💫

Do As I Can, Not As I Say: Grounding Language in Robotic Affordance——hoho

论文试图解决什么问题？

语言模型缺少真实环境的经验交互，使得它难以利用真实环境的数据进行决策。

（a significant weakness of language models is that they lack real-world experience, which makes it difficult to leverage them for decision making within a given embodiment）

另一方面，环境中的智能体如何抽取并利用语言模型的知识以指引物理操作？

（how can embodied agents extract and harness the knowledge of LLMs for pyhsically tasks?）

这是否是一个新的问题？

不是新问题。如在同一时间，也有其他学者如Huang，利用prompting engineering来抽取时序上的规划。

这篇文章要验证一个什么科学假设？

是否可以让LLM输出给机器人每个可执行动作的概率分布（possible），然后用另一个模型（affordance function）输出这些动作中成功完成任务的概率分布(useful)，如此一来，组合这两个概率分布，即是完成通过文本指令让机器人完成一个动作。

（the LLM describes the probability that each skill contributes to completing the instruction, and the affordance function describes the probability that each skill will succeed-combining the two provide the probability that each skill will perform the instruction succesfully）

有哪些相关研究？如何归类？谁是这一课题在领域内值得关注的研究员？

hoho_todo

论文中提到的解决方案之关键是什么？

核心思想：将LLM的输出通过价值函数落地到真实环境中执行任务

（The key idea of SayCan is to ground large language models through value functions——affordance functions that capture the log likelihood that a particular skiil will be able to succeed in the current state）

假设人类提供一条自然语言的指令i，如：How would you put an apple on the table？

机器人有动作空间$\prod$，每一条动作$\pi \in \prod$，

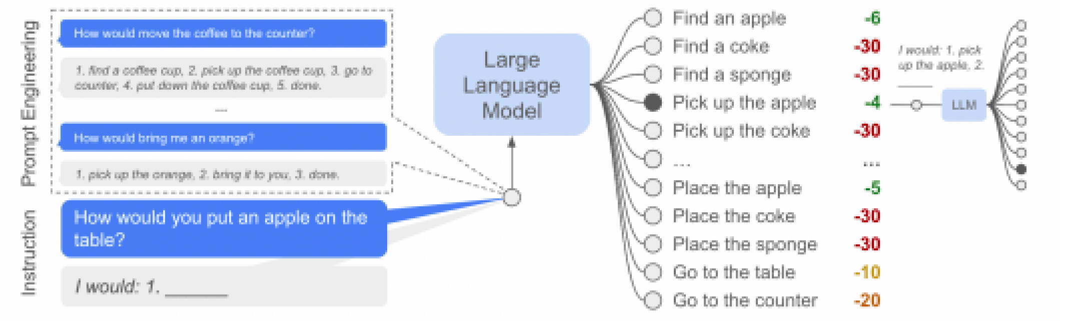
每个动作都有对应的自然语言描述 l_π $\in$ {Find an apple, Find a coke, Pick up the apple, Place the apple,…}

当前环境的状态s：机器人所见的图像

问题可以建模为：

1. LLM提供$p(l_\pi|i)$

(Say process。the probability that a skill’s textuatl label is a valid next step for the user’s instruction)



How would move the coffee to the counter?

1. find a coffee cup, 2. pick up the coffee cup, 3. go to counter, 4. put down the coffee cup, 5. done.

...

How would bring me an orange?

1. pick up the orange, 2. bring it to you, 3. done.

How would you put an apple on the table?

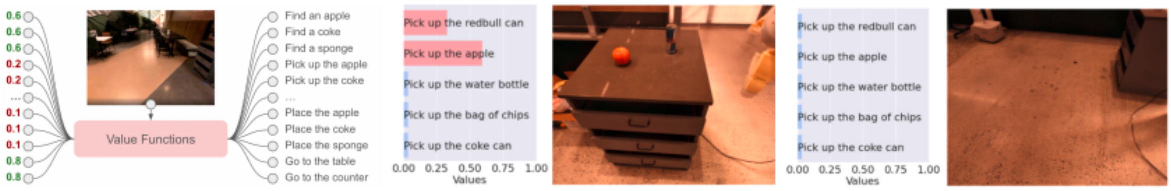
I would: 1. ______

I would: 1. pick up the apple, 2.

2. Affordance function提供$p(c_\pi|s,l_\pi)$

（Can process。the probability that skill π with textual label l_π successfully complete if executed from state s）

其中c_π为伯努利随机变量（Bernoulli random variable）$\in[0,1]$，表示完成的成功概率.



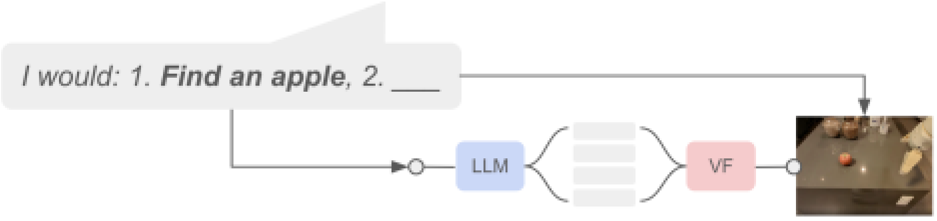
(a) (b) (c)

3. 容易知道整个任务完成的概率为$p(c_\pi|i,s,l_\pi)\propto p(l_\pi|i)p(c_\pi|l_\pi,s)$，于是选取的动作策略为

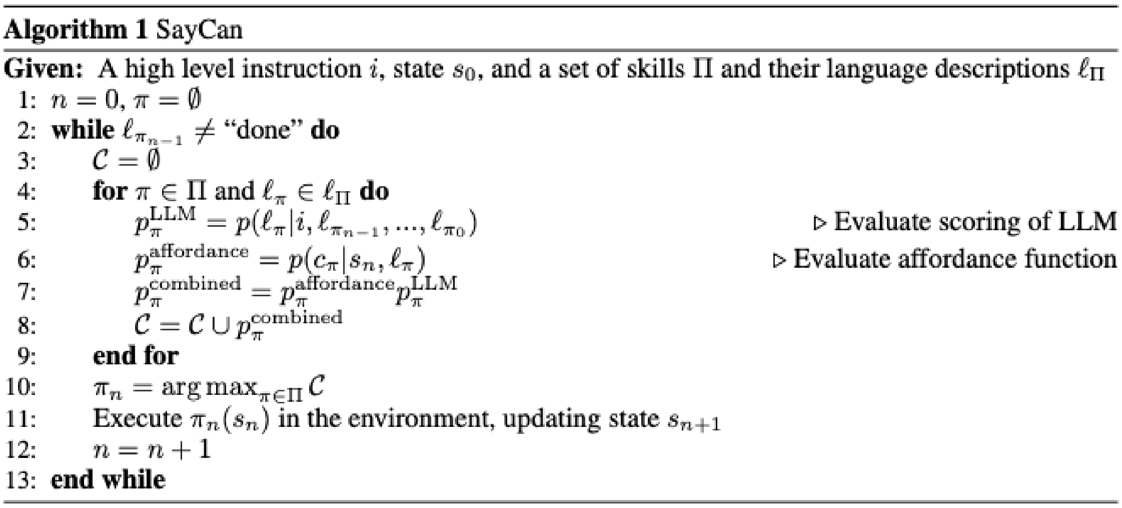
$$\pi=\arg\max_{\pi\in\prod}p(l_\pi|i)p(c_\pi|l_\pi,s)$$

(For each skill, the affordance function and the LLM probability are then multiplied together and ultimately the most probable skill is selected)

4. the next step，LLM会利用马尔可夫过程计算其输出概率$p_n^{LLM}=p(l_{\pi_n}|i,l_{\pi_{n-1}},l_{\pi_{n-2}}...l_{\pi_0})$



总的算法流程如下：



```
Algorithm 1 SayCan
Given: A high level instruction $i$, state $s_0$, and a set of skills $\Pi$ and their language descriptions $\ell_\Pi$
$n = 0, \pi = \emptyset$
while $\ell_{\pi_{n-1}} \neq$ “done” do
    $\mathcal{C} = \emptyset$
    for $\pi \in \Pi$ and $\ell_\pi \in \ell_\Pi$ do
        $p_\pi^{\text{LLM}} = p(\ell_\pi | i, \ell_{\pi_{n-1}}, ..., \ell_{\pi_0})$        ▷ Evaluate scoring of LLM
        $p_\pi^{\text{affordance}} = p(c_\pi | s_n, \ell_\pi)$        ▷ Evaluate affordance function
        $p_\pi^{\text{combined}} = p_\pi^{\text{affordance}} p_\pi^{\text{LLM}}$
        $\mathcal{C} = \mathcal{C} \cup p_\pi^{\text{combined}}$
    end for
    $\pi_n = \arg\max_{\pi \in \Pi} \mathcal{C}$
    Execute $\pi_n(s_n)$ in the environment, updating state $s_{n+1}$
    $n = n + 1$
end while
```

- 关于网络模型的架构

在机器人方面，需要训练其完成各种动作，本文对比了两种模型架构

首先是基于强化学习的MT-Opt（Mt-opt: Continuous multi-task robotic reinforcement learning at scale）：

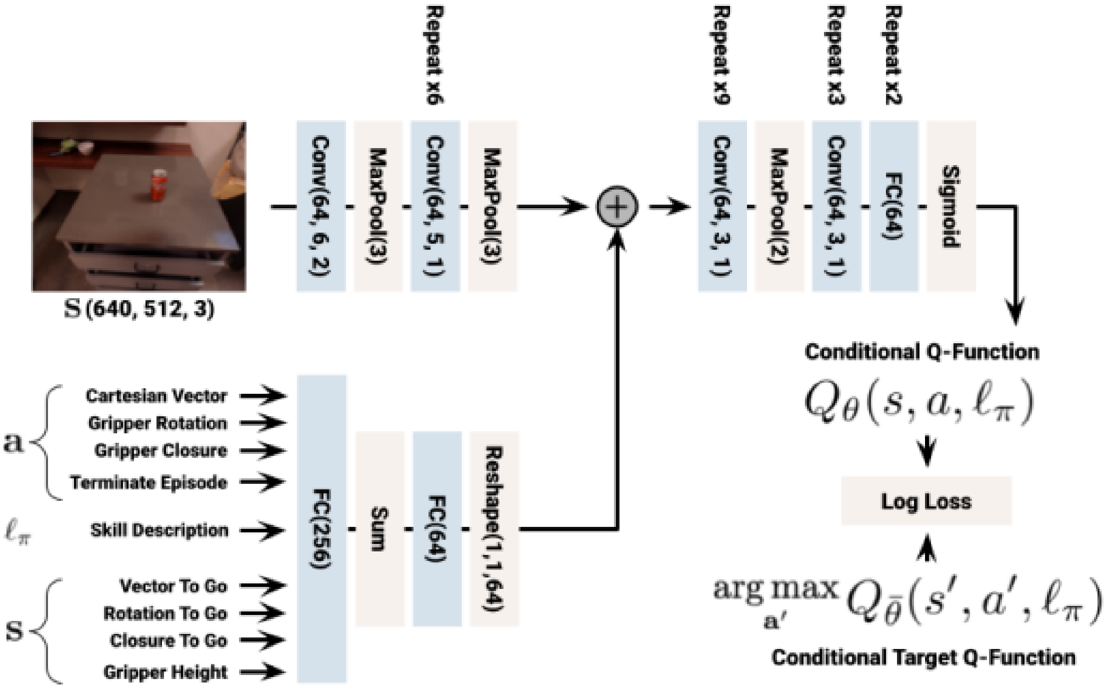


Figure 9: Network architecture in RL policy

另外是基于模仿学习的BC-Z：

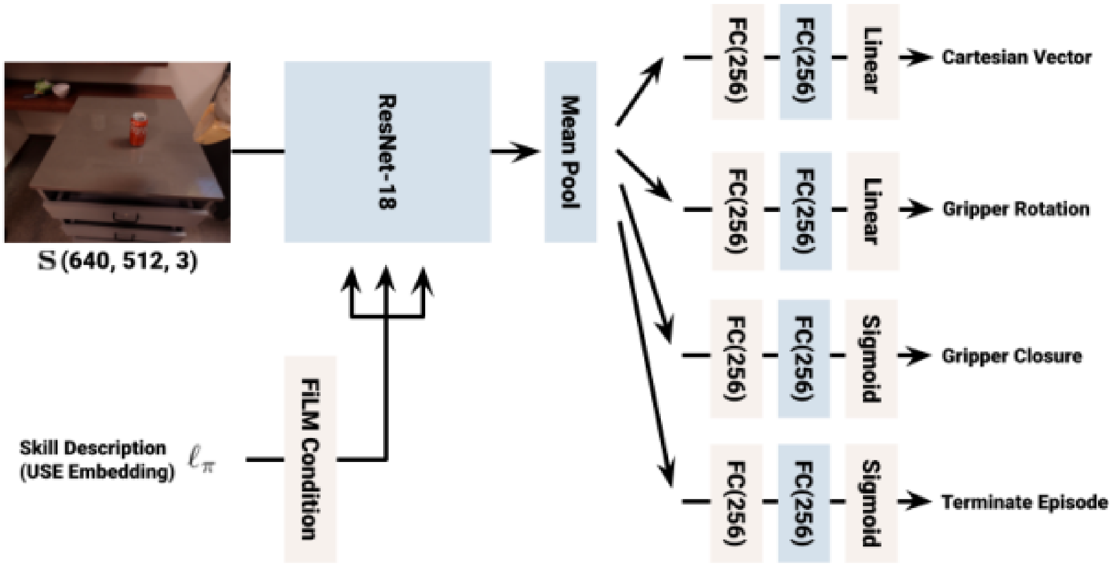


Figure 10: Network architecture in BC policy

作者认为BC-Z（Bc-z: Zero-shot task generalization with robotic imitation learning.）的性能表现更优

对于奖励函数，本文使用了简单稀疏奖励函数：当机器人根据指令描述最后完成动作，则奖励为1，否则为0.

（We utilize sparse reward functions with reward values of 1.0 at the end of an episode if the lan- guage command was executed successfully, and 0.0 otherwise）

对于语言模型使用的PaLM（Palm: Scaling language modeling with path- ways）

论文中的实验是如何设计的？

- 本文分别在真实环境和仿真环境中进行实验。

（We test our method in two environments: a real office kitchen and a mock environment mirroring the kitchen）

- 衡量标准两个：

规划成功率（plan success rate）：which measures whether the skills selected by the model are correct for the instruction, regardless of whether or not they actually successfully executed

执行成功率（execution success rate）：which measures whether the full PaLM-SayCan sys- tem actually performs the desired instruction successfully.

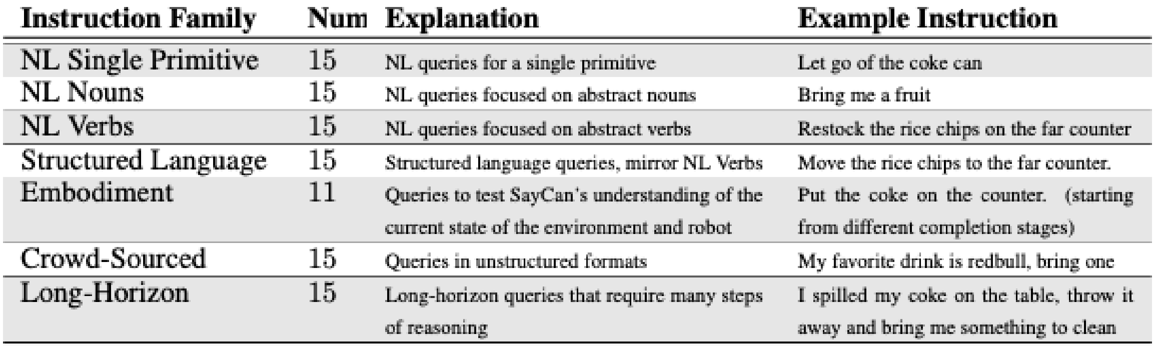
采用人工方式进行多数投票决定（majority voting）：We ask 3 human raters to indicate whether the plan/execution generated by the model can achieve the instruction, and if 2 out of 3 raters agree that the plan is valid, it is marked a success

用于定量评估的数据集是什么？代码有没有开源？

- 用于定量评估的数据集在实验环境中在线生成，采用人工方式进行评估：https://github.com/say-can/say-can.github.io/tree/main/data
- 代码开源：https://github.com/google-research/google-research/tree/master/saycan

论文中的实验及结果有没有很好地支持需要验证的科学假设？

本文将指令分为一个个族:

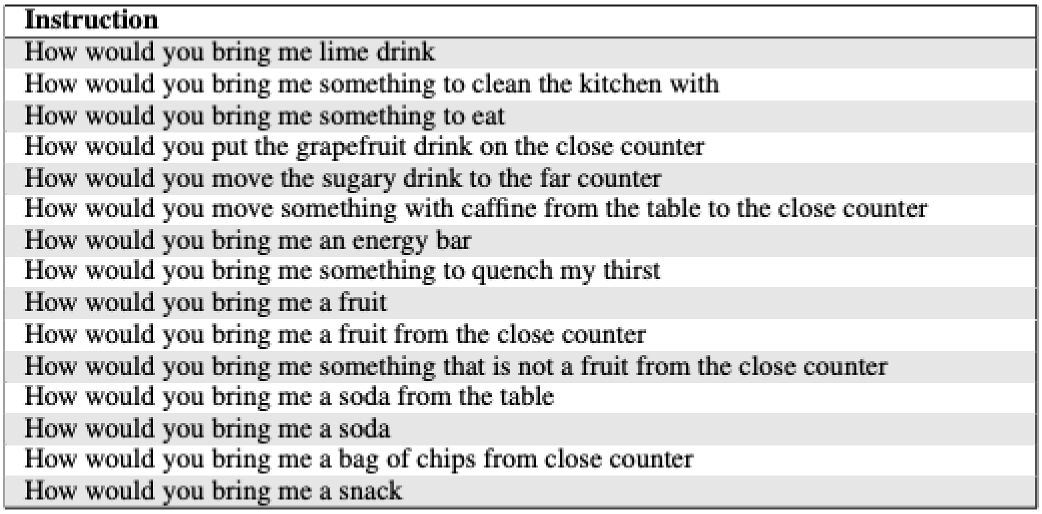


Instruction Family	Num	Explanation	Example Instruction
NL Single Primitive	15	NL queries for a single primitive	Let go of the coke can
NL Nouns	15	NL queries focused on abstract nouns	Bring me a fruit
NL Verbs	15	NL queries focused on abstract verbs	Restock the rice chips on the far counter
Structured Language	15	Structured language queries, mirror NL Verbs	Move the rice chips to the far counter.
Embodiment	11	Queries to test SayCan's understanding of the current state of the environment and robot	Put the coke on the counter. (starting from different completion stages)
Crowd-Sourced	15	Queries in unstructured formats	My favorite drink is redbull, bring one
Long-Horizon	15	Long-horizon queries that require many steps of reasoning	I spilled my coke on the table, throw it away and bring me something to clean

Table 1: **List of instruction family definitions:** We evaluate the algorithm on 101 instructions. We group the instructions into different families, with each family focusing on testing one aspect of the proposed method.

譬如NL Nouns即将名词进行同义词替换

（Given a natural language query that replaces a noun (typically an object or location) with a synonym）

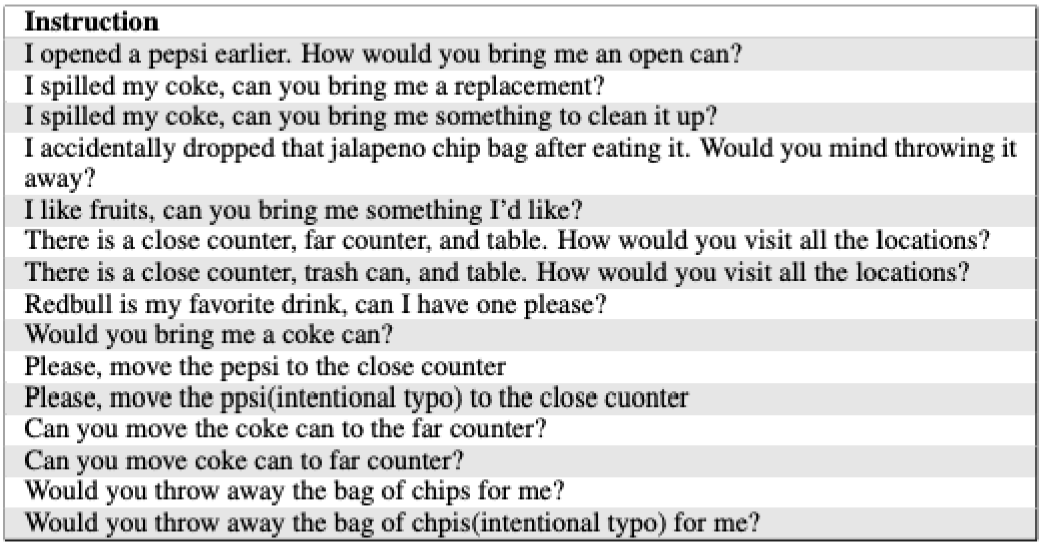


Instruction
How would you bring me lime drink
How would you bring me something to clean the kitchen with
How would you bring me something to eat
How would you put the grapefruit drink on the close counter
How would you move the sugary drink to the far counter
How would you move something with caffine from the table to the close counter
How would you bring me an energy bar
How would you bring me something to quench my thirst
How would you bring me a fruit
How would you bring me a fruit from the close counter
How would you bring me something that is not a fruit from the close counter
How would you bring me a soda from the table
How would you bring me a soda
How would you bring me a bag of chips from close counter
How would you bring me a snack

(c) NL Nouns

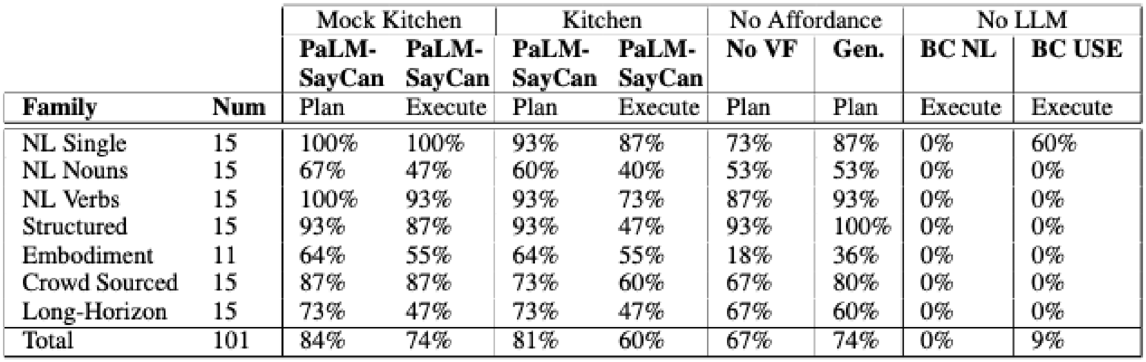
Crowd-Souce则有更加多的上下文语境

（giving humans a description of what occurred，and asking them what they would ask the robot to do）



Instruction
I opened a pepsi earlier. How would you bring me an open can?
I spilled my coke, can you bring me a replacement?
I spilled my coke, can you bring me something to clean it up?
I accidentally dropped that jalapeno chip bag after eating it. Would you mind throwing it away?
I like fruits, can you bring me something I'd like?
There is a close counter, far counter, and table. How would you visit all the locations?
There is a close counter, trash can, and table. How would you visit all the locations?
Redbull is my favorite drink, can I have one please?
Would you bring me a coke can?
Please, move the pepsi to the close counter
Please, move the ppsi(intentional typo) to the close cuonter
Can you move the coke can to the far counter?
Can you move coke can to far counter?
Would you throw away the bag of chips for me?
Would you throw away the bag of chpis(intentional typo) for me?

(f) Crowd-Sourced



		Mock Kitchen		Kitchen		No Affordance		No LLM	
		PaLM-SayCan	PaLM-SayCan	PaLM-SayCan	PaLM-SayCan	No VF	Gen.	BC NL	BC USE
Family	Num	Plan	Execute	Plan	Execute	Plan	Plan	Execute	Execute
NL Single	15	100%	100%	93%	87%	73%	87%	0%	60%
NL Nouns	15	67%	47%	60%	40%	53%	53%	0%	0%
NL Verbs	15	100%	93%	93%	73%	87%	93%	0%	0%
Structured	15	93%	87%	93%	47%	93%	100%	0%	0%
Embodiment	11	64%	55%	64%	55%	18%	36%	0%	0%
Crowd Sourced	15	87%	87%	73%	60%	67%	80%	0%	0%
Long-Horizon	15	73%	47%	73%	47%	67%	60%	0%	0%
Total	101	84%	74%	81%	60%	67%	74%	0%	9%

然后作者进行了一些消融实验：

- No VF: 不使用Value function，而只是最大化LLM生成概率模型$\pi = \arg\max_{\pi\in\prod} p(l_\pi|i)$

(remove the value functdion and choosing the maximum language score skill)

- Gen: 用生成式的LLM生成动作序列

(which uses the generative output of the LLM and then projects each planned skill to its maximal cosine similarity skill via USE embeddings)

- BC NL: In BC NL we feed the full instruction i into the policy (策略网络) – this approach is representative of standard RL or BC-based instruction following methods
- BC USE: we project the high-level instruction into the set of known language commands via the Universal Sentence Encoder (USE) embeddings [15] by embedding the instruction, all the tasks, and the combinatorial set of sequences tasks (i.e., we consider “pick coke can” as well as “1. find coke can, 2. pick coke can” and so on), and selecting the highest cosine similarity instruction.

这篇论文到底有什么贡献？

借助价值函数，可以有效的将语言模型的知识融入到决策过程当中，比仅使用LLM或仅使用强化学习性能更加好。

下一步呢？有什么工作可以继续深入？

目前对于整个动作序列，没有考虑中间步的动作是否成功（没用中间的奖励信号），仅仅只有在完成整个动作序列才收到奖励。

如

human: How would you put an apple on the table?

robot: 1). Go to the counter. 2). Find the apple, 3) Pick up the apple, 4) Go to the table, 5) place the apple. 然后才收到奖励1

如在第3步pick up the apple但失败了，然后机器人依然看到苹果的画面，模型给pick up the apple打很高分，但由于由于序列太长，生成pick up the apple的似然度很低:

$$p_n^{LLM} = p(l_{\pi_n} =' pick \quad up \quad the \quad apple'|i, l_{\pi_{n-1}}, l_{\pi_{n-2}}...l_{\pi_0})$$

动作概率变得不确定，最终采样到不同的动作，导致任务的失败。