International Journal of Research Studies in Science, Engineering and Technology
Volume 2, Issue 1, January 2015, PP 72-77
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Multilabel Image Classifier Using Active Learning

Sana Altaf Sayyed

MET's Institute of Engineering Bhujbal Knowledge City
M.E Computer.
sayyedsana53@gmail.com

Abstract: *Most active learning approaches select informative or representative unlabeled instances to query their labels for classification. Although several active learning algorithms have been proposed to combine the two criteria for query selection uses ad hoc in finding unlabeled instances that are both informative and representative. Informativeness measures the ability of an instance in reducing the uncertainty of a statistical model, where as representativeness measures if an instance well represents the overall input patterns of unlabeled data. The supervised machine learning techniques is applied to multilabel image classification problems. supervised learning, within the available data repository, only part of the data are labeled and utilized for training performances heavily rely on the quality of training images. The supervised learning techniques having hinders to large scale problems. High-order label correlation driven active learning is motivated by the virtue of leveraging label correlations to improve multi-label classification A high-order label correlation driven active learning approach that uses the iterative learning algorithm to choose the informative example-label pairs from which it learns so as to learn an accurate classifier with less annotation efforts.*

Keywords: *Active learning, multilabel classification, high-order label correlation.*

1. Introduction

The supervised learning techniques to image classification problems is that it is required large amount of labeled training images. The unlabeled images are easily available, where as annotation is expensive or time consuming. Active learning is worked in an iterative fashion. takes traditional binary myopic active learning as an example.

Fig1. *Internet images with multi-label characteristic. Pair-wise and higher order label correlations are manifest, and crucial to efficient image classification.*

The example is selected for each iteration with the highest informativeness score for annotation, while the classifier is retrained on the training dataset with new labeled example. One difficulty of binary myopic active learning is that at each iteration only one example is selected for annotation .Batch mode active learning is used to overcome the drawback of active learning .Binary classification mostly focus on an example is only associated with one label. Every example has multiple labels, thus the active learner has to select not only examples but also their labels for annotation. Fig. 1 shows typical images from Internet with beach theme is an example of real-world applications, such Internet image classification and retrieval, usually having multi-label characteristic.

In multi-label batch mode active learning is considered the factors proposed in high order active learning applicable to real-world scale image classification and retrieval, *e.g.* Internet image search engine. As Fig. 2 shows, images are crawled and sent to the engine (step 1), and active learning algorithm selects the most informative example-label pairs for human experts to annotate (step 2, 3, 4 5 and 6). High order active learning trains a classifier on the labeled training data (step 7). If the learned classifier can accurately classify images, the active learning is stopped. Otherwise, the active learning conducts another learning iteration from (step 2-7). Such iterative learning process continues until the obtained classifier meets the desired criteria. In case of real time images the user gives label to the images for that there is no need for annotation images are directly give to the active learning algorithm. The annotation should be accurate for the labeling on the basis of labels we are classifying the images.

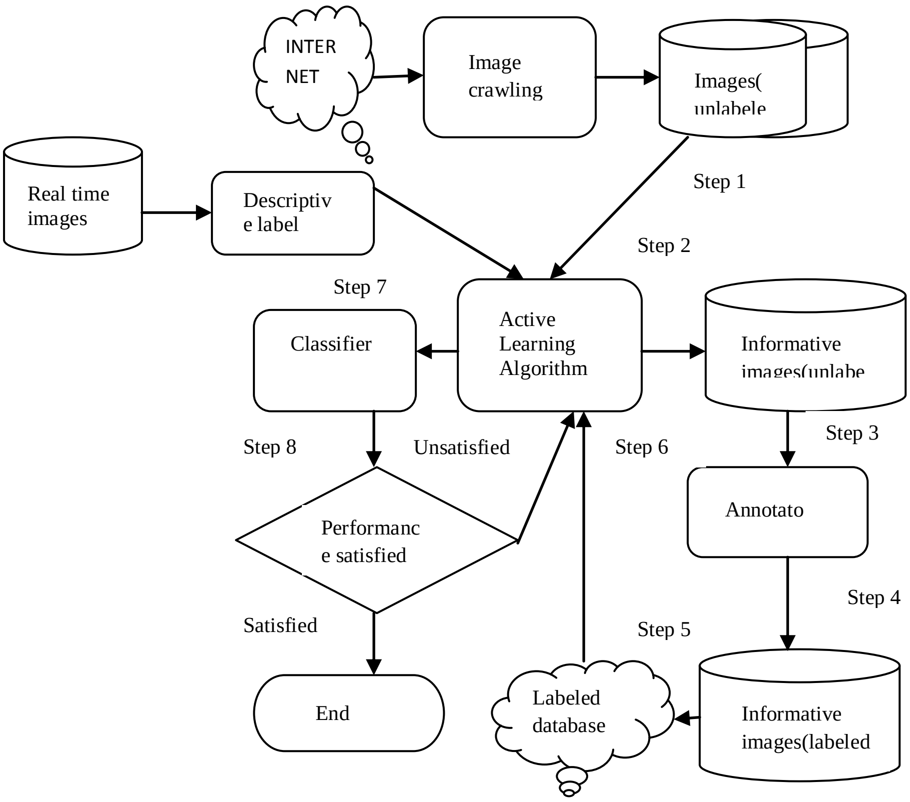

Fig. 2. *Framework for applying the proposed high order active learning real-world scale semantic image classification and retrieval, e.g. Internet image search engine.*

The proposed method is developed with the following consideration-

1. Images are collected and sent to the engine and active learning algorithm selects the most informative example-label pairs for human experts to annotate.
2. A Score function is defined to measure the informativeness of example-label pairs to measure the informativeness of example-label pairs.
3. Labels are usually dependent, and their inherent correlations are used for analyzing unknown labels from known labels we define cross-label uncertainty which gauges the disagreement between the mined label correlation and the label co-occurrence possibility.

2. LITERATURE SURVEY

The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data from which it learns. A large portion of the existing active learning techniques are designed for myopic active learning by C. Campbell and S. Tong [2][3]. Active learning has been extensively studied for a number of years, and researchers addressed it in a variety of ways including methods based on uncertainty sampling [2]. A large portion of the existing active learning techniques are designed for myopic active learning at each learning iteration only one example is selected for annotation, and the classifier is updated every time when a new annotated example becomes available discuss by B. Settles . In order to overcome the drawbacks of the myopic active learning, batch mode active learning is proposed and has attracted increasing attentions mentioned by K. Brinker [8]. K. Brinker performs batch mode selection by considering both the diversity and informativeness[4] . Z. Xu, K. Yu and V. Tresp also takes into account the diversity of the selected examples by querying cluster centroids that are close to the decision boundary[7]. S. C. Hoi and R. Jin have incorporated Fisher information to batch mode active learning for binary logistic regression. They also have tackled batch mode active learning under the semi-supervised learning setting[8].

For multi-label active learning, G. Qi, X. Hua, and Y. Rui, J. have utilized mutual information to measure the correlation between labels to achieve efficient learning. Y. Zhang has extended the value of Information framework to take into account label correlations for myopic active learning. For multi-label classification problems, One-vs-one or one-vs-all strategies methods are adopt to convert the original problem into a set of binary problems. Y. Guo, D. Schuurmans and S. C. Hoi tackle batch mode active learning under the semi-supervised learning setting[5][6].For multi-label active learning, G. Qi, X. Hua utilize mutual information to measure the correlation between labels to achieve efficient learning [9].

3. MULTI-LABEL BATCH MODE ACTIVE LEARNING

3.1 Problem Formulation

The multilable image classification proposed by considering the below notations .We use X to denote the feature space of examples. And also assume there is a label set Θ containing K different labels. The labels associated with an example $x \in X$ form a subset of Θ, which can be represented as a K-dimensional binary vector $Y = \{y_1, \ldots, y_k\}$, with 1 indicating that the example belongs to the corresponding concept and -1 otherwise . The current annotated labels of x can form a labeled example-label pair set $LP(x) = \{(x, y_i)|y_i$ is labeled$\}$, and the rest labels can form an unlabeled example-label pair set $UP(x) = \{(x, y_i)|y_i$ is unlabeled$\}$.Initially, the active learning algorithm is given with a small number of annotated example-label pairs, $L^0 = \{LP(x_1), \ldots, LP(x_N)\}$ and a large number of unlabeled example-label pairs, $U^0 = \{UP(x_1), \ldots, UP(x_N)\}$ [1].

Initial prediction models $P(y_j|x,\mathrm{w}^0_j)$, $1 \leq j \leq K$ can be obtained based on L^0 and U^0.w^0_j is given as the model parameter vector .For each learning iteration t, m unlabeled example-label pairs batch are considered as $S^t \subseteq U^{t-1}$ are selected for annotation. Let m be the predefined batch selection size. And the example-label pair sets are then updated as and g: $U^t = U^{t-1} - S^t$ and $L^t = L^{t-1} \cup S^t$. The updated prediction modes $P(y_j|x,\mathrm{w}^t_j)$ can be obtained on L^t and U^t . This process repeats upto the stop criterion is achieved. The goal is to search for the optimal selection St which leads to the best prediction models $P(y_j|x, \mathrm{w}^t_j)$ at each learning iteration of example.

3.2 Informative Example-Label Pairs Selection

For semi-supervised learning, unlabeled data are also considered for training set. Classifier is learned by simultaneously maximizing the likelihood of the labeled data and minimizing the label uncertainty of the unlabeled data. The objective function for this can be represented as below:

$$\sum_{i \in L} \mathrm{L}(\mathbf{x_i},\mathbf{y_j},\mathbf{w}) - \alpha \sum_{j \in U} \mathrm{UC}(\mathbf{x_j},\mathbf{w}) \tag{1}$$

In this α is a trade-off parameter for adjusting the relative influence of the labeled and unlabeled data. L and UC indicate likelihood and uncertainty functions respectively. A score function measuring the informativeness of selected examples can be defined as:

$$f(S) = \sum_{i \in L^{t-1} \cup S} \mathbf{L}(\mathbf{x_i},\mathbf{y_j},\mathbf{w^t}) - \alpha \sum_{j \in U^{t-1} - S} UC(\mathbf{x_j}, \mathbf{w^t}) \tag{2}$$

where $\mathbf{w}^t$ is the parameter vector learned on the updated dataset $L^{t-1} \cup S$. And the optimal selection S^* is the selection with the highest score. A score function for binary batch mode active learning by considering log likelihood on labeled data and adopting entropy as the uncertainty measure on unlabeled data:

$$f(S) = \sum_{i \in L^{t-1} \cup S} \log P(\mathbf{y}|\mathbf{x_i},\mathbf{w^t}) - \alpha \sum_{j \in U^{t-1} - S} H(\mathbf{y}|\mathbf{x_j}, \mathbf{w^t}) \tag{3}$$

This score function can be extended for MLBAL which measure the informativeness of example-label pairs:

$$f(S) = \sum_{i \in L^{t-1} \cup S} \log P(\mathbf{y}|\mathbf{x}_i,\mathbf{w}^t_r) \;-\; \alpha \sum_{j \in U^{t-1}-S} H(\mathbf{y}|\mathbf{x}_j, \mathbf{w}^t_s) \tag{4}$$

Where

$$H(\mathbf{y}_s|\mathbf{x}_j, \mathbf{w}^t_s) = -\sum_{ys=\pm 1} P(\mathbf{y}_s|\mathbf{x}_j, \mathbf{w}^t_s) \log P(\mathbf{y}_s|\mathbf{x}_j, \mathbf{w}^t_s) \tag{5}$$

measures the entropy of the unlabeled example-label pair*(x j , ys)*. $\mathbf{W}^t_s$ indicates model parameter vector obtained at iteration *t* for label *s*.

In addition to it, we consider the cross-label uncertainty which comes from the disagreement between the observed label correlation and the learned label prediction. For instance, in Fig. 1, we observe that image labels "*beach*" and "*ocean*" frequently co-occur, which indicates the beach and ocean labels are highly correlated. Then, the uncertainty between the two labels over an example image *x* appears if the predicted probabilities $P(y_{beach}|x)$ and $P(y_{ocean}|x)$ conflict with each other. .The two labels are highly correlated, the prediction for the label "*ocean*" can be regarded as a prediction for the label "*beach*" as well.

The prediction model disagreement from the label "*ocean*" to the label "*beach*" can be measured by the KL divergence $DK\ L(P(y_{beach}\ |x)\_P(y_{ocean}|x))$.Thus, the cross-label uncertainty of the unlabeled example label pair *(x, ybeach)* can be measured by the sum of the KL divergences from all its correlated labels. If we use $c(y_s)$ and $Cys = |c(y_s)|$ to denote all the correlated labels of *ys* and the number of the correlated labels respectively, the score function of selection can be redefined by taking into account cross-label uncertainty:

$$f(S) = \sum_{(x_j, y_s) \in U^{t-1}-S} \log P(\mathbf{y}_r|\mathbf{x}_i,\mathbf{w}^t_r) - \alpha \sum_{(x_i, y_r) \in L^{t-1} \cup S} \Big(H(\mathbf{y}_s|\mathbf{x}_j, \mathbf{w}^t_s) + 1/Cys \sum_{y_t \in c(y_s)} DK\ L(\mathrm{Py}_s \| \mathrm{Py}_t) \Big) \tag{6}$$

$$= \sum_{(x_i, y_r) \in L^{t-1} \cup S} \log P(\mathbf{y}_r|\mathbf{x}_i,\mathbf{w}^t_r) \;-\alpha \sum_{(x_j, y_s) \in U^{t-1}-S} 1/Cys \sum_{y_t \in c(y_s)} H(Py_s, Py_t) \tag{7}$$

$$DK\ L(Pys\|Pyt) = -\sum_{ys=\pm 1} P(\mathbf{y}_s|\mathbf{x}_j,\mathbf{w}^t_s) \log P(\mathbf{y}_s|\mathbf{x}_j,\mathbf{w}^t_s) / (\mathbf{y}_t|\mathbf{x}_j,\mathbf{w}^t_t) \tag{8}$$

$$H(Pys, Pyt) = -\sum_{ys=\pm 1} P(\mathbf{y}_s|\mathbf{x}_j,\mathbf{w}^t_s) \log P(\mathbf{y}_t|\mathbf{x}_j,\mathbf{w}^t_t) \tag{9}$$

Recall that KL divergence, *DK L(Pys||Pyt)*, is an asymmetric measure of the difference between two probability distributions *Pys* and *Pyt* . It increases with the discrepancy of *Pys* from *Pyt*. Cross entropy *H(Pys , Pyt) = H(Pys) + DK L(Pys||Pyt),* measures the average coding length of a variable generated by distribution *Pys* by using a coding scheme which is based on another distribution *Pyt.*

As mentioned before, some of the informative label correlations might involve more than two labels, and we call them high order label correlations. Such correlations are important to accurate label inferences.

For example, class label "*Apple*" can indicate either a fruit type or a computer brand. The inference from "*Apple*" to "*Mac*" based on their pair-wise correlation is weak and harmful to learning because of the ambiguous semantic meaning of apple. But if we consider higher order correlation, *e.g.*, correlation

among "*Apple*," "*Computer*" and "*Mac*," the inference becomes precise and helpful, *e.g.*, {"*Apple*," "*Computer*"} "*Mac*." "*Jaguar*" is another similar example, which can represent either a feline or a car brand. In order to incorporate high order label correlations, we now define an auxiliary compositional label. It is composed by one or several primary labels. Revisiting the previous example, we use *ys* , *yu* and *yv* to represent primary labels "*Mac*," "*Apple*" and "*Computer*" respectively, and assume the correlated labels of *ys* are *yu* and {*yu, yv* }. Two compositional labels *Yt*1 and *Yt*2 can be defined as:

$Yt1 = \{y_u\}$,

$Yt2 = \{y_u, y_v\}$.

The compositional label *Yt*1 is the primary label *yu* itself. The compositional label *Yt*2 is composed by two primary labels *yu* and *yv* , and it equals to 1 only when both *yu* and *yv* equal to 1. Then, the correlated labels of *ys* can be represented by compositional labels Yt_1 and Yt_2 , namely $c(y_s) = \{Yt_1, Yt_2\}$.We use $c(y_s)$ and $Cys = |c(y_s)|$ to represent all the correlated compositional labels of y_s and the number of the correlated compositional labels respectively. Now, with compositional label defined, the score function defined by Eq. 7 can be extended to incorporate high order label correlations:

$$f(s) = \sum_{(x_i, y_r) \in L^{t-1} \cup S} \log P(\mathbf{y_r}|\mathbf{x_i}, \mathbf{w^t_r}) - \alpha \sum_{(x_j, y_s) \in U^{t-1} - S} 1/Cys \sum_{y_t \in c(y_s)} H(Py_s, Py_t) \quad (10)$$

The prediction model *PYt* for the compositional label *Yt* can be leaned by treating the examples with all its primary labels as positive examples and the rest as negative examples.

4. CONCLUSION

Active learning turns practical for large scale real-world image classification with multilabel and batch mode characteristics considered and ingeniously handled .Such practical active learning based framework provides us the flexibility to control the balance between classification accuracy and annotation cost. Additionally, the quality control for crowd sourcing is another important factor for making the application of large scale active learning realistic. On the other hand, passive learning, which learns image classifiers only from the provided small amount of training data, has the lowest annotation cost, but suffers from insufficient classification accuracy. The proposed High order label correlation driven active learning offers us a flexible position between human annotation and passive learning. It provides us a principled efficient way to find a suitable balance between classification accuracy and annotation cost. Though the semi-supervised learning consider the unlabel images it require the human annotator. If we develop the system for unsupervised learning. It consider the unlabel images the annotation is done by itself. Also the system give the accurate and exact annotation of the images. The annotation should be proper for classification.

REFERENCES

[1] Bang Zhang,Yang Wang and Fang Chen,"Multilabel Image Classification via High-Orderb Label Correlation Driven Active Learning"Proc *IEEE Transaction on image processing*,vol.23,no.3,march 2014 .

[2] C. Campbell, N. Cristianini, and A. Smola, "Query learning with large margin classifiers," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 111–118.

[3] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2,pp. 45–66, Mar. 2002.

[4] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. 20th Int. Conf. Mach. Learn.*, vol. 3. 2003, pp. 59–66.

[5] Y. Guo and D. Schuurmans, “Discriminative batch mode active learning,” in *Proc. Adv. Neural Inf. Process.*, 2007, pp. 593–600.

[6] S. C. Hoi, R. Jin, J. Zhu, and M. R. Lyu, “Semi-supervised SVM batch mode active learning for image retrieval,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–7.

[7] Z. Xu, K. Yu, V. Tresp, X. Xu, and J. Wang, “Representative sampling for text classification using support vector machines,” in *Proc. Adv. Inf. Retr.*, vol. 3. 2003, pp. 393–407.

[8] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, “Batch mode active learning and its application to medical image classification,” in *Proc. 23rd ICML*, 2006, pp. 417–424.

[9] G. Qi, X. Hua, Y. Rui, J. Tang, and H. Zhang, “Two-dimensional active learning for image classification,” in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

AUTHORS’ BIOGRAPHY

Sana Altaf Sayyed received the B.E. degree in computer from the University of Pune, India. Currently appear Masters Degree in Computer from the University of Pune, India.