Learning Natural Language Interfaces with Neural Models

Li Dong (Microsoft Research; lidong1@microsoft.com)

DOI: 10.1145/3478369.3478375

Introduction

Language is the primary and most natural means of communication for humans. The learning curve of interacting with various services (e.g., digital assistants, and smart appliances) would be greatly reduced if we could talk to machines using human language. However, in most cases computers can only interpret and execute formal languages.

The research goal of my dissertation is to use neural models to build natural language interfaces which learn to map naturally worded expressions onto machine-interpretable representations. The task is challenging due to (1) structural mismatches between natural language and formal language, (2) the well-formedness of output representations, (3) lack of uncertainty information and interpretability, and (4) the model coverage for language variations. In this dissertation, we develop several flexible neural architectures to address these challenges.

The dissertation presents a framework based on encoder-decoder neural networks for natural language interfaces. Beyond sequence modeling, we propose a tree decoder to utilize the compositional nature and well-formedness of meaning representations, which recursively generates hierarchical structures in a top-down manner. To model meaning at different granularity levels, we present a structure-aware neural architecture which decodes semantic representations following a coarse-to-fine procedure.

The proposed neural models remain difficult to interpret, acting in most cases as a black box. We explore ways to estimate and interpret the model's confidence in its predictions, which we argue can provide users with immediate and meaningful feedback regarding uncertain outputs. We estimate confidence scores that indicate whether model predictions are likely to be correct. Moreover, we identify which parts of the input contribute to uncertain predictions allowing users to interpret their model.

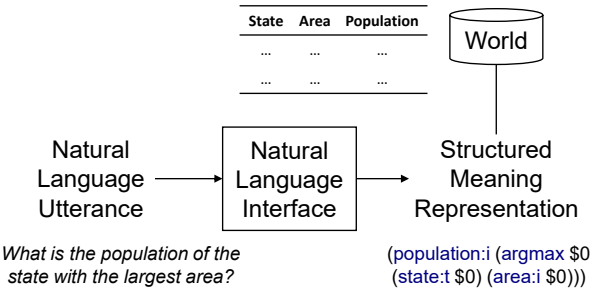


Figure 1: The goal of natural language interfaces is allowing users to interact with computers in human language. As shown by the example from the GEO dataset (Zelle & Mooney, 1996; Zettlemoyer & Collins, 2005), the model maps the input question to the λ-calculus meaning representation, and then execute it over the database to obtain the answer.

Model coverage is one of the major reasons resulting in uncertainty of natural language interfaces. Therefore, we develop a general framework to handle the many different ways natural language expresses the same information need. We leverage external resources to generate felicitous paraphrases for the input, and then feed them to a neural paraphrase scoring model which assigns higher weights to linguistic expressions most likely to yield correct answers. The model components are trained end-to-end using supervision signals provided by the target task.

Experimental results show that the proposed neural models can be easily ported across tasks. Moreover, the robustness of natural language interfaces can be enhanced by considering the output well-formedness, confidence modeling, and improving model coverage.

Neural Semantic Parsing

Semantic parsing is the task of translating text to a formal meaning representation such as logical forms or structured queries, which is one of the core components of natural language interfaces (as shown in Figure 1).

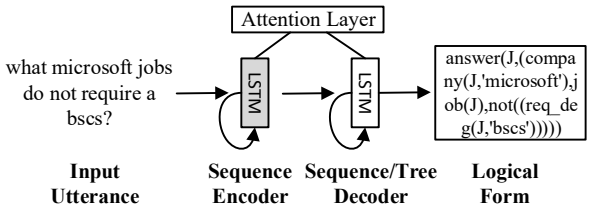


Figure 2: Input utterances and their logical forms are encoded and decoded with neural networks. An attention layer is used to learn soft alignments.

There has recently been a surge of interest in developing machine learning methods for semantic parsing, due in part to the availability of corpora containing utterances annotated with formal meaning representations. In order to predict the correct logical form for a given utterance, most previous systems rely on predefined templates and manually designed features, which often render the parsing model domain- or representation-specific.

In the dissertation, we aim to use a portable method to bridge the gap between natural language and logical form with minimal domain and linguistic knowledge (Dong & Lapata, 2016). The proposed framework (as shown in Figure 2) is portable as the models can be end-to-end trained by giving annotated data, namely, natural language utterances paired with their meaning representations. So we can easily adapt the models to different applications with minimal efforts. The structural gap between inputs and outputs is bridged by neural encoder-decoder networks augmented with attention mechanisms.

Constrained Decoding

After obtaining the output meaning representations from natural language interfaces, we usually need to execute them to obtain user intentions. Because the downstream executors only accept grammatical programs, it is beneficial to explicitly model the structure of predictions. The structural information of the output should be taken into consideration so that the models can generate well-formed meaning representations.

The fact that meaning representations are typically structured objects prompts efforts to develop neural architectures which explicitly account for their structure. In order to guarantee

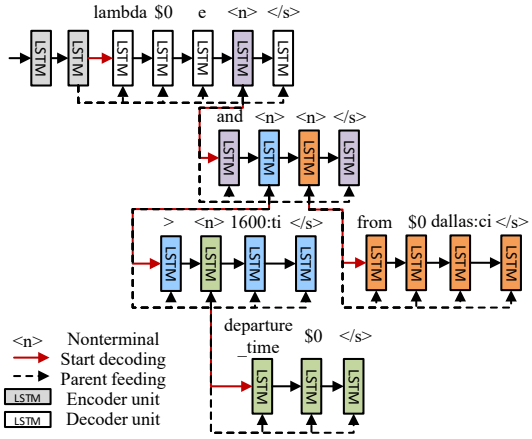


Figure 3: Sequence-to-tree model (Dong & Lapata, 2016) with a hierarchical tree decoder.

the well-formedness of the output we explicitly model the hierarchical and compositional nature of meaning representations, and thus develop a sequence-to-tree model (Dong & Lapata, 2016), and a coarse-to-fine decoding algorithm (Dong & Lapata, 2018).

In the first solution, the proposed tree decoder defines a placeholder to indicate nonterminal nodes as shown in Figure 3. Tree structures are recursively generated in a top-down, and left-to-right manner. The explicit modeling of hierarchical structures constrains results in the space of well-formed trees. In other words, ill-formed logical forms can be pruned from the candidate set. In order to model meaning at different levels of granularity, the second solution uses a structure-aware neural architecture to decode semantic representations from coarse to fine. The coarse meaning decoder first generates a rough sketch of the meaning representation, which omits low-level details, such as arguments and variable names. Then, the fine meaning decoder fills in missing details by conditioning on the input utterance and the sketch itself.

Confidence Modeling

Neural semantic parsing models map natural language text to a formal meaning representation (e.g., logical forms or SQL queries). However, despite achieving promising results, the neural semantic parsers remain difficult to interpret, acting in most cases as a black box, not providing any information about what

made them arrive at a particular decision. My dissertation explores ways to estimate and interpret the model's confidence in its predictions, which we argue can provide users with immediate and meaningful feedback regarding uncertain outputs (Dong, Quirk, & Lapata, 2018).

An explicit framework for confidence modeling would benefit the development cycle of neural semantic parsers which, contrary to more traditional methods, do not make use of lexicons or templates and as a result the sources of errors and inconsistencies are difficult to trace. Moreover, from the perspective of application, semantic parsing is often used to build natural language interfaces, such as dialogue systems. In this case it is important to know whether the system understands the input queries with high confidence in order to make decisions more reliably. For example, knowing that some of the predictions are uncertain would allow the system to generate clarification questions, prompting users to verify the results before triggering unwanted actions. In addition, the training data used for semantic parsing can be small and noisy, and as a result, models do indeed produce uncertain outputs, which we would like our framework to identify.

We categorize the causes of uncertainty into three types, namely *model uncertainty*, *data uncertainty*, and *input uncertainty* and design different metrics to characterize them. Furthermore, we propose a method based on backpropagation which allows to interpret model behavior by identifying which parts of the input contribute to uncertain predictions.

Query Paraphrasing

One of the challenges to build a robust natural language interface is model coverage. Due to the limited size of training data, it is challenging to handle the many different ways natural language expresses the same information need. As a result, small variations in semantically equivalent inputs may yield different results. For example, a hypothetical natural language interface must recognize that the questions "*who created microsoft*" and "*who started microsoft*" have the same meaning and that they both convey the `founder` relation in order to obtain the correct answer from a knowledge base. Moreover, one of the main causes of uncertainty is defined as data uncertainty, namely, uncertainty of predictions affected by the coverage of training data. If the pattern of input is unseen by the model on training data, it is difficult to predict reliable outputs. We leverage external resources to rewrite the natural language input during both training and test, so that model coverage can be increased by augmenting the original expression with its variations (Dong, Mallinson, Reddy, & Lapata, 2017).

Next Steps

This dissertation focuses on a single domain, and were trained on English utterances paired with their meaning representations. It is worth studying and exploring how to improve the proposed models' scalability in terms of supporting many different domains, languages, and supervision signals:

- **Cross-Domain Sharing.** For real-world applications (such as voice assistants), a system usually needs to handle many different domains. It is helpful to transfer and share knowledge across domains and meaning representations, especially when the data size of each domain is not large enough. We can share the common operators, predicates, and composition structures for similar examples.
- **Zero/Few-Shot Learning.** The training data size for a new domain is often small or even non-existent. It is valuable to conduct few-shot or zero-shot learning for a cold start, so that the model can be quickly adapted to a new similar domain.
- **Data Collection.** Model performance can usually be improved if more annotated data is fed to the model. However, sometimes it is difficult to directly annotate meaning representations for ordinary users. A new paradigm of training data collection is critical for the acceleration of model deployment.
- **Weakly Supervised Learning.** The models presented in this dissertation were trained on natural language utterances paired with their meaning representations. Given the data paucity of such parallel corpora, we can learn models from weak supervision signals (e.g., question-answer pairs),

which would reduce the annotation burden. Weakly supervised learning also provides a way to utilize online user feedback.

- **Multilingual Semantic Parsing.** Natural language interfaces should accept multiple languages, so users from different world regions can freely use their native language. A typical problem is that for some languages there is less data compared to others, which inevitably results in inferior semantic parsing performance.
- **Multi-Turn Interactions.** Sometimes users express their intentions in multiple utterances or update the requests according to the system responses. Users often tend to omit information which has been expressed in the conversation history. So the prediction should be conditioned on the current utterance as well as the interaction history.

References

Dong, L., & Lapata, M. (2016, August). Language to logical form with neural attention. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 33–43). Berlin, Germany.

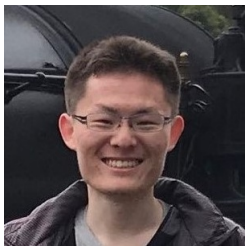
Dong, L., & Lapata, M. (2018). Coarse-to-fine decoding for neural semantic parsing. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 731–742). Association for Computational Linguistics.

Dong, L., Mallinson, J., Reddy, S., & Lapata, M. (2017, September). Learning to paraphrase for question answering. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 875–886). Copenhagen, Denmark: Association for Computational Linguistics.

Dong, L., Quirk, C., & Lapata, M. (2018). Confidence modeling for neural semantic parsing. In *Proceedings of the 56th annual meeting of the association for computational linguistics* (pp. 743–753). Association for Computational Linguistics.

Zelle, J. M., & Mooney, R. J. (1996). Learning to parse database queries using inductive logic programming. In *Proceedings of the 13th national conference on artificial intelligence* (pp. 1050–1055). Portland, Oregon.

Zettlemoyer, L., & Collins, M. (2005). Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the 21st conference on uncertainty in artificial intelligence* (pp. 658–666). Edinburgh, Scotland.



Li Dong is a Researcher at Microsoft Research, working on natural language processing, and representation learning. He received his PhD in School of Informatics at University of Edinburgh. Li's research has been recognized through the AAAI/ACM SIGAI Doctoral Dissertation Award Runner Up, the ACL-2018 Best Paper Honourable Mention, the AAAI-2021 Best Paper Runner Up, and fellowship from Microsoft.