Write Up

Marlon Garay

November 2024

1 Introduction

Using ZORRO as a test measure, we determine the robustness of a dataset in response to worst-case data uncertainty. By fine-tuning error injection and robustness models, we can determine the susceptibility of a dataset to misrepresenting the ground truth. This research aims to quantify how robust a dataset and also to develop a method to reverse engineer error uncertainty, capturing the ground truth as closely as possible.

2 Data Uncertainty

ZORRO is used as a test measure to make worst-case data uncertainty by injecting errors into the training data. The percentage of uncertain labels (*uncertain_pct*) with custom and multiple uncertainty radius are varied to simulate errors. These uncertainties are injected, reflecting how variations in the dataset can potentially lead to incorrect model outputs or misrepresentation of the underlying data.

3 Measure Robustness

For each combination of uncertainty percentage and radius, the robustness ratio is computed. This ratio measures how well the model performs even with errors introduced by the label uncertainty. The naive and subset approaches are used to measure robustness, focusing on both the overall dataset such as boundary indices of the data, which are more likely to be affected by uncertainty. This methodology assesses the susceptibility of the dataset to providing incorrect predictions, showing how the model's performance changes as a result of uncertain labels.

4 Quantifying Misrepresentation

To measure susceptibility to misrepresentation, the error injection parameters, such as uncertain percentage and radius, are fine-tuned to determine how vulnerable the model is to incorrect representations of the ground truth. This process helps identify whether the dataset accurately reflects real-world patterns, especially under high data uncertainty. The robustness model assesses how resilient the model is to this uncertainty, offering a quantitative measure of robustness that indicates the model's ability to maintain performance despite potential data flaws.

5 Heat map Visualization

The results are visualized as heatmaps. The robustness ratio is shown as a function of the uncertain percentage and the uncertainty radius, helping to reveal the threshold at which the model becomes unstable or overly sensitive to the injected uncertainty.

6 Inspiration from GOPHER

The approach takes concepts form Gopher, as both seek to find model performance issues caused by "problematic" data subsets. Gopher identifies data subsets that contribute to bias in the model. This code identifies subsets that degrade robustness due to label uncertainty. Both aim to evaluate issues in the data that impact model performance; however, Gopher focuses on bias and fairness; this code is more focused on robustness under uncertainty. This code focuses on ensuring the correctness of model predictions under uncertain data conditions.