The performance of the computer system largely depends on the availability of strongly representative visual features, able to characterize different visual properties of the images, and the use of effective algorithms for training classifiers for automatic image annotation.

A single image may contain different meanings organized in hierarchical semantics: hence, hierarchical multi-label classification (HMC) is strongly recommended for obtaining multi-label annotations. The task of multi-label classification is to assign multiple labels to each image. The assigned labels are a subset of a previously defined set or hierarchy of labels. HMC is used in various domains [6], such as text classification, scene and video classification, medical imaging and biological applications. One of the main issues involved in multi-label classification is the importance of detecting and incorporating the connections between the labels into the process of assigning multiple labels. A second and related issue is the additional complexity involved in learning multi-label classifiers, as compared to learning single-label classifiers.

In this paper, we present a HMC system for medical image annotation. This system consists of the two standard parts of image annotation systems, i.e., processing (feature extraction) and classification of images. The image processing part uses state-of-the-art approaches to convert an image to a set of numerical features extracted directly from the pixel values. The image classification part, which labels and groups the images, contains the main novelty of our approach: The labels can be organized in a hierarchy and an image can be labeled with more than one label (an image can belong to more than one group).

First, we generate four different types of descriptors suitable for X-ray medical images: raw pixel representation (RPR) [7], local binary patterns (LBP) [8], edge histogram descriptors (EHD) [9], and scale-invariant feature transform (SIFT) [10]. The features are generated using the medical X-ray images from the ImageCLEF2009 medical image annotation task [5]. Next, we use these features together with the annotations to train the classifiers. In particular, we use ensembles (bags and random forests) of PCTs for HMC and SVMs for single-label classification, the most widely used classifier in the area of image annotation. At the end, we assess the predictive performance of the classifiers using the hierarchical error measure (HEM) from ImageCLEF [5] and overall recognition rate (RR), commonly used for assessing the predictive performance over the database we use.

The main question that we address in our research is whether exploiting the semantic knowledge about the inter-class relationships among the image labels (organized in a hierarchical structure) can improve the predictive performance of a system for automatic image annotation. To this end, we compare the predictive performance of the ensembles of PCTs for HMC (that predict all labels simultaneously) to that of SVMs (each of them predicting a single label). We do this across all feature extraction techniques, thus evaluating the different feature extraction techniques and their use in HMC of medical X-ray images. Moreover, we investigate whether (and which type of) combination of feature extraction techniques yields better predictive performance. We consider low level (LL) and high level (HL) feature fusion/combination schemes [7].

To emphasize the generality of our approach, we have also tested it on the database of general images from the ImageCLEF@ICPR 2010 photo annotation task [11]. The images in this database are annotated with 53 visual concepts organized in a classification scheme with hierarchical structure, which we used to build ensembles of PCTs for HMC as classifiers. The 53 concepts include abstract categories (like partylife), the time of day (like day or night), persons (like no person visible, small or big group) and quality (like blurred or underexposed). A complete overview of the task is given by Nowak [11].

The remainder of the paper is organized as follows. In Section 2, we give an overview of related work. Section 3 introduces predictive clustering trees and their use for HMC. Section 4 describes the techniques for feature extraction from images. In Section 5, we explain the experimental setup for annotating medical images. The obtained results and a discussion thereof are given in Section 6. Section 7 describes the experiments in annotation of general images, as well as their results. Section 8 concludes the paper and points out some directions for further work.

## 2. Related work

In this section, we present some classification methods that are or can be used for image annotation. We begin by presenting the methods that are most widely used by the image annotation community. We then present some recent machine learning methods that can be used for hierarchical image annotation and discuss their relation to the method we propose.

Regardless of the number of visual concepts that have to be learned and their mutual connections, most of the present systems for annotation of general images (and medical images in particular) learn a separate model for each visual concept (label), i.e., they treat the classes as completely separate and independent (both visually and semantically). This means that multi-label classification problems are transformed into several binary classification problems. For example, the methods with high predictive performance at recent challenges/competitions in detection and annotation tasks (such as the PASCAL visual object classes challenge [12], the ImageCLEF medical image annotation task [13,5] and the ImageCLEF visual concept detection and annotation task [14]) perform multi-label classification by building binary classifiers for each label. The instances associated with particular label are in one class and the rest are in another class. For solving the binary classification problems, is common to use a SVM with a $\chi^2$ kernel [15]. This means that the increase of the number of labels used for annotation will linearly increase the complexity of such an approach.

To deal with a large number of labels/classes, many approaches combine binary classifiers using class hierarchies [16,17]. This results in a logarithmic increase of complexity as the number of labels increases. The class hierarchies can be automatically constructed through analysis of visual similarities: this can proceed top-down by recursive partitioning of the set of classes [18] or bottom-up by agglomerative clustering [19]. The hierarchies could also be found by exhaustive search or random sampling followed by cross-validation [20].

An alternative method for automatic construction of hierarchies is to query an external semantic network with class labels