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

Erin Gibson and Lane Lawley
University of Rochester
Department of Computer Science

Abstract

Unscoped Episodic Logical Forms (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 (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.

Unscoped Episodic Logical Forms (ULF) is a semantic representation of English derived from an intermediate step of the conversion of English sentences into Episodic Logic (EL). While EL is a richly specified formal logic suitable for language understanding, the current methods of deriving full EL formulas from text are not yet reliable enough for general use. ULF models EL's full predicate argument structure and type system, but omits some resolution steps necessary for full EL conversion, such as anaphora resolution, operator scope resolution, and word sense disambiguation. The result is a semantic representation with rich syntactic information, which facilitates conversions to and from surface English, and also rich predicate-argument and type information, which facilitates a variety of semantic inference tasks, including discourse inferences, natural logic-based monotonic inferences, and even event schema induction, after some post-processing for anaphora resolution.

Of course, generating natural language formulations from semantic representations is important for understanding instances of those representations that were not generated from any specific original text, e.g. machine-generated inferences. The Abstract Meaning Representation (AMR) project has received the most attention in recent semantic representation research, and research into English generation is no exception. Many techniques have been applied to AMR-to-English conversion, including phrase-based machine translation, graph convolutional encoders, and non-pre-trained transformer models.

This work applies, to the ULF semantic representation, the techniques of GPT-Too, an AMR-to-English converter based on a pre-trained autoregressive language model which is fine-tuned to generate pairs of AMRs and corresponding English sentences, in that order. GPT-Too's pre-trained transformer-based approach has achieved state of the art results for AMR to English generation. In applying this approach to ULF, which preserves a large amount of syntactic information, we find that the model requires far less training data, and generates sentences closer in meaning to the originals, than the same model architecture applied to AMR.

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.

Pre-trained language models seem to naturally excel as unsupervised, few-shot, multi-task learners. As such models are pre-trained on natural language text, it seems intuitive that transfer learning might happen more efficiently when fine-tuning on tasks utilizing representations with structural similarities to natural language. We hope, as larger models become easier to run, and as more datasets become available for semantic representations like ULF, that many tasks utilizing such representations will see increased performance from fine-tuned language models.