Make Knowledge Computable: Differentiable Neural Symbolic Reasoning

Ziniu Hu, Department of Computer Science, UCLA

1 Overview

The quest for making world knowledge computable has a long history in artificial intelligence. Earlier symbolic AI systems attempt to hard-code expert-level knowledge into programs. Though powerful to solve very complex questions[1], these systems are mostly domain specific, which require lots of human efforts to maintain the knowledge base and could not continuously learn from data. Over the past years, deep learning have shown great capacity to memorize a surprising amount of world knowledge. By training from massive corpora in an end-to-end manner, recent large-scale neural models can even outperform humans in many language and vision tasks, such as machine translation and image captioning. Yet, as knowledge is stored implicitly in the parameters of neural networks, existing neural models fail to handle many complex tasks that require reasoning over symbolic knowledge. For example, answering complex questions given an image/video requires inferring in-context and external knowledge of objects and their relationships, based on which to conduct compositional and logical reasoning; Synthesizing high-level hardware programs requires understanding structural and symbolic `C/C++` codes, predicting execution results and searching optimal programs in discrete space.

To address these issues, I propose a different direction: Instead of compiling the world knowledge statically into model parameters, I aim to model these symbolic knowledge in a more modular design, such that both neural models and symbolic AI module could understand the knowledge, compute and conduct reasoning. This vision is close to traditional **Neural-Symbolic AI systems** that integrate the two worlds. In these systems, the neural model focuses on parsing the input query x into symbolic programs z (such as SQL query or arithmetic circuit). Based on z, a symbolic module (such as numerical and logical solver) focuses on planning, deduction and reasoning, to generate the answer y. Despite the advantages of Neural-Symbolic AI systems, most previous studies in this line face a crucial obstacle: As symbolic modules are not differentiable, we cannot train the whole Neural-Symbolic model end-to-end, i.e., only using (x, y) pairs to train the whole system. Instead, most previous efforts require the annotation of the intermediate symbolic query z, to train the neural module. For many real-world applications, high-quality intermediate labels are arduous or even impossible to obtain, and the neural parser trained on limited training set of z cannot generalize well to different domains/distributions. This significantly limits the usage of previous Neural-Symbolic AI systems.

My ultimate research goal is to enable neural model to interact with symbolic reasoning module in a differentiable manner, and train such Neural-Symbolic model end-to-end without intermediate labels. To bring this vision about, I have conducted works on:

- **Designing Novel Reasoning Module**: design differentiable neural modules that can conduct symbolic reasoning, including knowledge graph reasoning [1, 2] and complex Logical inference [3].
- **Learning via Self-Supervision**: train the neural model via self-supervision from structural and symbolic knowledge base without additional annotation [4, 5, 6].
- **Generalizing across Domains**: the modular design of neural-symbolic system by its nature help to generalize better for Out-of-Distribution [7], Out-of-Vocabulary [8], cross-lingual [9] and cross-type [10].

Putting these pieces together, I am pursuing the ultimate vision to build end-to-end Neural-Symbolic system that has the capacity of reasoning, advancing to true human intelligence.

2 Prior Research Achievements

My vision is supported by my prior research, which has led to more than 20 research papers published in top Machine Learning venues (NeurIPS, ICLR, AAAI), Data Mining venues (KDD, WWW, WSDM), and Nature Language Processing venues (ACL, EMNLP). Notably, I received the **Best Full Paper Award** at WWW 2019, **Best Student Paper Award** at DLG-KDD 2020, and **Best Paper Award** at SoCal-NLP 2022. Many models I design have been integrated into machine learning libraries such as Pytorch-Geometric and DGL[2], utilized in many industrial products, including Google Youtube Shorts recommendation, Microsoft Graph, Facebook hate speech detection, Tiktok & Toutial search engine [11] and stock trend prediction service by Microsoft [12]. Collectively, my papers have been cited more than 1500 times since 2018, according to Google Scholar. The software tools I developed have received over 1300 stars in total on Github, and also served as core building blocks for many NSF research grants.

[1] `https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/breakthroughs/`

[2] They have become basic building blocks for modern models for structured and geometric data and are widely used in academia and industry. My proposed HGT [1] model is used as official tutorial in PyG.

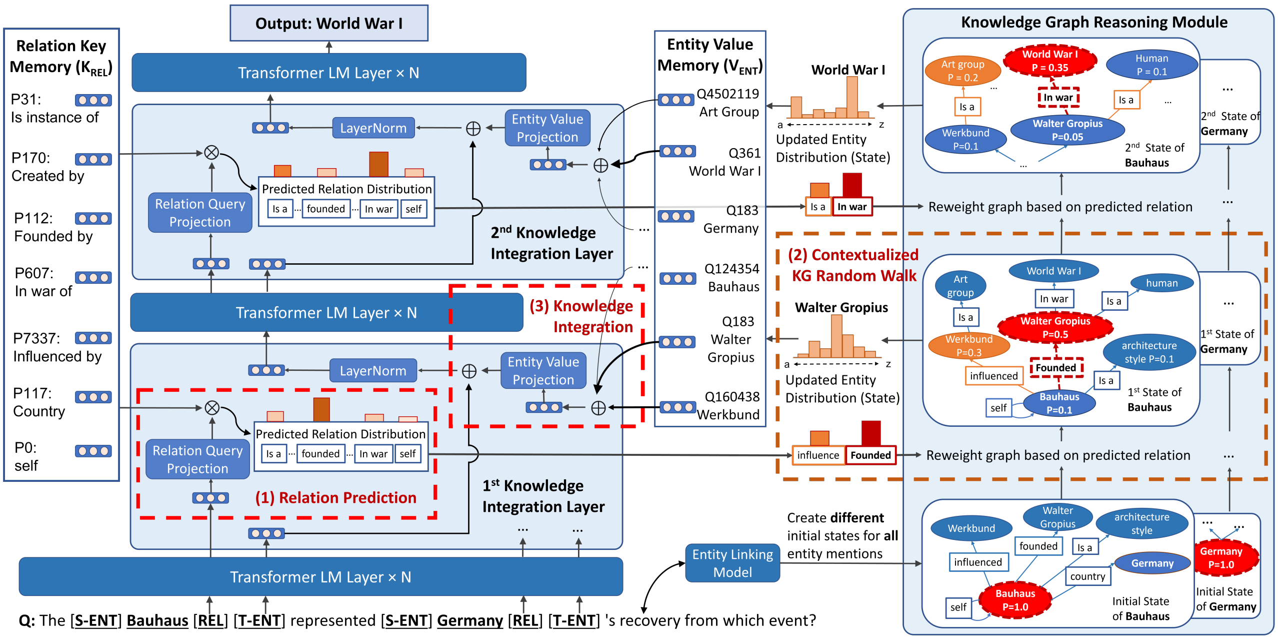

Figure 1: **OreoLM**: Integrating differentiable Knowledge Graph reasoning into neural Language Model. Three key procedures are highlighted in red dotted box: 1) *Relation Prediction*: predicts relation action for each entity mention. 2) *Contextualized KG Random Walk*: Based on the predicted relation, we re-weight each graph and conduct contextualized random walk to update entity distribution state. 3) *Knowledge Integration*: An weighted aggregated entity embedding is added into a placeholder token as retrieved knowledge.

2.1 Differentiable Symbolic Reasoning Module

The most important module to build an end-to-end Neural-Symbolic system is to make the symbolic reasoning step differentiable. My prior research focuses on **reasoning over Knowledge Graph** (KG), which consists of three central subtasks: 1) modeling heterogeneous KG via neural models; 2) reasoning over the knowledge graph to answer complex question; 3) answering first-order logical queries in embedding space.

HGT [1] addresses the first subtask to model complex multi-relational and large-scale knowledge graphs via Graph Neural Networks (GNNs). Before HGT, most existing GNNs are designed for homogeneous graphs, in which all nodes and edges belong to the same types, making them infeasible to represent more complex relational data, i.e., heterogeneous KG. To solve this problem, we propose Heterogeneous Graph Transformer (HGT) for modeling web-scale [13] heterogeneous graphs. We leverage the meta relation triplet to parameterize the weight matrices to calculate attention on each edge, empowering HGT to maintain dedicated representations for different nodes and edges. HGT can incorporate information from different types of high-order neighbors through messages passing across layers, so it can automatically learn "meta paths" which are essential for different downstream tasks. HGT significantly improves both benchmark performance and Microsoft Graph anomaly detection and Facebook hate speech detection service.

OreoLM [2] tackles the second subtask to conduct differentiable knowledge graph reasoning via graph walking, and is my first attempt to integrate such symbolic reasoning module with the neural language model (LM). The proposed knOwledge REasOning empowered Language Model (OreoLM) consists of a fully differentiable Knowledge Interaction Layer that could be inserted amid arbitrary Transformer layers as the interface for LMs to interact with KG. As illustrated in Figure 1, KIL sends relational instructions to guide KG reasoning and retrieve knowledge to solve the question. With the predicted relation, we conduct symbolic state transition for each reasoning path as walking over the graph. In this way, LM guides KG to walk towards the desired answer, while the retrieved knowledge improves LM. By adopting OreoLM to RoBERTa and T5, we show significant performance gain and state-of-the-art closed-book question answering performance and could summarize critical reasoning paths to interpret the model decision. OreoLM is being shifted to open-domain question-answering and summarizing system at Microsoft Azure Service.

FuzzQE [3] targets in the third and most challenging subtask of conducting first-order logical reasoning with conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$), illustrated in Figure 2. Our proposed Fuzzy Query Embedding (FuzzQE) borrow the idea of fuzzy logic as differentiable logical operators, which fully satisfy the axioms of logical operations and can preserve logical operation properties in vector space. In addition, our logical operations do not require learning any operator-specific parameters. We conducted experiments to show that even when our model is only trained with link prediction (without any complex query), it achieves better results than state-of-the-art logical query embedding models trained with extra complex query data.

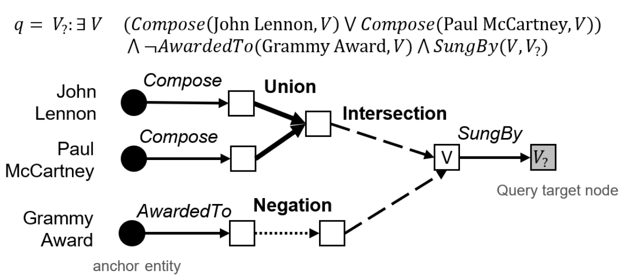

Figure 2: **FuzzQE** uses fuzzy logic as operators to build computational graph for answering First-Order Logic (FOL) Query over knowledge graph.

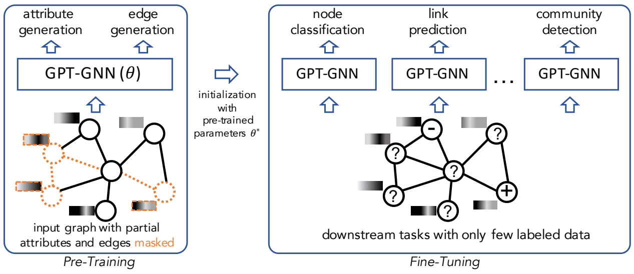

Figure 3: **GPT-GNN** introduce a self-supervised graph generation task to pre-train GNN, which could solve diverse downstream tasks with few labelled data.

2.2 Self-Supervised Learning from Structural Symbolic Knowledge

Recent deep learning is rapidly progressing from task-specific systems towards fundamental models that are trained from massive unlabeled datasets. To guide the proposed Neural-Symbolic models learning world knowledge without labeled data (especially intermediate annotations), it is ideal to also learn from unlabelled data. The challenges is that the complex reasoning module normally requires learning signal from structural knowledge instead of raw data like image and text corpus. To solve this issue, I propose several attempts to utilize the structural symbolic knowledge (e.g., knowledge graph) as self-supervision to pre-train neural models.

GPT-GNN [4] aims to capture the intrinsic structural and semantic properties of the graph so that it can generalize to any downstream tasks on this graph with a few fine-tuning steps, illustrated in Figure 3. To achieve this goal, we propose Generative Pre-Training of Graph Neural Networks (GPT-GNN), which models the graph distribution by learning to reconstruct the graph. To capture inherent dependency between node attributes and graph structure, we factorize the likelihood of graph generation into two steps with bridge edges as communication. GPT-GNN is used by Facebook to pre-train HGT [1] on billion-scale social media graph.

RGPT-QA [6] synthesizes a relational QA dataset covering a wide range of relations from both Wikidata triplets and Wikipedia hyperlinks. We then pre-train a QA model to infer the latent relations from the question, and then conduct extractive QA to get the target answer entity. RGPT-QA enhances the performance of popular QA models, especially on questions with long-tail relations.

ReVeaL [5] (Retrieval-Augmented Visual Language Pre-Training) transforms multimodal world knowledge into a key-value memory using neural representation learning, and then retrieve from it to answer knowledge-intensive queries. By decoupling the knowledge memorization from reasoning, we enable our model to leverage various external sources of multimodal knowledge (Wikipedia passages and images, the WikiData knowledge graph, Web image-text pairs and visual question answering data). ReVeaL achieves state-of-the-art performance in several knowledge-intensive Visual Question Answering and Image Captioning datasets.

Figure 4: **ReVeaL** augments a visual-language model with the ability to retrieve multiple knowledge from a diverse set of sources. Both retriever and generator are trained jointly in end-to-end manner.

2.3 Generalize across Domains via Modeling Symbolic Knowledge

One key advantage of symbolic reasoning over neural models is to better generalize to different distributions and domains. This is based on a key assumption called Sparse Mechanism Shift [14], stating that when transferring to a new domain with small distribution shifts, it tends to only change the ground-truth decision model in a local and sparse way. In other words, if we factorize the model into multiple modules, only a few modules need to be changed. For example, when doing math exams written in different languages, only the lower language understanding part need to get updated, while the mathematical reasoning module could generalize well among different domains. Therefore, by modeling data via a compositional and structural manner, we could regard each disentangled neural module to handle a specific functionality, and only change a particular module when we transfer to a new domain/distribution. Following this intuition, I conduct research to improve generalization for Out-of-Distribution, Out-of-Vocabulary and Cross-Lingual tasks.

MT-CRL [7] (Multi-Task Causal Representation Learning) aims to improve Out-of-Distribution generalization of multi-task learning (MTL) via regularizing spurious correlation. MT-CRL represents multi-task knowledge via disentangled neural modules, and learn which module is causally related to each task via MTL-specific graph invariant regularization. MT-CRL not only improves multi-task learning benchmark, but also is deployed in Google YouTube Shorts recommendation.

HiCE [8] (Hierarchical Context Encoder) aims to predict Out-Of-Vocabulary (OOV) word embedding based on corpus knowledge. We formulate OOV embedding learning as a few-shot regression task by predicting oracle embedding vectors (trained with abundant observations) based only on K contexts. Specifically, we use Model-Agnostic Meta-Learning (MAML) to adapt the model to the new corpus fast and robustly.

Emoji-LSA [9] improve cross-lingual sentiment classification performance. We utilize emojis, which are widely available in many languages, as a special symbolic knowledge to learn both the cross-language and language-specific sentiment patterns in different languages. The proposed method demonstrates the state-of-the-art performance on benchmark datasets, which are sustained even when sentiment labels are scarce.

3 Future Research Agenda

My ultimate research goal is to build a Neural-Symbolic Reasoning framework that has the ability to solve challenging tasks in many different research domains of computer science. This requires not only a powerful model design, but also deep understanding of diverse domain knowledge, making my research topic a interdisciplinary subject. During my Ph.D. studies, I have been fortunate to collaborate with researchers from diverse backgrounds, covering machine learning, data mining, NLP, computer vision and hardware infrastructure, from both academia and industry. This experience gives me a broad vision to continue this challenging but exciting research direction. In the future, I am excited to further improve my proposed neural-symbolic reasoning framework, as well as using it to solve most fundamental and significant challenges in other areas in computer science, such as program synthesis, hardware design, mathematical auto-proving and scientific discovery:

1. Towards More Expressive Differentiable Symbolic Reasoning Systems. My past studies have focused on symbolic reasoning over knowledge graphs. Outside this research domain, there exist many other interesting and powerful symbolic reasoning AI, including numerical reasoning, physics simulation, mathematical theorem prover, as well as many pre-defined APIs provided by industrial services. To enable integration of these symbolic reasoning capacities into neural models, I plan to build a more general interface to bridge the two worlds, supporting free interaction and backpropagation. Many challenges remain to be addressed, including how to properly model this heterogeneous and structural knowledge in a principled manner (ideally in a unified graph view), choose appropriate abstractions for reasoning procedure, and make reasoning differentiable. I am also interested in improving causal representation learning via modular design, and making the AI model capable of conducting causal inference and estimate uncertainty and risks.

2. Explore Program Synthesis via Neural-Symbolic Reasoning. Many fundamental tasks in computer science and artificial intelligence could be formalized as program synthesis. For example, dialogue chatbots require parsing human language into formal SQL programs; mathematical auto-proving requires transforming math equations; high-level synthesis of FPGA program requires compiling and latent execution of discrete `C/C++` programs. My past research on graph representation learning and symbolic reasoning is a natural solution for conducting program synthesis. Therefore, I am excited to apply my proposed neural-symbolic models to solve these interesting and challenging tasks. Take hardware synthesis as an example, I aim to represent symbolic program as latent variables, which we could execute via neural module to infer results. Based on it, we could search for the best program by optimizing the latent program to maximize output, in a differentiable manner.

3. Empower Scientific Discovery via Neural-Symbolic Reasoning. My proposed Neural-Symbolic models have already shown improvement in a wide range of artificial intelligence tasks. Outside of AI domains, many general scientific problems could also be abstracted as symbolic reasoning. For example, drug discovery and design require representing the molecule as geometric graphs; physics simulation requires understanding complex physic environments (represented with graph with particles, fluids, plasma as nodes, and their mutual interactions as edges). I am particularly interested in whether my proposed neural-symbolic AI models could be applied and benefit these fundamental scientific problems, helping building better scientific simulation tools. In addition, I am interested in utilizing the neural-symbolic systems to automatically discover world knowledge, including constructing domain-specific knowledge graph, discovering new Physics or Chemical governing laws from experiments, and identifying causal structures from real-world social data.

References

[1] **Ziniu Hu**, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous Graph Transformer. In *WWW 2020 (mostly cited paper)*.

[2] **Ziniu Hu**, Yichong Xu, Shuohang Wang, Ziyi Yang, Chengguang Zhu, Kai-Wei Chang, and Yizhou Sun. Empowering Language Models with Knowledge Graph Reasoning for Question Answering. In *EMNLP 2022 (best paper award in SoCal-NLP 2022)*.

[3] Xuelu Chen, **Ziniu Hu**, and Yizhou Sun. Fuzzy Logic based Logical Query Answering on Knowledge Graphs. In *AAAI 2022*.

[4] **Ziniu Hu**, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. GPT-GNN: Generative Pre-Training of Graph Neural Networks. In *KDD 2020 (Oral)*.

[5] **Ziniu Hu**, Alireza Fathi, David A Ross, Ahmet Iscen, Chen Sun, Zirui Wang, Kai-Wei Chang, Yizhou Sun, and Cordelia Schmid. ReVeaL: Retrieval-Augmented Visual Language Pre-Training with Multi-Source Multimodal Knowledge Memory.

[6] **Ziniu Hu**, Kai-Wei Chang, and Yizhou Sun. Relation-Guided Pre-Training for Open-Domain Question Answering. In *EMNLP-Finding 2021*.

[7] **Ziniu Hu**, Zhe Zhao, Xinyang Yi, Tiansheng Yao, Lichan Hong, Yizhou Sun, and Ed H Chi. Improving Multi-Task Generalization via Regularizing Spurious Correlation. In *NeurIPS 2022 (Spotlight)*.

[8] **Ziniu Hu**, Ting Chen, Kai-Wei Chang, and Yizhou Sun. Few-Shot Representation Learning for Out-Of-Vocabulary Words. In *ACL 2019*.

[9] Zhenpeng Chen, Sheng Shen, **Ziniu Hu**, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification. In *WWW 2019 (Best Full Paper Award)*.

[10] Minji Yoon, John Palowitch, Dustin Zelle, **Ziniu Hu**, Russ Salakhutdinov, and Bryan Perozzi. Zero-shot Transfer Learning within a Heterogeneous Graph via Knowledge Transfer Networks. In *NeurIPS 2022*.

[11] **Ziniu Hu**, Yang Wang, Qu Peng, and Hang Li. Unbiased LambdaMART: An Unbiased Pairwise Learning-to-Rank Algorithm. In *WWW 2019 (with US Patent)*.

[12] **Ziniu Hu**, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction. In *WSDM 2018 (with WIPO Patent),*.

[13] Difan Zou*, **Ziniu Hu***, Yewen Wang, Song Jiang, Yizhou Sun, and Quanquan Gu. Layer-Dependent Importance Sampling for Training Deep and Large Graph Convolutional Networks. In *NeurIPS 2019*.

[14] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proc. IEEE*, 109(5):612–634, 2021.