Active Learning Accounting for Model Uncertainty in the Rashomon Set International Conference on Machine Learning (ICML 2026)

Simon Nguyen [1] **Tyler McCormick** [1] **Cynthia Rudin** [2]

Abstract

ICML 2026 is in South Korean let's go bb

1. Introduction

Machine learning classification models seek to find the relationship between covariates and labels. Typically, a dataset consisting of both covariates and observed labels are used to train a model. However, in many applications collecting labels may be expensive. Such examples can be seen examples. In such scenarios, strategically determining which observations merit labeling will greatly improve the learning of covariate-label relationship.

Active learning allows researchers to make the most of constrained budgets by adaptively choosing which observations to label. The key task in active learning is choosing the most informative observations that will enhance the predictive quality of the model. One common metric of informativity is uncertainty (ref). Popular methods such as Marginal Sampling ref, Best-versus-Second Best (BvSB) (ref), MaxiMum Confidence Uncertainty (MCU) (ref), Multiple Peak Entropy (ref), and QBC (ref) are all popular metrics of predictive uncertainty. Yet, these methods only consider uncertainty within the model. From Madigan 1996 (ref), ignoring model uncertainty may lead to over-confident inference.

The current literature in active learning only consider uncertainty in model prediction, ignoring model uncertainty. Whereas the current active learning methods only consider the predicted label of the best performing model, there exists a multiplicity of good models. The phenomenon of a dataset admitting a multitude of different, near-equal models is termed as "Rashomon Effect" in Leo Breiman's seminal 2001 paper.

A deep ramification of the Rashomon Effect is predictive multiplicity, in which different models in the Rashomon set produce different, often opposing, label predictions. The machine learning community has only recently been realizing the detrimental effect of ignoring the Rashomon Effect. Some references commenting on the effect of ignoring the Rashomon effect eg. Rudin, 2024...

In this work, we propose an active learning algorithm that considers model uncertainty by accounting for the Rashomon Effect. We construct a Rashomon set of predictions

2. Active Learning Overview

We borrow notation from Liu et. al (2022) (ref). Say observation i is observef with data $(\mathbf{x}_i, y_i)$ for vector $\mathbf{x}_i$ in covariate space $\mathcal{X}$ and label y_i in output space $\mathcal{Y}$. The data is sent through a supervised learning model $F_\theta(\cdot) : X \to Y$ I think later down the road I might want to use $F_\theta^{(n)}$ or $F_{\theta^{(n)}}$ parameterized by θ. The goal of supervised learning is to learn θ that, in our instance, reduces the loss. Typically, the model parameter is learned from a training dataset a training dataset $D_{tr} = \{(\mathbf{x}_i, y_i)\}_{i=1}^I$ and the model is then tested by an independent dataset $D_{ts} = \{(\mathbf{x}_j, y_j)\}_{j=1}^J$.

Active learning seeks to adaptively and strategically choose which unlabelled observations should be sent to oracle labeling and then used in the supervised learning model. Denote the reservoir of unlabelled candidate observations as $D_{cdd}^{(n)} = \{(\mathbf{x}_k, y_k)\}_{k=1}^K$ with y_k initially unknown. A selector is the strategy used to select samples from $D_{cdd}^{(n)}$ to be labelled by expert knowledge. Denote the selector as $S_\psi(\cdot)$ parameterized by ψ. At each iteration $S_\psi(\cdot)$ will select a subset of observations, denoted $B^{(0)}$, from $D_{cdd}^{(n)}$ to send to oracle labeling and add to the training training set. The model is then retrained on the new training set and reparameterized as $F_{\theta^{(n)}}$. The process is repeated, gradually expanding the training set with informative observations, until a desired classification threshold is met. A typical active learning workflow is presented in Figure (2).

The choice of $B^{(0)}$ is chosen so as to find the observations that are most informative to improving F_θ. Following the work of Shu et. al (2019) (ref), we consider two metrics:

*Equal contribution [1]Department of Statistics, University of Washington, Seattle, USA [2]Departments of Computer Science, Durham, North Carolina, USA. Correspondence to: Simon Nguyen <simondn@uw.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

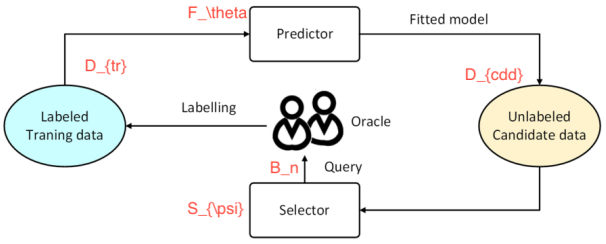

Figure 1. A standard active learning workflow.

uncertainty and representativeness.

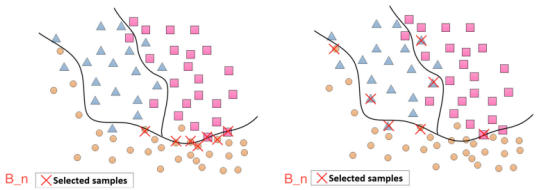

Figure 2. Figure (a) exemplifies a selector that recommends samples for oracle labeling soley based on uncertainty. Figure (b) exemplifies a selector that recommends samples for oracle labeling based on both unceratinty and representativeness.

3. Rashomon Theory Overview

4. Algorithm

5. Applications

Actually hmm these data sources are very clear cut:

- Youtube Spam Collection
- SMS Spam Collection
- Sundanese Twitter Dataset
- Sentiment Labelled Sentences

6. Questions

1. OHHHH so is the Rashomon set being constructed for
   (a) The set of predictions in the training data set from $F_\theta(\cdot)$
   (b) The set of observations that are recommended by the selector based on the similarity distance between D_{cdd} and $\hat{y}_{tr}$

Algorithm 1 Rashomon Active Learning

input $D_{cdd}^{(0)}$: Initial training data
output $F_\theta(\cdot)$: Model predictor

Initialize $D_{tr}^{(0)}$:
repeat
  Randomly select observations $B^{(0)}$ from $D_{cdd}^{(0)}$.
  Send $B^{(0)}$ for oracle labeling.
until At least one observation from each class label is represented.

Active Learning:
repeat
  Train $F_{\theta^{(n)}}$ on $D_{tr}^{(n)}$
  Predict labels for $\mathbf{x}_{tr}^{(n)}$ using $F_\theta(\cdot)$ to obtain $\hat{y}_{tr}^{(n)} = F_{\theta^{(n)}}(\mathbf{x}_{tr}^{(n)})$
  For each sample in $(\mathbf{x}_{tr}^{(n)}, \hat{y}_{tr}^{(n)})_{i=1}^{I}$, compute the uncertainty value probably Breaking Ties value ϵ_i:

$$\epsilon_i := \max_{c\in\mathcal{C}} \mathbb{P}(\hat{y}_{tr,i}^{n} = c|\mathbf{x}_{tr,i}^{(n)}) - \max_{c\in\mathcal{C}\setminus c^+} \mathbb{P}(\hat{y}_{tr,i}^{n} = c|\mathbf{x}_{tr,i}^{(n)})$$

  such that

$$c^+ := \arg\max_{c\in\mathcal{C}} \mathbb{P}(\hat{y}_{tr,i} = c|\mathbf{x}_{tr,i})$$

  Arrange ϵ_i in ascending order and select the top m observations with most predictive uncertainty: $(\tilde{\hat{y}}_{tr,i}^{(n)})_{i=1}^{m}$
  Compute similarity matrix between $(\tilde{\hat{y}}_{tr,i}^{(n)})_{i=1}^{m}$ and observations in $D_{cdd}^{(n)}$.
  Resample $B^{(n)}$ based on the top uncertainty while considering representation.
  Send $B^{(n)}$ to oracle labeling.
  Set $D_{tr}^{(n+1)} = D_{tr}^{(n)} \cup B^{(n)}$ and $D_{cdd}^{(n+1)} = D_{cdd}^{(n)} \setminus B^{(n)}$.
until Until $F_\theta(\cdot)$ is good lol

2. How are we going to measure uncertainity between models? Are we going to look at predictive multiplicity between models in the Rashomon set? Or are we going to some how combine the uncertainity metrics of each model in the Rashomon set?
3. Active learning paper with Rashomon adjustment
4. Weighting of uncertainty and diversity metrics?
5. Membership Query Synthesis/Data Augmentation query or Pool-Based Sampling (PBS)?

References

A. You *can* have an appendix here.