

Abstract

Unscoped Episodic Logical Forms^[1] (ULF) is a semantic representation for English sentences which captures semantic type structure, allows for linguistic inferences, and provides a basis for further resolution into Episodic Logic^[2] (EL). We present an application of pre-trained autoregressive language models to the task of rendering ULFs into English, and show that ULF's properties reduce the required training data volume for this approach when compared to AMR. We also show that the same system, when applied in reverse, performs well as an English-to-ULF parser.

I want to dance in my new shoes

English



ULF

```
(i.pro (pres want.v)
 (to dance.v
  (adv-a (in.p (my.d (mod-n
new.a)
  (plur
shoe.n))))))
```

Introduction

We apply the techniques of GPT-Too^[3], an AMR to English converter that finetunes a pretrained autoregressive language model to generate pairs of AMRs and corresponding English sentences, to generate English sentences from ULF, and vice versa. Shown below is the string format used in the fine-tuning datasets for English generation (top) and ULF parsing (bottom).

ULF <SEP> English <END>

English <SEP> ULF <END>

Language-Model-Based Parsing and English Generation for Unscoped Episodic Logical Forms

Training

We trained the Huggingface transformer gpt2-large model on a dataset of 1,378 hand-annotated ULF-English sentence pairs. The ULF-to-English model was fine-tuned to maximize the joint probability of the English tokens e_i and ULF tokens u_j as shown below.

$$P_{\text{GPT2}}(e|u) = \prod_{i=1}^n P_{\text{GPT2}}(e_i|u_{1:m}, e_{1:i-1})$$

$$P_{\text{GPT2}}(u) = \prod_{j=1}^m P_{\text{GPT2}}(u_j|u_{1:j-1})$$

Generation

We decoded the probability distributions and obtained output strings for both tasks using beam search, greedy sampling, nucleus (top- p) sampling, which truncates the infrequent tail of the probability distribution and samples from the top 90% of the probability mass, and thermal sampling, where, prior to weighted random token selection, the probability of each token is transformed according to a temperature parameter τ .

GPT-Too improved performance by parsing the top K sentences into AMR and scoring them, so we rescored with an external ULF parser and the metrics EL-Smatch and SemBLEU.

Evaluation

Scored generated English with: BLEU, chrF++, and METEOR
Scored generated ULF with: EL-Smatch and SemBLEU
All scores are out of 1.0

English Generation Results

Model	LDC2017T10 (AMR)			ULF		
	B	C	M	B	C	M
T	0.31	0.36	0.64	-	-	-
GT	0.30	0.25	0.59	-	-	-
GPT-Too	0.33	0.38	0.64	-	-	-
ulf2english	-	-	-	0.56	0.81	0.80
ulf-gpt2	-	-	-	0.72	0.84	0.80

Our model with beam search ties or outperforms the rule-based ulf2english tool, and significantly outperforms any AMR-to-English system.

English Generation: Decoding Methods Results

Model	BLEU	chrF++	METEOR
Greedy	0.67	0.81	0.76
Greedy (els)	0.67	0.81	0.76
Beam 10	0.72	0.84	0.80
Beam 10 (els)	0.72	0.84	0.80
Beam 15	0.72	0.84	0.80
Beam 15 (els)	0.72	0.84	0.80
Nucleus	0.68	0.82	0.77
Nucleus (els)	0.68	0.82	0.78
Temp 0.5	0.69	0.82	0.77
Temp 0.5 (els)	0.71	0.84	0.79
Temp 1.0	0.67	0.81	0.77
Temp 1.0 (els)	0.66	0.81	0.77

els = EL-Smatch rescoring added to generated log probability

- [1] Gene Louis Kim and Lenhart Schubert. 2019. A type-coherent, expressive representation as an initial step to language understanding. arXiv preprint arXiv:1903.09333.
- [2] Chung Hee Hwang and Lenhart K Schubert. 1993. Episodic logic: A situational logic for natural language processing. Situation Theory and its Applications, 3:303–338.
- [3] Manuel Mager, Ramon Fernandez Astudillo, Tahira Naseem, Md Arafat Sultan, Young-Suk Lee, Radu Florian, and Salim Roukos. 2020. GPT-too: A language-model-first approach for AMR-to-text generation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1846–1852, Online. Association for Computational Linguistics

Erin Gibson (egibson4@u.rochester.edu)
Lane Lawley (llawley@ur.rochester.edu)

University of Rochester
Department of Computer Science

Parsing Results

Model	SEM BLEU	EL-SMATCH
(Zhang et al., 2019) ^[4]	0.13	0.34
(Cai and Lam, 2020a) ^[5]	0.34	0.53
(Kim et al., 2021) ^[6]	0.47	0.60
ulf-gpt2	0.43	0.63

Our model slightly outperforms the ULF cache transition parser by the EL-SMATCH metric, and is slightly outperformed by it by the SEMBLEU metric. Both models significantly outperform the evaluated AMR parsers by both metrics.

Conclusion

In both tasks, we outperform not only comparable models for the AMR representation, but also existing models for the same tasks using ULF. Although our ULF training dataset ($N = 1,378$) is considerably smaller than the LDC2017T10 AMR dataset ($N = 36,521$), we achieve superior performance to all evaluated AMR-to-English and English-to-AMR models. This suggests that ULF's properties, perhaps most notably the syntactic information preserved in its type system, make it a compelling semantic representation for tasks where transduction to and from natural English is required.

- [4] Sheng Zhang, Xutai Ma, Kevin Duh, and Benjamin Van Durme. 2019. AMR parsing as sequence-to-graph transduction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 80–94, Florence, Italy. Association for Computational Linguistics.
- [5] Deng Cai and Wai Lam. 2020b. Graph transformer for graph-to-sequence learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 7464–7471.
- [6] Gene Louis Kim, Viet Duong, Xin Lu, and Lenhart Schubert. 2021. A transition-based parser for unscoped episodic logical forms. arXiv preprint arXiv:2103.08759.