An investigation of imitation learning algorithms for structured prediction

Andreas Vlachos,
Computer Laboratory,
University of Cambridge

Imitation learning:

- Learns controllers by observing demonstrations by humans
- Infers the reward function instead of requiring it
- Must account for the dependencies between actions

Structured prediction is similar to imitation learning:

- Complex output spaces are formed by inter-dependent actions
- Labeled data shows us the correct actions to perform a task, but not how to overcome mistakes
- SEARN (Daumé et al., 2009) casted structured prediction as imitation learning successfully.
- DAgger (Ross et al., 2011) achieved better stability

Question: How do they compare on a complex structured prediction task?

SEARN versus DAgger in training:

Input: training data S, optimal policy π^*, loss function ℓ, learning rate β
Output: Hypothesis H_N
Examples $E = \varnothing$
for $i=1$ **to** N **do**
    $p = (1-\beta)^{i-1}$
    current policy $\pi = p\pi^* + (1-p)H_{i-1}$
    SEARN: Examples $E = \varnothing$
    foreach s **in** S **do**
        Predict $\pi(s) = \hat{y}_{1:T}$
        foreach $\hat{y}_t$ **in** $\pi(s)$ **do**
            Extract features $\Phi_t = f(s, \hat{y}_{1:t-1})$
            foreach possible action y_t^j **do**
                SEARN: Predict $y'_{t+1:T} = \pi(s \mid \hat{y}_{1:t-1}, y_t^j)$
                DAgger: Predict $y'_{t+1:T} = \pi^*(s \mid \hat{y}_{1:t-1}, y_t^j)$
                Estimate cost $c_t^j = \ell(\hat{y}_{1:t-1}, y_t^j, y'_{t+1:T})$
            Add (Φ_t, c_t) to E
    Learn a classifier h_i from E
    SEARN: $H_i = \beta \sum_{k=1}^{i} \frac{(1-\beta)^{i-k}}{1-(1-\beta)^i} h_k$
    DAgger: $H_i = h_i$

Unless β=1, training is stochastic

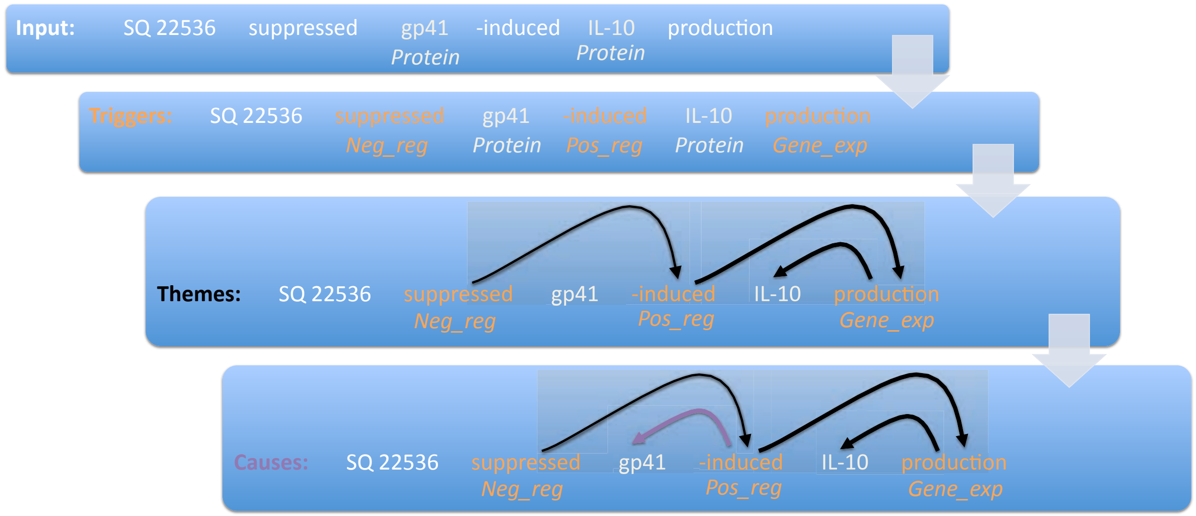
SEARN learns from the data of each iteration separately

Focused costing ameliorates stochastic cost estimation

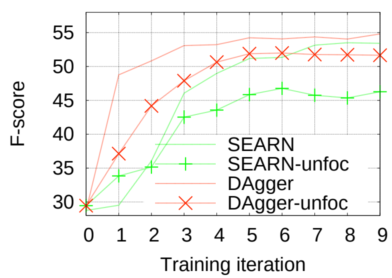
SEARN interpolates the newly learnt classifier with the previous one

DAgger uses the data from all iterations to learn the current hypothesis

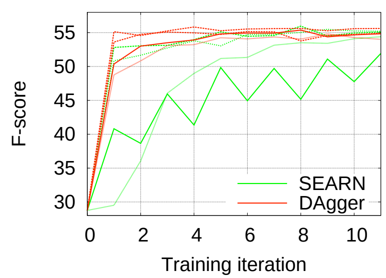
Decomposing biomedical event extraction:



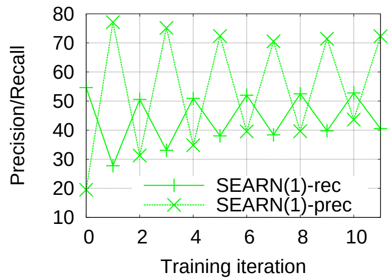
Experimental results:



Focused costing improves both SEARN and DAgger when training is stochastic. DAgger though is less sensitive.



DAgger is more stable than SEARN, converging in fewer iterations for a wide range of learning rates (0.1, 0.3, 0.7, 1), thus easier to use.



When training is deterministic, SEARN, exhibits an oscillating behavior. This is due to the high recall/low precision hypothesis learnt in the first iteration combined with the inflexible hypotheses combination.

References
Daumé III, Langford, and Marcu, 2009. Search-based structured prediction. Machine Learning, 75:297–325.
Ross, Gordon and Bagnell, 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of AISTATS*.
A. Vlachos and M. Craven, 2012. Biomedical event extraction from abstracts and full papers using search-based structured prediction. *BMC Bioinformatics*.

spacebook-project.eu