Neurocomputing 171 (2016) 826–836

Contents lists available at ScienceDirect

Neurocomputing

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

Multiple-instance learning based decision neural networks for image retrieval and classification

Yeong-Yuh Xu

Department of Computer Science and Information Engineering, HungKuang University, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 12 December 2014
Received in revised form
12 May 2015
Accepted 10 July 2015
Communicated by Shuiwang Ji
Available online 21 July 2015

Keywords:
Multiple instance learning
Content based image retrieval
Image classification
Neural networks
Multiple-instance learning based decision neural networks
MI-BDNN

ABSTRACT

The revolutionary Internet and digital technologies have spawned a need for technology that can organize abundantly available digital images for easy categorization and retrieval. Hence, content-based image retrieval (CBIR) has become one of the most active research areas for the last few decades. However, it is still an open issue to narrow down the gap between the high level semantics in the human minds and the low level features computable by machines. This paper proposes a multiple-instance learning based decision neural network (MI-BDNN) that attempts to bridge the semantic ga in CBIR. Multiple-instance learning (MIL) is a variation of supervised learning, where the training set is composed of many bags, and each bag contains many instances. If a bag contains at least one positive instance, it is labelled as a positive bag; otherwise, it is labelled as a negative bag. A novel discriminant function and learning schemes are employed in the MI-BDNN to learn the concept from the training bags. The proposed approach considers the image retrieval problem as a MIL problem, where a user's preferred image concept is learned by training MI-BDNN with a set of exemplar images, each of which is labelled as conceptual related (positive) or conceptual unrelated (negative) image. The MI-BDNN based CBIR system is developed, and the results of the experiments showed that MI-BDNN can successfully be used for real image retrieval and classification problems.

1. Introduction

Since the proliferation of mobile devices embedded high quality digital cameras is becoming the convergent platform for personal sensing, computing, and communication, digital content is undergoing an explosion in recent years. Image retrieval systems have recently taken on a new importance with its emerging applications to search, browse, and retrieve images from large content warehouses. To search for images, a user may provide query keyword, image file/link, or click on some images, and image retrieval systems will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc. Textual information about images can be easily searched using existing technology, but requires humans to manually describe every image in the database. This is impractical for a large-scale database, and possible to miss images that use different synonyms in their descriptions. Hence, during the last several decades, there has been a dramatic proliferation of research concerned with content-based image retrieval (CBIR) [1,2], where the search will analyze the actual image contents (e.g., colors, shapes, textures) rather than the metadata such as keywords, tags, and/or descriptions associated with the image.

The challenge of CBIR is how to narrow down the gap between the high level semantics in the human minds and the low level features computable by machines. There are two key issues in the challenge. The first is what similarity between two images means to the user. Since the "perceptual similarity" depends upon the application, the person, and the context of usage, the machine has to learn them on-line with user in the loop. This user-assisted learning is known as the relevance feedback learning [3]. The second is what portion of each image is important to the user. Because there is natural ambiguity as to what portion of each image is important to the user, CBIR can be considered as a Multiple-Instance Learning (MIL) problem [4].

MIL is first analyzed by Dietterich et al. [5]. They investigated the drug activity prediction problem, trying to predict that whether a new molecule was qualified to make some drug, through analyzing a collection of known molecules. In MIL problems, the training set is composed of many bags, each of which contains many instances. If a bag contains at least one positive instance, it is labelled as a positive bag; otherwise, it is labelled as a negative bag. The labels of the training bags are known, but that of the training instances are unknown. The task is to learn the concept from the training set for correctly labelling unseen bags. To solve the MIL problems, Dietterich proposed three

E-mail address: yyxu@sunrise.hk.edu.tw

http://dx.doi.org/10.1016/j.neucom.2015.07.024
0925-2312/

axis-parallel rectangle (APR) algorithms to search the appropriate axis-parallel rectangles constructed by the conjunction of the features extracted from molecules. After Dietterich et al., numerous MIL algorithms have been developed, such as Diverse Density [4], Bayesian-kNN and Citation-kNN algorithms [6], and EM-DD algorithm [7], and successfully applied to many applications [8–14]. In the MIL-based CBIR application, each image is labelled either a positive or negative. A positive image contains a set of regions (instances) in which at least one region is conceptual to the user. On the other hand, all of the regions in a negative image are not conceptual to the user. There has been many research exploring the application of Multiple-Instance Learning (MIL) techniques to CBIR [8,15,9]. More works on MIL can be found in [16], and a good survey of the MIL approaches is provided by [17].

The robustness, adaptation, and ability to automatically learn from examples make neural network approaches attractive and exciting for MIL. Ramon and De Raedt [18] presented a neural networks framework for MIL. Zhang and Zhou proposed a multi-instance neural network named BP-MIP [19,20], which extended the popular BP [21] algorithm with a global error function defined at the level of bags instead of at the level of instances. How to construct a neural model structure is crucial for successful recognition. All the above neural networks for MIL are based on the all-class-in-one-network (ACON) structure, where all the classes are lumped into one super-network. The supernet has the burden of having to simultaneously satisfy all the teachers, so the number of hidden units tends to be large.

Since it is well known that neural networks are very robust and adaptive, for the applications which undergo many variation factors such as CBIR, neural networks seem to be a good remedy to solve the problems. The motivation of this work is to construct an effective as well as efficient neural model structure for MIL to tackle CBIR. In this paper, a multiple-instance learning based decision neural network (MI-BDNN) is proposed, and an MI-BDNN based image retrieval system [22] is built and verified by using the SIVAL data set [15].

Compared to the other MIL-based CBIR approach [8,15,9], MI-BDNN has two properties as follows. First, based on the neural networks approach, MI-BDNN learns the underlying rules from the given collection of representative images and can be adapted from user feedbacks. In other words, the query results of MI-BDNN can be fine-tuned via the relevance feedback learning. Second, based on the one-class-in-one-network (OCON) structure [23], MI-BDNN devotes one of its subnets to the representation of each image concept to be recognized. This kind of structure will be easy to maintain. An addition or a deletion of one or more subnets is sufficient to take care of any change of image concepts. In contrast, the neural networks based on the ACON structure would have to involve a global updating. Two characteristics make the MI-BDNN suitable for content-based image retrieval and classification. These characteristics are summarized as follows:

- Inherited the modular structure from its predecessor in [24], the proposed MI-BDNN has the merits of both the OCON neural networks structure and statistical approaches. Pandya and Macy [25] compared the performance between the ACON and OCON structures, and observed that OCON model achieves better training and generalization accuracies. Besides, the discriminant function of the proposed MI-BDNN is in a form of probability density, which yields high accuracy rates compared to other approaches, as discussed in Section 4.
- The proposed MI-BDNN adopts a hybrid locally unsupervised (LU) and globally supervised (GS) learning rule. During the locally unsupervised (LU) learning phase, each subnet is trained individually, and no mutual information across the classes may be utilized. After the LU phase is completed, the training enters the globally supervised (GS) learning phase, where teacher information is introduced to reinforce or anti-reinforce the decision boundaries obtained during LU phase. The details of the LUGS learning rule are described in Section 2.2. Compared with its predecessor [24], in which each subnet is just trained individually, the proposed MI-BDNN with the LUGS learning schema has the advantages that it is able not only to approximate the decision region of each image class locally and efficiently, but also to fine-tune the decision boundaries between classes globally and precisely. This characteristic makes the proposed MI-BDNN more suitable for image classification, as discussed in Section 4.

It has to be noted that the locally unsupervised learning in the proposed approach is not purely unsupervised to some extent. In the MI-BDNN, the locally unsupervised learning in each subnet is to find hidden structure in "unlabeled" training data. Although the positive or negative labels given to each image in the MIL-based CBIR application could be referred as a kind of supervised information, we say that the locally learning is "unsupervised" since the structure labels are not given to each training example during the locally learning phase.

The reminder of this paper is organized as follows. In the next section, the proposed MI-BDNN and its learning scheme are introduced. Then, in Section 3, the image representation for MIL is described. Experimental results are presented and discussed in Section 4. Finally, Section 5 draws the conclusions.

2. Multiple-instance learning based decision neural networks

The MI-BDNN, following the MIL constraint, is proposed to model concepts. The structure of the MI-BDNN for an m concepts MIL problem is depicted in Fig. 1, where the MI-BDNN consists of m subnets, and each subnet is designated to represent one concept. In order to model m concepts, the discriminant functions in all subnets are designed to capture the nature of MIL, and will be trained by the LUGS two-phase learning [26]: the locally unsupervised (LU) learning and the globally supervised (GS) learning. During the LU phase, each subnet is trained individually

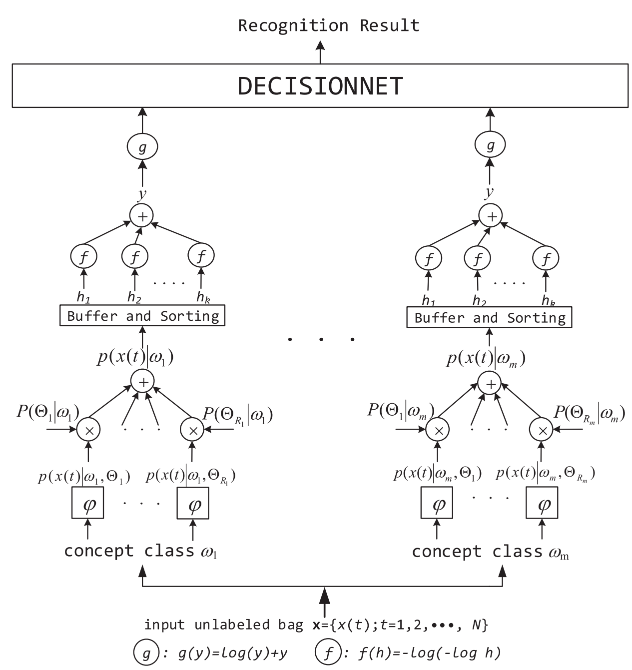

Fig. 1. Structure of MI-BDNN. Each subnet is designated to recognize one concept.