Training Semantic Parsers

Pradeep Dasigi
Carnegie Mellon University / Allen Institute for Artificial Intelligence

Got Supervision?

x_i: flights from Dallas leaving after 4 in the afternoon

y_i: (lambda $0 e
    (and
     (>(departure_time $0) 1600:ti)
     (from $0 dallas:ci)))

$$D = \{x_i, y_i\}_{i=1}^{N}$$

Task: Given x_{N+k} find y_{N+k}

Fully supervised

Got Supervision?

x_i: flights from Dallas leaving after 4 in the afternoon

y_i: (lambda $0 e
(and
(>(departure_time $0) 1600:ti)
(from $0 dallas:ci)))

$$D = \{x_i, y_i\}_{i=1}^{N}$$

Task: Given x_{N+k} find y_{N+k}

Fully supervised

x_i: Which athlete was from South Korea after 2010?

~~y_i: ((reverse athlete)~~
~~(and~~
~~(nation south_korea)~~
~~(year ((reverse date) (>= 2010)))~~

z_i: Kim Yu-Na

w_i:

Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea	2010–2014	2
Tenley Albright	United States	1952-1956	2

$$D = \{x_i, w_i, z_i\}_{i=1}^{N}$$

Task: Given x_{N+k}, w_{N+k} find y_{N+k} such that $[\![y_{N+k}]\!]^{w_{N+k}} = z_{N+k}$

Weakly supervised

Got Supervision?

x_i: flights from Dallas leaving after 4 in the afternoon

y_i: (lambda $0 e
(and
(>(departure_time $0) 1600:ti)
(from $0 dallas:ci)))

x_i: Which athlete was from South Korea after 2010?

~~y_i: ((reverse athlete)~~
~~(and~~
~~(nation south_korea)~~
~~(year ((reverse date) (>= 2010)))~~

- In the fully supervised setting, the problem is similar to most other structured prediction problems
- We will focus on the weakly supervised setting

z_i: Kim Yu-Na

w_i:

Athlete	Nation	Olympics	Medals
Kim Yu-na	South Korea	2010-2014	2
Tenley Albright	United States	1952-1956	2

$$D = \{x_i, y_i\}_{i=1}^N$$

Task: Given x_{N+k} find y_{N+k}

$$D = \{x_i, w_i, z_i\}_{i=1}^N$$

Task: Given x_{N+k}, w_{N+k} find y_{N+k} such that $[\![y_{N+k}]\!]^{w_{N+k}} = z_{N+k}$

Fully supervised

Weakly supervised

Three common training methods

- Maximum Marginal Likelihood
- Structured Learning Methods
- Reinforcement Learning Methods

And some hybrid approaches..

Maximum Marginal Likelihood

- Given $D = \{x_i, w_i, z_i\}_{i=1}^{N}$
- We want to optimize $\max_{\theta} \prod_{x_i, z_i \in D} p(z_i | x_i; \theta)$
- But the semantic parser defines a distribution over logical forms.
- So we marginalize over logical forms that yield z_i

$$\max_{\theta} \prod_{x_i, w_i, z_i \in D} \sum_{y_i \in Y \,|\, [\![y_i]\!]^{w_i} = z_i} p(y_i | x_i; \theta)$$

- Y could be the set of all valid logical forms, if we are using constrained decoding during training
- Even then, the summation could be intractable!

MML: Approximating $\mathbf{Y}$

- Perform heuristic search
- Search may be bounded, by length or otherwise
- $\mathbf{Y}$ is approximated as a subset of retrieved logical forms

Two options for search:

Online Search	Offline Search
Search for consistent logical forms during training, as per model scores	Search for consistent logical forms before training
Candidate set changes as training progresses	Candidate set is static
Less efficient	More efficient

Structured Learning Methods

- More commonly used with traditional semantic parsers
  - Eg. Margin based models and Latent variable structured perceptron (Zettlemoyer and Collins 2007)
- Typically involve heuristic search over the state space like MML methods
- Unlike MML, can use arbitrary cost function
- Training typically maximizes margins or minimizes expected risks

Reinforcement Learning Methods

- Comparison with MML:
  - Like MML Y is approximated
  - Unlike MML, the approximation is done using sampling techniques
- Comparison with structured learning methods
  - Like structured learning methods, the reward function can be arbitrary
  - Unlike structured learning methods, reward is directly maximized
- Training typically uses policy gradient methods

  Example from Liang et al., 2017, using REINFORCE

$$\max_{\theta} \sum_{x} \mathbb{E}_{P_{\theta}(a_{0:T}|x)}[R(x, a_{0:T})]$$

These methods could complement each other!

- Bridging MML and RL
  - Guu et al., EMNLP 2017
- Bridging RL and structured learning methods
  - Iyyer et al., ACL 2017

Summary

- Three general directions for training weakly supervised semantic parsers
  - MML gets around weak supervision by marginalizing over logical forms that yield correct denotations
  - Structured learning methods can incorporate global constraints into learning
  - RL approaches let us define arbitrary reward functions to get around weak supervision
- These approaches can be complementary
- Bridging them is an interesting research direction

Papers using MML

Online search

Berant, Chou, Frostig, and Liang. 2013. Semantic parsing on freebase from question-answer pairs. In EMNLP.

Guu, Pasupat, Liu, and Liang. 2017. From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood. In ACL.

Goldman, Latcinnik, Naveh, Globerson, and Berant. 2018. Weakly-supervised Semantic Parsing with Abstract Examples. In ACL

Offline search

Krishnamurthy, Dasigi, and Gardner. 2017**.** Neural Semantic Parsing with Type Constraints for Semi-Structured Tables. In EMNLP.

Papers using structured learning methods

Yih, Chang, He, and Gao. 2015. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base.In ACL.

Berant and Liang. 2015. Imitation Learning of Agenda-based Semantic Parsers. In TACL

Xiao, Dymetman, and Gardent. 2016. Sequence-based Structured Prediction for Semantic Parsing. In ACL.

Iyyer, Yih, and Chang. 2017. Search-based Neural Structured Learning for Sequential Question Answering. In ACL.

Papers using Reinforcement Learning

Andreas, Rohrbach, Darrell, and Klein. 2016. Learning to Compose Neural Networks for Question Answering. In NAACL.

Liang, Berant, Le, Forbes, and Lao. 2017. Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision. In ACL.

Guu, Pasupat, Liu, and Liang. 2017. From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood. In ACL.

Zhong, Xiong, and Socher. 2017. Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning. ArXiv preprint.