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

Debanjan Goswami and Shayok Chakraborty

Department of Computer Science, Florida State University

ABSTRACT

In this Supplemental File, we include: (*i*) number of samples in the training, unlabeled and test sets for each dataset; (*ii*) a parameter sensitivity study; and (*iii*) a schematic diagram of our binary query active learning algorithm.

1. NUMBER OF SAMPLES IN THE TRAINING, UNLABELED AND TEST SETS

The number of samples in the initial training set L, unlabeled set U and test set for each dataset used in our experiments, are detailed in Table 1. The table also shows the query budget (number of (image-label) pairs queried in each active learning iteration) for each dataset.

Dataset	Training	Unlabeled	Test	Query Budget
NUS-WIDE	1,500	3,000	1,000	200
MIML	600	1,000	400	40
COCO	1,200	1,600	800	60

Table 1. Number of samples in the initial training, unlabeled and test sets, and query budget for all the datasets.

2. PARAMETER SENSITIVITY STUDY

In this section, we studied the effects of the weight parameters λ and β on the active learning performance. λ was used as a weight parameter to govern the relative importance of the confidence and the redundancy terms in the objective function (Equation (5) in the main paper):

$$\min_{Q,w} \quad (QC) + \lambda w^\top R w$$

The results on the NUS-WIDE dataset with $\lambda = \{0, 0.025, 0.5\}$ are depicted in Figure 1. We note that a high value of λ such as $\lambda = 0.5$ depicts comparatively inferior performance. Thus, putting too much emphasis on the redundancy term adversely affects the learning performance. Further, $\lambda = 0$ also depicts inferior learning performance. This shows that completely removing the redundancy term also adversely affects the learning performance, due to the duplication of information in the queried data. Thus, placing adequate importance on the redundancy term is necessary in improving the active learning performance. $\lambda = 0.025$ depicts the best performance in this experiment.

The parameter β was used as the weight of the label correlation matrix μ in the computation of the confidence matrix C (Equation (3) in the main paper):

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

The results on the NUS-WIDE dataset with $\beta = \{0, 0.25, 0.5, 1\}$ are depicted in Figure 2. We note that $\beta = 0$ depicts the worst performance. This shows that ignoring the label correlations while actively querying the (image-label) pairs may adversely affect the learning performance. The performance improves with non-zero values of β, which enables our model to appropriately exploit the label correlations, and avoid querying duplicate labels.

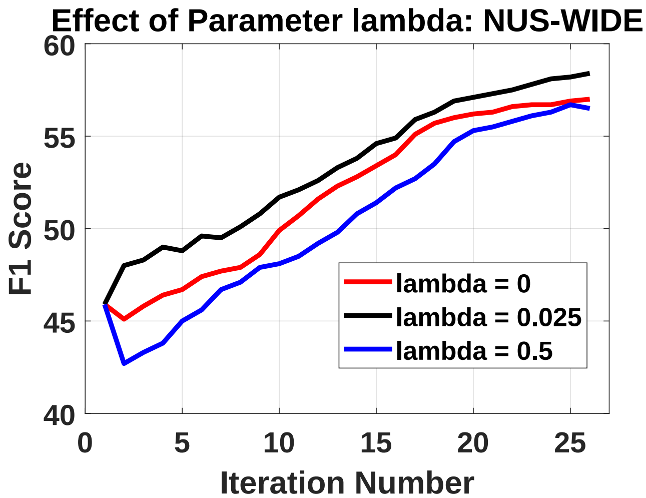

Figure 1. Effect of the weight parameter λ on the NUS-WIDE dataset. Best viewed in color.

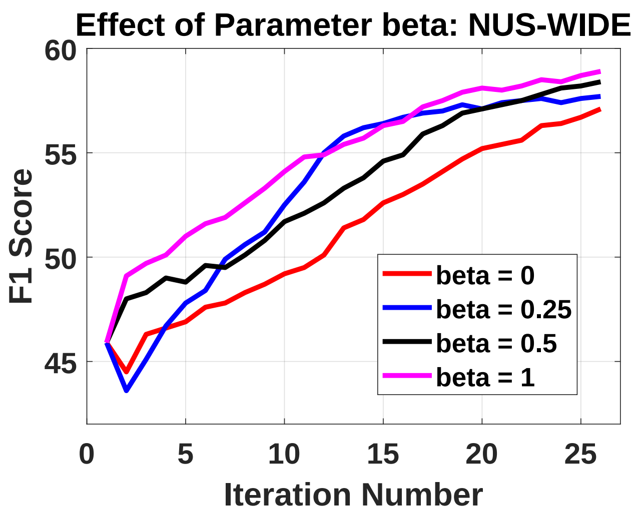

Figure 2. Effect of the weight parameter β on the NUS-WIDE dataset. Best viewed in color.

3. SCHEMATIC DIAGRAM OF OUR METHOD

A schematic diagram of our proposed algorithm is depicted in Figure 3. A deep neural network is first trained on the initial training set. In each active learning iteration, our proposed framework uses the predictions of the deep model on the unlabeled samples to select a batch of k (image-label) pairs to pose the binary queries. This is achieved by solving a binary matrix Q using a linear programming (LP) solver (detailed in Section 3 of the main paper). The binary annotations are obtained from the labeling oracles and the newly annotated images and labels are appended to the training set. The deep neural network is updated and the process is continued iteratively until a stopping criterion is satisfied.

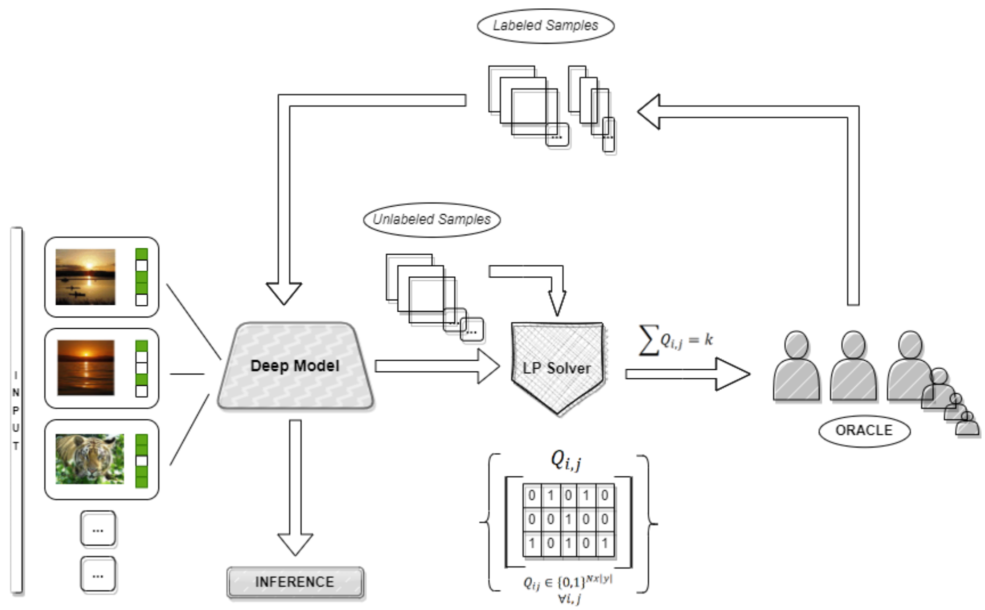

Figure 3. Schematic diagram of our binary query active learning algorithm.