



蘇州大學

SOOCHOW UNIVERSITY

常识与事件

養天地正氣
法古今完人

楊永清題

张国梁
2021.11.24

- 常识本身
- 常识怎么来?
- 如何判断系统有常识推断能力?

没有复杂模型，放心食用

常识是什么： 模糊的、抽象的、广泛的

大家也知道，但是也说不明白常识到底是什么。

因为这个问题的答案就是一个常识。

常识的重要性：

常识很重要已经成为共识，

那么逻辑上来说到底为什么重要？

“我能跳起来摸着飞机”

-> 我能跳多高；飞机能飞多高

-> 我摸不着

“我今天中午想吃恐龙”

-> “今天泛指21世纪头20年”； 恐龙已于6500万年前全部灭绝

-> 我吃不到

人为什么可以一瞬间就能判断以上两个问题，依赖的是：

1. 概念化 (conceptualization)：人类心理上对于事物和类别的抽象表征。
2. 组合性 (compositionality)：人类具备概括能力的关键，它指的是从已认知的成分中理解和产生新的组合的能力。

比如：“白天鹅”与“黑天鹅”。

你知道“白”意味着什么（知道了“颜色”这一概念），也知道“天鹅”意味着什么（知道了“动物”这一概念，以及“天鹅”这个子群归属于“动物”之后）。

因此即使曾经从未见过“黑天鹅”这种组合，也应该能够把这二者联系起来。即你也知道黑天鹅是什么，能大概想像出样子。

因此对于“概念化”和“组合性”，人是具有两种能力的：

1. 概念解耦 (concept disentanglement)
2. 组成推理 (compositional inference)

这两种能力让人拥有了推理能力。

但由于组合空间是巨大的，如何保证组合推理的可靠性就显得至关重要了，而这就直接依赖于“常识”。

如果想让计算机也拥有更强的推理能力，那么必须依靠丰富的常识知识。

使用常识知识嵌入到各个下游任务中，比如指代消解、问答、事件预测等，都能提高效果。

然而据估计，典型的成年人所知道的常识性知识公理多达100,000,000条。在 NLI (natural language inference) 应用中，常识知识的缺乏正是目前的一大瓶颈。

- 常识本身
- 常识怎么来?
- 如何判断系统有常识推断能力?

常识怎么来？



1. 爬。

典型的常识叙述：

“人能呼吸”；“狗有四条腿”；“用电脑需要电”；“吃饭了就不饿”

常识描述其实就是废话！

因此，很少这样的文本能拿来爬（新华字典还行）。

2. 标。

优点：噪音数据少，可用性强

缺点：慢、累、少。

那么有没有什么好的办法来获取常识？

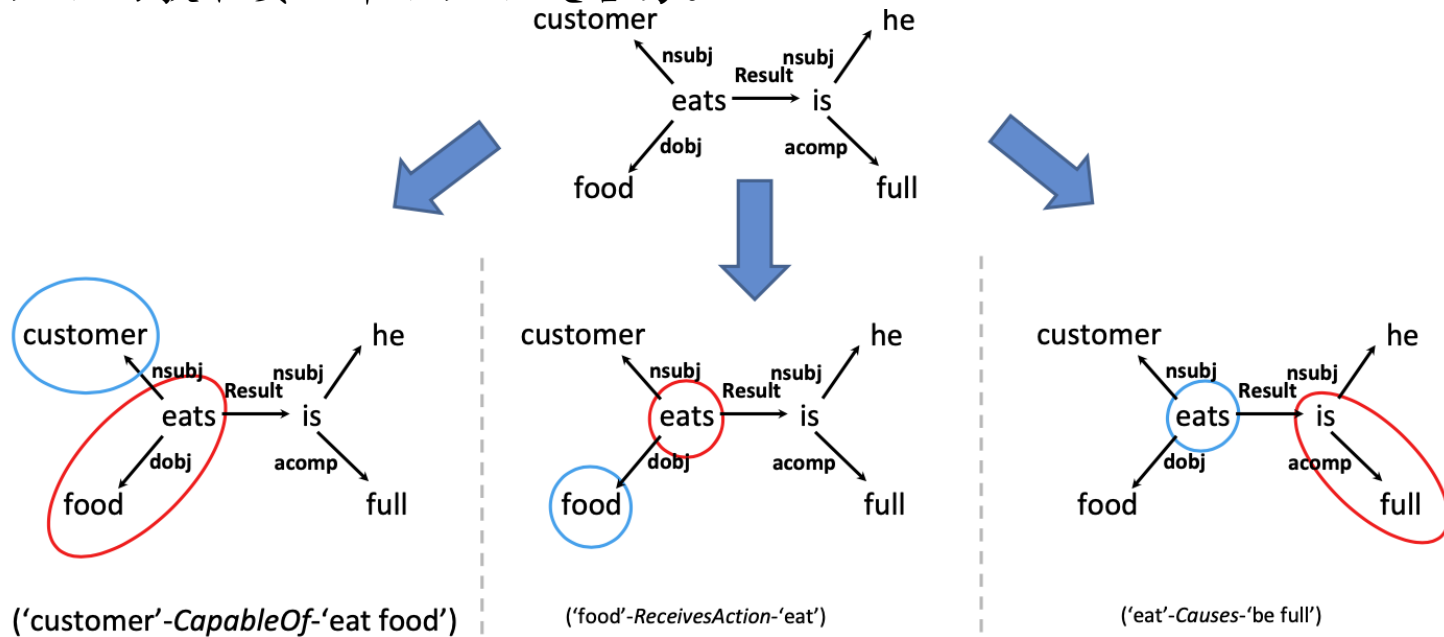
常识怎么来？

通过事件来获取常识知识

为什么可以这么做：

- 事件叙述本身是包含很多隐式常识的。
- 事件是可以转化为常识知识的。
- 事件知识的获取要比常识知识更容易。

“我中午吃了5碗鸡汤面。”



Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. IJCAI 2020.

常识怎么来?



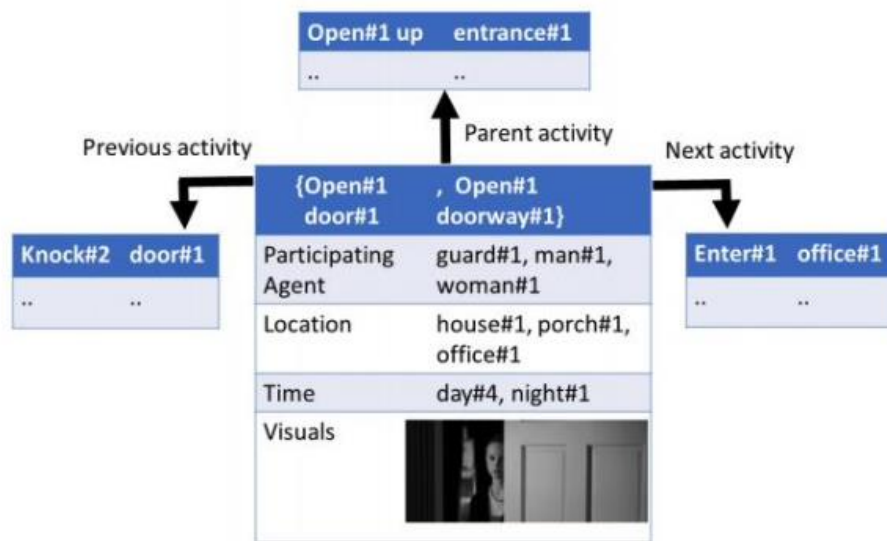
从Event Knowledge转化而来

Knowlywood (Tandon et al., 2015)

什么类型的事件更能具有常识信息: 560 个影视剧脚本
构建方法:

Node: Verb + Object

Edge: Temporal Relation



“Knock door”->“open up entrance”->“enter office”

Scripts

- Movies
- TV Series
- Sitcoms
- Novels



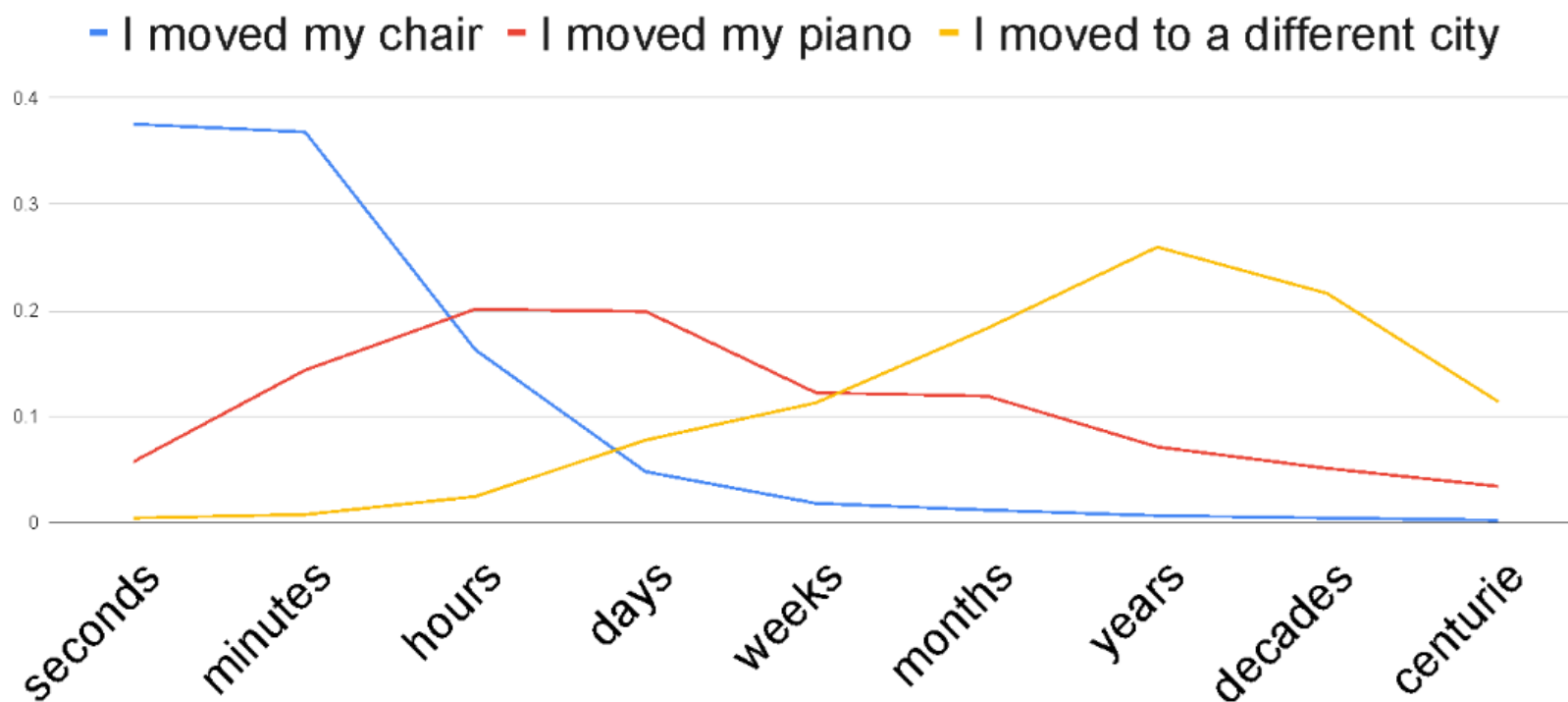
常识怎么来?



从预训练模型中来

TacoLM (Zhou et al., 2020)

事件时间常识

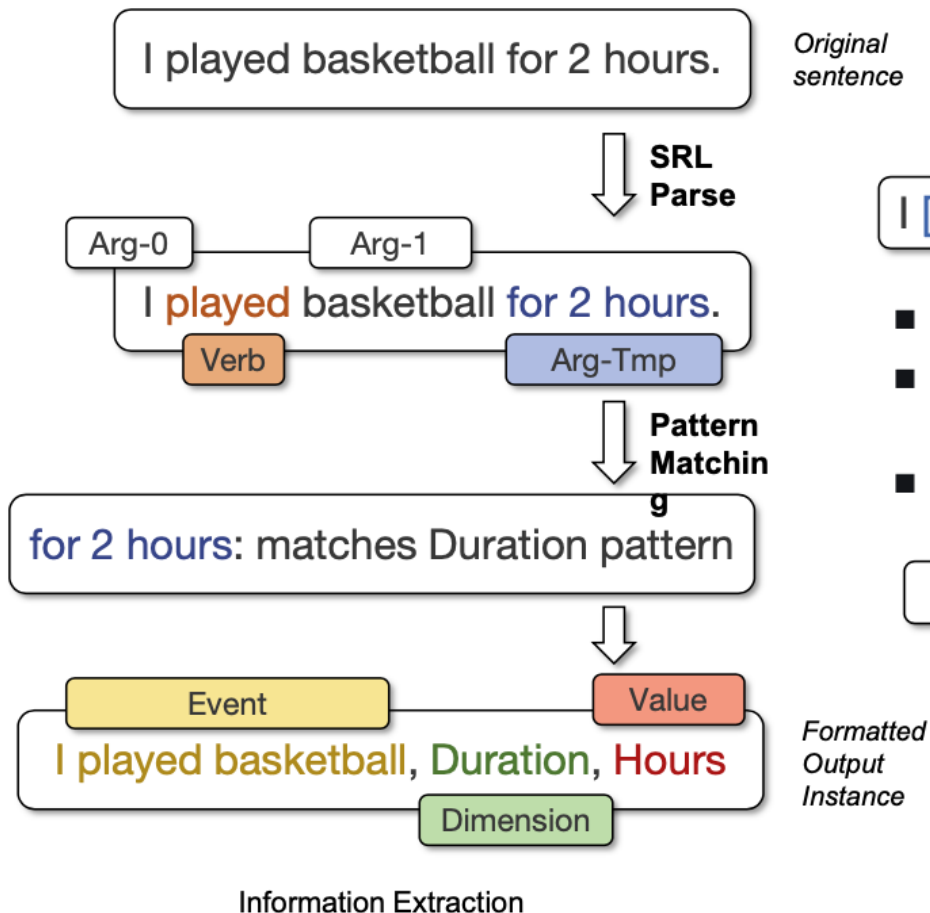


Ben Zhou, Qiang Ning, Daniel Khashabi,, and Dan Roth. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020.

常识怎么来?



从预训练模型中来



I [M] played basketball [SEP] [M] [DUR] [HRS]

- Baseline Model: Pre-trained BERT-base
- Main objective: mask some tokens and recover them
- How we mask:
 - With some probability, mask temporal

I [M] played basketball [SEP] [M] [DUR] [MASK]

- 常识本身
- 常识怎么来?
- 如何判断系统有常识推断能力?

PIQA: Reasoning about Physical Commonsense in Natural Language

Q1: 在不使用刷子涂眼影的情况下，我应该用棉签还是牙签？

Q2: 要使用矿泉水瓶将蛋黄和蛋清分离你应该怎么做？

- a. 对着蛋黄挤压矿泉水瓶，然后再松开，就能产生吸力吸走蛋黄。
- b. 将矿泉水瓶放在蛋黄上，不断推动，就能够产生吸力吸走蛋黄。



To separate egg whites from the yolk using a water bottle, you should...

a. **Squeeze** the water bottle and press it against the yolk. **Release**, which creates suction and lifts the yolk.

b. **Place** the water bottle and press it against the yolk. **Keep pushing**, which creates suction and lifts the yolk.





PIQA: <https://yonatanbisk.com/piqa/>

数据来源:

一个指导使用日常材料、烘焙、修车、烹饪等的网站。

通过改变一些关键词使之生成对应的错误选项。

数据集由 16,000 多个训练的 QA 对组成，正确和不正确解决方案所使用的单词之间至少有 85% 的重叠。

Rank	Model	Accuracy
	Human Performance (Bisk et al. '20)	94.9
1	DeBERTa-xxlarge Alibaba Group ICBU Tech	83.5
2	GPT-3 OpenAI	82.8*
3	Anonymous Anonymous	79.0
4	RoBERTa-Large Baseline lr 1e-5, 8 ep, 4/batch/GPU (4x V100), max seq 150	77.1
5	Zero-shot GPT-XL self-talk with GPT-medium AI2	69.5
6	OpenAI GPT Baseline	69.2
7	Zero-shot GPT-XL with COMET AI2	68.4
8	BERT-Large Baseline lr 1e-5, 8 ep, 6/batch/GPU (4x V100), max seq 150	66.8

Token-level:

Sense Making (SM): 测试是否能区分有意义的和无意义的陈述

- I work 8 hours a day .
- I work 25 hours a day.

Conjunction Acceptability (CA) : 测试是否有逻辑常识性 (通过连词)

They broadcast an announcement, but a subway came into the station and I couldn't hear it.

They broadcast an announcement, before a subway came into the station and I couldn't hear it .

Winograd Schema Challenge (WSC) : 测试代词解析能力

Paul tried to call George on the phone, but he wasn't successful.

Who is he? Candidate: A. Paul B. George

Sentence-level:

Sense Making with Reasoning (SMR) : 测试是否具有原因推理能力

Can is usually made of gold .

- A. gold is too bright to make cans,*
- B. gold is too soft to make cans*
- C. gold is too expensive to make cans*

SWAG : 测试是否具有事件推理能力

On stage, a woman takes a seat at the piano.

- A. sits on a bench as her sister plays with the doll*
- B. smiles with someone as the music plays*
- C. is in the crowd, watching the dancers*
- D. nervously sets her fingers on the keys.*

- 常识本身
- 常识怎么来?
- 如何判断系统有常识推断能力?
- 如何应用常识
- 如何提高系统的常识推断能力



END
THANKS !