PRISM: A Rich Class of Parameterized Submodular Information Measures for Guided Subset Selection

Suraj Kothawade, Vishal Kaushal, Ganesh Ramakrishnan, Jeff Bilmes, Rishabh Iyer

THE UNIVERSITY OF TEXAS AT DALLAS

UNIVERSITY of WASHINGTON

INTRODUCTION

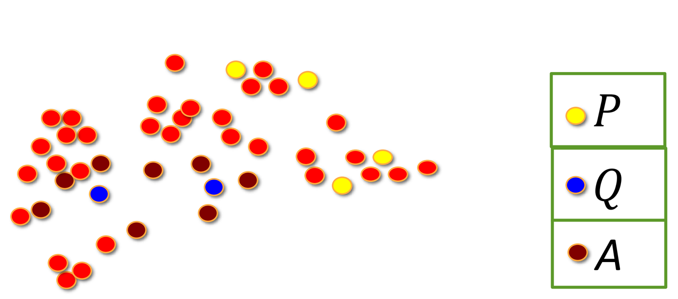
- The task of guided subset selection is to select subsets of a specific kind.
- Guided subset selection is useful for **targeted learning** and **guided summarization**.
- We propose parameterized submodular information measures which can be used to target a certain slice of data that is critical for such applications.
- We empirically demonstrate the performance of guided data subset selection for targeted learning - improving the performance on an image classification task for imbalanced datasets, and for various flavors of summarization.

PROBLEM FORMULATION

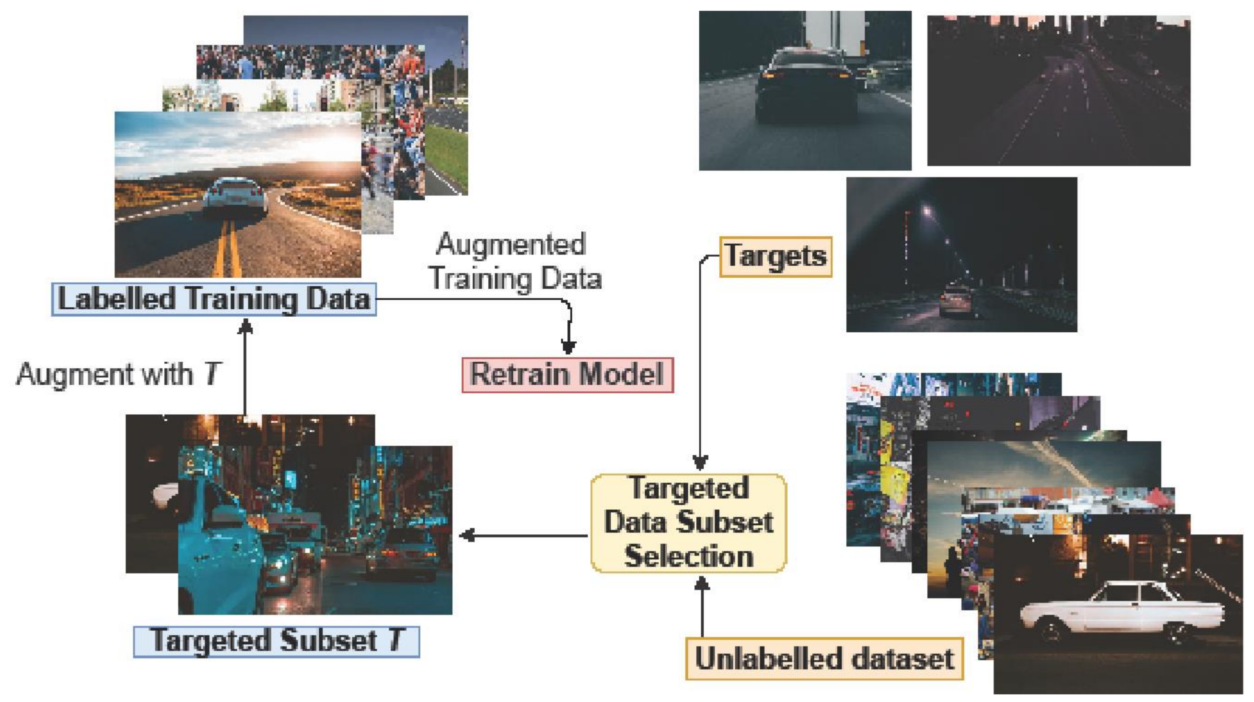
Goal: To select a "guided/targeted" data subset for improving data imbalance or accuracy of the task DNN.

The Submodular Conditional Mutual Information (SCMI) is defined as $I_f(A;Q|P) = f(A \cup P) + f(Q \cup P) - f(A \cup Q \cup P) - f(P)$. It jointly models the similarity between A and Q, and their dissimilarity with P.

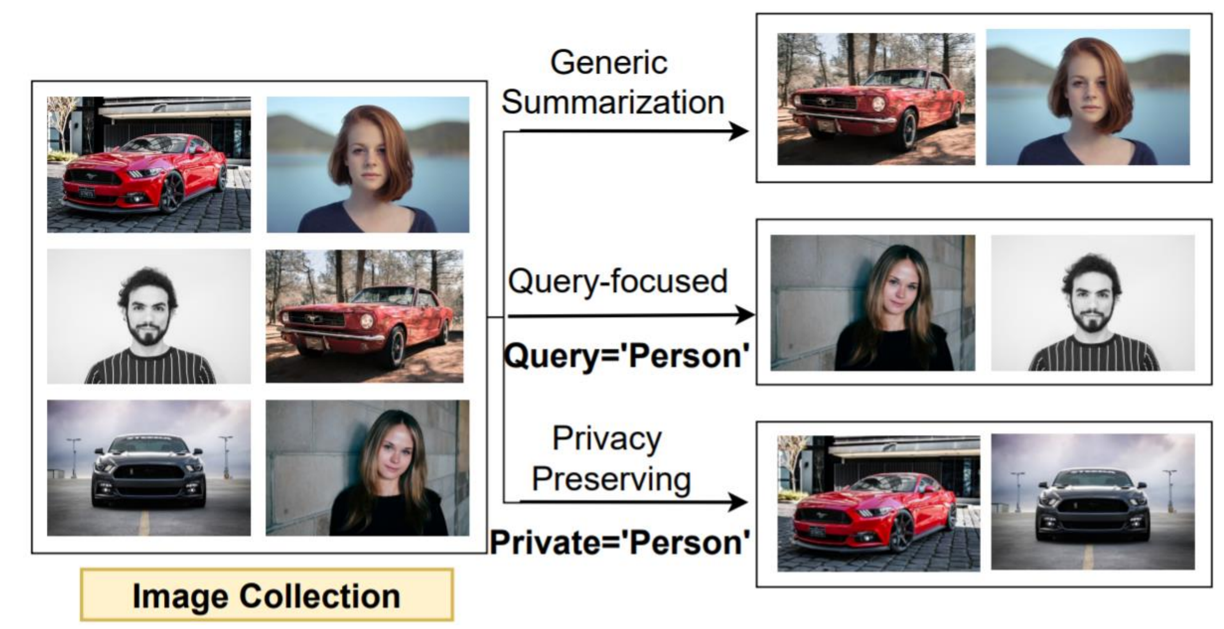
- Q is from an auxiliary set V' different from the ground set V.
- For guided subset selection, V is the source set of data instances and the target is a subset of data points (validation set or the specific set of examples of interest).
- Define $f : 2^{V \cup V'} \rightarrow \Re$.
- Although f is defined on $V \cup V'$, discrete optimization is only defined on $A \subseteq V$.
- To find an optimal subset we maximize $I_f(A;Q|P)$ using a greedy strategy.



Applications of Guided Subset Selection

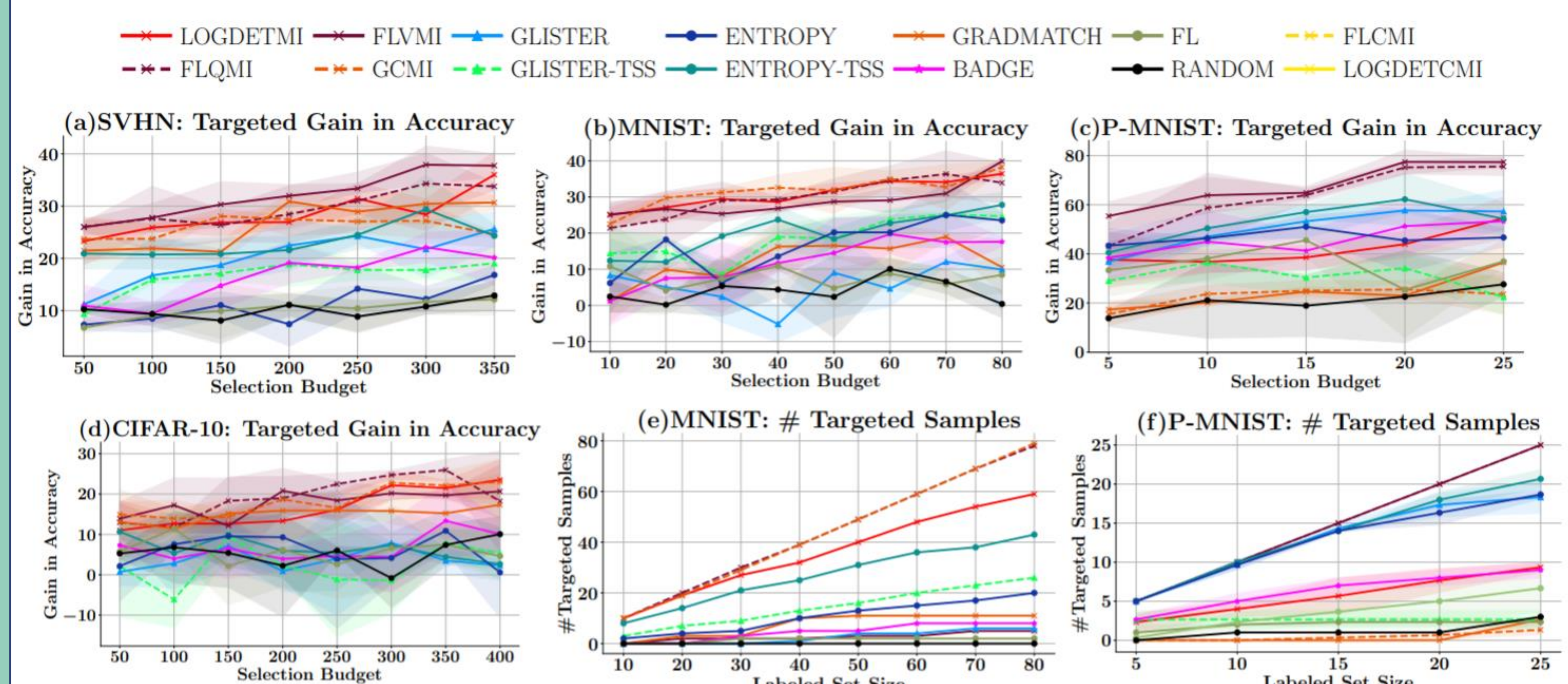


Targeted Learning

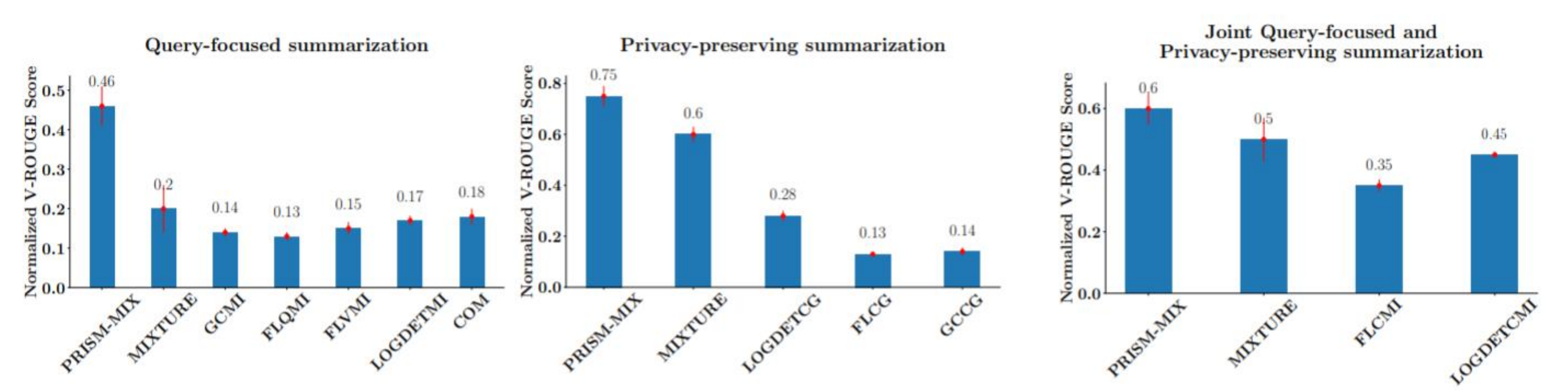


Guided Summarization

RESULTS



MI functions consistently outperform all baselines by ≈20-30% in terms of average accuracy on target classes.



PRISM-MIX outperforms other techniques on all flavors of summarization due to joint learning of mixture weights and internal parameters.

Targeted Learning

Given: Initial Labeled set of Examples: E, large unlabeled dataset: U, A target subset/slice where we want to improve accuracy: T, Loss function L for learning

1. Train model with loss L on labeled set E and obtain parameters θ_E.
2. Compute the gradients $\{\nabla_{\theta_E} L(x_i, y_i), i \in U\}$ and $\{\nabla_{\theta_E} L(x_i, y_i), i \in T\}$.
3. Using the gradients, compute the similarity kernels and define the submodular function f and diversity function g.
4. $\hat{A} \leftarrow \max_{A \subseteq U, |A| \leq K} I_f(A;T) + \gamma g(A)$
5. Obtain the labels of elements in A^*: $L(\hat{A})$
6. Train a model on the combined labeled set $E \cup L(\hat{A})$

CONCLUSIONS

- We presented PRISM, a rich class of functions for guided subset selection.
- PRISM allows to model a broad spectrum of semantics across query-relevance, diversity, query-coverage and privacy-irrelevance.
- We demonstrated its effectiveness in targeted learning as well as in guided summarization.
- In our paper, we showed that PRISM has interesting connections to several past work, further reinforcing its utility.
- Through experiments on targeted learning and guided summarization for diverse datasets, we empirically verified the superiority of PRISM over existing methods.

PAPER

SCAN ME

Get the paper for more technical details and results: https://arxiv.org/pdf/2103.00128.pdf