Partial Label Learning meets Active Learning: Enhancing Annotation Efficiency through Binary Questioning

Shivangana Rawat, Chaitanya Devaguptapu, Vineeth N Balasubramanian
cs20mtech12001@iith.ac.in, email@chaitanya.one, vineethnb@iith.ac.in
Indian Institute of Technology Hyderabad, Fujitsu Research India

Abstract

Supervised learning is a powerful method in machine learning, but the acquisition of labeled data can be costly. To address this challenge, two techniques, namely active learning (AL) and partial label learning (PLL), have been developed to decrease the expenses associated with annotation in supervised learning. In this context, we present a novel approach that combines AL and PLL techniques, aiming to enhance the efficiency of annotation.

Motivation

Active learning (AL) is a strategic approach aimed at optimizing the annotation budget by carefully selecting and labeling the most informative samples. In our research, we propose a hypothesis that suggests an enhancement in annotation efficiency can be achieved by posing more detailed and specific questions to the oracle.

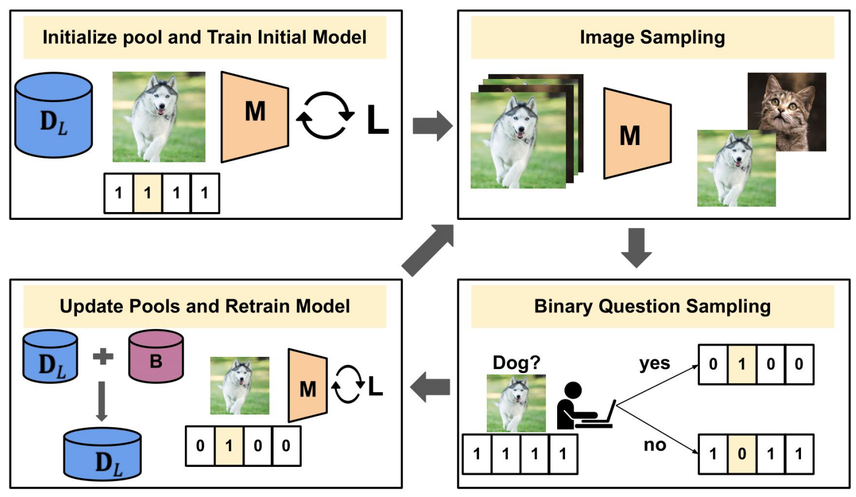
Methodology

The process involves presenting the oracle with an (image, class) pair and asking the question, "Does the specific image belong to the given class?" The oracle provides a binary response of either “yes” or “no”.

- **When the binary question receives a "yes" response, we identify the true label, thereby finding a single label within our candidate set for that image which becomes the true label.**
- **When the binary question receives a "no" response, we can eliminate that particular class from the candidate set associated with the image.**

Binary Questions are strategically chosen that have a higher likelihood of receiving a "yes" response using model feedback.

Partial Label Learning strategies are used to utilize the partial labels generated as a result of “no” responses.



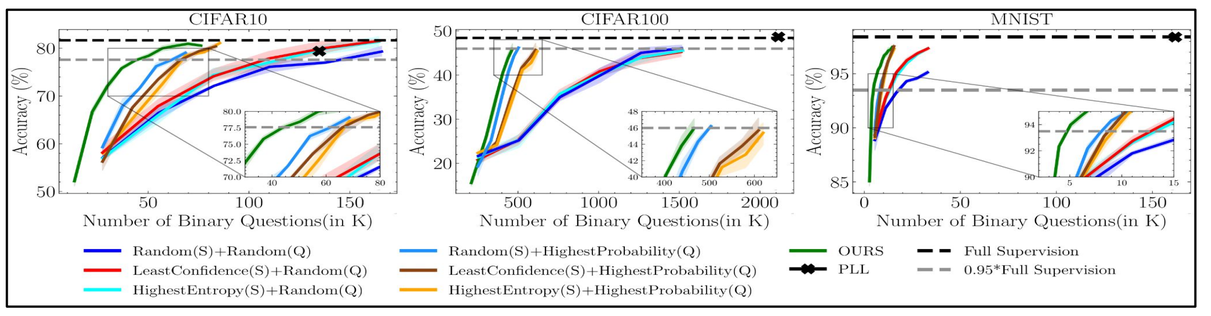
Experimental Details

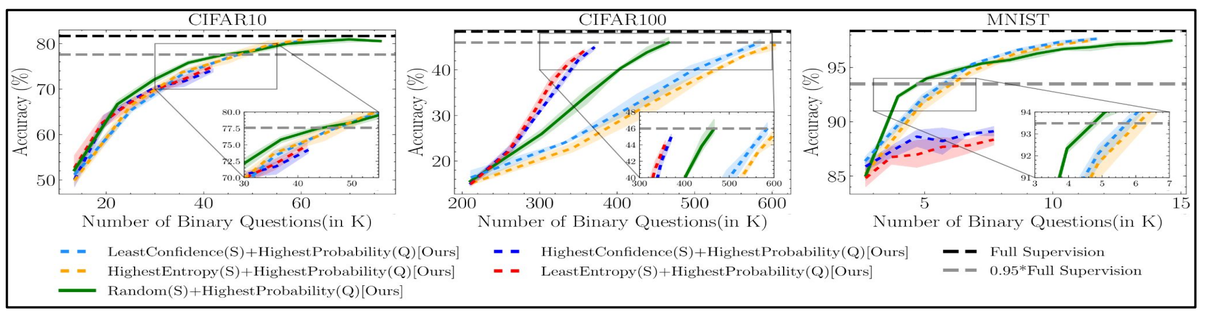
- **Image Sampling Strategies : Random, Confidence, Entropy**
- **Question Sampling Strategies: Random, Highest Confidence**
- **Datasets: CIFAR10, CIFAR100, MNIST**
- **Evaluation Criterion: Top-1 accuracy vs the number of binary questions required to achieve it.**
- **Learning from Partial Labels: PRODEN**

Results

Our method outperforms all baseline approaches.

Selecting easy samples initially improves performance for datasets with large number of classes.





Key Contribution: **We provide a novel integration of active learning and partial label learning and introduce a binary questioning mechanism to reduce the annotations costs without compromising on the accuracy.**