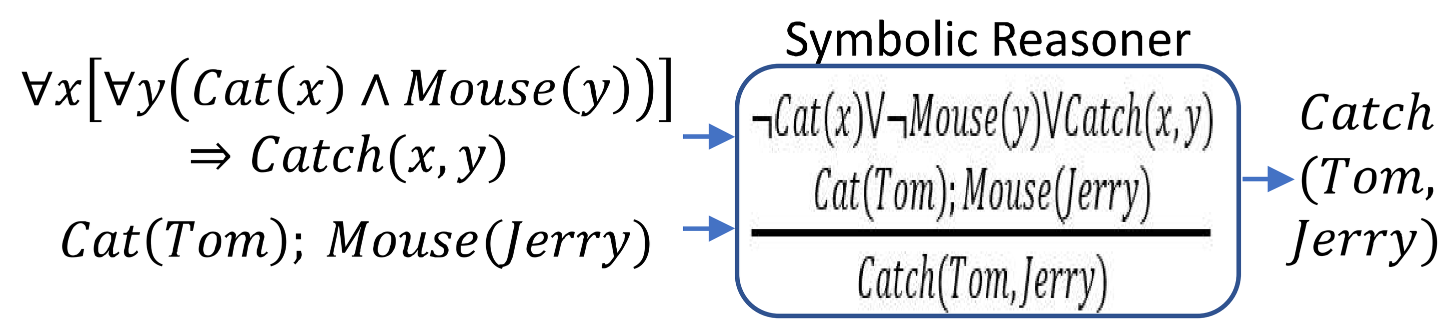
Logical Reasoning over Natural Language as Knowledge Representation

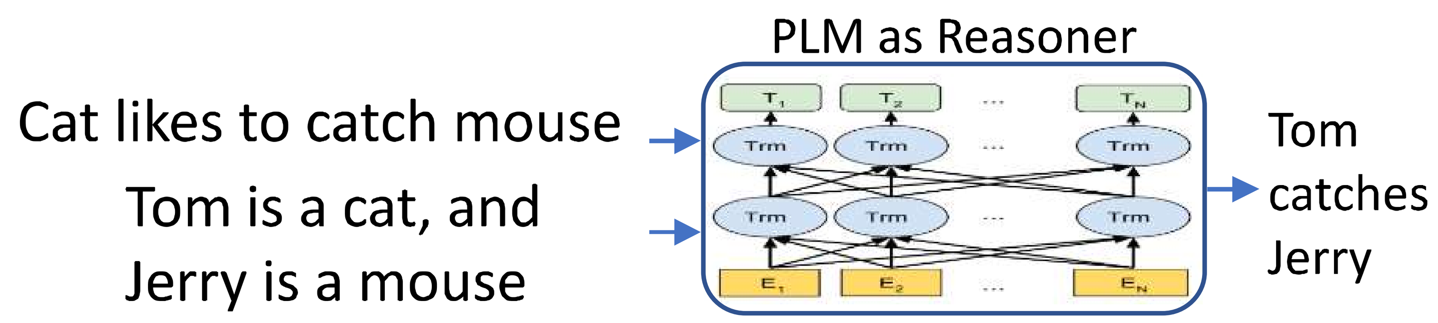
Zonglin Yang [1] Xinya Du [2] Rui Mao [1] Jinjie Ni [1] Erik Cambria [1]

[1]Nanyang Technological University [2]University of Texas at Dallas

A New Paradigm for Logical Reasoning



(a) Formal language as knowledge representation and symbolic reasoner



(b) Natural language as knowledge representation and PLM as reasoner

Figure 1. Comparison between the previous paradigm which uses formal representation and symbolic reasoner, and the new paradigm which uses natural language as knowledge representation and PLM as reasoner.

Definition and Categorization of Logical Reasoning

The definition and categorization of logical reasoning are handled by philosophy research. Overall it is still under debate for the clear definition of categorization of logical reasoning. Here we adopt [2]'s view.

Specifically, an argument consists of premises and a conclusion. If the premises can provide conclusive support to the conclusion, i.e., if the premises of the argument were all true, it would be impossible for the conclusion of the argument to be false, then the argument is a deductive argument. Otherwise the argument is an inductive or abductive argument. Both inductive and abductive reasoning provide explanations about the world but their explanation differ in the degree of generality. More specifically, the distinction between inductive and abductive conclusions strictly parallels the dichotomy *extension* vs. *intension*, or *generality* vs. *informativeness*. In other words, an inductive conclusion extends or generalizes to unobserved individuals, while an abductive one provides more specific information (e.g., unobserved properties) about existing specific individuals.

For example, if a white ball is found in a bag, inductive reasoning might lead to the conclusion that "all balls in this bag are white", while abductive reasoning might lead to the conclusion that "someone put the white ball into this bag".

Our main focus is to understand language model's logical reasoning ability through the three subtypes of logical reasoning to provide finer analysis and avoid ambiguity. Therefore we focus on papers that specialized on one (or more) of the three subtypes of logical reasoning (instead of only mention "reasoning" but no further specification or focus).

Highlights

- We summarize the three previously separately investigated reasoning types (in NLP) together with a new overall concept, logical reasoning over natural language (LRNL) as knowledge representation, and provide the first survey of LRNL. For each reasoning type, we review existing tasks, datasets, and methods for each task.
- We provide the first precise and non-ambiguous definition and categorization for logical reasoning in NLP (criteria borrowed from philosophy literature).
- We list advantages, challenges, and future directions for LRNL, from the aspect of the historical development of logical reasoning (especially the previous paradigm of computational methods for logical reasoning, the formal-language-based methods).

Advantages of LRNL

1. **Over Formal Language** (1) brittleness; (2) knowledge-acquisition bottleneck; (3) inability to handle raw data (e.g. natural language); (4) sensitivity to label errors; (5) failure to capture the semantic meaning of formal language symbols.
2. **Over Existing Nero-Symbolic (NeSy) Systems** (1) knowledge-acquisition bottleneck; (2) scalability.
3. **Over E2e Methods** (1) more explanability; (2) more controllability; (3) less catastrophic forgetting.

Deductive Reasoning

Existing tasks:

1. hypothesis classification: given background theories, to answer whether a particular hypothesis is correct.
2. proof generation: given background theories, to answer whether a particular hypothesis is correct and to give a proof tree from theories to hypothesis.
3. implication enumeration: given background theories, to reach to as much conclusions as possible.

Most papers in the field focus on proof generation task.

Method	Generation based	Inference w/ hypothesis	Stepwise	Proof direction	Heuristic search	Verifier	Human-authored realistic proof	Stage
PRover [10]	✗	✓	✗	N/A	N/A	✗	✗	1
multiPRover [11]	✗	✓	✗	N/A	N/A	✗	✗	1
EntailmentWriter [4]	✓	✓	✗	N/A	N/A	✗	✓	1
ProofWriter [14]	✓	✗	✓	→	✗	✗	✗	2
EVR [7]	✓	✗	✓	←	✗	✗	✗	2
IBR [8]	✗	✓	✓	←	✓	✗	✗	2
IRGR [9]	✓	✓	✓	→	✓	✗	✓	2
SI [3]	✓	✗	✓	→	✓	✗	✗	2
FaiRR [12]	✓	✗	✓	→	✓	✗	✗	2
MetGen [6]	✓	✗	✓	Both	✓	✗	✓	2
SCSearch [1]	✓	✗	✓	→	✓	✗	✓	2
ADGV [13]	✓	✗	✓	Both	✓	✓	✓	3
NLProofS [16]	✓	✓	✓	→	✓	✓	✓	3
Entailer [15]	✓	✓	✓	←	✓	✓	✓	3
Teachme [5]	✓	✓	✓	←	✓	✓	✗	3

Table 1. Methods for Proof Generation task. "Generation based" means whether $proof$ is created by generative inference model, otherwise is by utilizing embeddings to classify nodes and edges of $proof$. "Inference w/ $hypothesis$" means whether $hypothesis$ is provided during inference. → and ← denote forward/backward stepwise proof generation. "Heuristic seach" with ✗ means exhaustive search.

Current methods for the proof generation task roughly consist of three stages. In each stage, one key new technique is considered and developed. In stage 1, PLMs are used for forming $proof$ in one inference step. In stage 2, modular-based, stepwise frameworks are developed to create $proof$ (each module is usually implemented with a single PLM). In stage 3, a verifier is added as a new module to make sure that each reasoning step reflects the belief of PLMs.

Inductive Reasoning

Inductive reasoning is a less explored field, mainly because of lack of established connection between current development in NLP and (philosophical) definition of inductive reasoning. Existing tasks are rule classification and rule generation.

The most notable work in the field is <Language Models as Inductive Reasoners> [17], which establish the connection between the three fundamental requirements of inductive reasoning in philosophy literature, and use LLMs to evaluate generated rules from the three requirements.

The three requirements are "rule should be consistent with input facts", "rule should reflect the reality", and "rule should be more general than input facts".

Abductive Reasoning

Existing tasks for abductive reasoning can be summarized as explanation classification, and explanation generation w/o and w/ theory.

Most papers in the field focus on explanation classification. Explanation classification methods in general lack a progressive relationship within each other. However, many methods share one thing in common: they in general introduce knowledge via various ways, including retrieve structural or non-structural knowledge through prompt, multi-task, and another round of pretraining. Other methods include explore additional loss term, and treating the task as a ranking problem.

Challenges of LRNL & Possible Future Directions

Reliable Verifier on Reasoning Steps

Many state-of-the-art methods use verifier to check the correctness of generated reasoning results. However, the current verifiers only reflect the internal beliefs of PLMs. It is doubtful whether PLMs have obtained the knowledge for verification.

Better Automatic Evaluation Metrics

It is generally difficult to automatically evaluate generative reasoning implications, especially with realistic and not synthetic datasets. The difficulty mainly lies in that the same semantic meaning can be expressed with diversified forms, and that different conclusions might be all acceptable (especially in abductive and inductive reasoning).

Understanding the Internal Mechanism of PLMs for Reasoning

Until now research works only focus on investigating whether the input/output behaviors of PLMs can be used to simulate a reasoner or complete reasoning tasks. However, it is still a challenging open research question to understand the internal mechanism of PLMs for reasoning.

More Impacts on (NLP) Applications

These advantages of LRNL make it possible (but might still challenging) to deal with many (NLP) applications such as medical diagnosis and legal NLP tasks, since many medical and legal problems could be seen as pure logical reasoning problems with very large rule base (e.g., medical knowledge and laws).

Probabilistic Inference

In reality, pure deductive reasoning has not always been used. When people include "likely" in their expressions, uncertainty is introduced, which makes the reasoning process probabilistic; In addition, inductive reasoning and abductive reasoning are by default non-monotonic reasoning. This uncertainty aspect has not been focused in current research.

Reasoning with Incomplete Information

The current proof generation task requires all necessary premises provided to create a proof tree. Only one work [13] focuses on proof generation with the incomplete information task. However, the task they adopt only overlooks one premise, while in reality more might be missing.

Inductive Reasoning on Web Corpora

Currently, the dataset for rule generation tasks in inductive reasoning provides manually selected facts [17]. However, to best leverage a system's ability in handling natural language, it should be able to work on raw web corpora to induce rules, which leads to a more challenging task of inductive reasoning on web corpora.

Interactions between Reasoning Types

Multiple reasoning types can be used together for complex tasks. Existing works only utilize deductive reasoning with abductive reasoning to create a proof tree. However, many other collaborations are possible, such as using inductive reasoning to collect a (large) rule base, which is to be used as the theory base for deductive reasoning.

Contact Information

LRNL Bench: `https://github.com/ZonglinY/LRNL-bench` (to come)
Twitter: `https://twitter.com/Yang_zy223`
Email: zonglin.yang@ntu.edu.sg
Personal website: `https://zongliny.github.io`