Transformer Semantic Parsing

Gabriela Ferraro[1,2] & Hanna Suominen[2,1,3]

Data61, Commonwealth Scientific and Industrial Research Organisation, Australia
Research School of Computer Science, The Australian National University, Australia
Department of Future Technologies, University of Turku, Finland
gabriela.ferraro@data61.csiro.au hanna.suominen@anu.edu.au

Problem

- Transform sentences into meaning representation
- Input sentence:
  Show me tomorrow flights from ci0 to ci1
- Output logic formule:
  `(lambda $0 e (and (flight $0 (from $0 ci0) (to $0 ci1) (tomorrow $0)))`

Model

- Self-attention neural semantic parser with Transformer (see Vaswani et al. (2017)

Two Data Sets from Different Domains

- `ATIS`: queries from a flight booking system
- `GEO`: queries about US geographical information

Difficulty Levels

- `anonymised variables`
- `non-anonymised variables (non-anon)`
- `question split` (`quest-split`, based on input sequences)
- `query split` (based on the output sequence to ensure a more diverse set of logic formulae)

Data set	Input	Output
ATIS	ground transport in ci0	(lambda $0 e (and GT) (to_city ci0)))
ATIS non-anon quest-split	ground transport in Denver	(lambda $0 e (and (GT $0) (to_city $0 denver)))
GEO	how many citizen in s0	(population:i s0)
GEO non-anon quest-split	how many citizens in Alabama	(population alabama)
GEO non-anon query-split	how many citizens in Boulder	(population boulder)

No. of Training (Train), Development (Dev), and Test (Test) Examples

Data set	Train	Dev	Test
ATIS	4,434	491	448
ATIS non-anon	4,029	504	504
GEO	600	0	280
GEO non-anon quest-split	583	15	279
GEO non-anon query-split	543	148	186

Input (I) and Output (O) Vocabulary Sizes.
Entity anonymisation had a bigger impact on the vocabulary size of the I than the O.

Data set	I vocab	O vocab
ATIS	166	433
ATIS non-anon	444	887
GEO	51	120
GEO non-anon quest-split	141	243
GEO non-anon query-split	149	254

Experimental Results and Conclusions

- Transformer was competitive with other state-of-the-art models and outperformed strong baselines in some settings.

Model Comparison Using the Accuracy [%] on Anononymised Test Sets

Model	ATIS	GEO
Statistical Baselines		
ZC07 (Zettlemoyer and Collins, 2007)	84.6	86.07
TISP (Zhao and Huang, 2015)	84.2	**88.9**
Neural Baselines		
Seq2Seq + Attention (Dong and Lapata, 2016)	84.15	84.6
Seq2Tree + Attention (Dong and Lapata, 2016)	86.9	87.1
ASN (Rabinovich et al., 2017)	85.3	85.7
ASN + Attention (Rabinovich et al., 2017)	85.9	87.1
coarse2fine (Dong and Lapata, 2018)	87.7	88.2
Our Neurals		
Bi-GRU	85.93	86.42
Transformer	**87.95**	86.78

- Our model became the new state-of-the-art on `ATIS` with its accuracy of 87.95%.
- However, the best result on `GEO` was by a statistical semantic parser called Type-Driven Incremental Semantic Parsing (TISP).

Model Comparison Using the Accuracy [%] on Non-Anononymised Test Sets

Model	ATIS NA	GEO NA quest split	GEO NA query split
Seq2Seq + Attention (Dong and Lapata, 2016)	72.02	67.39	41.94
coarse2fine (Dong and Lapata, 2018)	**79.1**	72.4	52.69
Bi-GRU	73.41	72.4	56.45
Transformer	75.99	**75.27**	**63.98**

- Results with anonymisation were always better than without.
- `GEO query-split` was harder than `GEO question-split`.
- Our model learnt token attributes as opposed to only one-to-one mappings from an input to output.

Difficult Examples

- Considerable difference between the length of input sentences and their corresponding logical formulae resulted in 0 accuracy.

Data set	Input	Output
ATIS	fare code fb0 what doe that mean	fb0
	what type of plane is a ac0	ac0
GEO	what is the average population per square km in s0	density:i s0
	what is the length of the r0 in s0	len:i r0