Neurocomputing 196 (2016) 95–106

Contents lists available at ScienceDirect

Neurocomputing

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

Practice makes perfect: An adaptive active learning framework for image classification

Zhipeng Ye, Peng Liu, Jiafeng Liu, Xianglong Tang, Wei Zhao*

Harbin Institute of Technology, No. 92 West Dazhi Street, Harbin 150001, China

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form
4 January 2016
Accepted 31 January 2016
Available online 9 March 2016

Keywords:
Active learning
Entropy-based sampling
Image classification
Bag-of-visual-words
Cognitive model

ABSTRACT

Active learning is an effective method for iteratively selecting a subset of images from an unlabeled dataset. One of the most widely used active learning strategies is uncertainty sampling. However, traditional sampling strategies do not take the category of samples into consideration, and the selected images do not reflect the desired training distribution, leading to the result that additional labeling work needs to be done. To deal with these problems, from the aspect of visual perception, we improve the traditional entropy-based uncertainty sampling strategy by introducing a certainty measurement estimated by a bag-of-visual-words (BoVW). The Rescorla–Wagner perceptive model is utilized to quantify the stop criterion. This method differs from previous approaches that treated sampling and classifying process separately: we treat the learning process as a uniform model by proposing a new evolving sample selection method that uses the unified negative-accelerated learning principle and takes category distribution into consideration. A classifier is trained to provide category distributions for the sampling process to improve its sampling performance and reduce additional annotation costs for the human annotator. During the training process, weights for both modules are adaptively initialized by the inner similarity of sample set measured by structural similarity (SSIM), and dynamically adjusted according to the learning process of the human. In addition to the regular tests that are commonly utilized by traditional sampling methods, the transfer test, based on transfer learning theory, is utilized to further evaluate the performance of different sampling strategies. Experimental results on real world datasets show that our active sampling framework outperforms both baseline and state-of-the-art adaptive active learning strategies.

1. Introduction

Image classification is an important problem in computer vision and machine learning. Because of the explosive growth of imaging techniques, especially the fast popularization of smart devices, automatically obtaining semantic image information has become indispensable and has a wide range of applications in real-world information systems. A noteworthy problem in practice is that it is relatively easy to get unlabeled data, but labeling is expensive. Traditional supervised learning methods do not capture this distinction, motivating the field of active learning, which aims to train an accurate prediction model with minimum cost by labeling the most informative instances.

A typical active learning framework consists of two parts: a learning engine and a sample selection engine. In each iteration, the learning engine trains a model to predict the labels of unlabeled samples based on the training set. The sample selection engine selects the most effective unlabeled samples based on a certain strategy. For each annotation iteration, a subset of instances is carefully chosen for annotation by an annotator and the remaining instances remain unlabeled. The goal of active learning is to train a classifier with the labeling instances selected by active sampling methods. As summarized in Fig. 1, for active learning, there are three main categories: (i) membership query synthesis [1], (ii) stream-based selective sampling [2], and (iii) pool-based sampling [3]. Membership query synthesis is reasonable for many problems; however, this comes at the cost of discarding the distribution of samples. It is also time-consuming and awkward for a human annotator. The stream-based selective sampling strategy scans each unlabeled instance from source data one at a time for the learner to query. Although it overcomes the shortcoming of the membership query synthesis strategy, a threshold for the quantity of information still needs to be tuned, which is not adaptive for different scenarios. The pool-based

* Corresponding author.
E-mail address: zhaowei@hit.edu.cn (W. Zhao).

http://dx.doi.org/10.1016/j.neucom.2016.01.091
0925-2312/

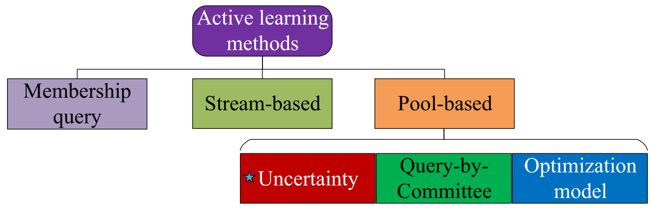

Fig. 1. Main active learning scenarios. Uncertainty sampling is the approach used in this study.

sampling method assumes that both a small set of labeled data and a large pool of unlabeled data are available, and it has been successfully used in image retrieval and classification [4,5].

For the most widely used pool-based sampling method, there are three main developed query strategies [6]: (i) uncertainty sampling, (ii) query-by-committee, and (iii) optimization models. Uncertainty sampling is fit for most classifiers with good precision and generalization ability [7] according to different forms of uncertainty measurement functions. Of these, information entropy [8] is the most commonly used. Query-by-committee is a well-known strategy that seeks a few ambiguous instances to find the best model with a small set of hypotheses. Boosting and bagging methods are commonly used to build up a committee [9]. The Kullback–Leibler divergence [10] is mostly used to evaluate the divergence of each member in the committee [11]. An optimization model is a general query strategy containing a set of query optimization functions, including “expected gradient length” (EGL) [12], loss minimization [13], and variance reduction [14].

Traditional entropy-based methods only consider the uncertainty of samples and ignore their category distribution, which restricts the quality of sampling. Thus, in this paper, we focus on obtaining a more accurate classifier by improving the quality of sampling and reducing the re-annotation work for the human annotator. The negative-accelerated adaptive active sampling with a bag-of-visual-words (NAASB) framework is proposed by introducing a certain measurement into the traditional strategy to deal with this problem. The proposed sampling framework contains two parallel and interactive subtasks, image classifier training and sample selection:

(a) *Image classifier training*: A labeled set is utilized to train an image classifier and adaptively initialize the parameters of NAASB according to the features of different categories. Then, for each iteration, samples are selected using an active learning strategy and annotated to train and improve the performance of the image classifier. The stop criterion for the iteration is measured by the performance of the classifier according to the negative-accelerated principle.

(b) *Sample selection*: In every round of active learning, the certainty output of the classifier is introduced to take category distribution into consideration for entropy-based active sampling strategy. Meanwhile, the selected samples are annotated by the image classifier to reduce the work of the human annotator.

The rest of the paper is organized as follows. Section 2 summarizes the related work. In Section 3, we describe and explain the proposed NAASB framework in detail. Experimental results are given in Section 4. Finally, we offer the conclusions in Section 5.

2. Related work

In this section, we review related methods, including uncertainty sampling strategy in active learning, the bag-of-visual-words (BoVW) method in image classification, and the negative-accelerated learning principle provided by the Rescorla–Wagner model.

2.1. Uncertainty sampling strategy

Uncertainty sampling [15] is one of the most common frameworks that employ probabilistic models to evaluate the potential of instances. The prediction result of a single instance is represented by a vector, whose elements are the posterior probability with respect to each class label. Generally, the informativeness of a sample can be assessed by using the uncertainty of the instances based on models trained from the current labeled sample set. If the uncertainty of a sample is high, it implies that the current model does not have sufficient knowledge to classify the sample, and it is helpful for this sample to be contained in the training set to improve the underlying model. The uncertainty metric is designed based on individual instance importance for an accurate model construction. Applying a traditional uncertainty criterion means that the most uncertain samples should be selected. This strategy is straightforward for both probabilistic and non-probabilistic learning models, and can be divided into three main substrategies [16]: least confidence [17], sample margin [18], and entropy [8]. When using a probabilistic model for binary classification, uncertainty sampling simply queries the instances whose posterior probability of being positive is nearest to 0.5 [3], while for non-probabilistic classifiers, classifiers such as decision trees [15] and nearest-neighbors [19] are utilized by allowing each neighbor to vote for the class label of the selected instance. The uncertainty criterion has been widely explored because of its simplicity. The complexity of the uncertainty sampling strategy is $O(n)$, where n is the scale of the unlabeled pool. The deficiency of the least confidence and sample margin strategies is that they ignore the output label distributions for the remaining classes, while entropy is a typical metric that represents the uncertainty over the whole output prediction distribution.

2.2. Image classification techniques

Image classification has been an active research field over the past several decades. There have been extensive research efforts to develop effective visual object recognizers, such as object and scene classification [20,21], video event detection [22], and action recognition [23]. Nowadays, bag-of-features (BoF) [24] based classification is one of the most classic and commonly used approaches in image retrieval and scenario classification, whose simplicity and effectiveness have been tested over the years. Inspired by BoF, BoVW [25] was proposed to classify images. Visual words are first obtained by k-means clustering local features, commonly described by difference of Gaussians (DoGs) or Scale-invariant feature transforms (SIFTs), and the clustering centers are considered to be the base of the BoVW to represent the whole image for training and testing. However, there are several drawbacks to BoVW: (i) spatial relationships are ignored between image patches during the construction of the visual vocabulary [26], (ii) the hard-assignment strategy of k-means does not necessarily generate an optimized visual vocabulary [27], and (iii) semantics are ignored during the clustering process [28].

Many studies have been carried out to solve these problems. Chai [29] utilized foreground segmentation to improve classification performance on weakly annotated datasets. To address the fact that spatial information among local features was ignored, spatial pyramid matching (SPM) [30] was proposed to make use of spatial information for object and scene categorization. In addition, there are several ways to generate a rich descriptive visual vocabulary. Wang presented a simple but effective coding scheme called locality-constrained linear coding (LLC) in place of the vector quantization (VQ) coding in traditional SPM to improve the categorization performance [31]. Germert introduced ambiguity into visual words to improve their descriptive ability [32].