Importance of Negative Sampling in Weak label learning

Ankit Shah*, Fuyu Tang, Zelin Ye, Rita Singh, Bhiksha Raj

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

aps1, bhiksha@andrew.cmu.edu

Introduction

- **Weak-label learning**: A challenging task involves learning from data "bags" containing both positive and negative instances, with only the bag labels known.
- **Importance of Sampling**: The pool of negative instances is typically larger than the positive instances, making the selection of the most informative negative instances crucial for performance.
- **Open Problem**: The selection strategy for negative instances in weak-label learning hasn't been extensively studied.
- **Our Contribution**: We introduce several sampling strategies of negative instances for weak-label learning, showcasing its importance in achieving improved classification performance and computational efficiency.

Problem Illustration

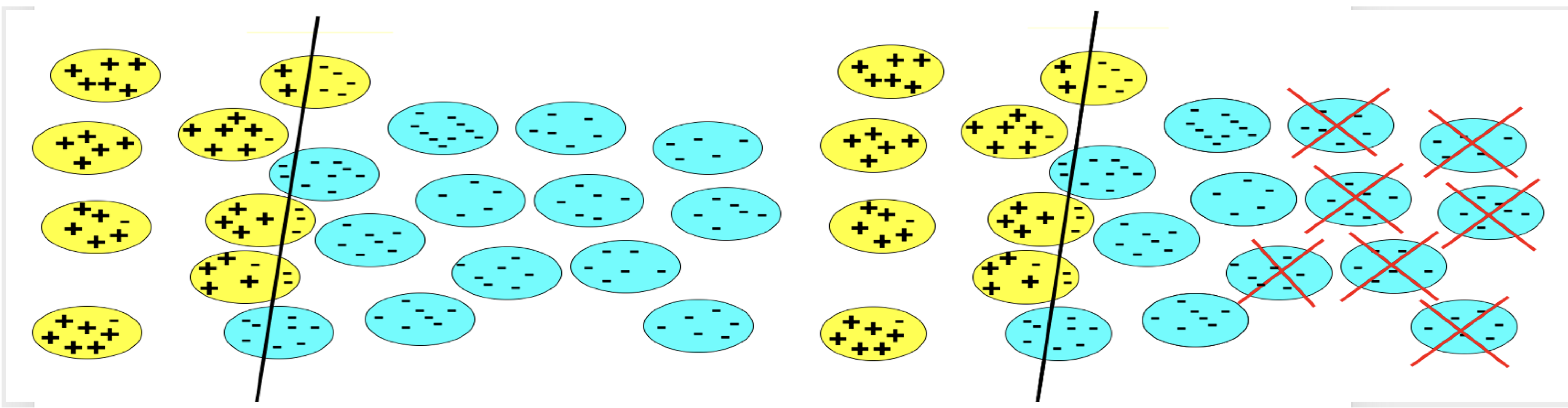

Figure 1: Illustration of the negative sampling bags which contribute to the decision boundary a) Random sampling considers all bags and b) Selective sampling strategy automatically figures bags not important, thus not sampled for training.

DataSet and Negative Sampling Methods

- Test our methods on two datasets: CIFAR-10 and AudioSet. CIFAR-10 for image classification, containing 60,000 32x32 color images in 10 classes. AudioSet is a large-scale dataset of manually annotated audio events, spanning a wide range of real-world sounds.
- Random Sampling: Randomly select k bags with uniform distribution at each epoch. Each negative bag has an equal chance of being selected. This strategy is the baseline for all the sampling strategies.
- Gradient Embedding: A strategy that uses gradient information to determine the importance of negative samples.

$$g_b = \frac{\partial}{\partial \theta_{\text{out}}} \text{Loss}(f(x;\theta), \hat{y}(b_i)) \tag{1}$$

$$S_G^- = \underset{b_i \in D_{B^-}}{\text{argmax}} \|g_b\|_2 \tag{2}$$

- Entropy-based Approach: This method selects negative samples based on the uncertainty or entropy of their predictions

$$S_E^- = \underset{b_i \in D_{B^-}}{\text{argmax}} - \sum P_\theta(\hat{y}_c|b_i) \log(P_\theta(\hat{y}_c|b_i)) \tag{3}$$

- SVM-based Margin Strategy: A strategy that leverages the margin concept from Support Vector Machines to select negative samples.

$$S_M^- = \underset{b_i \in D_{B^-}}{\text{argmin}} |P_\theta(\hat{y}_{positive}|b_i) - P_\theta(\hat{y}_{negative}|b_i)| \tag{4}$$

- BADGE Strategy: The strategy combines uncertainty and diversity by picking bags according to gradient embedding as centers of k clusters in the pool. Bag with the largest gradient embedding are selected and K-MEANS++ helps incorporate more diverse sample.

Results

- For CIFAR-10 and CIFAR-100 for gradient embedding method doesn't improve over than traditional random sampling whereas BADGE, SVM-Margin, and KL-prob consistently outperform. Negative sampling strategies show improvement in Audioset in majority of classes amongst top 40 classes.

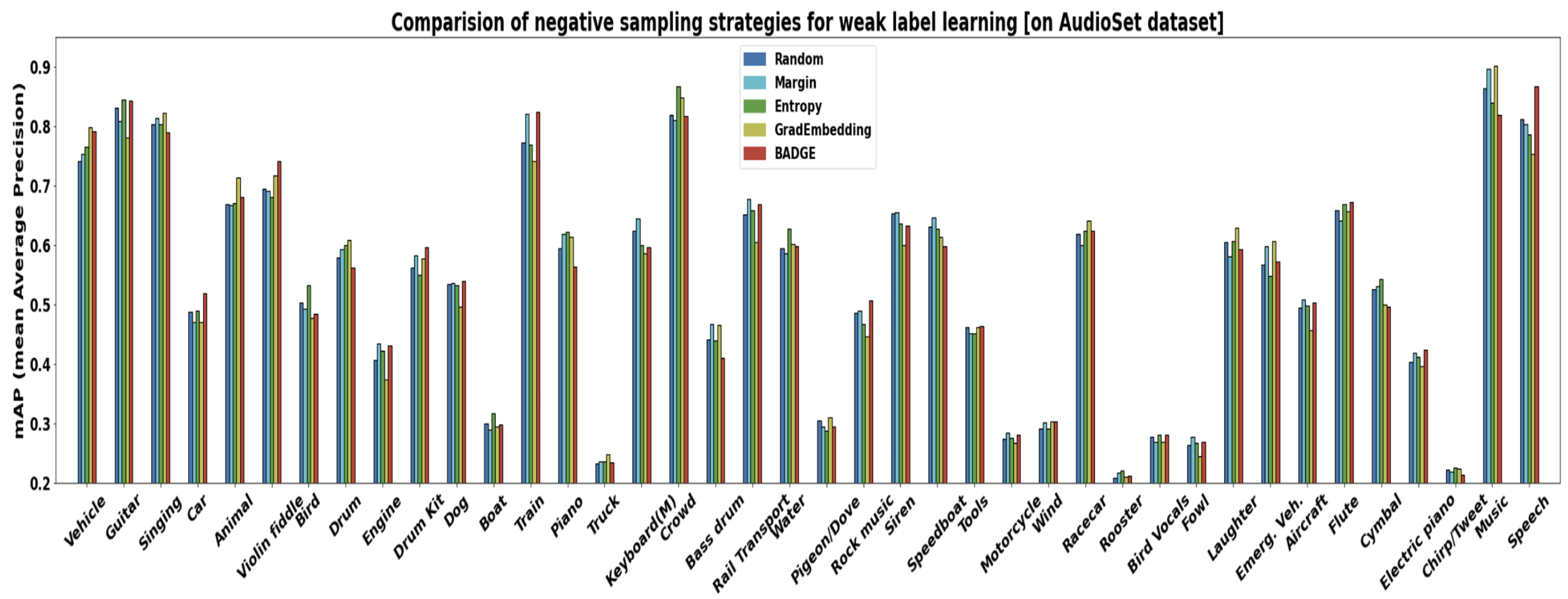

Figure 2: Comparison of negative sampling strategies for weak label on AudioSet

Conclusions

- Gradient Embedding and BADGE Sampling consistently outperform Random, Margin and Entropy based sampling. They quantify similarities and differences between positive and negatives allowing for selection of informative samples for weak label classification
- All the proposed sampling strategies provide an overall improvement of at least 40% of all classes on AudioSet.