

Case ID:M11-109P

Published: 7/18/2011

Inventors

Shayok Chakraborty

Vineeth Nallure

Balasubramanian

Sethuraman Panchanathan

Contact

Shen Yan
shen.yan@skysonginnovations.
com

Adaptive Batch Mode Active Learning for Evolving a Classifier

Due to the tremendous increase in the amount of digital data, effective large scale data classification is playing an increasingly important role. In active learning algorithms a ?classifier? is used to classify the unlabeled data. To ensure reliable performance of the classifier, it must be trained using a number of labeled examples, or the classifier?s ?training set.? Such systems often rely on humans to manually label the training set. It is impractical for human beings to hand-label large datasets, so in order to optimize the labeling effort associated with training data classifiers, active learning algorithms have been implemented which select only the promising and exemplar instances for manual labeling. Current methods utilize the pool-based strategy which only labels a single datum at a time after which the classifier is retrained, but this is time consuming and inefficient.

Researchers at Arizona State University have developed a new technology that incorporates batch mode active learning systems. This method selects a batch of unlabeled data points simultaneously from a given body of unlabeled data as opposed to the pool based method which selects only one at a time. The classifier is retrained once after every batch of data points is selected and labeled. The selection of multiple instances facilitates parallel labeling increasing efficiency and productivity. The proposed technology improves on current models by simultaneously solving for both the batch size as well as the specific data batch to be classified. The batch size and data are determined based on projected improvements in the classifier?s efficiency in classifying unlabeled data.

Potential Applications

- Video analytics
- Face recognition software
- Video processing
- Medical imaging
- Video surveillance and security
- Gaming establishments

Benefits and Advantages

- Utilizes batch mode active learning
- Dynamically adapts to the complexity of data stream
- Intelligently chooses the optimal batch size
- Specific batch chosen to increase classifier efficiency

