Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2645–2655
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

SVM-based active feedback in image retrieval using clustering and unlabeled data

Rujie Liu[a,*], Yuehong Wang[a], Takayuki Baba[b], Daiki Masumoto[b], Shigemi Nagata[b]

[a] Fujitsu Research and Development Center, Co. Ltd., Beijing 100025, PR China
[b] Fujitsu Laboratories, Kawasaki, Japan

Received 14 February 2007; received in revised form 24 August 2007; accepted 28 January 2008

Abstract

In content-based image retrieval, relevance feedback is studied extensively to narrow the gap between low-level image feature and high-level semantic concept. However, most methods are challenged by small sample size problem since users are usually not so patient to label a large number of training instances in the relevance feedback round. In this paper, this problem is solved by two strategies: (1) designing a new active selection criterion to select images for user's feedback. It takes both the informative and the representative measures into consideration, thus the diversities between these images are increased while their informative powers are kept. With this new criterion, more information gain can be obtained from the feedback images; and (2) incorporating unlabeled images within the co-training framework. Unlabeled data partially alleviates the training data scarcity problem, thus improves the efficiency of support vector machine (SVM) active learning. Systematic experimental results verify the superiority of our method over existing active learning methods.

Keywords: Image retrieval; Relevance feedback; Clustering; SVM; Unlabeled data

1. Introduction

The success of content-based image retrieval (CBIR) is largely limited due to the gap between low-level features and high-level semantic concepts. To bridge this gap, relevance feedback (RF) initially developed in text retrieval was introduced into CBIR since the 1990s and big success was achieved [1,2] RF focuses on the interactions between the user and the search engine by letting the user provide feedback regarding the retrieval results, e.g. by labeling images returned as either positive or negative; from this feedback, the engine is refined and improved results are returned to the user.

Many RF methods have been developed in recent years along the path from heuristic based techniques to optimal learning algorithms. Early RF approach belongs to the family of "query point movement" (QPM) category, for which the task of the search engine consists in, at each retrieval round, finding a better query point and adjusting the weights of various features to adapt to the user's perception [3–5]. However, all these methods make strong assumption that the target class has an elliptical shape, which greatly limits their performance.

Later on, researchers begin to look at this problem more systematically by formulating it into a learning, classification, or density estimation problem. In Ref. [6], the assumption regarding the shape of the target class was removed, and a Parzen window method is used instead to estimate the distribution of positive and negative samples. Discriminant EM method [8] casts image retrieval as a transductive learning problem by combining unlabeled examples in supervised learning to achieve better classification, however, this scheme has the potential difficulty in computational expenses, especially when the database is large. Based on the observation that "all positive examples are alike and each negative example is negative in its own way", Zhou and Huang [1] proposed biased discriminant analysis (BDA) and its kernel form to find the transformed space where the positive examples cluster while the negative ones scatters away. To handle the singularity problem caused by small sample size, Tao [7] designed direct kernel BDA scheme, where

* Corresponding author. Tel.: +86 10 59691504; fax: +86 10 59691000.
E-mail address: rjliu@cn.fujitsu.com (R. Liu).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.01.023

direct discriminant analysis is used to replace the regularization method used in BDA.

Recent work on RF often relies on support vector machines (SVMs). Compared with other learning algorithms, SVM appears to be a good candidate for several reasons: generalization ability; without restrictive assumptions regarding the data; fast learning and evaluation for RF; flexibility, e.g., prior knowledge can be easily used to tune its kernels, etc. While most of the existing work using SVMs for RF concentrates on 2-class SVMs [9–11], 1-class SVMs were also used in order to learn from positive examples only [12]. In addition, different ensemble methods were also proposed to combine a number of SVM classifiers to improve the classification performance [13,14].

At first sight, above methods can be directly applied to incrementally learn user's perception through the user's feedback information. However, all these learners are challenged by small sample size problem. This is because few users will be so patient to label a large number of training instances in the RF process. To make most use of the limited training data, how to choose images for the user to label is a crucial issue in minimizing the amount of interactions between the user and the learner required to reach good results. Two strategies are commonly used to address the training data scarcity problem: (1) active learning, or selective sampling and (2) exploiting unlabeled data.

Active learning studies the strategy for the learner to actively select samples to query the user for labels, in order to achieve the maximal information gain in decision making. Cox et al. [15] used entropy minimization in search of the set of images that, once labeled, will minimize the expected number of future iteration. Tong and Chang [9] proposed the SVM_{Active} algorithm for applications in text classification and image retrieval. According to their criterion, good requests should maximally reduce the size of the version space, which can be approximately achieved by selecting the points near the SVM boundary. Later on, more efforts were made to improve the power of Tong's SVM_{Active} solution. In Ref. [10], the concept of mean version space was proposed, where both the size of the version space and the posterior probabilities were taken into consideration, and it was claimed that this criterion was tailored for each specific learning task thus can maximally shrink the version space. Angle-diversity algorithm [16] tried to select batches of new training examples by balancing the distance from samples to the classification hyperplane and the diversity among these samples.

Learning from unlabeled data has become a hot topic over the past few years for solving the problem of insufficient training data. It aims to enhance classification accuracy with the help of unlabeled data, and some promising results have been reported. Transductive learning is one of the popular methods in this area, which levers the unlabeled data near the labeled ones to increase the negative-labeled pool. Some methods use a generative model for the classifier and employ EM scheme to model the label or parameter estimation process [8], while others construct a graph on the examples such that the minimum cut on the graph yields an optimal labeling of the unlabeled examples [17] Another prominent achievement in exploiting unlabeled data is the co-training paradigm, which trains two classifiers separately on two distinct views and uses the predictions of each classifier on unlabeled examples to augment the training set of another classifier [18].

In this paper, we aim to improve SVM_{Active} by following two strategies:

(1) Enhance its active selection criterion by incorporating a representative measure. To achieve the maximal information gain, we argue that the selected images must be informative given the current estimation of the target and the redundancy between these images has to be low. SVM_{Active} criterion returns only the most ambiguous images for labeling, however, there may exist much redundancy between these examples. In such case, these images are too similar to each other, and the feature space represented by these samples is generally very small, thus, the SVMs trained on these examples are usually biased and unstable, especially during the first round of the iterations.
To solve the redundancy problem, we propose the representative criterion to maximize the information capacity of the returned images. In each retrieval round, the images that are near the classification boundary are clustered by an unsupervised learning process and one image is selected from each cluster for labeling, where the selection is carried out based on the combination of informative and representative measures of these images. With this dynamic clustering process, diversities between the selected images are increased and the feature space covered by them is enlarged, thus, more information gain can be obtained from these returned images.

(2) Exploit unlabeled data within the co-training framework. Distinct views required by co-training can be provided by different learning algorithms, e.g., by using SVM and adaptive weighted Euclidean distance model. In each feedback round, a QPM analogous procedure is adopted to build a simple Euclidean distance model based on all the images labeled till now, and its most confident negative examples are used to augment the training pool of the SVM. Here, two conservative mechanisms are employed, that is, the number of negative examples used in co-training is very few and they are only temporarily used. In this way, the influence of the possible mistakes made by the Euclidean distance model can be largely limited. Experimental results verify that this strategy is very effective both in alleviating the performance degradation during the first feedback round, and in enhancing SVM's classification power in the late stages.

The rest of this paper is organized as follows. In Section 2, we firstly introduce the SVM_{Active} method, and then, we incorporate our representative measure to generate a combined criterion. Section 3 presents the usage of unlabeled data within the co-training framework. The system framework is illustrated in Section 4. The experimental results are shown in Section 5, followed by the conclusion in Section 6