This WACV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Active Batch Sampling for Multi-label Classification with Binary User Feedback

Debanjan Goswami Shayok Chakraborty
Department of Computer Science, Florida State University

Abstract

Multi-label classification is a generalization of multi-class classification, where a single data sample can have multiple labels. While deep neural networks have depicted commendable performance for multi-label learning, they require a large amount of manually annotated training data to attain good generalization capability. However, annotating a multi-label data sample requires a human oracle to consider the presence/absence of every single class individually, which is extremely laborious. Active learning algorithms automatically identify the salient and exemplar instances from large amounts of unlabeled data and are effective in reducing human annotation effort in inducing a machine learning model. In this paper, we propose a novel active learning framework for multi-label learning, which queries a batch of (image-label) pairs and for each pair, poses the question whether the queried label is present in the corresponding image; the human annotators merely need to provide a binary feedback ("yes/no") in response to each query, which involves much less manual work. We pose the image and label selection as a constrained optimization problem and derive a linear programming relaxation to select a batch of (image-label) pairs, which are maximally informative to the underlying deep neural network. Our extensive empirical studies on three challenging datasets corroborate the potential of our method for real-world multi-label classification applications.

1. Introduction

Multi-label learning has attracted significant research attention in recent years, where each data sample can have multiple classes associated with it [42]. For instance, classifying the contents of a natural scenery image is a multi-label problem, as a single image can have multiple class labels (such as sunset, ocean, mountains etc.) associated with it. Deep neural networks have revolutionized the field of computer vision and have depicted state-of-the-art results in a variety of applications, including multi-label learning [16]. However, deep models are extremely data hungry and require a large amount of labeled training data to furnish good generalization capability. While gathering unlabeled data is cheap and easy, annotating them with class labels is an expensive process in terms of time, labor and human expertise. The problem is even more severe for multi-label classification, since the human annotator [1] has to meticulously inspect the presence/absence of every single class to annotate a given data sample. Hence, developing a strategy to minimize the human annotation effort in a multi-label problem is of immense practical importance.

Active Learning (AL) algorithms automatically identify the most informative samples from large amounts of unlabeled data and are instrumental in reducing human annotation effort in inducing a machine learning model [20]. In a typical *serial query* based setup, the learner queries a single sample in each AL iteration and the model is updated after every individual query. This results in frequent model updates, which can be computationally expensive, particularly for deep neural networks. Further, serial query AL can only utilize a single human annotator at any given point of time; in crowdsourcing platforms like AMT, multiple annotators can provide labels to queried samples simultaneously. To alleviate these challenges, *batch mode active learning (BMAL)* techniques have been developed which query a batch of unlabeled samples in each AL iteration. Batch mode AL has been successfully used in a variety of computer vision applications, such as image recognition [37], semantic segmentation [1], object detection [3] and image regression [19] among others.

In this paper, we propose a novel AL framework for multi-label classification. Our algorithm identifies a batch of exemplar images, together with a label for each, and poses the question: "does image x_i contain label y_j?"; the human annotator has to merely provide a binary answer *"yes / no"*. Providing such binary feedback is much more efficient and less laborious than annotating all the labels in a given sample. We derive a criterion based on informativeness, diversity, and label correlation to quantify the usefulness of every (image-class) pair. The active image and label selection problem is then solved through a single integrated framework to derive a batch of informative pairs for manual

[1] we use the terms annotator, oracle, labeler and user interchangeably in this paper

annotation. Although validated on image data in this work, the proposed framework is generic and can be used in any application involving multi-label data, such as text mining, music and audio analysis among others.

The rest of the paper is organized as follows: we present a survey of related techniques in Section 2; our framework is detailed in Section 3; the results of our experiments are presented in Section 4 and we conclude with discussions in Section 5.

2. Related Work

In this section, we present a survey of active learning in general, followed by a survey of active learning for multi-label classification.

Active Learning (AL): AL is a well-researched topic in the machine vision literature. Uncertainty sampling, in a variety of forms, is the most commonly used strategy for AL, where samples with the highest prediction uncertainties are queried for their labels [11, 13]. With the advent and popularity of deep neural networks, *deep active learning (DAL)* has become popular, where the goal is to query informative unlabeled samples for manual annotation and simultaneously learn discriminating feature representations using a deep neural network. Common DAL techniques include a task agnostic scheme which learns a loss prediction function to predict the loss value of an unlabeled sample and queries samples accordingly [37], a technique which decomposes the training loss for DAL into learning the network parameters and active sampling through alternating optimization [22], an AL strategy based on temporal output discrepancy that queries samples based on the discrepancy of outputs given by the models at different optimization steps during training [9] and an AL framework that queries unlabeled samples that can provide the most positive influence on model performance [17]. Techniques based on adversarial learning have depicted particularly impressive performance in DAL [6, 24, 46].

AL for Multi-label Classification: Active learning has also been studied in the context of multi-label classification [29]. Existing methods can be broadly categorized into two groups: (i) methods that query all the labels of a given unlabeled sample; and (ii) methods that query (sample-label) pairs. Several criteria have been studied to identify the informative samples in the first category, such as uncertainty sampling [12], uncertainty and diversity [2], uncertainty and representativeness [38], expected loss reduction [34] and prediction inconsistency [21] among others. However, these methods query all the labels of a given unlabeled sample and can be costly especially when the number of labels is large. Further, these methods ignore the intrinsic relationships embedded in the label distribution, which can be useful in formulating the sampling strategy of multi-label active learning [43].

Techniques in the second category query the informative (sample-label) pairs of the form (x^*, y^*) and pose the binary question whether the label y^* is present in the unlabeled sample x^* or not. These methods can further reduce the annotation cost (since they only need to acquire part of labels of an unlabeled sample) and can also exploit the correlations among the different labels to drive the AL process. They have attracted significant research attention over the years [7,8,10,18,28,30,31,35,36,39,40,43,44]. We present an overview of some of the techniques here. A two dimensional AL strategy was proposed by Qi *et al.* [18], which minimized a multi-label Bayesian error bound to identify the informative (image-label) pairs. Guo *et al.* [7] and Wu *et al.* [28] exploited low-rank feature representation with an informativeness criterion to mine label correlations for multi-label AL. Wu *et al.* [31] also exploited the label dependency with the input features and used the informativeness of each (image-label) pair for active sampling. A semi-supervised multi-label AL framework was proposed by Wu *et al.* [30] which utilized classification prediction information, label correlation information, and example spatial information to select unlabeled (image-label) pairs for annotation. Huang *et al.* proposed *QUIRE* [8], a min-max AL strategy that combined informativeness and representatives for active (image-label) selection. Huang and Zhu also proposed *AUDI* [10], which exploited both uncertainty and diversity in the instance space as well as the label space, and actively queried (image-label) pairs. Yu *et al.* recently proposed a cost effective MLAL framework called *CMAL*, which queries (subexample-label) pairs, instead of the (example-label) pairs [38]. It operates in two steps, where it first selects the most informative (example-label) pairs by leveraging uncertainty, label correlation and label space sparsity; it then greedily queries the most probable positive (subexample-label) pairs of the selected (example-label) pair in a multi-instance learning setup.

Although these methods have demonstrated promising performance, most of them query only a single (image-label) pair in each AL iteration and the underlying model is updated after every individual query [7,18,28,30,35,36,44]. As mentioned before, frequent model updates can be computationally expensive, especially for deep neural networks. Further, querying a single (image-label) pair cannot leverage the presence of multiple labeling oracles (who can simultaneously label samples) resulting in a wastage of available resources. A few methods are designed to query k (image-label) pairs, by querying the optimal pair repeatedly k times [8, 10, 31, 32]. However, they do not consider the redundancy among the queried data, resulting in sub-optimal performance (as validated in our empirical studies). Moreover, most of the existing multi-label AL techniques use SVMs or Binary Relevance k Nearest Neighbors (BRkNNs) as the underlying classification

model [7, 12, 28, 30–32, 38–40]; we employed deep neural networks as the base model in our experiments, which are largely unexplored for the multi-label AL problem. Our contributions in this paper can be summarized as follows:

(i) We propose a novel active sampling framework to simultaneously query batches of (image-label) pairs for manual annotation in a multi-label setup. We incorporate label correlations and sample redundancy in our framework to avoid querying duplicate images / labels.

(ii) We pose the image and label selection as a constrained optimization problem and derive a linear programming relaxation to solve the same.

(iii) We conduct extensive experiments to study the performance of our framework using state-of-the-art deep neural network architectures.

We now describe our framework.

3. Proposed Framework

3.1. Problem Setup

Consider a multi-label learning problem where we are given a labeled training set L and an unlabeled set U, with $|L| \ll |U|$. Let N denote the number of unlabeled images, $N = |U|$. Also, let Y denote the set of labels in the dataset. The samples in L are fully annotated with all the $|Y|$ labels. Let θ denote the deep neural network trained on L. We are given a query budget k and a parameter Y_{max}, which denotes the maximum number of labels that can be queried per image in a given AL iteration (to ensure that the queries are distributed across a large number of images). Our objective is to select a batch of k (image-label) pairs (x_i, y_j), and for each pair, pose the question: "does image x_i contain the label y_j ?", such that the user annotation augments maximal information to the deep learning model.

In order to identify the optimal set of images and labels to be queried, we need a function to quantify the utility score of a batch of (image-label) pairs. We used a function based on class presence uncertainty, label correlation and image redundancy for this purpose. The first criterion ensures that we query those (image-label) pairs where there is maximal uncertainty regarding the presence of the given label in the given image; the second criterion exploits the correlation among the labels of a given image in formulating the active query strategy; the redundancy criterion ensures that we query a diverse set of images in our batch and avoid duplicate image queries. These are detailed below.

Computing Class Presence Uncertainty: Let p_{ij} denote the probability that image x_i contains the label y_j (computed using the current deep neural network θ). We used Shannon's entropy to compute the prediction uncertainty of the presence of label y_j in image x_i:

$$H_{ij} = p_{ij} \log p_{ij} + (1 - p_{ij}) \log(1 - p_{ij}) \qquad (1)$$

A high value of H_{ij} denotes high uncertainty of the deep model in predicting the presence of label y_j in sample x_i, and thus, a more useful pair for active query.

Computing Label Correlations: The labels in a multi-label dataset usually share a correlation among each other, that is, information about the presence / absence of a particular label in a particular sample often provides relevant information about the presence / absence of the other labels in the same sample. Appropriately exploiting the label correlations can potentially result in more efficient active sampling. Common methods of estimating label correlations include mutual information [18], association rule mining [41] etc. However, as noted by Wu *et al.*, these methods can only identify positive correlations of simultaneous appearances of the labels [30]. Real-world applications often exhibit negative correlations, where the presence of one label decreases the probability of the presence of another label in a pair of labels, which is not modeled by many of the correlation estimation methods. To account for this, we utilized the chi-square statistic to estimate the correlations among the labels, as it considers all possible positive and negative combinations of labels in a pair. It is very easy to compute and has been used in previous multi-label learning research with promising results [28, 30]. Let μ_{ij} denote the correlation between label y_j of sample x_i with the other labels of sample x_i, which can be estimated from the labeled data L. A high value of this term denotes that label y_j of sample x_i is highly correlated with the other labels of sample x_i; hence, querying label y_j of this sample will reveal maximal information about the other labels of x_i. A high value of μ_{ij} is thus desirable from an active query perspective. To compute μ_{ij}, we proceed as follows. We first compute a contingency table for each label pair y_i and y_j:

	y_j	$\overline{y_j}$
y_i	A	B
$\overline{y_i}$	C	D

Table 1. Contingency table for the label pair y_i and y_j.

The label correlations calculated by chi-square estimation can be defined as follows [30]:

$$M_{ij} = \frac{AD - BC}{\sqrt{(A+B)(A+C)(C+D)(B+D)}} \qquad (2)$$

Note that the chi-square statistic values are symmetric, i.e. $M_{ij} = M_{ji}$. When $M_{ij} < 0$, the relationship between y_i and y_j is negative; otherwise, the relationship is positive. $M_{ij} = 0$ indicates that there is no relationship between

these two labels. The higher the value of M_{ij} is, greater is the correlation between the two labels.

Now, let Y denote the set of labels in the dataset. Consider a sample x_i for which some of the labels are known and the others are unknown. For our active label query (detailed next), we would like to compute the correlation between an unknown label of this sample and the other unknown labels. Let μ_{ij} denote the correlation between the unknown label j and the other unknown labels of sample x_i. It is computed as the average of the label correlations of this unknown label with other unknown labels in the same sample [30]:

$$\mu_{ij} = \begin{cases} \frac{1}{l_u} \sum_{k=1}^{|Y|} |M_{jk}|.sign(y_k \in UL(x_i)), & \text{if } \; l_u > 1 \\ 0, & \text{if } \; l_u = 1 \end{cases}$$

where l_u is the number of labels of sample x_i for which the values are not yet known, $UL(x_i)$ is the set of labels of sample x_i for which the values are not yet known and $sign(.)$ is a *sign* function whose value is 1 of the predicate inside the function evaluates to true and 0 otherwise.

Given H_{ij} and μ_{ij}, we computed a confidence matrix $C \in \Re^{|Y| \times N}$, where $C(j,i)$ denotes the confidence of the deep model in predicting the presence of class y_j in image x_i, while also considering the correlation between label y_j and the other labels of each unlabeled image x_i. Since we would like to maximize both H_{ij} and μ_{ij} in our active queries, we computed a weighted summation of the two terms, and inverted that to form our confidence matrix C (high entropy corresponds to low confidence and vice versa):

$$C(j,i) = \frac{\alpha}{H_{ij} + \beta \mu_{ij}} \quad i = 1, \ldots N, \; j = 1, \ldots |Y| \quad (3)$$

where α and β are constants.

Computing Image Redundancy: Since our method is designed to query a batch of (image-label) pairs (to utilize multiple labeling oracles and to avoid frequent model updates), it is important to consider data redundancy, so that duplicate images are not queried. We computed a redundancy matrix $R \in \Re^{N \times N}$, where $R(i,j)$ denotes the redundancy between images x_i and x_j in the unlabeled set. The cosine similarity was used to quantify the redundancy between a pair of samples; negative values were replaced with 0, so that R contains only non-negative entries:

$$R(i,j) = max(0, cos(\mathcal{F}(x_i), \mathcal{F}(x_j))) \quad (4)$$

where $cos(\mathcal{F}(x_i), \mathcal{F}(x_j)) = \frac{\mathcal{F}(x_i)^\top \mathcal{F}(x_j)}{||\mathcal{F}(x_i)||.||\mathcal{F}(x_j)||}$, and $\mathcal{F}(x)$ denotes the deep feature representation of image x. A low value of $R(i,j)$ implies that images x_i and x_j have low redundancy between them. Note that, the label correlation term quantifies the relationship among the labels for a given image, whereas the redundancy term quantifies the redundancy between a pair of images. Cosine similarity has been previously used to compute similarity in AL research, with promising results [5]. Depending on the application, other metrics can also be used to compute the uncertainty, correlation and redundancy terms.

3.2. Active Sampling Framework

Given C and R, our objective is to query a batch of (image-label) pairs such that in each pair, the deep model has low confidence in predicting the presence of the given class in the given image, the given label is highly correlated with the other labels in the given image, and the queried images have minimal redundancy among them. We define a binary matrix $Q \in \{0,1\}^{N \times |Y|}$, where each row corresponds to an unlabeled image and each column corresponds to a label. A value of 1 in a row denotes that the image should be selected for annotation, and the position(s) of 1 in a particular row of Q denote the label(s) that should be used to pose the binary queries for this image. We also define a binary vector $w \in \{0,1\}^{N \times 1}$ where $w_i = 1$ denotes that image x_i is selected for annotation, and $w_i = 0$ denotes that it is not selected. The active selection of (image-label) pairs can thus be posed as the following optimization problem:

$$\begin{aligned} \min_{Q,w} \quad & \text{Tr}(QC) + \lambda w^\top R w \\ s.t. \quad & \langle Q, E \rangle = k \\ & (Q.e)_i \leq Y_{max}, \forall i \\ & w_i = \min(1, (Q.e)_i), \forall i \\ & w_i, Q_{ij} \in \{0,1\}, \forall i,j \end{aligned} \quad (5)$$

where $\lambda > 0$ is a weight parameter governing the relative importance of the two terms, E is a matrix of size $N \times |Y|$ (same size as Q) with all entries 1, e is a vector of size $|Y| \times 1$ with all entries 1, k is the labeling budget, $\langle ., . \rangle$ denotes the inner product operator and Tr denotes the trace of a matrix. The first term in the objective function denotes that the deep model has low confidence in predicting the presence of the selected labels in the corresponding selected images and that the selected labels are highly correlated with the other labels of the selected images; the second term ensures that the selected images have minimal redundancy among them. The first constraint denotes the total number of queries posed by Q is equal to the specified budget; the second constraint ensures that the number of 1s in each row of Q is less than or equal to Y_{max}, that is, the number of queries posed for each image is less than or equal to the pre-specified limit Y_{max}; the third constraint denotes that w_i is equal to 1 if there is at least one entry with value 1 in row i of Q (image x_i is selected for annotation), and w_i

is equal to 0 if all the entries in row i of Q have value 0 (image x_i is not selected); the fourth constraint denotes that w is a binary vector and Q is a binary matrix. We now discuss an efficient strategy to solve this optimization problem, as presented in the following theorem.

Theorem 1. *The optimization problem defined in Equation (5) can be expressed as an equivalent linear programming (LP) problem.*

Proof. We simplify the definition of w in the third constraint and rewrite the optimization problem as:

$$\begin{aligned} \min_{Q,w} \quad & \mathrm{Tr}(QC) + \lambda w^\top R w \\ s.t. \quad & \langle Q, E \rangle = k \\ & (Q.e)_i \leq Y_{max}, \forall i \\ & Q_{ij} \leq w_i, \forall i, j \\ & w_i, Q_{ij} \in \{0, 1\}, \forall i, j \end{aligned} \tag{6}$$

The constraint $Q_{ij} \leq w_i, \forall i, j$ denotes that if row i in Q has at least one entry as 1, then w_i has to be 1. If row i in Q has all entries as 0, then w_i is free to be 0 or 1. However, we are solving a minimization problem with $w^\top R w$ in the objective, and R has only non-negative entries; this criterion will force w_i to be equal to 0, as that will result in a better (lower) value of the objective. This shows that the constraint $w_i = \min(1, (Q.e)_i), \forall i$ in Equation (5) is equivalent to the linear constraint $Q_{ij} \leq w_i, \forall i, j$ in Equation (6).

The first term in the objective function can be expressed as a linear term: $\mathrm{Tr}(QC) = \sum_{i,j} C_{ij}.Q_{ji}$. Also, let $d_{ij} = w_i.w_j$. Clearly, D is a binary matrix of size $N \times N$ with all entries 0 or 1. The second term in the objective can then be written as: $w^\top R w = \sum_{i,j} d_{ij}.r_{ij}$
The optimization problem can thus be expressed as:

$$\begin{aligned} \min_{Q,w,D} \quad & \sum_{i,j} C_{ij}.Q_{ji} + \lambda \sum_{i,j} d_{ij}.r_{ij} \\ s.t. \quad & \sum_{i,j} Q_{ij} = k \\ & d_{ij} = w_i.w_j, \forall i, j \\ & (Q.e)_i \leq Y_{max}, \forall i \\ & Q_{ij} \leq w_i, \forall i, j \\ & w_i, Q_{ij}, D_{ij} \in \{0, 1\}, \forall i, j \end{aligned} \tag{7}$$

Now, we attempt to express the quadratic equality $d_{ij} = w_i.w_j, \forall i, j$ as a linear term. The quadratic equality implies that d_{ij} equals 1 only when both w_i and w_j are 1 and equals 0 otherwise. This can be expressed as the linear inequality $w_i + w_j \leq 1 + 2d_{ij}, \forall i, j$. From the inequality, we note that when both w_i and w_j are 1, d_{ij} is forced to be 1. When w_i and w_j are both 0, or one of them is 0 and the other one is 1, d_{ij} is free to be 0 or 1. Using the same argument as before, we note that we are solving a minimization problem with $\sum_{i,j} d_{ij}.r_{ij}$ in the objective and R has only non-negative entries; thus, the nature of the problem will force d_{ij} to be 0 as it will produce a lower value of the objective. Replacing the quadratic equality with the linear inequality, we express the optimization problem as follows:

$$\begin{aligned} \min_{Q,w,D} \quad & \sum_{i,j} C_{ij}.Q_{ji} + \lambda \sum_{i,j} d_{ij}.r_{ij} \\ s.t. \quad & \sum_{i,j} Q_{ij} = k \\ & w_i + w_j \leq 1 + 2d_{ij}, \forall i, j \\ & (Q.e)_i \leq Y_{max}, \forall i \\ & Q_{ij} \leq w_i, \forall i, j \\ & w_i, Q_{ij}, D_{ij} \in \{0, 1\}, \forall i, j \end{aligned} \tag{8}$$

In this optimization problem, both the objective function and the constraints are linear in the variables Q, w and D. It is thus a linear programming (LP) problem.

□

We vectorize the variables Q, w and D, append them one below the other and express the objective function and the constraints in terms of this new variable. The integer constraints are then relaxed into continuous constraints and the problem is solved using an off-the-shelf LP solver. After obtaining the continuous solution, we recover the integer solution of our variable of interest Q, using a rounding approach where the k highest entries in Q are reconstructed as 1 and the other entries as 0, observing the constraints. The pseudo-code of our algorithm, for one active learning iteration, is outlined in Algorithm 1.

3.3. Computational Considerations

Computing the redundancy matrix R (Equation (4)) involves quadratic complexity. We first note that R needs to be computed only once in our framework (before the start of the AL iterations). Moreover, the theory of random projections can be used to reduce the computational overhead. Random projections have been successfully used to speed up computations, where an original data matrix $A \in \Re^{m \times D}$ is multiplied by a random projection matrix $X \in \Re^{D \times d}$ to obtain a projected matrix $B \in \Re^{m \times d}$ in the lower dimensional space d: $B = \frac{1}{\sqrt{d}} AX$, where $d \ll min(m, D)$ [27]. We plan to study this as part of our future research.

Further, Sridhar *et al.* [25] proposed an algorithm to solve large-scale LP problems and showed that we can recover solutions of comparable quality by rounding an approximate LP solution instead of the exact one. These approximate LP solutions can be computed efficiently by applying a stochastic-coordinate-descent method to a

Algorithm 1 The Proposed Multi-label Active Learning Algorithm with Binary User Feedback

Require: Labeled training set L, unlabeled set U, query budget k, parameters α, β, Y_{max} and λ, a deep neural network architecture for multi-label classification

1: Train the deep model on the training set L
2: Compute the prediction entropy matrix H (Equation (1))
3: Compute the label correlation matrix μ (detailed in Section 3.1)
4: Compute the confidence matrix C using α, β, H and μ (Equation (3))
5: Compute the redundancy matrix R (Equation (4))
6: Solve the LP problem in Equation (8) after relaxing the integer constraints
7: Round the solution to derive the matrix Q
8: Select the unlabeled images and the corresponding labels to pose the binary queries based on the entries in Q
9: Update the deep model with the user response to the binary queries (detailed in Section 4)

quadratic-penalty formulation of the LP. A parallel version of the algorithm was also proposed, which is suitable for execution on multi-core, shared-memory architectures. In their empirical studies, the authors reported computational speedup by a factor of 2.8 to 9.0 (time taken by an off-the-shelf LP solver divided by the time taken by their method), with corresponding solution quality of 1.04 and 1.21 (ratio of the solution objective obtained by this method to that by an off-the-shelf LP solver) for solving LP minimization problems, similar to the one in this paper. Thus, this method has the potential to substantially reduce the computation time, without sacrificing too much on the solution quality. We plan to explore this framework to further improve the computation time of our algorithm, as part of future work. Please refer to [25] for further details about this algorithm, its convergence analysis and worst-case complexity bounds.

4. Experiments and Results

Datasets: We used three challenging datasets to study the performance of our algorithm: NUS-Wide [4], MIML [45] and COCO [14]. All these datasets are widely used as benchmarks in multi-label learning research.

Experimental Setup: Each dataset was divided into three parts: an initial training set (where all the images were completely annotated with all the labels), an unlabeled set and a test set. Each algorithm queried k (image-label) pairs in each AL iteration (where k is a pre-specified query budget); the label information of these queried pairs was obtained from the human annotators and were then appended to the training set. The deep CNN was updated, and its performance was evaluated on the held-out test set. The process was continued iteratively until a stopping condition was satisfied (taken as 25 iterations in this work). The goal was to study the improvement in performance on the test set with increasing label queries.

The query budget k was taken as 200 for NUS-WIDE, 40 for MIML and 60 for the COCO dataset. The parameters λ, Y_{max}, α and β were taken as 0.025, 5, 1 and 0.1 respectively based on preliminary experiments. The F1-score was used as the evaluation metric as commonly done in multi-label learning research [28, 30, 31]. All the results were averaged over three runs to rule out the effects of randomness.

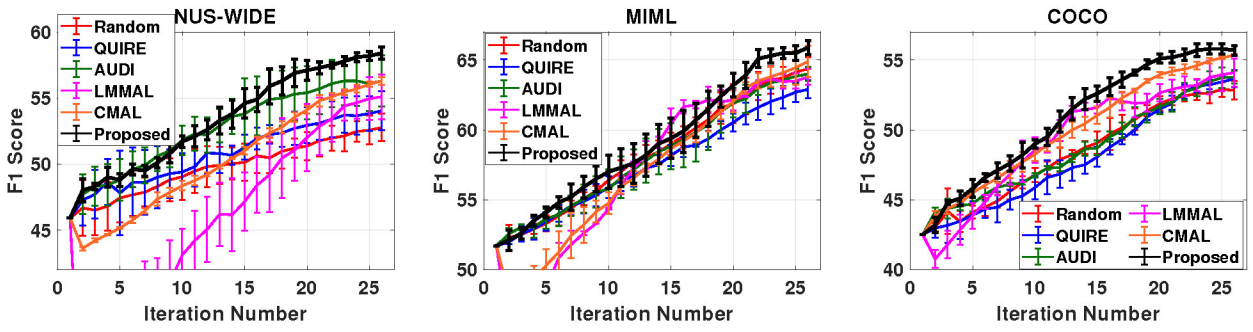
Implementation Details: We used the ResNeXt-50 [33] deep model, pretrained on the ImageNet-1k dataset, as the underlying Convolutional Neural Network (CNN) architecture in our experiments [2]. The input images were scaled and normalized to a fixed size of 256×256 pixels and fed into the CNN. The features extracted were fed into one dropout layer, followed by 1 fully connected layer with $|Y|$ neurons. We used the *Adam* optimizer with a learning rate of 0.0001 and a batch size of 32, and the network was trained for a maximum of 35 epochs in each active learning iteration. We used the binary cross entropy loss to train the deep CNN, due to its promising performance in multi-label classification [15]:

$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{|Y|}[y_{ij}\log(\sigma(f_{ij}))+(1-y_{ij})\log(1-\sigma(f_{ij}))] \quad (9)$$

where n denotes the number of training samples, $|Y|$ denotes the number of labels in the dataset, y_{ij} denotes the ground truth information as to whether label y_j is present in sample x_i ($y_{ij} = 1$) or not ($y_{ij} = 0$), f_{ij} denotes the output of the CNN corresponding to label y_j of sample x_i and σ denotes the sigmoid activation function.

Comparison Baselines: It is well-established in the multi-label AL literature that querying (image-label) pairs results in much better performance than querying all the labels of a given image [29]. We therefore used five multi-label AL techniques which query (image-label) pairs as comparison baselines in our work: (i) *Random Sampling*, which queries a batch of k (image-label) pairs at random; (ii) *QUIRE*, which selects (image-label) pairs for annotation by simultaneously considering informativeness and representativeness through a min-max optimization framework [8]; (iii) *AUDI*, which selects (image-label) pairs for query based on uncertainty and diversity in the instance space, as well as the label space [10]; (iv) *LMMAL*, which trains a low-rank mapping matrix to identify the relation

[2] https://pytorch.org/hub/pytorch_vision_resnext/



(a) NUS-WIDE (b) MIML (c) COCO

Figure 1. Active Learning performance comparison. Query budget = 200. Best viewed in color.

between the feature space and the label space of a certain multi-label dataset (to exploit full label correlation), and uses uncertainty sampling to query the informative (image-label) pairs [7]; and (v) *CMAL*, which is a very recently proposed cost effective query strategy that first selects the most informative (example-label) pairs by leveraging uncertainty, label correlation and label space sparsity, and then greedily queries the most probable positive (subexample-label) pairs of the selected (example-label) pair [38]. Both *QUIRE* and *AUDI* are widely used baselines in multi-label AL research. *LMMAL* and *CMAL* are two recently proposed methods for multi-label AL. The same deep network architecture was used for all the baseline methods, for fair comparison. We used the macro F1 score as our evaluation metric, as commonly done in multi-label AL research [29].

4.1. Active Learning Performance

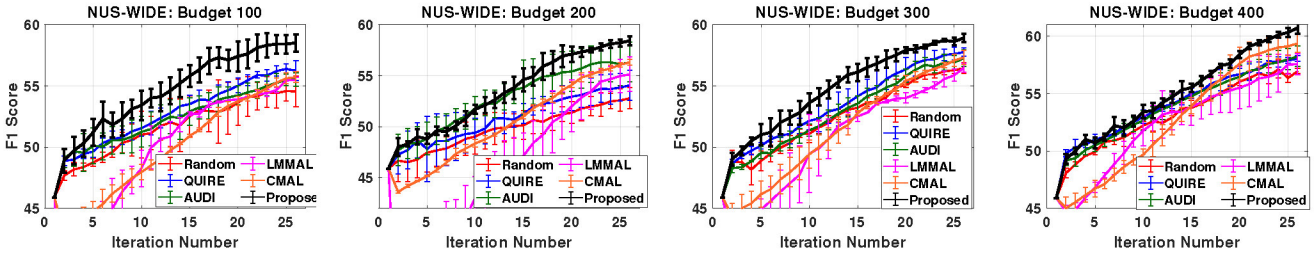
The active learning performance results are depicted in Figure 1. In each graph, the x-axis denotes the AL iteration number and the y-axis denotes the F1 score on the test set. *Random Sampling* depicts erratic performance and attins poor F1 score for the NUS-WIDE and COCO datasets. Among the other methods, *AUDI* seems to outperform *QUIRE* for the three datasets. *CMAL* is not consistent in its performance; it sometimes depicts very good performance, as in the COCO dataset, where it achieves the best F1 score among the baselines; however, its performance is much worse for the NUS-WIDE dataset. The same observation holds for *LMMAL*, which outperforms *Random, QUIRE* and *AUDI* on the COCO dataset, but depicts the worst performance on the NUD-WIDE dataset. All these methods query k (image-label) pairs (in a given AL iteration) by repeatedly querying the optimal pair k times; they do not consider the redundancy in the queried data. This shows that myopically querying multiple (image-label) pairs repeatedly may produce sub-optimal results, due to the redundancy among the queried samples. The proposed method comprehensively outperforms all the baselines across all the three datasets. In almost all the active learning iterations, it depicts the highest F1 score compared to all the baselines, for all the datasets. It also attains the highest F1 score after 25 AL iterations for all the three datasets. Our method explicitly models data redundancy and label correlation in its sampling framework; it thus avoids duplicate sample or label queries and depicts much improved performance consistently. Further, since our method queries batches of (image-label) pairs simultaneously, it can utilize the presence of multiple labeling oracles and can also avoid frequent model updates. These results unanimously corroborate the potential of our framework to identify batches of informative (image-label) pairs, and substantially reduce human annotation effort in training a deep neural network for multi-label classification applications.

4.2. Study of Query Budget

The goal of this experiment was to study the effect of query budget k on AL performance. The results on the NUS-WIDE dataset for query budgets 100, 200, 300 and 400 are shown in Figure 2. Our framework consistently outperforms the baselines across all budgets, showing its usefulness across different query budgets. It attains the highest F1 score for most of the AL iterations compared to all the baselines, across all query budgets. This result is particularly significant from a practical standpoint, where the available query budget is dependent on time, resources and other constraints of an application, and is different for different applications.

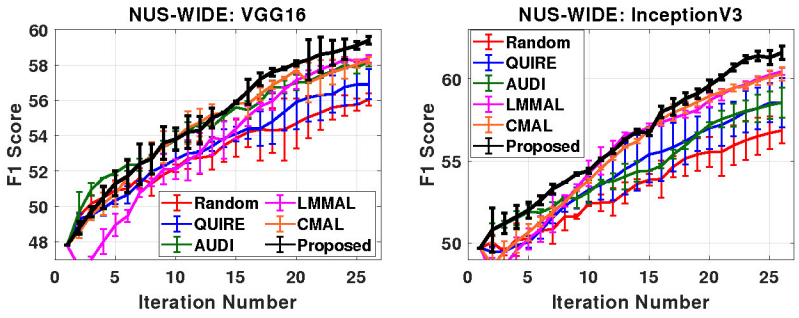
4.3. Study of Network Architecture

In this experiment, we studied the effect of the underlying deep network architecture on AL performance. We used the VGG-16 [23] and InceptionV3 [26] architectures due to their popularity in computer vision. The results on the NUS-WIDE dataset, with query budget 200, are shown in Figure 3. Our framework once again depicts impressive performance for both the network architectures and attains



(a) Budget 100 (b) Budget 200 (c) Budget 300 (d) Budget 400

Figure 2. Study of query budget on the NUS-WIDE dataset. Best viewed in color.



(a) VGG16 as Base Model (b) InceptionV3 as Base Model

Figure 3. Study of network architecture on the NUS-WIDE dataset. Query budget = 200. Best viewed in color.

the highest F1 score after 25 AL iterations. This shows the robustness of our framework to the underlying network architecture.

4.4. Ablation Study

An ablation study of the proposed method on the NUS-WIDE dataset is shown in Figure 4. We compare our method against two cases: (i) $\lambda = 0$, that is, without using the redundancy term R in the objective function (Equation (5)); and (ii) $\beta = 0$, that is, without using the label correlation term in computing the confidence matrix (Equation (3)). We note that, removing either the redundancy term or the label correlation term adversely affects the performance of our method. This shows that it is important to consider both sample redundancy (to avoid querying duplicate images) and label correlations (to avoid querying duplicate labels within a particular image) in our MLAL algorithm for querying (image-label) pairs.

Further results on parameter sensitivity analysis are included in the Supplemental File, due to space constraints.

5. Conclusion and Future Work

In this paper, we proposed a novel batch mode active learning framework for multi-label learning, which poses only binary queries to the human annotators. We posed the (image-label) pair selection as a binary integer programming problem and derived a linear programming relaxation

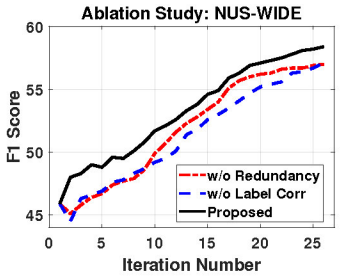


Figure 4. Ablation study on the NUS-WIDE dataset. Best viewed in color.

to identify the images and the corresponding labels which will augment maximal information to the underlying deep neural network. Our empirical evaluations demonstrated the promise and potential of our method in reducing human annotation effort in training a deep neural network for multi-label classification. As part of future work, we plan to study the performance of our framework on other multi-label learning applications, such as text mining, bioinformatics etc.

6. Acknowledgment

This research was supported in part by the National Science Foundation under Grant Number: IIS-2143424 (NSF CAREER Award).

References

[1] A. Casanova, P. Pinheiro, N. Rostamzadeh, and C. Pal. Reinforced active learning for image segmentation. In *International Conference on Learning Representations (ICLR)*, 2020. 1

[2] S. Chakraborty, V. Balasubramanian, and S. Panchanathan. Optimal batch selection for active learning in multi-label classification. In *ACM Multimedia Conference (ACM MM)*, 2011. 2

[3] J. Choi, I. Elezi, H. Lee, C. Farabet, and J. Alvarez. Active learning for deep object detection via probabilistic modeling. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 1

[4] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng. Nus-wide: A real-world web image database from national university of singapore. In *ACM Conference on Image and Video Retrieval (CIVR)*, 2009. 6

[5] C. Coleman, E. Chou, J. Katz-Samuels, S. Culatana, P. Bailis, A. Berg, R. Nowak, R. Sumbaly, M. Zaharia, and I. Yalniz. Similarity search for efficient active learning and search of rare concepts. In *AAAI Conference on Artificial Intelligence*, 2022. 4

[6] M. Ducoffe and F. Precioso. Adversarial active learning for deep networks: a margin based approach. In *International Conference on Machine Learning (ICML)*, 2018. 2

[7] A. Guo, J. Wu, V. Sheng, P. Zhao, and Z. Cui. Multi-label active learning with low-rank mapping for image classification. In *IEEE International Conference on Multimedia and Expo (ICME)*, 2017. 2, 3, 7

[8] S. Huang, R. Jin, and Z. Zhou. Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(10):1936 – 1949, 2014. 2, 6

[9] S. Huang, T. Wang, H. Xiong, J. Huan, and D. Dou. Semi-supervised active learning with temporal output discrepancy. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 2

[10] S. Huang and Z. Zhou. Active query driven by uncertainty and diversity for incremental multi-label learning. In *IEEE International Conference on Data Mining (ICDM)*, 2013. 2, 6

[11] A. Joshi, F. Porikli, and N. Papanikolopoulos. Scalable active learning for multiclass image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(11):2259 – 2273, 2012. 2

[12] X. Li and Y. Guo. Active learning with multi-label svm classification. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2013. 2, 3

[13] X. Li and Y. Guo. Adaptive active learning for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 2

[14] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, L. Zitnick, and P. Dollar. Microsoft COCO: Common objects in context. In *arXiv:1405.0312v3*, 2015. 6

[15] J. Liu, W. Chang, Y. Wu, and Y. Yang. Deep learning for extreme multi-label text classification. In *ACM SIGIR Conference on Information Retrieval*, 2017. 6

[16] W. Liu, H. Wang, X. Shen, and I. Tsang. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(11):7955 – 7974, 2022. 1

[17] Z. Liu, H. Ding, H. Zhong, W. Li, J. Dai, and C. He. Influence selection for active learning. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 2

[18] G. Qi, X. Hua, Y. Rui, J. Tang, and H. Zhang. Two-dimensional active learning for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 2, 3

[19] H. Ranganathan, H. Venkateswara, S. Chakraborty, and S. Panchanathan. Deep active learning for image regression. In *Deep Learning Applications, Springer*, 2020. 1

[20] P. Ren, Y. Xiao, X. Chang, P. Huang, Z. Li, B. Gupta, X. Chen, and X. Wang. A survey of deep active learning. *ACM Computing Surveys*, 54(9), 2021. 1

[21] O. Reyes, C. Morell, and S. Ventura. Effective active learning strategy for multi-label learning. *Neurocomputing*, 273:494–508, 2018. 2

[22] C. Shui, F. Zhou, C. Gagne, and B. Wang. Deep active learning: Unified and principled method for query and training. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020. 2

[23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 7

[24] S. Sinha, S. Ebrahimi, and T. Darrell. Variational adversarial active learning. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[25] S. Sridhar, V. Bittorf, J. Liu, C. Zhang, C. Re, and S. Wright. An approximate, efficient solver for LP rounding. In *Neural Information Processing Systems (NeurIPS)*, 2013. 5, 6

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 7

[27] S. Vempala. The random projection method. In *Americal Mathematical Society*, 2004. 5

[28] J. Wu, A. Guo, V. Sheng, P. Zhao, Z. Cui, and H. Li. Adaptive low-rank multi-label active learning for image classification. In *ACM Multimedia Conference (ACM MM)*, 2017. 2, 3, 6

[29] J. Wu, V. Sheng, J. Zhang, H. Li, T. Dadakova, C. Swisher, Z. Cui, and P. Zhao. Multi-label active learning algorithms for image classification: Overview and future promise. *ACM Computing Surveys*, 53(2):1 – 35, 2020. 2, 6, 7

[30] J. Wu, C. Ye, V. Sheng, J. Zhang, P. Zhao, and Z. Cui. Active learning with label correlation exploration for multi-label image classification. *The Institution of Engineering and Technology (IET)*, 11(7):577 – 584, 2017. 2, 3, 4, 6

[31] J. Wu, S. Zhao, V. Sheng, J. Zhang, C. Ye, P. Zhao, and Z. Cui. Weak-labeled active learning with conditional label dependence for multilabel image classification. *IEEE Transactions on Multimedia*, 19(6):1156 – 1169, 2017. 2, 3, 6

[32] J. Wu, S. Zhao, V. Sheng, P. Zhao, and Z. Cui. Multi-label active learning for image classification with asymmetrical conditional dependence. In *IEEE International Conference on Multimedia and Expo (ICME)*, 2016. 2, 3

[33] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 6

[34] B. Yang, J. Sun, T. Wang, and Z. Chen. Effective multi-label active learning for text classification. In *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2009. 2

[35] C. Ye, J. Wu, V. Sheng, P. Zhao, and Z. Cui. Multi-label active learning with label correlation for image classification. In *IEEE International Conference on Image Processing (ICIP)*, 2015. 2

[36] C. Ye, J. Wu, V. Sheng, S. Zhao, P. Zhao, and Z. Cui. Multi-label active learning with chi-square statistics for image classification. In *ACM International Conference on Multimedia Retrieval (ICMR)*, 2015. 2

[37] D. Yoo and I. Kweon. Learning loss for active learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2

[38] G. Yu, X. Chen, C. Domeniconi, J. Wang, Z. Li, Z. Zhang, and X. Zhang. Cmal: Cost-effective multi-label active learning by querying subexamples. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 34(5):2091 – 2105, 2022. 2, 3, 7

[39] B. Zhang, Y. Wang, and F. Chen. Multilabel image classification via high-order label correlation driven active learning. *IEEE Transactions on Image Processing (TIP)*, 23(3):1430 – 1441, 2014. 2, 3

[40] B. Zhang, Y. Wang, and W. Wang. Batch mode active learning for multi-label image classification with informative label correlation mining. In *IEEE Workshop on the Applications of Computer Vision (WACV)*, 2012. 2, 3

[41] B. Zhang, Y. Wang, and W. Wang. Batch mode active learning for multi-label image classification with informative label correlation mining. In *IEEE Workshop on the Applications of Computer Vision (WACV)*, 2012. 3

[42] M. Zhang and Z. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 26(8):1819 – 1837, 2014. 1

[43] Y. Zhang. Multi-task active learning with output constraints. In *AAAI Conference on Artificial Intelligence*, 2010. 2

[44] S. Zhao, J. Wu, V. Sheng, C. Ye, P. Zhao, and Z. Cui. Weak labeled multi-label active learning for image classification. In *ACM Multimedia Conference (ACM-MM)*, 2015. 2

[45] Z. Zhou and M. Zhang. Multi-instance multi-label learning with application to scene classification. In *Neural Information Processing Systems (NeurIPS)*, 2006. 6

[46] J. Zhu and J. Bento. Generative adversarial active learning. In *Advances of Neural Information Processing Systems (NeurIPS) Workshops*, 2017. 2