Research Statement

Zhanming (Allan) Jie
ByteDance AI Lab
allan@bytedance.com
Last update: April 11, 2022

Abstract

We have been working extensively on building efficient structured models for fundamental natural language processing tasks such as named entity recognition and semantic parsing (Jie, 2020). We proposed various hypergraph structures and developed efficient dynamic-programming algorithms to perform inference. Our research efforts demonstrated that the hypergraph-based models have a strong capability and expressivity in extracting structures from text. We further developed a structured prediction framework that is able to perform efficient inference for almost all DAG-based hypergraphs. On top of that, we are currently focusing on developing structured models that are capable to perform structured symbolic reasoning and achieve the *system 2* ability (Kahneman, 2011; Bengio et al., 2021). As structures are expensive to annotate in natural language, we also introduce the future direction towards using self-supervised learning for extracting structures.

1 Structured Prediction

We built structured models for the task of named entity recognition (NER) (Jie et al., 2017; 2019; Jie & Lu, 2019; Xu et al., 2021; Yang et al., 2020) and semantic parsing (Jie & Lu, 2014; 2018). We focused on these two tasks as both of them require semantic-level understanding from natural language. Motivated by the shallow semantics existing in the syntactic dependency trees, we built *dependency-guided* models for NER (Tjong Kim Sang & De Meulder, 2003) and extracted meaningful structures in semantic parsing (Zelle & Mooney, 1996). The special property that named entities often form sub-tree structures in syntactic dependencies allows us to efficiently identify the possible candidates of named entities in the graphical model design. We design a *dependency-guided model* (DGM) (Jie et al., 2017) based on the semi-Markov CRF (Lafferty et al., 2001; Sarawagi & Cohen, 2004) model with linear complexity on average while achieving comparable or better performance than the semi-Markov CRF. Furthermore, we proposed the *dependency-guided LSTM-CRF* models (Jie & Lu, 2019; Xu et al., 2021) to fully exploit the dependency structured information such as high-order dependencies and dependency labels. In semantic parsing, we proposed the *dependency-based hybrid tree* model (Jie & Lu, 2014; 2018) to learn the latent dependency trees (Zhou et al., 2021) which is able to represent the semantic structure and capture the alignment with natural language. Motivated by the Eisner's algorithm in dependency parsing (Eisner, 2000), we design a hypergraph that allows us to perform dynamic programming and efficiently extract the latent dependency-based hybrid tree from natural language. The experiments show that dependency-based hybrid trees are more robust in capturing the language-specific characteristics such as word-ordering compared to the constituency-based trees (Lu, 2014). Our research efforts demonstrate that the hypergraph-based models are expressive and capable of learning meaningful structures from natural language.

1.1 Efficient Hypergraph Inference Framework

In practice, we need to design efficient inference algorithms for different hypergraph structures (e.g., linear-chain or tree structures). Such algorithms often cost a significant amount of efforts to implement. In order to make the inference process more generic, we have developed and continue to work on a structured prediction framework (Lu, 2017)[1] that is able to help us perform inference on

[1] https://github.com/sutd-statnlp/statnlp-neural

almost all arbitrary structures. The underlying inference algorithm is a generalized version of the *inside-outside* algorithm (Baker, 1979) where we developed the extension and applied to general directed acyclic hypergraph structures. As the linear-chain CRF is the most commonly used architecture in practice, we also developed the efficient framework[2] for the research community to use. Our framework is able to improve the experimental complexity from $\mathcal{O}(N)$ to $\mathcal{O}(\log N)$ following parallel-scan algorithm (Blelloch, 1990; Rush, 2020). Along this line of work, we are interested in building more efficient algorithms for general hypergraph structures. For example, typical constituency parsing or other structured parsing tasks require efficient inference algorithm for tree-based structures. How to efficiently incorporate the inference procedure into our frameworks remains as a technical challenge. The Torch-Struct (Rush, 2020) framework implemented many architectures but the underlying codebase is specific to some particular structures. Our purpose is to allows developers or researchers to only focus on the structure design rather than the inference procedure.

2 Structured Symbolic Reasoning

W also focus on the reasoning ability of structured models (Jie et al., 2022) where we perform reasoning over the structures. Specifically, we try to answer the question that "*How can we make use of the expressiveness in graphical models to perform symbolic reasoning?*". Our goal is to build an interpretable model that gives explainable predictions under the scope of modern neural architectures. We proposed a deductive system that allows us iteratively construct partial structures and perform reasoning over these structures. We adopt the math word problem (MWP) solving as the downstream task as it essentially evaluate the reasoning ability of our proposed model. To make iterative and explainable predictions, we carefully design the "*expression chain*" structures using neural module networks (Andreas et al., 2016; Gupta et al., 2020). We then perform reasoning (such as *addition* and *subtraction*) among the mathematical quantities by forming mathematical expressions. Our experiments demonstrate that the underlying deductive system presents explainable step during the MWP solving procedure.

The underlying deductive system design can also be applied to other tasks that consider multi-step reasoning, such as HotpotQA (Yang et al., 2018). The reasoning happens between states (e.g., partial structures) where we apply differentiable actions to evolve the previous state to a new state (Guo et al., 2018). The resulting step-by-step process is similar to the concept of "*reasoning chain*" (Jhamtani & Clark, 2020; Saha et al., 2020) and diffusion models (Sohl-Dickstein et al., 2015; Weng, 2021). For example, we are interested in how models exploit the relations among the rationales (i.e., relevant sentences) in reading comprehension. Such relations could be latent and we need to build latent-variable models (Jie & Lu, 2018) to find such relations. Besides, we can also apply the deductive system to dependency parsing where the tree generation happens in a top-down and step-by-step manner. As for diffusion models in DALL-E-2 (Ramesh et al., 2021), the deductive system is also able to sample from the intermediate state and continue to generate new states.

3 Self-Supervised Learning for Structures

The success of Transformers (Vaswani et al., 2017) architecture brings our attention to self-supervised learning. Transformer-based models with self-supervised pre-training are able to achieve state-of-the-art performance in downstream NLU tasks. In the recent remark by Meta Research[3]:

> "*We believe that self-supervised learning is one of the most promising ways to build such background knowledge and approximate a form of common sense in AI systems.*"

A huge amount of research efforts have demonstrated that the current language models with standard random masking pre-training strategy has a strong capability in understanding the natural language as well as the linguistic structures behind it. However, the secret behind remains a mystery to us as we cannot explicitly extract or model the meaning or the linguistic structures through self-supervised learning. Such phenomena inspire us to explore the direction that extracting structures from the self-supervised perspective. For example, Hu et al. (2021) shows that we can perform grammar induction after pre-training with a parsing specific objective function. The essential idea is to design an elegant

[2] `https://github.com/allanj/pytorch_neural_crf`

[3] https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/

pre-training objective function to incorporate the structured prior knowledge of the downstream task, such as semantic and syntactic structures.

REFERENCES

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Neural module networks. In *Proceedings of CVPR*, 2016.

James K Baker. Trainable grammars for speech recognition. *The Journal of the Acoustical Society of America*, 65(S1):S132–S132, 1979.

Yoshua Bengio, Yann Lecun, and Geoffrey Hinton. Deep learning for ai. *Communications of the ACM*, 64(7):58–65, 2021.

Guy E Blelloch. Prefix sums and their applications. 1990.

Jason Eisner. Bilexical grammars and their cubic-time parsing algorithms. In *Advances in probabilistic and other parsing technologies*, pp. 29–61. Springer, 2000.

Chu Guo, Zhanming Jie, Wei Lu, and Dario Poletti. Matrix product operators for sequence-to-sequence learning. *Physical Review E*, 98(4):042114, 2018.

Nitish Gupta, Kevin Lin, Dan Roth, Sameer Singh, and Matt Gardner. Neural module networks for reasoning over text. In *Proceedings of ICLR*, 2020.

Xiang Hu, Haitao Mi, Zujie Wen, Yafang Wang, Yi Su, Jing Zheng, and Gerard de Melo. R2d2: Recursive transformer based on differentiable tree for interpretable hierarchical language modeling. In *Proceedings of ACL*, 2021.

Harsh Jhamtani and Peter Clark. Learning to explain: Datasets and models for identifying valid reasoning chains in multihop question-answering. In *Proceedings of EMNLP*, 2020.

Zhanming Jie. *Leveraging Dependency Trees for Structured Prediction*. PhD thesis, Singapore University of Technology and Design, 2020.

Zhanming Jie and Wei Lu. Multilingual semantic parsing : Parsing multiple languages into semantic representations. In *Proceedings of COLING*, 2014.

Zhanming Jie and Wei Lu. Dependency-based hybrid trees for semantic parsing. In *Proceedings of EMNLP*, 2018.

Zhanming Jie and Wei Lu. Dependency-guided lstm-crf for named entity recognition. In *Proceedings of EMNLP*, 2019.

Zhanming Jie, Aldrian Obaja Muis, and Wei Lu. Efficient dependency-guided named entity recognition. In *Proceedings of AAAI*, 2017.

Zhanming Jie, Pengjun Xie, Wei Lu, Ruixue Ding, and Linlin Li. Better modeling of incomplete annotations for named entity recognition. In *Proceedings of NAACL*, 2019.

Zhanming Jie, Jierui Li, and Wei Lu. Learning to reason deductively: Math word problem solving as complex relation extraction. In *Proceedings of ACL*, 2022.

Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML*, pp. 282–289, 2001.

Wei Lu. Semantic parsing with relaxed hybrid trees. In *Proceedings EMNLP*, 2014.

Wei Lu. A unified framework for structured prediction: From theory to practice. In *Proceedings of EMNLP: Tutorial Abstracts*, 2017.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditiona limage generation with clip latents. 2021. URL https://cdn.openai.com/papers/dall-e-2.pdf.

Alexander M Rush. Torch-struct: Deep structured prediction library. In *Proceedings of ACL: System Demonstrations*, pp. 335–342, 2020.

Swarnadeep Saha, Sayan Ghosh, Shashank Srivastava, and Mohit Bansal. Prover: Proof generation for interpretable reasoning over rules. In *Proceedings of EMNLP*, 2020.

Sunita Sarawagi and William W Cohen. Semi-markov conditional random fields for information extraction. In *Proceedings of NeurIPS*, pp. 1185–1192, 2004.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of ICML*, 2015.

Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of NAACL*, 2003.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of NeurIPS*, 2017.

Lilian Weng. What are diffusion models? *lilianweng.github.io*, 2021. URL https://lilianweng.github.io/posts/2021-07-11-diffusion-models/.

Lu Xu, Zhanming Jie, Wei Lu, and Lidong Bing. Better feature integration for named entity recognition. In *Proceedings of NAACL*, pp. 3457–3469, 2021.

Xuefeng Yang, Benhong Wu, Zhanming Jie, and Yunfeng Liu. Overview of the nlpcc 2020 shared task: Autoie. In *Proceedings of NLPCC*, 2020.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of EMNLP*, 2018.

John M Zelle and Raymond J Mooney. Learning to parse database queries using inductive logic programming. In *Proceedings of the national conference on artificial intelligence*, 1996.

Yuxiang Zhou, Lejian Liao, Yang Gao, Zhanming Jie, and Wei Lu. To be closer: Learning to link up aspects with opinions. In *Proceedings EMNLP*, 2021.