

自然语言处理: Part 3

自然语言上的逻辑推理

- 数据集
- 证明生成方法

知识表示与推理回顾

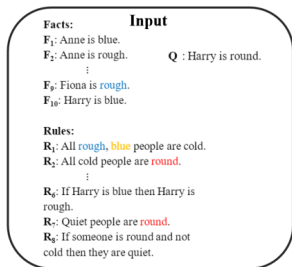
- 知识表示研究如何用符号语言表示人类的知识。
- 然而人类智能活动所需要知识通常不是显式表示的。
- 知识推理研究如何从显式表示的知识中计算得到隐式表示的知识。
- 前面我们学习了如何用谓词逻辑来表示人类知识，
- 如何用归结从显式表示的知识中计算得到隐式表示的知识。

自然语言上的逻辑推理

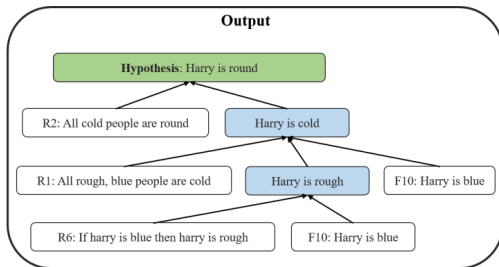
- 然而人类知识通常是用自然语言表示的
- 虽然语义解析，即如何把自然语言自动转换为形式化语言，近年来取得了进展，但仍然是一个困难的任务
- 因而希望直接在自然语言上进行逻辑推理
- 与深度学习的“黑盒”计算相比，自然语言上的逻辑推理所产生的用自然语言表示的推理过程可以为机器的决策提供解释。

举例

推理所需要知识是显式给定的，输出是推理序列或证明树



(a) 自然语言知识库



(b) 证明树，树的每个节点为自然语言句子

RuleTaker数据集 (IJCAI-20)

- 由美国的艾伦人工智能研究所提出
- 自动生成，知识库由事实和规则组成，无证明树
- 数据集分为不同推理深度的五个部分
- 实验表明大规模语言模型在该数据集上表现良好

(Input Facts:) Alan is blue. Alan is rough. Alan is young.
Bob is big. Bob is round.
Charlie is big. Charlie is blue. Charlie is green.
Dave is green. Dave is rough.

(Input Rules:) Big people are rough.
If someone is young and round then they are kind.
If someone is round and big then they are blue.
All rough people are green.

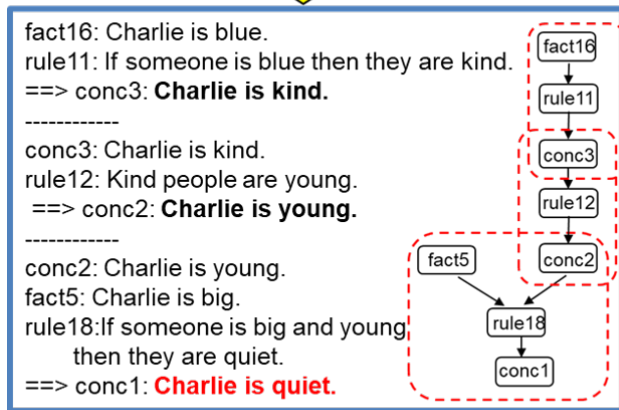
Q1: Bob is green. True/false? **[Answer: T]**

Q2: Bob is kind. True/false? **[F]**

Q3: Dave is blue. True/false? **[F]**

ProofWriter 数据集 (ACL Findings-21)

在RuleTaker 的基础上增加了自然语言表述的证明树



EntailmentBank 数据集 (EMNLP-21)

- 基于美国高中生科学课程考试题和相关知识，人工编写了回答考试题所需的证明树
- 数据集中自然语言句子存在模糊性和歧义性，证明树中含有复杂的推理规则和推理结构，这些为证明生成任务带来更大的难度

Question: How might eruptions affect plants?

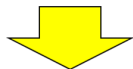
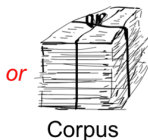
Answer: They can cause plants to die

Hypothesis

H (hypot): Eruptions can cause plants to die

Text

sent1: eruptions emit lava.
sent2: eruptions produce ash clouds.
sent3: plants have green leaves.
sent4: producers will die without sunlight
sent5: ash blocks sunlight.



Entailment Tree

H (hypot): Eruptions can cause plants to die

int1: Eruptions block sunlight.

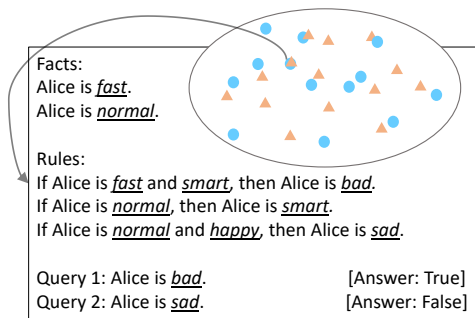
sent4: producers will die without sunlight.

sent2: eruptions produce ash clouds.

sent5: ash blocks sunlight.

SimpleLogic 数据集 (2022)

- 由UCLA提出
- 知识库由命题逻辑的事实和规则组成
- 生成不同分布的问答推理数据集
- 实验发现现有的预训练语言模型通过挖掘统计特征来完成推理任务，而非学习到正确的推理算法



FOLIO 数据集 (2022)

- 由耶鲁大学等单位提出的人工标注数据集
- 对自然语言语句进行了一阶逻辑标注
- 相比自动构建的数据集，该数据集中的句子有更加丰富的词汇、自然语言表述和逻辑组合，这些为语义解析和直接在自然语言上进行推理增加了难度。

A FOLIO example based on the Wild Turkey Wikipedia page: https://en.wikipedia.org/wiki/Wild_turkey

NL premises

1. There are six types of wild turkeys: Eastern wild turkey, Osceola wild turkey, Gould's wild turkey, Merriam's wild turkey, Rio Grande wild turkey, and the Ocellated wild turkey.
2. Tom is not an Eastern wild turkey.
3. Tom is not an Osceola wild turkey.
4. Tom is also not a Gould's wild turkey, or a Merriam's wild turkey, or a Rio Grande wild turkey.
5. Tom is a wild turkey.

NL Conclusions -> Labels

- A. Tom is an Ocellated wild turkey. -> True
- B. Tom is an Eastern wild turkey. -> False
- C. Joey is a wild turkey. -> Unknown

FOL Premises

1. $\forall x(\text{WildTurkey}(x) \rightarrow (\text{Eastern}(x) \vee \text{Osceola}(x) \vee \text{Goulds}(x) \vee \text{Merriams}(x) \vee \text{Riogrande}(x) \vee \text{Ocellated}(x)))$
2. $\neg(\text{WildTurkey}(\text{tom}) \wedge \text{Eastern}(\text{tom}))$
3. $\neg(\text{WildTurkey}(\text{tom}) \wedge \text{Osceola}(\text{tom}))$
4. $\text{WildTurkey}(\text{tom}) \rightarrow \neg(\text{Goulds}(\text{tom}) \vee \text{Merriams}(\text{tom}) \vee \text{Riogrande}(\text{tom}))$
5. $\text{WildTurkey}(\text{tom})$

FOL conclusions -> Labels

- A. $\text{Ocellated}(\text{tom}) \rightarrow \text{True}$
- B. $\text{Eastern}(\text{tom}) \rightarrow \text{False}$
- C. $\text{WildTurkey}(\text{joey}) \rightarrow \text{Unknown}$

ProntoQA (ICLR-2023)

- 由纽约大学提出的自动生成数据集
- 要求语言模型根据提示生成证明，以判定模型是否真正学习到推理。
- 实验表明现有的语言模型在少样本设定中能正确完成单步推断，但不擅长于证明规划

Q: Each cat is a carnivore. Every carnivore is not herbivorous. Carnivores are mammals. All mammals are warm-blooded. Mammals are vertebrates. Every vertebrate is an animal. Animals are multicellular. Fae is a cat. True or false: Fae is not herbivorous. — context
A: Fae is a cat. Cats are carnivores. Fae is a carnivore. Every carnivore is not herbivorous. Fae is not herbivorous. — query
True — chain-of-thought
— label

carnivore:食肉的, herbivorous:食草的, vertebrate: 脊椎动物

证明生成方法概述

- 早期研究一次性调用语言模型生成整个证明树。
- 后续研究提出迭代调用语言模型来生成证明树，使模型更容易学习并泛化到长证明。
- 基本思想是：使用语言模型来生成单步推理，并采用符号搜索算法把单步推理串联成证明树。
- 然而现有方法在较困难的数据集上的性能仍然较差。

一次性生成方法举例

- 将知识库句子和猜想逐句拼接为自然语言段落，作为生成式预训练模型的输入，
- 然后要求预训练的自然语言生成模型一次性输出线性化的证明描述语段，
- 最后根据语段构建证明树。

Oyvind Tafjord, Bhavana Dalvi, Peter Clark. ProofWriter: Generating Implications, Proofs, and Abductive Statements over Natural Language. In Findings of the Association for Computational Linguistics (ACL/IJCNLP Findings 2021)

一次性生成方法的缺点

- 忽略了证明树内部事实、规则和中间结论之间的因果关系，
- 通过挖掘数据特征来学习证明生成，
- 并非真正学习到证明生成，
- 导致模型缺乏可信度和泛化性。

迭代生成方法举例

- 使用前向链（forward-chaining）方法
- 系统的主干，表示为SI，包括两个微调的语言模型，一个用于选择，一个用于推理。
- 系统还有另外两个微调的语言模型
- 一个用来终止推理过程，并返回所需格式的答案。
- 另一个是评价当前推理步骤的质量的价值函数，用于推理迹上的束搜索

Antonia Creswell, Murray Shanahan: Faithful Reasoning Using Large Language Models. CoRR abs/2208.14271 (2022)

选择-推理

- 给定一个问题和一个由若干陈述组成的上下文，模型产生回答问题的一系列演绎推理步骤。
- SI主干将每个推理步骤分成两部分。
- 首先，给定问题，从上下文中选择一组语句。
- 然后，计算选择的语句的一个逻辑推论，并将其添加到上下文中。
- 对SI进行多次迭代以产生推理轨迹。
- 最终推理被用来回答问题。