CS754 Assignment-2

Saksham Rathi, Ekansh Ravi Shankar, Kshitij Vaidya

Declaration: The work submitted is our own, and we have adhered to the principles of academic honesty while completing and submitting this work. We have not referred to any unauthorized sources, and we have not used generative AI tools for the work submitted here.

Question 3

Research Paper Chosen

We have chosen the research paper Neural Group Testing to Accelerate Deep Learning by authors **Weixin Liang and James Zou** from Stanford University. The paper was published on 9th May 2021 (a copy attached in this folder).

The specific ML/DS problem targeted in the paper

The paper addresses the computational challenges associated with deep neural networks, particularly the high inference costs when processing large datasets. The authors propose a method to accelerate neural network inference by testing groups of samples simultaneously, thereby reducing the number of forward passes required. This approach is especially beneficial in tasks like image moderation, where the goal is to detect rare inappropriate images efficiently. They found that neural group testing can group up to 16 images in one forward pass and reduce the overall computation cost by over 73%.

Some finer details

Groups of samples that test negative are ruled out, which saves testing many people individually. If a group tests positive, samples in that group are then retested adaptively.

Unknown Signal Vector: Represents the set of input samples, each labeled as either containing the target feature (e.g., an inappropriate image) or not.

Pooling Matrix: Defines how input samples are combined into groups for testing. Each row corresponds to a group test, and each column represents an input sample. An entry of '1' indicates the inclusion of a sample in a particular group, while '0' indicates its exclusion.

Measurement Vector: Contains the outcomes of the group tests. A positive result indicates that at least one sample in the group possesses the target feature, while a negative result signifies that none of the samples in the group have the feature.

Algorithm

Neural Group Testing modifies a deep neural network to test multiple samples in one forward pass while preserving accuracy. The goal is to determine whether at least one positive sample exists in a batch, rather than classifying each sample individually. It has two key components:

- **Neural Group Testing Network Design** Given an individual testing network $\phi : X \to Y$ that classifies samples individually, the goal is to construct a group testing network Φ that takes a set of M samples, $X = \{x_1, x_2, \ldots, x_M\}$, and predicts a positive label if at least one sample in the group is positive:

$$y = \max_{1 \le m \le M} y_m$$

where y_m is the individual ground-truth label for sample x_m. Three designs for modifying the network are proposed

  - Design 1: Pixel Merge
    - Inspired by Mixup, this method averages multiple images at the pixel level:

    $$x_{\text{merged}} = \frac{1}{M} \sum_{m=1}^{M} x_m$$

    - The model is fine-tuned to predict a positive label if any original sample was positive.
    - It is computationally simple but less effective due to information loss.

  - Design 2: Feature Merge - Instead of merging at the pixel level, feature representations are extracted using the first T layers of the network:

    $$\phi(x_m) = f^{(T)} \circ f^{(T-1)} \circ \cdots \circ f^{(1)}(x_m)$$

    - The aggregated feature is then passed through the remaining network layers:

    $$\Phi(X) = \rho \left(\text{aggregate} \left(\{\phi(x_m)\}_{m=1}^{M} \right) \right)$$

    - The aggregation step (e.g., max-pooling, averaging) ensures permutation invariance.

  - Design 3: Tree Merge - A hierarchical structure is used, recursively merging feature representations in a tree structure. - Each level aggregates features of two or more samples before passing them through deeper layers. - This reduces computation significantly while maintaining accuracy.

- Neural Group Testing Algorithms To efficiently schedule testing, different group testing strategies are explored:

  - Algorithm 1: Two-Round Testing - Initial tests are performed on groups of size M. - If a group is positive, individual samples within it are tested.

  - Algorithm 2: Multi-Round Testing - If a group is positive, it is recursively divided into smaller subgroups until individual samples are tested.

  - Algorithm 3: One-Round Testing - Each sample is tested in two overlapping groups, and results are decoded based on group responses.