Active Imitation Learning with Noisy Guidance

Kianté Brantley,¹ Amr Sharaf,¹ Hal Daumé III ¹,²

¹ University of Maryland, ² Microsoft Research
Structured Prediction Problems

*for example, Named Entity Recognition:*

| Word      | Label |
|-----------|-------|
| After     | O     |
| completing| O     |
| his       | O     |
| Ph.D.     | O     |
| ,         | O     |
| ……        | ……    |
Problem:

- Can we design an algorithm to reduce expert annotation cost for structure prediction problems?
Imitation Learning

Expert Demonstrator: (Annotator)

Named Entity Recognition

Input: After completing his Ph.D., Ellis worked at Bell Labs from 1969 to 1972 on probability theory.

Prediction: 0

- states input combined with policy’s previous prediction
- actions o, per, org, misc, loc

training set: \( D = \{(state, actions)\} \) from expert \( \pi^* \)

goal: learn policy \( \pi_\theta (s) \rightarrow a \)
Imitation Learning using DAgger

Initialize Dataset $D$
Initialize $\hat{\pi}_1$
for $i = 1$ to $N$ do
    $\pi_i = \beta_i \pi^* + (1 - \beta_i) \hat{\pi}_i$
    Sample T-step trajectory $\pi_i$
    Get dataset $D_i = \{(s, \pi^*(s))\}$
    Aggregate dataset $D \leftarrow D \cup D_i$
    Train classifier $\hat{\pi}_{i+1}$ on $D$

Named Entity Recognition

Pro:
- The policy is able to learn from its own state distribution.

Stéphane Ross, Geoff J. Gordon, and J. Andrew Bagnell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In AI-Stats.
Imitation Learning using DAgger

Initialize Dataset \( D \)
Initialize \( \hat{\pi}_1 \)
for \( i = 1 \) to \( N \) do
\[
\pi_i = \beta_i \pi^* + (1 - \beta_i) \hat{\pi}_i
\]
Sample T-step trajectory
Get dataset \( D_i = \{(s, \pi^*(s))\} \)
Aggregate dataset \( D \leftarrow D \cup D_i \)
Train classifier \( \hat{\pi}_{i+1} \)

Con:
- For every state that we visited we queried an expert for the optimal action.

Named Entity Recognition

After completing his Ph.D., Ellis worked at Bell Labs from 1969 to 1972 on probability theory.

Stéphane Ross, Geoff J. Gordon, and J. Andrew Bagnell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In AI-Stats.
Initialize Dataset $D$
Initialize $\hat{\pi}_1$

for $i = 1$ to $N$ do

$$\pi_i = \beta_i \pi^* + (1 - \beta_i) \hat{\pi}_i$$

Sample T-step trajectory

for $t = 1$ to $T$

set $\hat{p}_t = \pi_\theta(y_t^1 \mid s_t)$

draw Bernoulli variable $Z_t$ of parameter $b + |\hat{p}_t|$

if $Z_t = 1$

Get dataset $D_t = \{(s_t, \pi^*(s_t))\}$

Aggregate dataset $D \leftarrow D \cup D_t$

Train classifier $\hat{\pi}_{i+1}$ on $D$

Question:

☐ Can reduce expert queries even further?
Our Approach: LeaQI
(Learning to Query for Imitation)

Key Ideas:
- We assume access to a noisy heuristic function
- Use a disagreement classifier to decide if we should query the expert or the heuristic function
- Train the disagreement classifier using the Apple Tasting framework
After completing his Ph.D., Ellis worked at Bell Labs from 1969 to 1972 on probability theory.

**One-Sided Feedback Learning**

**Named Entity Recognition**

**Heuristic Function**

- Learn difference classifier to predict when a Heuristic and Expert disagree
- **Difference classifier only gets feedback** when it predicts disagree and we query the expert
- **Difference classifier does not get feedback** when it predicts agree and we query the heuristic function
- **We use an Apple Tasting algorithm to reduce false negatives** in the difference classifier predictions
| Experiment Details |
|-------------------|

| NER               | Keyphrase                      | POS               |
|-------------------|--------------------------------|-------------------|
| Language          | English                        | English           |
| Dataset           | CoNLL’03                       | English           |
| Heuristic         | Gazeteer                       | Modern Greek      |
| Huer. Quality     | P88%, R27%                     | SemEval 2017 Task 10 |
|                   |                                | Universal Dependencies |
|                   |                                | Unsupervised model |
|                   |                                | Dictionary Wiktionary |
|                   |                                | 67% acc           |
Experiment Results

Q1
Active vs Passive

Q2
Heuristic as features vs Policy
Thank you!