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
……	…..

Expert

Structured Predi

for example, Nam

Word	
After	
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

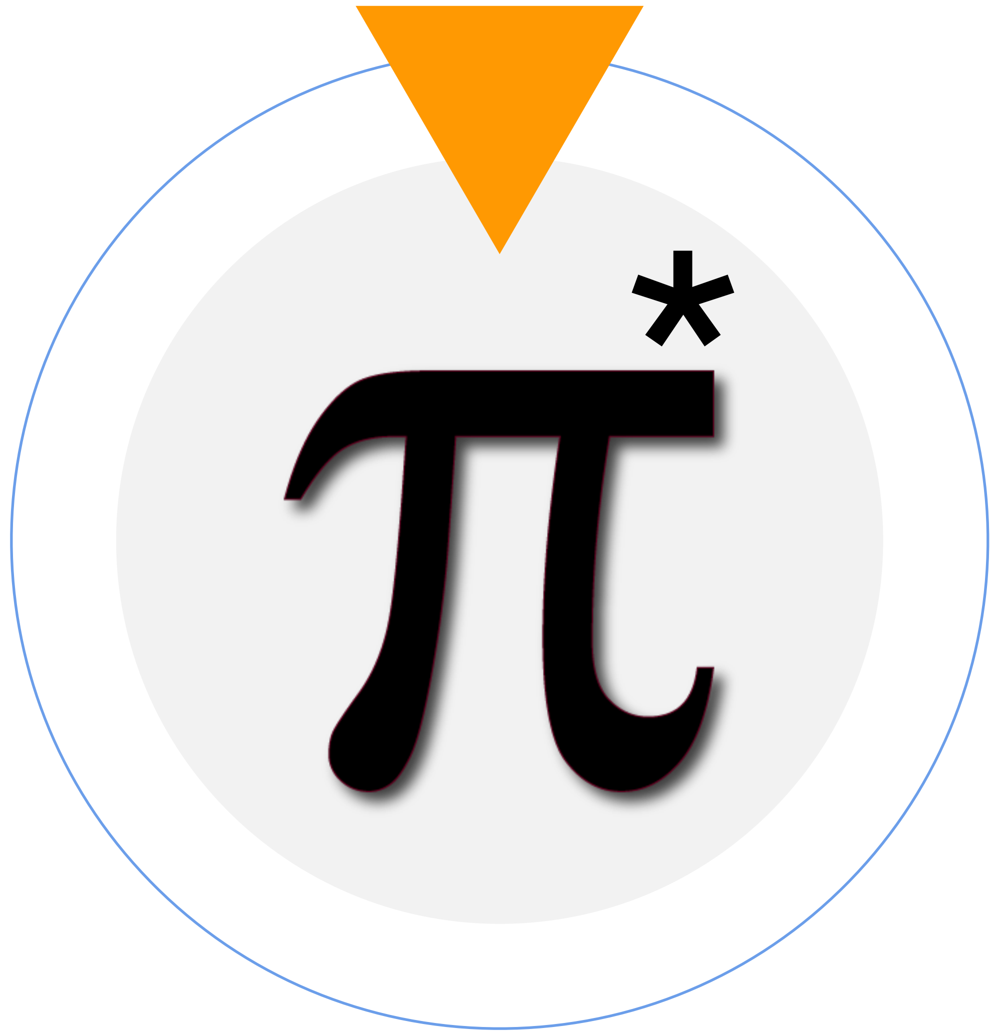
Input: After completing his Ph.D. , Ellis worked at Bell Labs from 1969 to 1972 on probability theory..

Prediction: o

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

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

goal: learn policy π_θ(s) -> 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 trajector

Get dataset $D_i = \{(s,$

Aggregate dataset D

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

Named Entity Recognition

his Ph.D., Ellis worke

O O O PER O

O O O PER O

Pro:

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

Stéphane Ross, Geoff J. Gordon, and J. Andrew Bag- nell. 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 trajector

Get dataset $D_i = \{(s,$

Aggregate dataset D

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

Named Entity Recognition

his Ph.D., Ellis worke

O O O PER O

O O O PER O

Con:

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

Stéphane Ross, Geoff J. Gordon, and J. Andrew Bag- nell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In AI-Stats.

Active Learning with DAgger

Initialize Dataset D

Initialize $\hat{\pi}_1$

for $i = 1$ to N do

- $\pi_i = \beta_i \pi^* + (1 - \beta_i)\hat{\pi}$
- Sample T-step traject
- 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

gnition

leting his Ph.D., Ellis worke

O O O O PER O

PER O

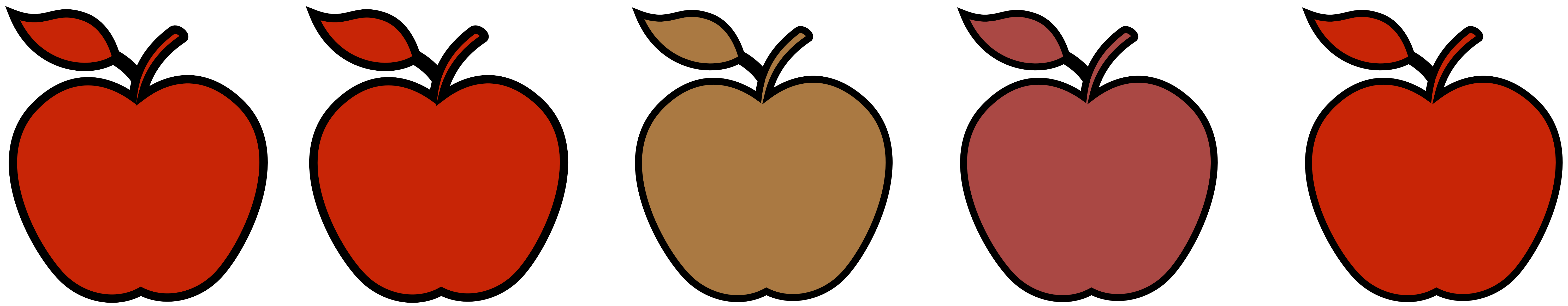
Question:

- [ ] 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



One-Sided Feedback Learning

Named Entity Recognition

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

π

Heuristic Function

Gazette: O O O

π^* O O O

Heuristic Function

- Noisy, bias and cheap

LeaQI One-Side Feedback Problem

- 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 classier predictions

Experiment Details

	NER	Keyphrase	POS
Language	English	English	Modern Greek
Dataset	CoNLL’03	SemEval 2017 Task 10	Universal Dependencies
Heuristic	Gazeteer	Unsupervised model	Dictionary Wiktionary
Huer. Quality	P88%, R27%	P20%, R44%	67% acc

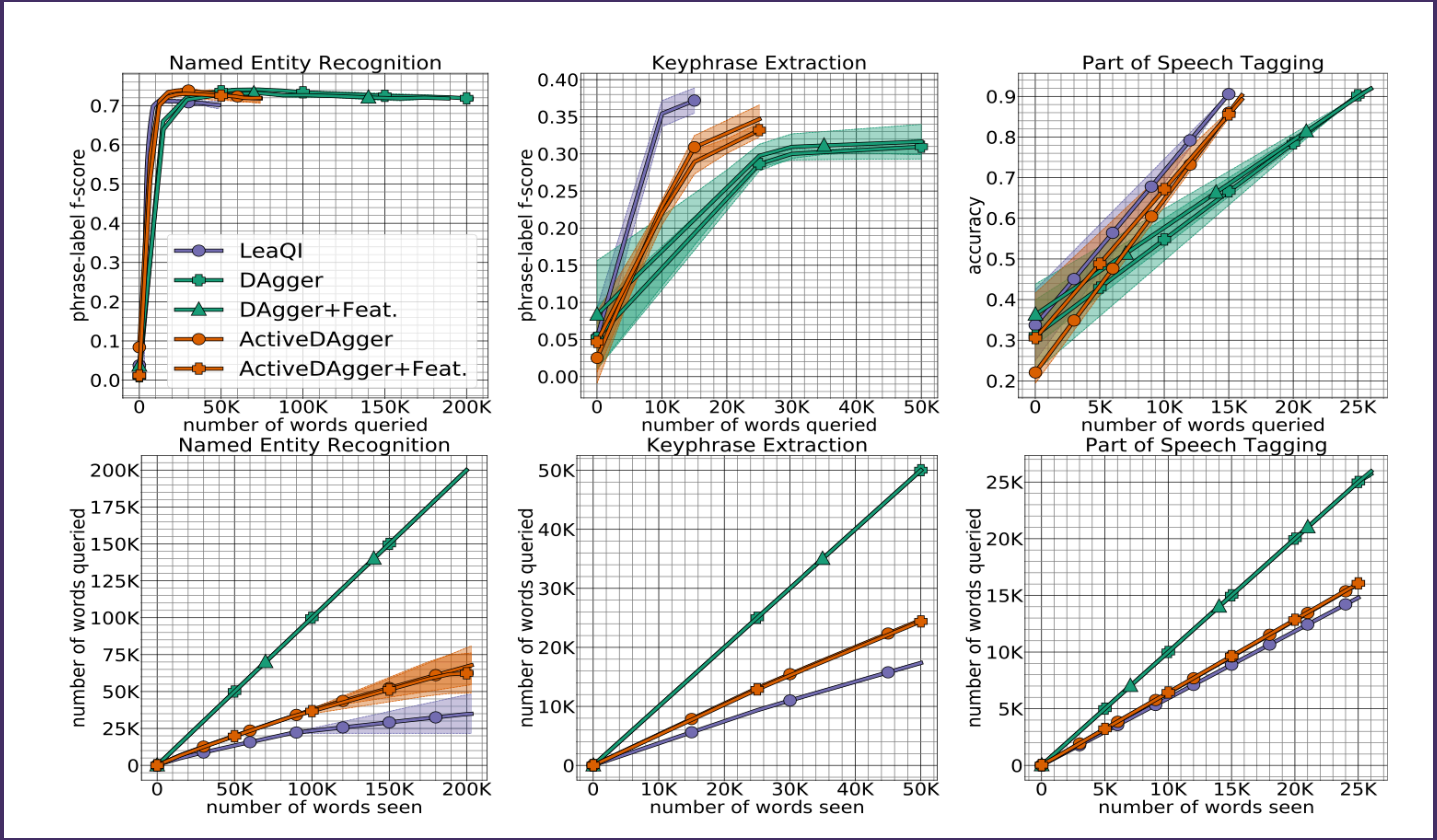
Q1

Active vs Passive

Q2

Heuristic as features vs Policy

Experiment Results




Thank you!