Pattern Recognition Letters 38 (2014) 38–45

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Fast and efficient visual codebook construction for multi-label annotation using predictive clustering trees

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

[a] Faculty of Computer Science and Engineering, Ss Cyril and Methodius University, Rugjer Boshkovikj 16, MK-1000 Skopje, Macedonia
[b] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 10 January 2013
Available online 6 November 2013

Communicated by Eckart Michaelsen

Keywords:
Automatic image annotation
Visual codebook construction
Predictive clustering trees
Multi-label classification

ABSTRACT

The bag-of-visual-words approach to represent images is very popular in the image annotation community. A crucial part of this approach is the construction of visual codebook. The visual codebook is typically constructed by using a clustering algorithm (most often *k*-means) to cluster hundreds of thousands of local descriptors/key-points into several thousands of visual words. Given the large numbers of examples and clusters, the clustering algorithm is a bottleneck in the construction of bag-of-visual-words representations of images. To alleviate this bottleneck, we propose to construct the visual codebook by using predictive clustering trees (PCTs) for multi-label classification (MLC). Such a PCT is able to assign multiple labels to a given image, i.e., to completely annotate a given image. Given that PCTs (and decision trees in general) are unstable predictive models, we propose to use a random forest of PCTs for MLC to produce the overall visual codebook. Our hypothesis is that the PCTs for MLC can exploit the connections between the labels and thus produce a visual codebook with better discriminative power. We evaluate our approach on three relevant image databases. We compare the efficiency and the discriminative power of the proposed approach to the literature standard – *k*-means clustering. The results reveal that our approach is much more efficient in terms of computational time and produces a visual codebook with better discriminative power as compared to *k*-means clustering. The scalability of the proposed approach allows us to construct visual codebooks using more than usually local descriptors thus further increasing its discriminative power.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Many of the popular methods for image annotation are using a bag-of-visual-words to represent the visual content of an image (Nowak and Dunker, 2010; Everingham et al., 2012; van de Sande et al., 2010). The basic idea of this approach is first to sample a set of key-points (i.e., local regions) from the image using some method (e.g., densely, randomly or using a key-point detector). A local visual descriptor, such as the scale-invariant feature transform (SIFT) descriptor and normalized pixel values, is then calculated for each key-point. The resulting distribution of local descriptors is then quantified against a pre-specified visual codebook. The visual codebook converts the distribution of local descriptors into a histogram. The main issues that need to be considered when applying this approach include sampling of the key-points, selection of the type of local descriptor, building the visual codebook (set of visual words) and assignment of the local descriptors to visual words from the visual codebook.

1.1. Motivation

The discriminative power of the visual codebook determines the performance of the annotation. However, the construction of the visual codebook and the assignment of the local descriptors to the visual words from the visual codebook is very often a bottleneck in image annotation (Philbin et al., 2007). This is mainly because *k*-means clustering, which is computationally expensive, is the most widely used method for visual codebook construction. The computational cost of the *k*-means algorithm is even more pronounced on complex datasets with a large number of images (van de Sande et al., 2010). Moreover, *k*-means clustering is a un-supervised learning algorithm and thus does not exploit the information contained in the label annotations.

To resolve these issues, Moosmann et al. (2008) and Uijlings et al. (2009) propose to use supervised tree-based machine learning methods to construct the visual codebook. These methods are able to exploit the image annotations in the context of single-label classification, i.e., for images annotated with only a single label.

* Corresponding author. Tel.: +389 2 3099 159.
E-mail addresses: ivica.dimitrovski@finki.ukim.mk (I. Dimitrovski), Dragi.Kocev@ijs.si (D. Kocev), suzana.loshkovska@finki.ukim.mk (S. Loskovska), Saso.Dzeroski@ijs.si (S. Džeroski).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.10.016

This means that the proposed approaches can handle image databases with relatively simple images with single annotations. However, in reality, the images are annotated with multiple labels: An image depicting some street will probably also depict buildings, cars, trees and people.

1.2. Contribution of the research

In this paper, we present a method for fast and efficient construction of visual codebooks that is able to use images with multiple annotations/labels. We use predictive clustering trees (PCTs) for multi-label classification (MLC) (Blockeel et al., 1998) to decrease the time needed to construct the visual codebook and, at the same time, to improve its discriminative power. PCTs are capable of annotating an instance with multiple labels simultaneously and thus exploiting the interactions that may occur among the different labels.

Although most visual codebooks are built without using the labels, our approach uses the available annotations and the interactions among them to guide the visual codebook construction process. The PCTs are trained to perform multi-label classification: The visual codebook is then constructed by assigning a distinct visual word to each leaf of the tree. The PCTs can be considered as a data-driven and a semantic approach to visual codebook construction because they rely on the available image annotations (labels).

The main research questions that we are addressing in this manuscript are as follows. First, we investigate whether the proposed method for codebook construction is more efficient and scalable than the literature standard (i.e., *k*-means). Next, we test the discriminative power of the obtained visual codebook and compare it to the discriminative power of the codebook obtained using *k*-means. Furthermore, we investigate the influence of the number of selected key-points used to obtain the visual words on the discriminative power of the codebook. Finally, we examine the influence of the number of PCTs for MLC used for codebook construction on the codebook's discriminative power.

1.3. Organization of the paper

The remainder of this paper is organized as follows. In Section 2, we present related work. Section 3 introduces predictive clustering trees and their extension for multi-label classification. In Section 4, we explain the experimental setup. The obtained results and a discussion thereof are given in Section 5. Section 6 concludes the paper.

2. Related work

Many studies have shown that the bag-of-visual-words approach exhibits an impressive performance for image annotation problems (Everingham et al., 2012; Nowak, 2010; Nowak and Huiskes, 2010). A crucial step in the bag-of-visual-words approach is the codebook construction. The visual codebook can be constructed using unsupervised or supervised machine learning methods.

The unsupervised methods are most widely used by the image annotation community. Typically, the visual codebook is constructed by applying *k*-means clustering to the key-points, i.e., the local descriptors (e.g., SIFT) extracted from the images (van de Sande et al., 2010; Lowe, 2004). The *k*-means algorithm has two serious limitations when applied to the image annotation problem (Jurie and Triggs, 2005). First, it works with small visual codebooks, i.e., with only thousands of visual words, while many datasets may have tens of thousands of visual words. Second, it constructs more clusters close to the most frequently occurring features. These limitations can be addressed with the hierarchical *k*-means (HKM) approach (Nister and Stewenius, 2006) and radius-based clustering (van Gemert et al., 2010).

The supervised machine learning methods for constructing visual codebooks use the label annotations of the images to guide the construction of the visual codebooks. They can use single-label or multi-label methods. With the first type of methods, a visual codebook is constructed for each label separately and then these are aggregated over all possible labels. The second type of methods constructs a single visual codebook valid for all labels. Our approach belongs to the second type of methods.

Uijlings et al. (2009) and Chatfield et al. (2011) have recently published detailed overviews of bag-of-words methods for creation of visual codebooks. These surveys compare several state-of-the-art methods for un-supervised and supervised creation of visual codebooks used in the context of single-label classification or multi-class classification. The comparison of tree-based and *k*-means methods for visual codebook construction reveals that the tree-based methods are more efficient than methods based on *k*-means. However, the improvement of the computational efficiency comes with the price of decreasing the discriminative power of the codebook (i.e., the classifier using the tree-based codebook produced worse annotations). The methods from the literature have, so far, considered the construction of visual codebooks in the context of single-label classification/annotation or multi-class classification/annotation. Another method, proposed by Wojcikiewicz et al. (2010), constructs a separate codebook for each possible label and then reconciles the several codebooks using agglomerative information bottleneck. The experimental evaluation over a single image database with small number of labels per image showed small increase of performance as compared to classical *k*-means clustering. All of these methods, in the process of codebook construction, do not consider the label dependencies, i.e., that if an image is annotated with the label *cloud*, it will probably be also annotated with the label *sky*.

In order to alleviate all of these issues, we propose here a method for constructing a visual codebook for multi-label classification/annotation problems that can explore the existing connections between the labels. This is due to the fact that we construct a single predictive model that is valid for the complete label space. Moreover, decision trees in combination with random forest ensemble methods have not yet been considered for building visual codebooks for a large number of visual concepts (labels), where an image can be annotated with multiple labels (multi-label image annotation problems). To this end, we propose to use random forest of PCTs for MLC, considering that there are many real-life challenging multi-label image annotation problems (Nowak and Dunker, 2010; Nowak, 2010; Nowak and Huiskes, 2010).

3. Visual codebook construction using predictive clustering trees

3.1. The task of multi-label classification (MLC)

The problem of single-label classification is concerned with learning from examples, where each example $\mathbf{x} \in \mathcal{X}$ ($\mathcal{X}$ denotes the domain of descriptive variables for the examples) is associated with a single label λ_i from a finite set of disjoint labels $\mathcal{L} = \{\lambda_1, \lambda_2, \ldots, \lambda_Q\}, Q > 1$. For $Q > 2$, the learning problem is referred to as *multi-class classification*. On the other hand, the task of learning a mapping from an example $\mathbf{x} \in \mathcal{X}$ to a set of labels $\mathcal{Y} \subseteq \mathcal{L}$ is referred to as a *multi-label classification*. In contrast to multi-class classification, alternative labels in multi-label classification are not assumed to be mutually exclusive: multiple labels may be associated with a single example, i.e., each example can be a member of more than one class. Labels in the set $\mathcal{Y}$ are relevant

Download English Version:

https://daneshyari.com/en/article/533889

Download Persian Version:

https://daneshyari.com/article/533889

Daneshyari.com