



Building end-to-end QA model with reward based objective

TEAM REINFORCE (Prashant and Eti)



Baseline Model(SearchQA)

1. Concatenate all the passages of a question together and send them as 1 paragraph
2. Use BiDAF model trained on SquadQA to predict results on SearchQA

	Accuracy(EM)(Ours)
Bidirectional LSTM + char embeddings + Reduced Vocabulary size	27.8



Architecture for Passage Ranking

- Concatenate query and sentence embeddings together
- Pass them through a MLP followed by softmax to obtain $p(s = s_l | x, d)$
- Initially train using the technique of Distant Supervision and then use RL for further training

$$X = [Q; S_l]$$

$$Z = W(MLP(X))$$

$$p(s = s_l | q, d) = softmax(Z)$$



REINFORCE Algorithm

- Objective Function: Maximize expected reward

$$J(\theta) = \sum_{p_k \in \text{passages}} p_{\theta}(p = p_k | \text{query}, \text{passages}) R_{\theta}(p_k)$$

Where R is the reward that the agent gets after selecting passage p_k

(calculated from BiDAF model)

$$R_{\theta}(p_k) = \log p_{\theta}(y = y^* | \text{query}, p_k)$$



REINFORCE Algorithm

- Update rule: Gradient of the objective function is approximated with a sample

$$\nabla J(\theta) \approx \nabla \log p_{\theta}(y|\hat{p}, \text{query}) + \log p_{\theta}(y|\hat{p}, \text{query}) \cdot \nabla \log p_{\theta}(\hat{p}|\text{query}, \text{passages})$$

- Sampling introduces variance in the algorithm and makes it quite unstable to train
- Options that could help: Curriculum learning, distant supervision, DAGGER like training approach; Baseline approach



Distant Supervision

- Use to make training stable
- Simple heuristic : Treat first sentence with the answer as the golden label and train the passage ranking model for few epochs
- Transition to RL based objective slowly over epochs



This week: To Do

- Baseline RL Model
- Add support for curriculum learning and DAGGER like training approach



Timeline

March 03 - March 12	March 13 - March 30	March 31 - April 15	April 16 - April 30
Implementation and training of own BiDAF model in dynet	<ul style="list-style-type: none">Perform experiments on SQuAD dataset- Analyze model on SearchQA Dataset	<ul style="list-style-type: none">- Analyze model on SearchQA Dataset- Implement the passage ranking functionality	Error Analysis and further improvement