Active Learning with SVM

Jun Jiang
City University of Hong Kong, Hong Kong

Horace H. S. Ip
City University of Hong Kong, Hong Kong

INTRODUCTION

With the increasing demand of multimedia information retrieval, such as image and video retrieval from the Web, there is a need to find ways to train a classifier when the training dataset is combined with a small number of labelled data and a large number of unlabeled one. Traditional supervised or unsupervised learning methods are not suited to solving such problems particularly when the problem is associated with data in a high-dimension space. In recent years, many methods have been proposed that can be broadly divided into two groups: **semi-supervised** and **active learning** (AL). Support Vector Machine (SVM) has been recognized as an efficient tool to deal with high-dimensionality problems, a number of researchers have proposed algorithms of Active Learning with SVM (ALSVM) since the turn of the Century. Considering their rapid development, we review, in this chapter, the state-of-the-art of ALSVM for solving classification problems.

BACKGROUND

The general framework of AL can be described as in Figure 1. It can be seen clearly that its name – **active learning** – comes from the fact that the learner can improve the classifier by actively choosing the “optimal” data from the potential query set *Q* and adding it into the current labeled training set *L* after getting its label during the processes. The key point of AL is its sample selection criteria.

AL in the past was mainly used together with neural network algorithm and other learning algorithms. Statistical AL is one classical method, in which the sample minimizing either the variance (D. A. Cohn, Ghahramani, & Jordan, 1996), bias (D. A. Cohn, 1997) or generalisation error (Roy & McCallum, 2001) is queried to the oracle. Although these methods have strong theoretical foundation, there are two common problems limiting their application: one is how to estimate the posterior distribution of the samples, and the other is its prohibitively high computation cost. To deal with the above two problems, a series of **version space based AL** methods, which are based on the assumption that the target function can be perfectly expressed by one hypothesis in the version space and in which the sample that can reduce the volume of the version space is chosen, have been proposed. Examples are query by committee (Freund, Seung, Shamir, & Tishby, 1997), and SG AL (D. Cohn, Atlas, & Ladner, 1994). However the complexity of version space made them intractable until the version space based ALSVMs have emerged.

The success of SVM in the 90s has prompted researchers to combine AL with SVM to deal with the semi-supervised learning problems, such as distance-based (Tong & Koller, 2001), RETIN (Gosselin & Cord, 2004) and Multi-view (Cheng & Wang, 2007) based ALSVMs. In the following sections, we summarize existing well-known ALSVMs under the framework of **version space theory**, and then briefly describe some mixed strategies. Lastly, we will discuss the research trends for ALSVM and give conclusions for the chapter.

VERSION SPACE BASED ACTIVE LEARNING WITH SVM

The idea of almost all existing heuristic ALSVMs is explicitly or implicitly to find the sample which can reduce the volume of the **version space**. In this section, we first introduce their theoretical foundation and then review some typical ALSVMs.

Figure 1. Framework of active learning

```
Initialize Step: An classifier h is trained on the initial labeled training set L
step 1:    The learner evaluates each data x in potential query set Q (subset of or whole
           unlabeled data set U) and query the sample x* which has lowest EvalFun(x, L,
           h, H) to the oracle and get its label y*;
step 2:    The learner update the classifier h with the enlarged training set {L + ( x*,
           y*)};
step 3:    Repeat step 1 and 2 until stopping training;

Where
      ➢   EvalFun(x, L, h, H): the function of evaluating potential query x (the lowest
          value is the best here)
      ➢   L: the current labeled training set
      ➢   H: the hypothesis space
```

Version Space Theory

Based on the Probability Approximation Correct learning model, the goal of machine learning is to find a consistent classifier which has the lowest generalization error bound. The Gibbs generalization error bound (McAllester, 1998) is defined as

$$\varepsilon_{Gibbs}(m, P_H, z, \delta) = \frac{1}{m}\left(\ln\left(\frac{1}{P_H(V(z))}\right)\right) + \ln\left(\frac{em^2}{\delta}\right)$$

where P_H denotes a prior distribution over hypothesis space H, $V(z)$ denotes the version space of the training set z, m is the number of z and δ is a constant in [0, 1]. It follows that the generalization error bound of the consistent classifiers is controlled by the volume of the version space if the distribution of the version space is uniform. This provides a theoretical justification for version space based ALSVMs.

Query by Committee with SVM

This algorithm was proposed by (Freund et al., 1997) in which $2k$ classifiers were randomly sampled and the sample on which these classifiers have maximal disagreement can approximately halve the **version space** and then will be queried to the oracle. However, the complexity of the structure of the version space leads to the difficulty of random sampling within it. (Warmuth, Ratsch, Mathieson, Liao, & Lemmem, 2003) successfully applied the algorithm of playing billiard to randomly sample the classifiers in the SVM version space and the experiments showed that its performance was comparable to the performance of **standard distance-based ALSVM** (SD-ALSVM) which will be introduced later. The deficiency is that the processes are time-consuming.

Standard Distance Based Active Learning with SVM

For SVM, the **version space** can be defined as:

$$V = \left\{w \in W \mid \|w\| = 1,\ \ y_i(w \bullet \Phi(x_i)\ ) > 0,\ \ i = 1,...,m\right\}$$

where $\Phi(.)$ denotes the function which map the original input space X into a high-dimensional space $\Phi(X)$, and W denotes the parameter space. SVM has two properties which lead to its tractability with AL. The first is its duality property that each point w in V corresponds to one hyperplane in $\Phi(X)$ which divides $\Phi(X)$ into two parts and vice versa. The other property is that the solution of SVM w^* is the center of the **version space** when the version space is symmetric or near to its center when it is asymmetric.

Based on the above two properties, (Tong & Koller, 2001) inferred a lemma that the sample nearest to the

Related Content

Modeling Malaria with Multi-Agent Systems

Fatima Rateb, Bernard Pavard, Narjes Bellamine-BenSaoud, J.J. Mereloand M.G. Arenas (2005). *International Journal of Intelligent Information Technologies (pp. 17-27).*

www.irma-international.org/article/modeling-malaria-multi-agent-systems/2381

Artificial Immune Systems for Anomaly Detection in Ambient Assisted Living Applications

Sebastian Bersch, Djamel Azzi, Rinat Khusainovand Ifeyinwa E. Achumba (2013). *International Journal of Ambient Computing and Intelligence (pp. 1-15).*

www.irma-international.org/article/artificial-immune-systems-for-anomaly-detection-in-ambient-assisted-living-applications/101949

Customer Segmentation Using K-Means Algorithm

Debabrata Datta, Anal Acharya, Kwanan Mondal, Meghna Mondaland Tanushree Sarkar (2025). *Navigating Organizational Behavior in the Digital Age With AI (pp. 213-250).*

www.irma-international.org/chapter/customer-segmentation-using-k-means-algorithm/364231

Adoption of Virtual Reality and Augmented Reality in the Hotel Industry

Abhinav Kumar Shandilyaand Dilip Kumar (2024). *Hotel and Travel Management in the AI Era (pp. 1-18).*

www.irma-international.org/chapter/adoption-of-virtual-reality-and-augmented-reality-in-the-hotel-industry/356240

Use of Fuzzy Set Theory in DNA Sequence Comparison and Amino Acid Classification

Subhram Das, Soumen Ghosh, Jayanta Paland Dilip K. Bhattacharya (2017). *Emerging Research on Applied Fuzzy Sets and Intuitionistic Fuzzy Matrices (pp. 235-253).*

www.irma-international.org/chapter/use-of-fuzzy-set-theory-in-dna-sequence-comparison-and-amino-acid-classification/171908