Pattern Recognition 47 (2014) 3106–3119

Contents lists available at ScienceDirect

Pattern Recognition

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

Active learning with multi-criteria decision making systems

Ran Wang [a,b], Sam Kwong [a,b,*]

[a] Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong
[b] City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, China

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form
26 September 2013
Accepted 18 March 2014
Available online 27 March 2014

Keywords:
Active learning
Multiple-instance learning
Multi-criteria decision making
Support vector machine

ABSTRACT

In active learning, the learner is required to measure the importance of unlabeled samples in a large dataset and select the best one iteratively. This sample selection process could be treated as a decision making problem, which evaluates, ranks, and makes choices from a finite set of alternatives. In many decision making problems, it usually applied multiple criteria since the performance is better than using a single criterion. Motivated by these facts, an active learning model based on multi-criteria decision making (MCMD) is proposed in this paper. After the investigation between any two unlabeled samples, a preference preorder is determined for each criterion. The dominated index and the dominating index are then defined and calculated to evaluate the informativeness of unlabeled samples, which provide an effective metric measure for sample selection. On the other hand, under multiple-instance learning (MIL) environment, the instances/samples are grouped into bags, a bag is negative only if all of its instances are negative, and is positive otherwise. Multiple-instance active learning (MIAL) aims to select and label the most informative bags from numerous unlabeled ones, and learn a MIL classifier for accurately predicting unseen bags by requesting as few labels as possible. It adopts a MIL algorithm as the base classifier, and follows an active learning procedure. In order to achieve a balance between learning efficiency and generalization capability, the proposed active learning model is restricted to a specific algorithm under MIL environment. Experimental results demonstrate the effectiveness of the proposed method.

1. Introduction

In classification, the use of active learning technique [11,2] is an effective method of selecting a small number of training samples from a large unlabeled dataset. It iteratively measures the informativeness of unlabeled samples, selects the one with the highest preference level, makes query on its label, and updates the training set. The goal is to learn an accurate model by making as few queries as possible. Applying active learning to classification can efficiently reduce the effort for manual labeling and data redundancy.

Currently, most existing active learning strategies are based on single criterion. However, it is commented that the integration of multiple criteria is likely to give better performance than each single one. In [44], a multi-criteria based learning strategy is proposed for named entity recognition, which measures the informativeness of unlabeled samples by the weighted-sum of normalized uncertainty and diversity. This measurement tries to maximize the contribution of the selected samples. However, since the ranges of the criteria are quite different, the weighted-sum of their normalized values tends to weaken some important information when the weights are not well set, and may fail to reflect the sample's priority with respect to each criterion.

Multi-criteria decision making (MCDM) [56] is a well known branch of decision making, which evaluates a finite set of alternatives on the basis of two or more criteria. MCDM tasks include choosing the best alternative from the set of candidates, or sorting the alternatives into a preference preorder. Under some assumptions, the decision maker is required to maximize an utility function [50], or indicate the priority between any two alternatives with respect to each criterion. Due to the generality and universality, MCDM has been applied to various areas such as operational research [40,53,27], evolutionary multi-objective optimization [19,5,45,14], knowledge discovery [8], and preference modeling [18]. It is noteworthy that the main task in active learning is to measure the preference levels of unlabeled samples, which might be well handled by a MCDM system. In this paper, an active learning model with MCDM system is proposed. By investigating the priority between any two unlabeled samples, a preference preorder is determined for each criterion. The dominated indices and dominating indices of a sample in the preference preorders are then defined and computed to reflect its informativeness. Since the preference preorders are evaluated

* Corresponding author.
E-mail address: CSSAMK@cityu.edu.hk (S. Kwong).

http://dx.doi.org/10.1016/j.patcog.2014.03.011
0031-3203/

uniformly and independently, this strategy is likely to provide accurate priority information.

On the other hand, multiple instance learning (MIL) is a task that aims to train classification models from structured data. In MIL, the samples are grouped into bags, and each sample is taken as an instance of its bag. A bag is positive if at least one of its instances is positive, and is negative only if all of its instances are negative [36,37]. MIL was first introduced to drug activity prediction [15], and then applied to real-world machine learning scenarios such as content-based image retrieval (CBIR) [10] and document classification [43]. In CBIR and document classification, the images or documents are taken as the bags, and the segmented regions or short passages are their instances. The key step in MIL is to identify the instances that are responsible for the final decisions in each bag [9,42,29].

Many successful algorithms for training MIL classifiers have been developed [4,16,54], however, the problem of how to collect sufficient training bags has not been properly addressed. One solution to this problem is applying multiple-instance active learning (MIAL), which is able to select and label the most informative bags from numerous unlabeled ones. Currently, MIAL has been realized with three types of strategies: (1) instance-level learning; (2) bag-level learning; and (3) mixed mode. The distinction among them is lying in the query objects, which could be instances, bags, and their combinations. By analyzing the heuristic optimization of MIL algorithms, it is observed that applying MCDM system to MIAL is likely to achieve a good trade-off between the learning efficiency and the generalization capability. Thus, the proposed active learning model is further restricted to a specific algorithm under MIL environment.

The rest of this paper is organized as follows: in Section 2, background knowledge and motivation of this work are given. In Section 3, the active learning model with MCDM system is proposed. In Section 4, the proposed learning model is refined to a specific algorithm under MIL environment. In Section 5, experimental comparisons are conducted to show the feasibility and effectiveness of the proposed method. Finally, conclusions and future work directions are given in Section 6.

2. Backgrounds and motivation

2.1. Basic definitions in MCDM

Generally, a MCDM system consists of two phases [53]: (1) information input and construction; (2) aggregation and exploitation. In the first phase, a typical MCDM problem [24,6] is mathematically modeled as a decision matrix defined in Definition 1.

Definition 1 (*decision matrix*). Given a MCDM problem with n distinct alternatives $\mathbf{A}_1, \dots, \mathbf{A}_n$ and m criteria $Cr_1, \dots, Cr_m$, where the level of achievement of $\mathbf{A}_i$ $(i = 1, \dots, n)$ with regard to Cr_k $(k = 1, \dots, m)$ is denoted by $Cr_k(\mathbf{A}_i)$, then

$$\begin{bmatrix} Cr_1(\mathbf{A}_1) & Cr_2(\mathbf{A}_1) & \dots & Cr_m(\mathbf{A}_1) \\ Cr_1(\mathbf{A}_2) & Cr_2(\mathbf{A}_2) & \dots & Cr_m(\mathbf{A}_2) \\ \vdots & \vdots & \ddots & \vdots \\ Cr_1(\mathbf{A}_n) & Cr_2(\mathbf{A}_n) & \dots & Cr_m(\mathbf{A}_n) \end{bmatrix}$$

is called the decision matric of this problem.

In the second phase, the aggregation and the exploitation are carried out on the basis of the decision matrix constructed in the first phase, as well as the decision maker's willingness. Generally, $Cr_k(\mathbf{A}_i)$ could be either cardinal [38] or ordinal [13,12,17]. If the decision matrix is composed of cardinal numbers, the system is required to maximize an utility function $u(\mathbf{A}_i) = \sum_{k=1}^{m} w_k Cr_k(\mathbf{A}_i)$, where w_k is the weight of Cr_k. While for preference modeling problems, the decision matrix is usually composed of ordinal numbers. In this case, the system is required to indicate that, regarding each criterion Cr_k, which of the following four relations holds for any two distinct alternatives $\mathbf{A}_i, \mathbf{A}_j$:

1. $\mathbf{A}_i \succ \mathbf{A}_j$: $\mathbf{A}_i$ is preferred to $\mathbf{A}_j$;
2. $\mathbf{A}_i \prec \mathbf{A}_j$: $\mathbf{A}_j$ is preferred to $\mathbf{A}_i$;
3. $\mathbf{A}_i ? \mathbf{A}_j$: $\mathbf{A}_i$ is incomparable to $\mathbf{A}_j$;
4. $\mathbf{A}_i \approx \mathbf{A}_j$: $\mathbf{A}_i$ is indifferent to $\mathbf{A}_j$.

Holding these relations, all the alternatives could be arranged into some preference preorders as defined in Definition 2.

Definition 2 (*preference preorder*). Given a MCDM problem with n distinct alternatives $\mathbf{A}_1, \dots, \mathbf{A}_n$ and m criteria $Cr_1, \dots, Cr_m$, where the level of achievement of $\mathbf{A}_i$ $(i = 1, \dots, n)$ with regard to Cr_k $(k = 1, \dots, m)$ is denoted by $Cr_k(\mathbf{A}_i)$. If for Cr_k there exists $Cr_k(\mathbf{A}_1^*) \succ \approx Cr_k(\mathbf{A}_2^*) \succ \approx \dots \succ \approx Cr_k(\mathbf{A}_n^*)$, where $\mathbf{A}_1^* \neq \mathbf{A}_2^* \neq \dots \neq \mathbf{A}_n^* \in \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n\}$, then the order $\mathbf{A}_1^*, \mathbf{A}_2^*, \dots, \mathbf{A}_n^*$ is called a preference preorder of $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$ with regard to Cr_k, denoted by $\mathcal{P}_k$.

It is obvious that the anterior alternative in $\mathcal{P}_k$ gets more priority and importance with respect to Cr_k. Based on the preference preorders, the following definitions could be further introduced.

Definition 3 (*k-th dominated index and k-th dominating index*). Given that $\mathcal{P}_k$ is a preference preorder of alternatives $\mathbf{A}_1, \dots, \mathbf{A}_n$ with regard to Cr_k $(k = 1, \dots, m)$, then the k-th dominated index and the k-th dominating index of $\mathbf{A}_i$ $(i = 1, \dots, n)$, denoted by $\psi_k^{\succ}(\mathbf{A}_i)$ and $\psi_k^{\prec}(\mathbf{A}_i)$, are respectively defined as

$$\psi_k^{\succ}(\mathbf{A}_i) = \sum_{j = 1, \dots, n, j \neq i} d(\succ, R_{ij}^{(k)}), \tag{1}$$

and

$$\psi_k^{\prec}(\mathbf{A}_i) = \sum_{j = 1, \dots, n, j \neq i} d(\prec, R_{ij}^{(k)}). \tag{2}$$

where $d(\cdot, \cdot)$ is a distance metric, $R_{ij}^{(k)} \in \{\succ, \prec, \approx, ?\}$ is the relation of $\mathbf{A}_i$ and $\mathbf{A}_j$ with regard to Cr_k.

Definition 4 (*dominated index and dominating index*). Given that $\mathcal{P}_1, \dots, \mathcal{P}_m$ are the preference preorders of alternatives $\mathbf{A}_1, \dots, \mathbf{A}_n$ with regard to $Cr_1, \dots, Cr_m$, then the dominated index and the dominating index of $\mathbf{A}_i$ $(i = 1, \dots, n)$, denoted by $\psi^{\succ}(\mathbf{A}_i)$ and $\psi^{\prec}(\mathbf{A}_i)$, are respectively defined as

$$\phi^{\succ}(\mathbf{A}_i) = \sum_{k=1}^{m} w_k \psi_k^{\succ}(\mathbf{A}_i), \tag{3}$$

and

$$\phi^{\prec}(\mathbf{A}_i) = \sum_{k=1}^{m} w_k \psi_k^{\prec}(\mathbf{A}_i), \tag{4}$$

where w_k is the weight of Cr_k.

The dominated index and the dominating index of $\mathbf{A}_i$ respectively reflect the degrees of $\mathbf{A}_i$ being dominated by others and dominating others in $\mathcal{P}_1, \dots, \mathcal{P}_m$. Obviously, in a MCDM problem, an alternative with lower dominated index and higher dominating index is preferred.

2.2. Active learning

In active learning, the learner is required to measure the preference levels of unlabeled samples and select the best one from a large dataset iteratively, which is supposed to be well