Multi-label Batch Mode Active Learning via High-Order Label Correlation

Miss. Sana A. Sayyed [1], Prof. R. P. Dahake[2]

Department of Computer Engineering, MET Institute of Engineering Bhujbal Knowledge City, Nashik, Savitribai Phule Pune University

Abstract—**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. Most active learning approaches select informative or representative unlabeled instances. These instances used to query their labels for classification. Labeled data are 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 choose the informative example-label pairs from which it learns so as to learn an accurate classifier with less annotation efforts. A high-order label correlation driven active learning is proposed to achieve the exact match of the requested images. Here in this we have to focused on accurate annotation of the images and domain specific multilabel image classification.**

Index Terms—**Active learning, high-order label correlation. multilabel classification.**

I. INTRODUCTION

The supervised learning techniques to image classification problems are 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. It takes traditional binary myopic active learning as an example. The example is selected for each iteration with a 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. The images can be labelled as "Mountain" and "Wave" respectively. The classification over these images is known as multi-class image classification.

We have worked with accurate annotation of the images and domain specific multilabel image classification. In the implementation of multi-label batch mode active learning problem with multi-label and batch mode factors are considered, as the proposed high-order label correlation driven active learning makes active learning applicable to real-world scale image classification and retrieval. A score function is defined to measure the informativeness of example-label pairs and based on this image retrival accurately obtained.

Fig. 1 Multilabel-label characteristic images of real world .

It actively selects the most informative unlabeled examples for labeling at each learning iteration, so as to gradually learn more accurate classifiers with less annotation efforts.With selected examples annotated and included into training dataset, classifier is updated. Within all the possible selections, the optimal selection leads to an updated classifier which maximizes the likelihood of the labeled data and minimizes the uncertainty of the unlabeled data. The proposed method considers not only pair-wise but also higher order label correlations.

II. LITERATURE SURVEY

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 is proposed by C. Campbell. 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 [3].

ISSN: 2278 – 7798

In order to overcome the drawbacks of the myopic active learning, batch mode active learning is proposed by K. Brinker. It selects a batch of informative examples for annotation at each learning iteration. K. Brinker performs batch mode selection by considering both the diversity and informativeness of the selected examples [4] . Z. Xu and et al. also have performed batch mode selection by taking into account the diversity of the selected examples by querying cluster centroids that are close to the decision boundary of the selected examples in SVM framework. [7].

S. C. Hoi et al. have tackled batch mode active learning under the semi-supervised learning setting[8]. But most of them focus on binary classification problems in which an example is only associated with one label. Multi label active learning is associated with many labels. For multi-label active learning, G. Qi and et al. have utilized mutual information to measure the correlation between labels to achieve efficient learning and mutual information to measure the correlation between labels to achieve efficient learning for multi-label active learning [9].In myopic active learning and batch mode active learning only one example is associated with only one label. To overcome this problem we multi-label batch mode active learning .As we can further extends the work of [1], It selects a batch of informative examples instead of a single example for annotation at each learning iteration. It takes informative label correlations, by defining their cross-label uncertainty. Multilabel images are classified and as per user query images are retrival.

III SYSTEM OVERVIEW

Batchmode active learning allows the learner to query instances in groups, which is better suited to parallel labeling environments.

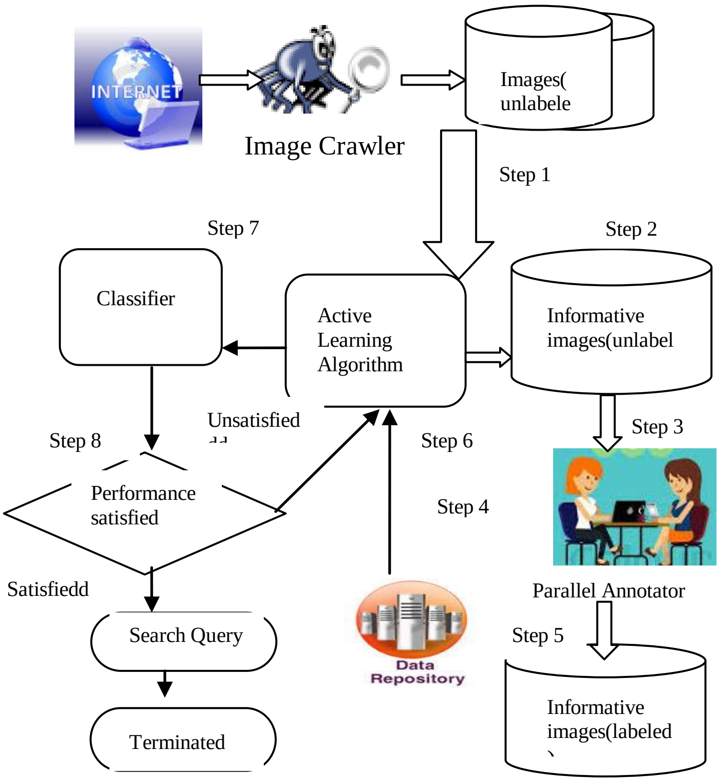

Fig. 2 System architecture of Multi-label Batch Mode Active learning.

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 Hoal (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 (step7). If the learned classifier can accurately classify images, the active learning is stopped. The proposed method is developed with the following steps.

a :- Examples are collected from user, Ineternet. The examples are in the unlabel from these images are sent to the engine.

b :- Active learning algorithm selects the most informative example-label pairs for human experts to annotate . The unlabel examples is annotated and convert into the label examples.

c :- Learned classifier can accurately classify images and based on classification XML file is generated.

d :- A score function is defined to measure the informativeness of example-label pairs to measure the informativeness of example-label pairs.

e :- Labels are usually dependent, and their inherent correlations are useful for inferring 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 from the learned classification mode.

f :- An informative correlation can be found from one association rule, and its informativeness is measured by the support and confidence of the rule.

g :- After applying the score function on the labelled images.The optimal iteration is selected and then the Score function optimization is done.the optimal example-label pairs selection *S* should have the highest score.

IV IMPLEMENTATION STRATEGY

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- which can be represented as a K-dimensional binary vector Y = { $y1, \dots, yK$ } with1 indicating that the example belongs to the corresponding concept and -1 otherwise. Which can be represented as a K- 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 The current annotated labels of x can form a labeled Initial prediction models $P(y_j|x,w^0_j)$, $1 \leq j \leq K$ 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 $U^t = U^{t-1} - S^t$ and $L^t = L^{t-1} \cup S^t$. The updated prediction modes $P(y_j|x,w^t_j)$ can be

ISSN: 2278 – 7798

obtained on L^t and U^t . The goal is to search for the optimal selection *St* which leads to the best prediction models $P(y_j|x, w^t_j)$ at each learning iteration of example.

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 in Eq 1:

$$= \sum_{i\in L} \mathcal{L}(\mathbf{x_i},\mathbf{y_j},\mathbf{w}) - \alpha \sum_{j\in U} \mathcal{UC}(\mathbf{x_j},\mathbf{w}) \qquad \textbf{Eq.(1)}$$

Active learning goes one step further.It actively selects the informative unlabeled examples for labeling at each iteration.

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

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 given in Eq 3:

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

This score function can be extended for Multi Label Batch mode Active Learning (MLBAL) which measure the informativeness of example-label pairs:

$$f(S) = \sum_{i\in L^{t-1}\cup S} \log P(\mathbf{y}|\mathbf{x_i},\mathbf{w^t_r}) - \alpha \sum_{j\in U^{t-1}-S} H(\mathbf{y}|\mathbf{x_j},\mathbf{w^t_s}) \qquad \textbf{Eq.(4)}$$

where

$$H(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) = -\sum_{y_s=\pm1} P(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) \log P(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) \qquad \textbf{Eq.(5)}$$

Where H measures the entropy of the unlabeled example-label pair(x_j,y_s). $\mathbf{w}^t_s$ indicates model parameter vector obtained at iteration *t* for label *s* as in Eq. 4 and Eq. 5.

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 "*water*" and "*mountain*" frequently co-occur, which indicates the mountain and water labels are highly correlated. Then, the uncertainty between the two labels over an example image *x* appears if the predicted probabilities $P(y_{mountain}|x)$ and $P(y_{water}|x)$ conflict with each other. .The two labels are highly correlated, the prediction for the label "*water*" can be regarded as a prediction for the label "*mountain*" as well. The prediction model disagreement from the label "*water*" to the label "*mountain*" can be measured by the KL divergence $D_{KL}(P(y_{mountain}|x)_P(y_{water}|x))$is metioned in Eq 6.Thus, the cross-label uncertainty of the unlabeled example label pair $(x, y_{mountain})$ can be measured by the sum of the KL divergences from all its correlated labels. If we use $c(y_s)$ and $C_{ys}=|c(y_s)|$ to denote all the correlated labels of y_s and the number of the correlated labels respectively, the score function of selection can be redefined by taking into account cross-label uncertainty :

$$f(S) = \sum_{(x_i,y_r)\in L^{t-1}\cup S} \log P(\mathbf{y_r}|\mathbf{x_i},\mathbf{w^t_r}) - \alpha \sum_{(x_j,y_s)\in U^{t-1}-S} \Big(H(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) + 1/C_{ys} \sum_{y_t\in c(y_s)} D_{KL}(P_{ys}\|P_{yt}) \Big) \qquad \textbf{Eq.(6)}$$

where

$$D_{KL}(P_{ys}\|P_{yt}) = -\sum_{ys=\pm1} P(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) \log P(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) / (\mathbf{y_t}|\mathbf{x_j},\mathbf{w^t_t}) \qquad \textbf{Eq.(7)}$$

$$H(P_{ys}, P_{yt}) = -\sum_{ys=\pm1} P(\mathbf{y_s}|\mathbf{x_j},\mathbf{w^t_s}) \log P(\mathbf{y_t}|\mathbf{x_j},\mathbf{w^t_t}) \qquad \textbf{Eq.(8)}$$

Recall that KL divergence, $D_{KL}(P_{ys}\|P_{yt})$, is an asymmetric measure of the difference between two probability distributions P_{ys} and P_{yt} in Eq.7. It increases with the discrepancy of P_{ys} from P_{yt}. Cross entropy $H(P_{ys}, P_{yt}) = H(P_{ys}) + D_{KL}(P_{ys}\|P_{yt})$, measures the average coding length of a variable generated by distribution P_{ys} by using a coding scheme which is based on another distribution P_{yt}.. The correlated labels of y_s can be represented by compositional labels Y_{t_1} and Y_{t_2}, namely $c(y_s) = \{Y_{t_1}, Y_{t_2}\}$.We use $c(y_s)$ and $C_{ys} = |c(y_s)|$ 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. 6 can be extended to incorporate high order label correlations as shown in Eq. 9:

$$f(S) = \sum_{(x_i,y_r)\in L^{t-1}\cup S} \log P(\mathbf{y_r}|\mathbf{x_i},\mathbf{w^t_r}) - \alpha \sum_{(x_j,y_s)\in U^{t-1}-S} 1/C_{y_s} \sum_{y_t\in c(y_s)} H(P_{y_s}, P_{y_t}) \qquad \textbf{Eq.(9)}$$

The prediction model P_{Y_t} for the compositional label Y_t can be leaned by treating the examples with all its primary labels as positive examples and the rest as negative examples. Multilabel images are classified and as per user query images are retrival.

V. EXPERIMENT AND RESULT

For multi-label active learning, utilize mutual information to measure the correlation between labels to achieve efficient learning. The experiment is performed on natural scene images of COREL dataset. It takes images for experiment on the basis of there informativeness they are classified into 20 different category. The statistic about different labels can be found in Fig.3. HoAL consistently performed *Pair-wise* verifies that wise utilization of high order label correlations can help achieving more efficient and accurate multi-label image classification.

ISSN: 2278 – 7798

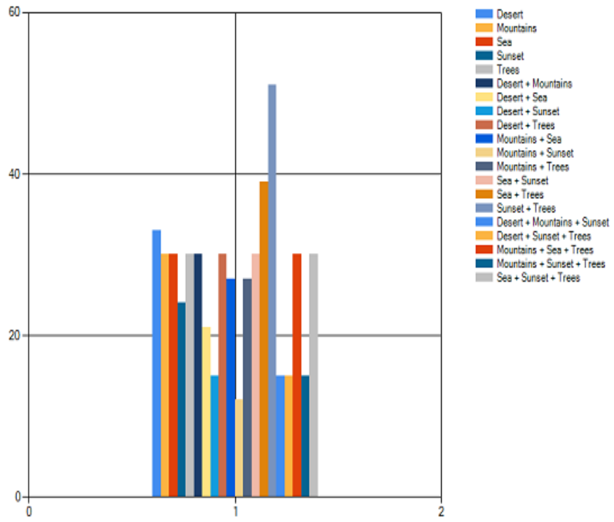

Fig. 3 Graphical representation of classification of images

The effect of batch selection size and the effect of high order label correlations on multilabel batch mode active learning An initial labeled example-label pair set is randomly selected and provided for training initial prediction models. As we have referred the various performance evaluation techniques as mentioned in [1]. The average precision as shown in Fig. 4 and average recall as shown in Fig.5 on the set of example label pair.

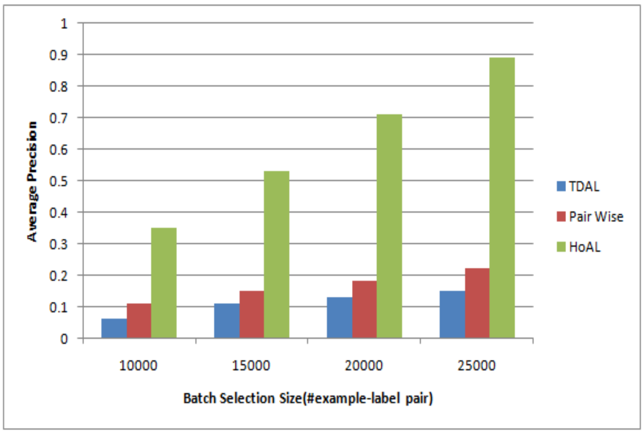

Fig .4 Average precision for example-label pairs with different batch selection sizes.

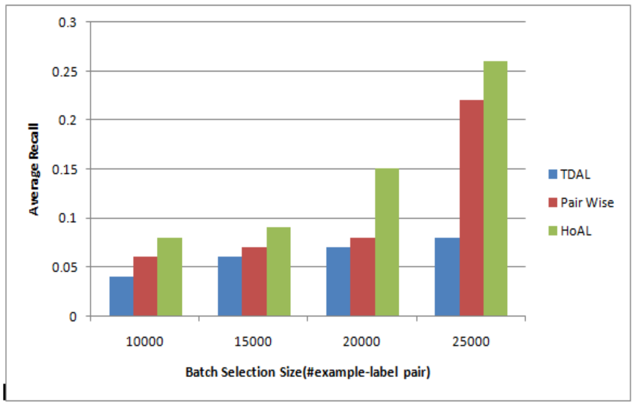

Fig .4 Average precision for example-label pairs with different batch selection sizes.

VI. CONCLUSION

Active learning turns practical for large scale real world image classification with multilabel instead of one label and batch mode characteristics are considered. Parallel annotation is implemented for batch mode active learning, such practical active learning based framework provides us the flexibility to control the balance between classification accuracy and annotation cost. The proposed High order label correlation driven active learning classify the multilabel images. Classified images are retrieved as per user query.

REFERENCES

[1] Bang Zhang,Yang Wang and Fang Chen, IEEE,*Multilabel Image Classification via High-Orderb Label Correlation Driven Active Learning*, Int. J.Document Anal. Recognit. 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., Recognit.,2002, pp. 111 - 118.

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

[4] K. Brinker, *Incorporating diversity in active learning with support vector machines*, Proc. Int. Conf.Mach. Learn.Recognit.,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.

[10] G. Qi, X. Hua, Y. Rui, J. Tang, T. Mei, and H. Zhang, *Correlative multi-label video annotation*, in Proc. 15th Int. Conf. Multimedia, 2007, pp.17–26.

[11] B. Zhang, Y. Wang, and W. Wang, *Batch mode active learning for multilabel image classification with informative label correlation mining*, in Proc. IEEE WACV, Jan. 2012, pp. 401–407.

[12] Z. Zhou and M. Zhang, *Multi-instance multi-label learning with applicationto scene classification*, in Proc. 20th Annu. Conf. NIPS, 2006,pp. 1–8.

ISSN: 2278 – 7798

Author Profile

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

Prof. R. P. Dahake She is currently working as Assistant Professor in Department of Computer Engineering,MET's IOE Bhujbal Knowledge City, Nasik, Maharashtra,India. She is Ph.D. scholar in Savitribai Phule Pune University, Pune Maharashtra.She has presented papers at National and International conferences and also published papers in National and International Journals on various aspects of Computer Engineering and Networks. Her areas of interest include Image Processing, Computer Networks Security , Cloud Computing and Embedded Systems.

ISSN: 2278 – 7798