Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence

Dynamic Batch Mode Active Learning via L1 Regularization

Shayok Chakraborty, Vineeth Balasubramanian and Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC)
Arizona State University
schakr10@asu.edu, vineeth.nb@asu.edu, panch@asu.edu

Abstract

We propose a method for dynamic batch mode active learning where the batch size and selection criteria are integrated into a single formulation.

Introduction

The generation of humongous amounts of data in today's digital world together with the scarcity of human labor associated with labeling the individual data points, has spurred research in the field of *active learning*. Active learning algorithms automatically select the exemplar data instances from an unlabeled set and thereby reduce human annotation effort in training a classifier. Conventional methods of active learning have focused on the *pool-based* strategy where the classifier selects one point at a time from an unlabeled pool and is updated after the selection of every new point (Tong and Koller 2000; Cohn, Ghahramani, and Jordan 1996). This may be inefficient due to the time involved in frequently retraining the classifier; also this scheme is unable to exploit the possible presence of multiple labeling agents, as it selects a single point in every iteration. To overcome these limitations, batch mode active learning (BMAL) algorithms, which attempt to select a batch of points simultaneously from an unlabeled set, have been proposed in recent years (Guo and Schuurmans 2008; Hoi et al. 2006; Hoi, Jin, and Lyu 2009). Such techniques are of paramount importance in applications based on video data due to the inherent redundancy among the captured frames. Due to its wide usage, we focus on face based biometric recognition as the exemplar application and explain our dynamic batch mode active learning framework.

An ideal BMAL system can be conceptualized as consisting of two main components: (i) to decide the batch size (number of data points to be queried from a given unlabeled set of points) and (ii) to select the most appropriate data points from the unlabeled pool once the batch size has been determined. Both these steps are critical in ensuring maximum generalization capability of the learner with minimum human labeling effort, which is the primary objective in any active learning application. However, most of the existing efforts on batch mode active learning are based on greedy heuristics, where a batch of points is selected to maximize a heuristic score function. Also, all the existing methods of batch mode active learning require the batch size to be supplied as an input to the algorithm. In a biometric recognition application, it is impractical to decide on a number without any knowledge of the data stream in question. The batch size should depend on the quality and variability of the images in the unlabeled stream and also on the level of confidence of the current classifier on the unlabeled images.

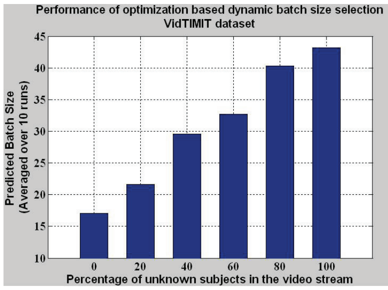
In this work, we propose a numerical optimization based strategy to simultaneously decide the batch size as well as identify the informative points to be selected for manual annotation, through a single framework. Our method has the same complexity as the state-of-the-art static BMAL technique, where the batch size is pre-specified by the user.

Mathematical Formulation

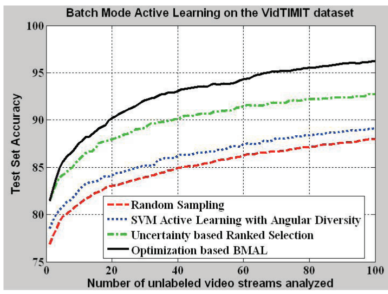
Consider a BMAL problem which has a current labeled set L_t and a current classifier w^t trained on L_t. The classifier is exposed to an unlabeled video U_t at time t. The objective is to select a batch B from the unlabeled stream in such a way that the classifier w^{t+1}, at time $t+1$, trained on $L_t \cup B$ has maximum generalization capability. An efficient strategy to ensure this condition is to minimize the entropy of the updated learner on the remaining $|U_t - B|$ images. Considering the specific challenges of face based biometric data, it may also be useful to select points from the sparsely populated regions of the unlabeled set. These points lie away from the main stream of points and may furnish valuable information (e.g. an informative visage made momentarily by a subject). Thus, a logical strategy to guide the point selection process, is to select a batch of points so as to maximize the following score function:

$$f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) \qquad (1)$$

The first term denotes the distance of every selected point from other points in the unlabeled set while the second term quantifies the entropy (uncertainty) of the updated model on the unselected images. It is intuitive that the value of this objective improves as more data instances are selected for manual labeling. Since the batch size m is unknown and flexible, the obvious solution to this problem is to select *all the images* in the data stream for manual annotation. However, this



(a) Dynamic Batch Size Selection



(b) Comparison of BMAL techniques

Figure 1: Dynamic Batch Mode Active Learning on the VidTIMIT dataset.

incurs huge labeling effort and defeats the basic purpose of active learning. To prevent this, we modify the objective by imposing a penalty on the batch size as follows:

$$\tilde{f}(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m \quad (2)$$

The penalty term ensures that only the relevant, informative images (for which the density and entropy terms outweigh the penalty term) get selected in the batch.

We define a binary vector M with $|U_t|$ entries, where the entry M_i denotes whether the unlabeled point x_i will be included in the batch ($M_i = 1$) or not ($M_i = 0$). With this definition, the batch size m equals the zero-norm of the vector M. Replacing the zero-norm by its tightest convex approximation, the one norm, we derive an equivalent optimization problem in terms of the L1 regularization:

$$\max_{M} \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 \sum_j M_j \quad (3)$$

subject to the constraint: $M_j \in [0, 1]$. We relax the constraint and solve for a local optimum using the Quasi Newton method. Thus, solving for a single vector M enables us in identifying the batch size m ($=\sum_j M_j$) and the data points to be selected for manual annotation (by setting the top m entries in M as 1 to greedily recover the integer solution). A similar dynamic batch selection strategy can be developed with any other objective function deemed suitable for a given application.

Experiments and Results

For the sake of brevity, we present results obtained only on the VidTIMIT face dataset, which represents challenging real world conditions.

Experiment 1: Dynamic Batch Size Selection

25 subjects were selected and randomly partitioned into a "known" group with 20 subjects and an "unknown" group with the remaining 5 subjects. The base classifier was trained on the known subjects. Unlabeled video streams (each containing 100 frames) were then presented to the learner for batch selection and the proportions of unknown subjects in the video streams were gradually increased from 0% to 100% in steps of 20%. The batch sizes decided by the learner were noted for each of the video streams. The results are shown in Figure 1(a) and corroborate the fact that, with increasing proportions of unknown subjects, the learner decides on a larger batch size. Thus, the framework enables the active learner to automatically adjust itself to the complexity level of the data through the chosen batch size.

Experiment 2: Active Learning Performance

The proposed BMAL scheme was compared against the heuristic BMAL techniques. Here, unlabeled video streams were presented to the learner and a batch of points was queried using the different techniques. The selected points were then appended to the training set and tested on a pre-determined test set, containing about 5000 images, spanning all the 25 subjects. The purpose was to study the growth in accuracy of the learner on the same test set with increasing size of the training set. To facilitate fair comparison, the batch size for each video stream was decided using the framework and the same number was supplied as an input to each of the other strategies. The results are shown in Figure 1(b). The graphs depict that the proposed scheme succeeds in selecting the salient and prototypical data points from unlabeled sets and attains a given level of generalization accuracy with the least number of labeled examples.

Discussion

The proposed algorithm is flexible and it is easy to modify the objective in situations where multiple sources of information (e.g. audio and video) or contextual information are available. Our future work will mainly include intelligent selection of the weight parameters λ_1 and λ_2 and handling scaling issues of the proposed algorithm.

References

Cohn, D.; Ghahramani, Z.; and Jordan, M. 1996. Active learning with statistical models. *JAIR*.

Guo, Y., and Schuurmans, D. 2008. Discriminative batch mode active learning. In *NIPS*.

Hoi, S. C. H.; Jin, R.; Zhu, J.; and Lyu, M. R. 2006. Batch mode active learning and its application to medical image classification. In *ICML*.

Hoi, S.; Jin, R.; and Lyu, M. 2009. Batch mode active learning with applications to text categorization and image retrieval. *IEEE TKDE*.

Tong, S., and Koller, D. 2000. Support vector machine active learning with applications to text classification. *JMLR*.