“Hariram, Veeramani”
Review for Paper Titled:

“Dataset Reduction Technology based on Mutual Information for Black-box Attacks”

1- Summary

This work implements an optimized algorithm to generate a particular subset of data samples from a huge dataset wherein the redundancy among the dataset samples given by ensemble Mutual Information Metric is minimized and would also be sufficient for performing Black-box attacks.
To accomplish Dataset Reduction- this work models dataset as an undirected weighted complete graph with n vertices and proposes an approximation solution for the NP hard problem of finding an induced graph with k vertices (n<k) of which the sum of edge weights is minimal. This in turn translates to the criteria of the minimized dataset needed for less expensive query efficient, effective/powerful Black Box-attacks.
Specifically, Data Reduction Technique proposed here reduces the dataset size& mutual information among the data samples while retaining the accuracy of the generated attacker’s rival model independent of the Model’s domain or dataset under consideration.
The Proposed Technique outperforms previously reported works both in terms of Attack Accuracy and Transferability.

2- Novelty/Contribution

The idea of modelling the problem statement in the form of a NP Hard Graph problem and arriving at a new metric DRMI proposed here is what I think is the key element of novelty here. Though, there have been many data reduction techniques/ similarity metrics reported previously such as Correlation Matrix (CMAL), Class Probability of Prediction (CPB) and Traces of Activated Neurons (TRACE), DRMI gives superior performance than the rest of the techniques.

3- Evaluation

Evaluation is carried on for different datasets and explored for different Dataset Degrees and DRMI exhibits superior performance while operating on different datasets. For instance, both on comparatively simpler datasets such as CIFAR10 and much more complex Datasets like ImageNet. Also, the Key result is DRMI’s reduction technology works well, and it keeps up the performance even when the attacker’s dataset does not match the distribution of the target training dataset

4- Unique Strength/Weakness

Many adversaries execute Black-box attacks by relying on the querying capabilities, but the queries are in general limited and costly, the proposed Data reduction technology would open the door of myriads of attacks to operate in a query efficient way by exploiting redundancies both in the dataset and in Queries.