Learning to Maximize Mutual Information for Dynamic Feature Selection

DISCLAIMER: Summarized by AI

Problem they are trying to solve / Purpose of method

Problem:
Traditional feature selection methods often rely on a static approach, selecting a fixed set of features for all data instances. However, this is inefficient when different samples require different subsets of features to achieve accurate predictions. Existing dynamic feature selection (DFS) methods frequently rely on heuristic policies or require labeled data to learn effective feature selection strategies, which can limit their generalizability.

Purpose of the Method:
The paper introduces a method that formulates DFS as a problem of learning to acquire features that maximize the mutual information (MI) between selected features and labels. By leveraging mutual information, the method can generalize across tasks and does not require labeled data at test time, aiming to be both efficient and adaptive.

How does it differ from other methods?

- Unlike previous DFS approaches that often require task-specific supervision or learn greedy selection policies, this method learns a generalizable feature acquisition policy by maximizing mutual information.
- Prior work typically does not optimize mutual information directly and often lacks the ability to adapt selection strategies dynamically across different inputs.
- The proposed method is **unsupervised at test time**, generalizes across tasks, and does not rely on fixed feature ordering or hard-coded selection rules.

How the method works

Simple Overview:
The method proposes a mutual information-driven framework for dynamic feature selection. It learns a policy that sequentially selects features to acquire based on how much mutual information they provide about the target label, conditioned on already observed features.

Detailed Explanation:

- The feature acquisition process is modeled as a sequential decision-making task.

- At each step, the policy selects the next feature that provides the most additional information about the label.
- A differentiable objective is used to maximize the **conditional mutual information (CMI)** between the selected features and the label.
- A variational bound is introduced to approximate and optimize mutual information in a tractable way.
- The method is trained with unlabeled data and generalizes to new tasks at test time without requiring labels.