法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

第七节课：对话系统从原理到应用

- 具体领域案例
- 强化学习
- 工业界bots

本节内容

- 订餐机器人案例研究
- 对话系统的强化学习
- Api.ai为例看工业界chatbot

参考文献

- 订餐机器人案例研究
  - A network-based end-to-end trainable task-oriented dialogue system (2016)
- 对话系统的强化学习
  - Interactive reinforcement learning for task-oriented dialogue management (2016)
  - Deep Reinforcement Learning for Dialogue Generation (2016)
  - Deal or No Deal? End-to-end learning for negotiation dialogues (2017)
- 工业界的chatbot

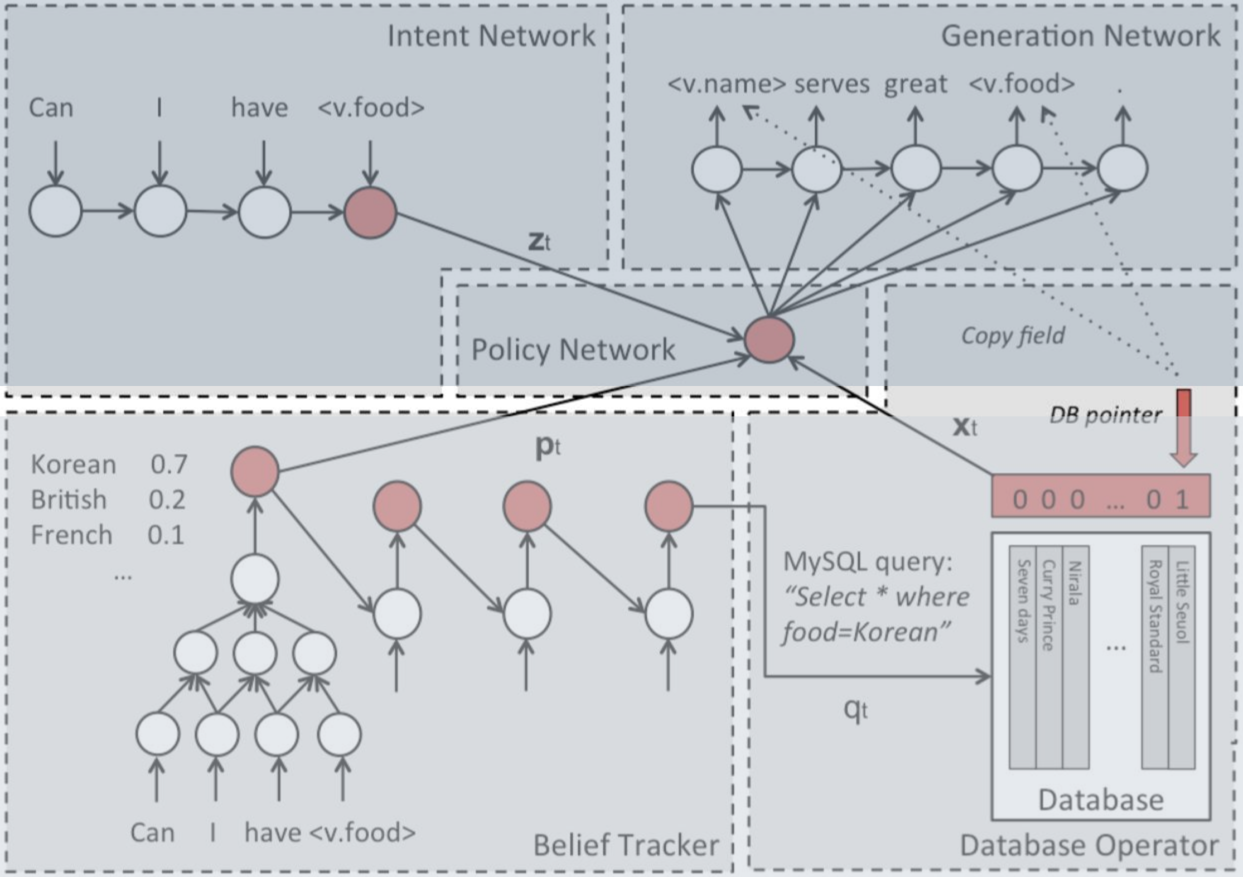
综合多个module的Ent-to-End模型

订餐机器人案例

- A network-based end-to-end trainable task-oriented dialogue system (2016)
  - 关键词：Ent-to-end, sequence-to-sequence mapping, policy, goal, Wizard-of-Oz framework, multi-module
  - 模型； 数据； 效果

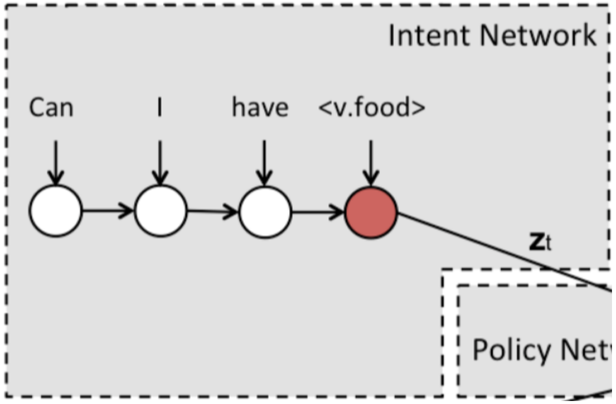
Seq2seq模型在封闭领域应用

类似于简单的SEQ2SEQ模型的结构



为SEQ2SEQ提供有用的信息并保持对话过程的一致性

一轮对话的流程



模型接受用户的输入序列(input sequence), 转化为两个内部表示

1. 由intent network生成的表示用户意图的分布式表达（**distributed representation**）

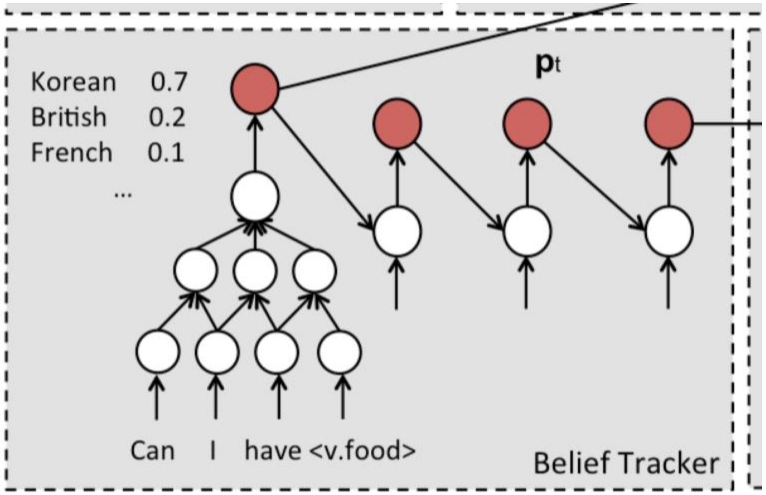
Intent network可以类比成seq2seq模型里面的encoder

一轮对话的流程

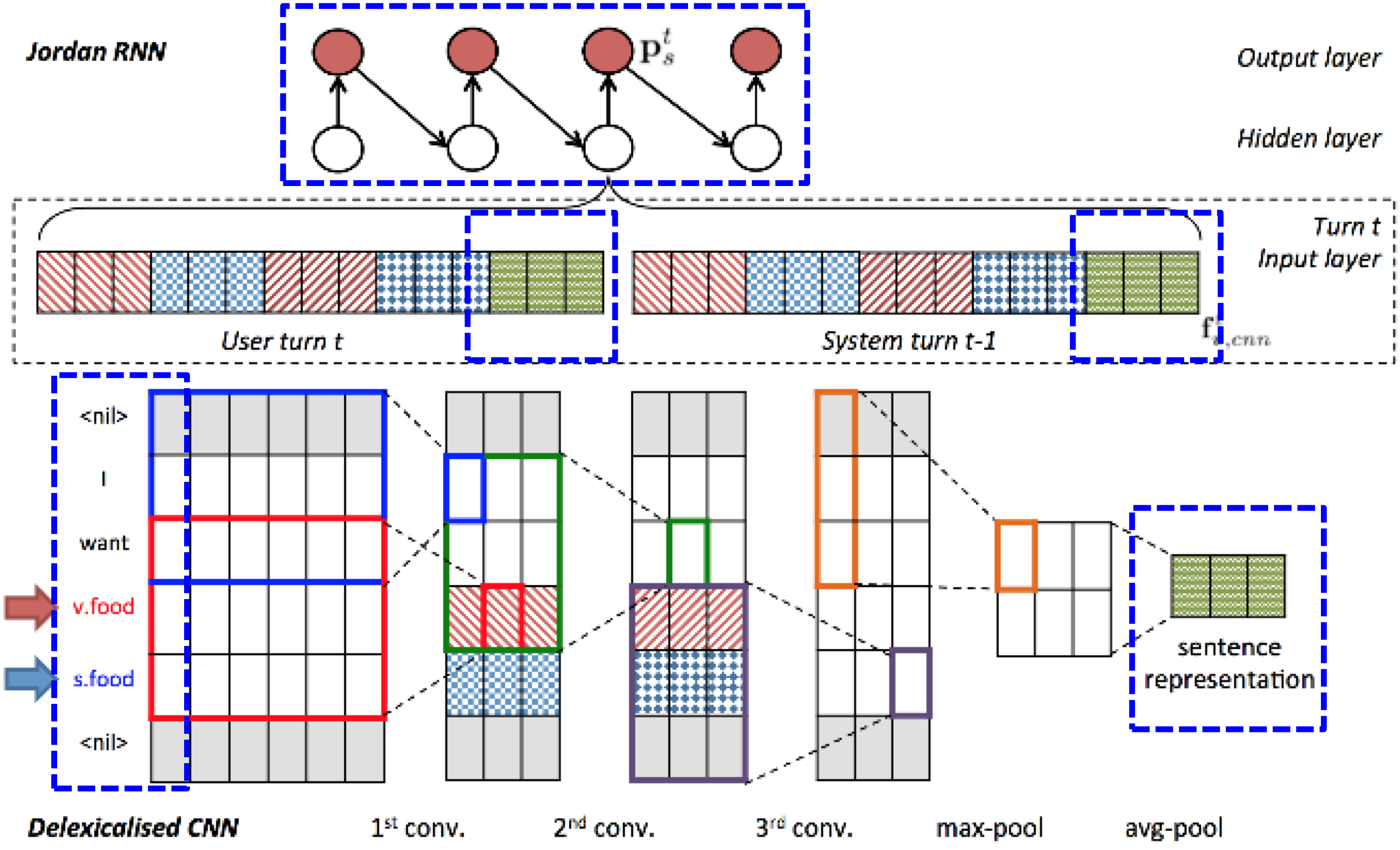
模型接受用户的输入序列(input sequence)，转化为两个内部表示

1. 由intent network生成的表示用户意图的分布式表达（distributed representation）

2. 由一组belief trackers生成一个slot-value pair的概率分布，叫做（belief state）



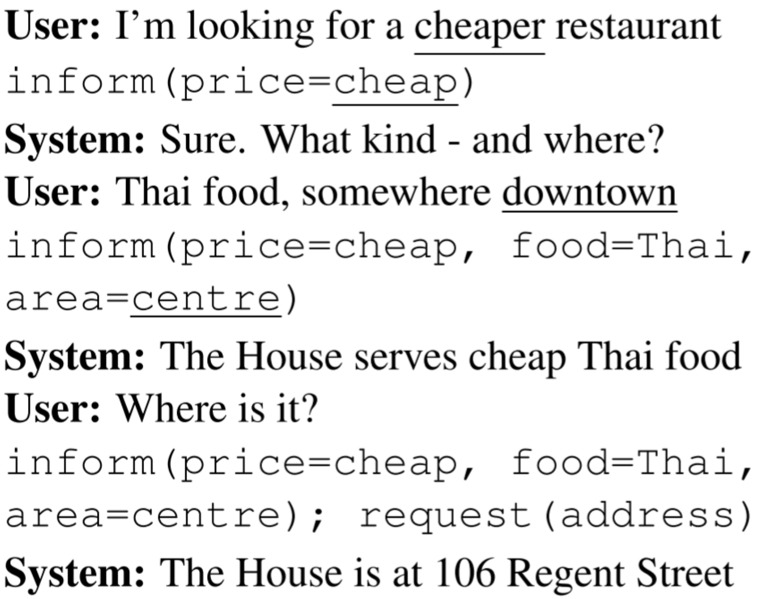
一轮对话的流程



Belief tracker（也叫Dialogue *State* tracker）的实现细节😱

对话系统中的belief tracker

- Belief tracker 是对话系统的一个核心部分，用来记录对话过程中用户的目标信息
- (One of the core components of modern spoken dialogue systems is the belief tracker, which estimates the user’s goal at every step of the dialogue.——neural belief tracker: data driven dialogue state tracking 2016)



User: I’m looking for a cheaper restaurant

`inform(price=cheap)`

System: Sure. What kind - and where?

User: Thai food, somewhere downtown

`inform(price=cheap, food=Thai, area=centre)`

System: The House serves cheap Thai food

User: Where is it?

`inform(price=cheap, food=Thai, area=centre); request(address)`

System: The House is at 106 Regent Street

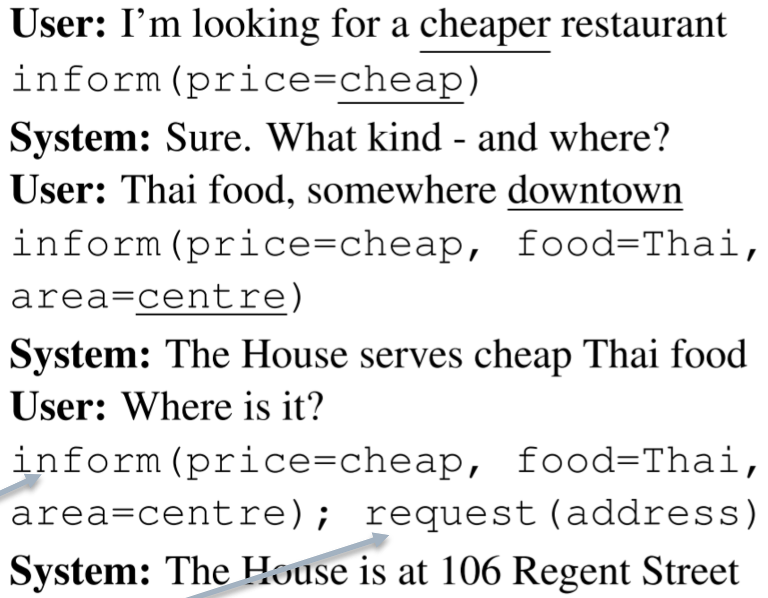
对话系统中的belief tracker

□ 对话系统有一个domain ontology，表示系统能够处理的用户意图的集合

□ Domain ontology 定义一系列slot,和每个slot可以包含的数值/选项

□ 系统需要从对话记录（包含用户输入和系统的回复）中发现并记录

1. 用户的目标和信息（goal, informable slots)
2. 用户的对搜索结果的问题（requests)



User: I’m looking for a cheaper restaurant

`inform(price=cheap)`

System: Sure. What kind - and where?

User: Thai food, somewhere downtown

`inform(price=cheap, food=Thai, area=centre)`

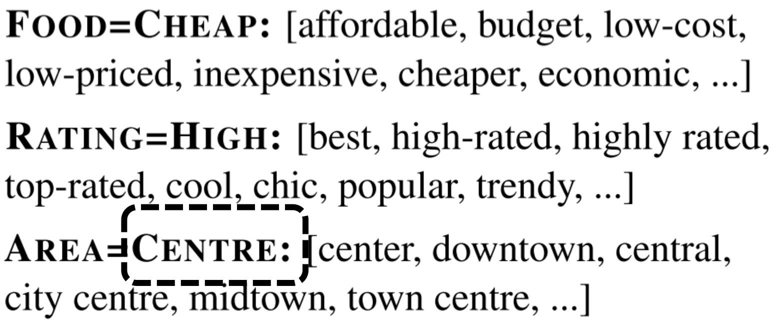
System: The House serves cheap Thai food

User: Where is it?

`inform(price=cheap, food=Thai, area=centre); request(address)`

System: The House is at 106 Regent Street

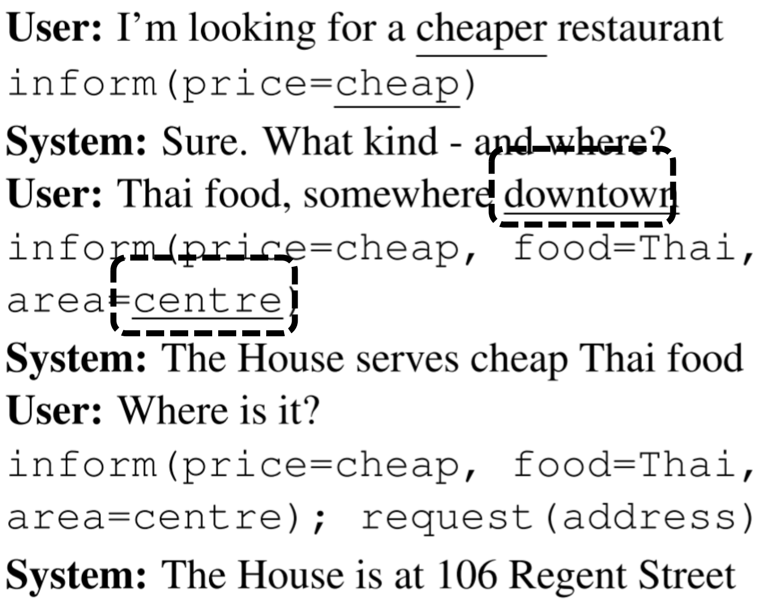
对话系统中的belief tracker



FOOD=CHEAP: [affordable, budget, low-cost, low-priced, inexpensive, cheaper, economic, ...]

RATING=HIGH: [best, high-rated, highly rated, top-rated, cool, chic, popular, trendy, ...]

AREA=CENTRE: [center, downtown, central, city centre, midtown, town centre, ...]



User: I'm looking for a cheaper restaurant

`inform(price=cheap)`

System: Sure. What kind - ~~and where?~~

User: Thai food, somewhere downtown

`inform(price=cheap, food=Thai, area=centre)`

System: The House serves cheap Thai food

User: Where is it?

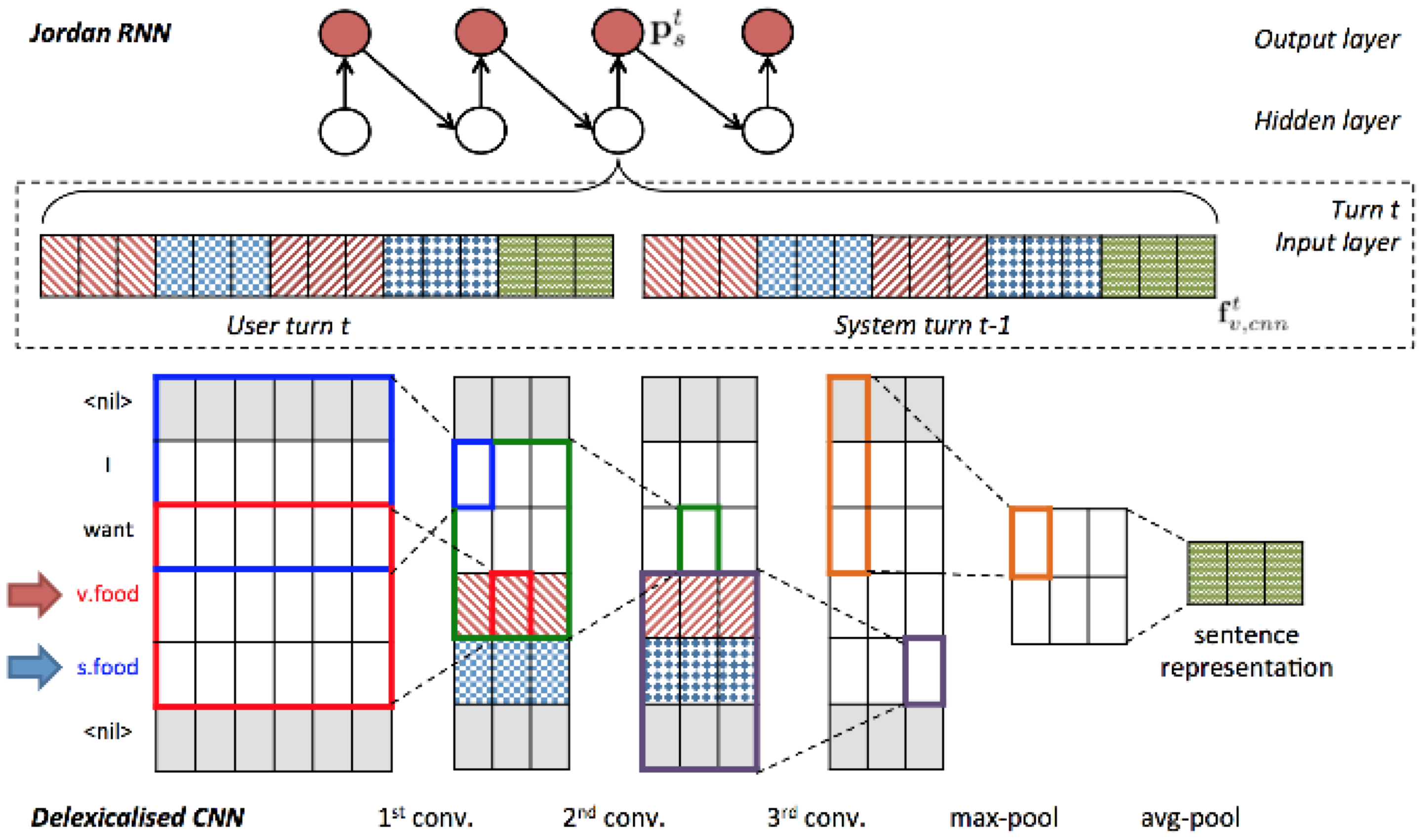
`inform(price=cheap, food=Thai, area=centre); request(address)`

System: The House is at 106 Regent Street

一个使用semantic dictionary做**delexicalization** 的例子，一些意义相似的词语被转化为一个标准形式

也可以使用表示semantic similarity的词向量参数简介达到这样的效果

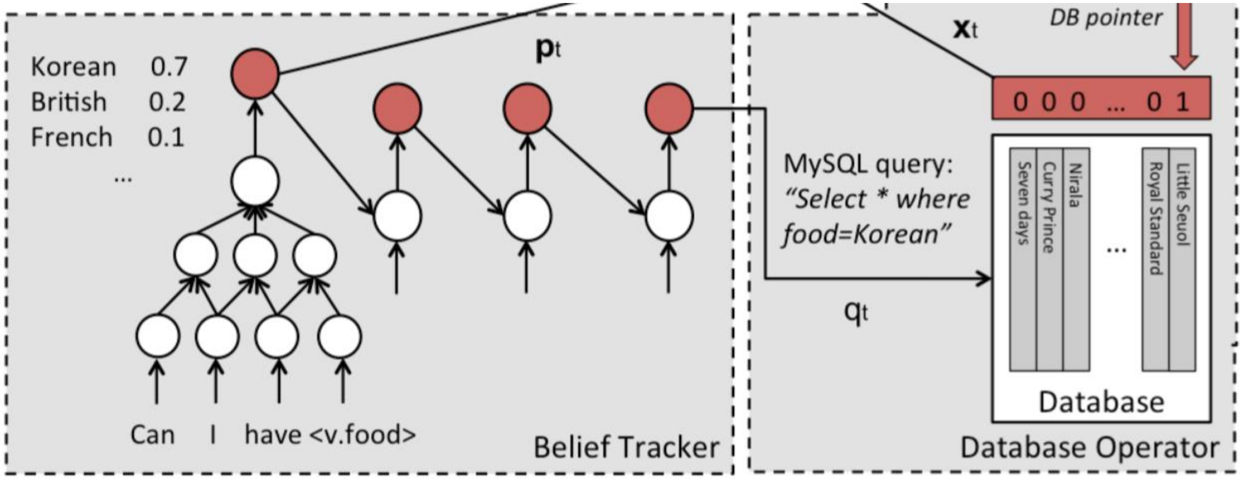
一轮对话的流程



Belief tracker（也叫Dialogue *State* tracker）的实现细节

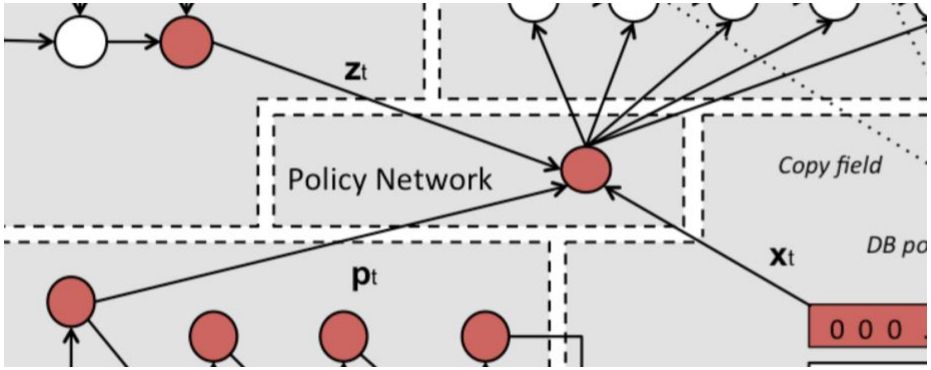
一轮对话的流程

基于belief tracker的输出，产生query，在数据库中寻找对应的条目



一轮对话的流程

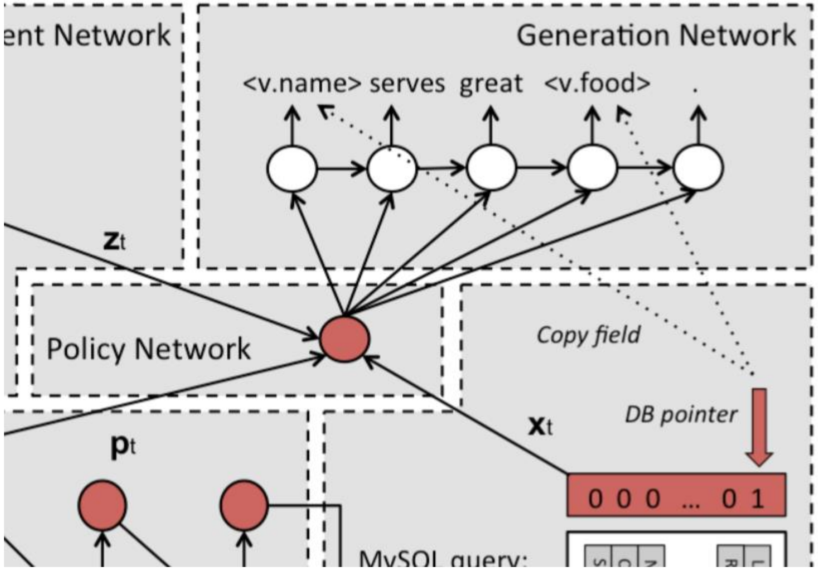
将intent network, belief tracker, DB query综合起来生成输出状态



$\boldsymbol{o}_t = \tanh(W_{zo}\boldsymbol{z}_t + W_{po}\boldsymbol{p}_t + W_{xo}\boldsymbol{x}_t)$

$\boldsymbol{p}_t = [p_{s1}^t, p_{s2}^t \dots p_{sK}^t]$ 综合所有的slot-value pair

一轮对话的流程



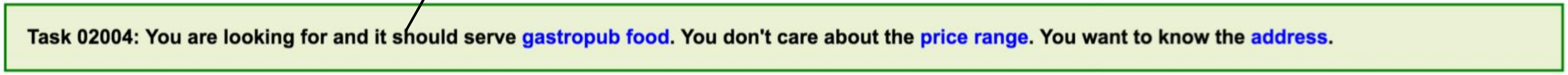
RNN decoder生成最终回复

数据收集流程

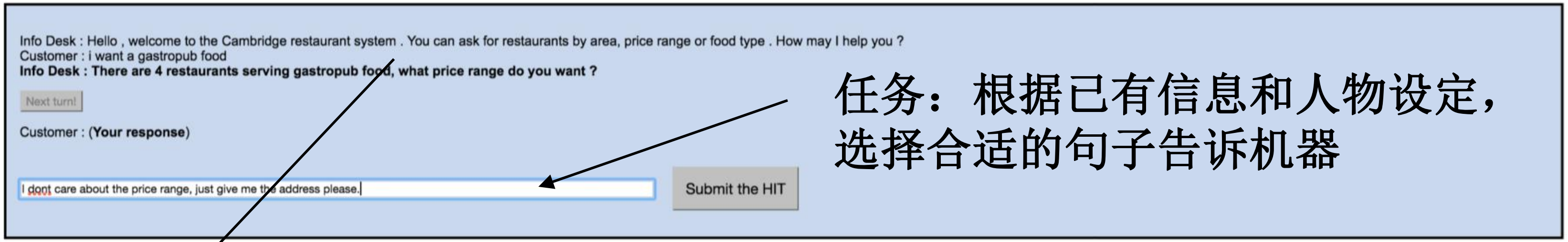
- 使用统计方法构建聊天系统的最大瓶颈之一是收集合适的数据
- 使用奥兹巫师实验的方式搜集数据
  - 在人机交互领域，一个奥兹巫师实验是一个研究实验，在这个实验中，受试者与一个计算机系统进行交互，这个系统被认为是自主的，但实际上是由一个看不见的人类操作或部分操作的。

数据收集流程

任务：你想要寻找一个酒吧类餐厅，不关心价格范围，需要知道地址



Task 02004: You are looking for and it should serve gastropub food. You don't care about the price range. You want to know the address.



Info Desk : Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area, price range or food type . How may I help you ?
Customer : i want a gastropub food
Info Desk : There are 4 restaurants serving gastropub food, what price range do you want ?

Customer : (**Your response**)

I dont care about the price range, just give me the address please.

任务：根据已有信息和人物设定，选择合适的句子告诉机器

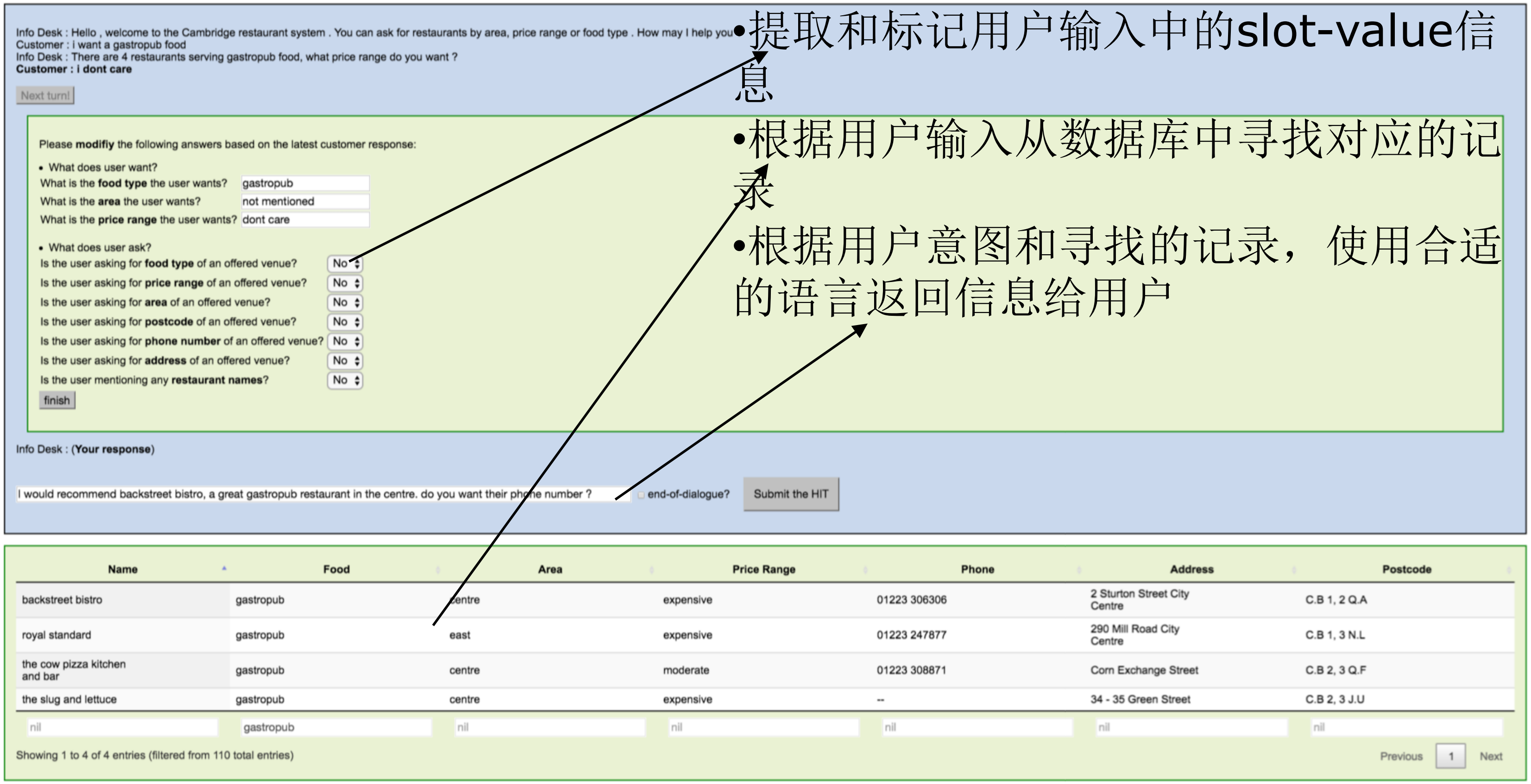
已有信息：

- Wizard：你好，欢迎来到剑桥餐馆系统，你可以根据地址，价格和食物种类询问餐馆
- User：我想找一家酒吧类餐厅
- Wizard：有四家酒吧类餐厅，请问你选择什么价格范围的餐厅？

数据收集流程

假装你是机器（wizard），

- 提取和标记用户输入中的slot-value信息
- 根据用户输入从数据库中寻找对应的记录
- 根据用户意图和寻找的记录，使用合适的语言返回信息给用户



Info Desk : Hello , welcome to the Cambridge restaurant system . You can ask for restaurants by area, price range or food type . How may I help you
Customer : i want a gastropub food
Info Desk : There are 4 restaurants serving gastropub food, what price range do you want ?
Customer : i dont care

Please **modifiy** the following answers based on the latest customer response:

- What does user want?

What is the **food type** the user wants? gastropub
What is the **area** the user wants? not mentioned
What is the **price range** the user wants? dont care

- What does user ask?

Is the user asking for **food type** of an offered venue? No
Is the user asking for **price range** of an offered venue? No
Is the user asking for **area** of an offered venue? No
Is the user asking for **postcode** of an offered venue? No
Is the user asking for **phone number** of an offered venue? No
Is the user asking for **address** of an offered venue? No
Is the user mentioning any **restaurant names**? No

Info Desk : (**Your response**)

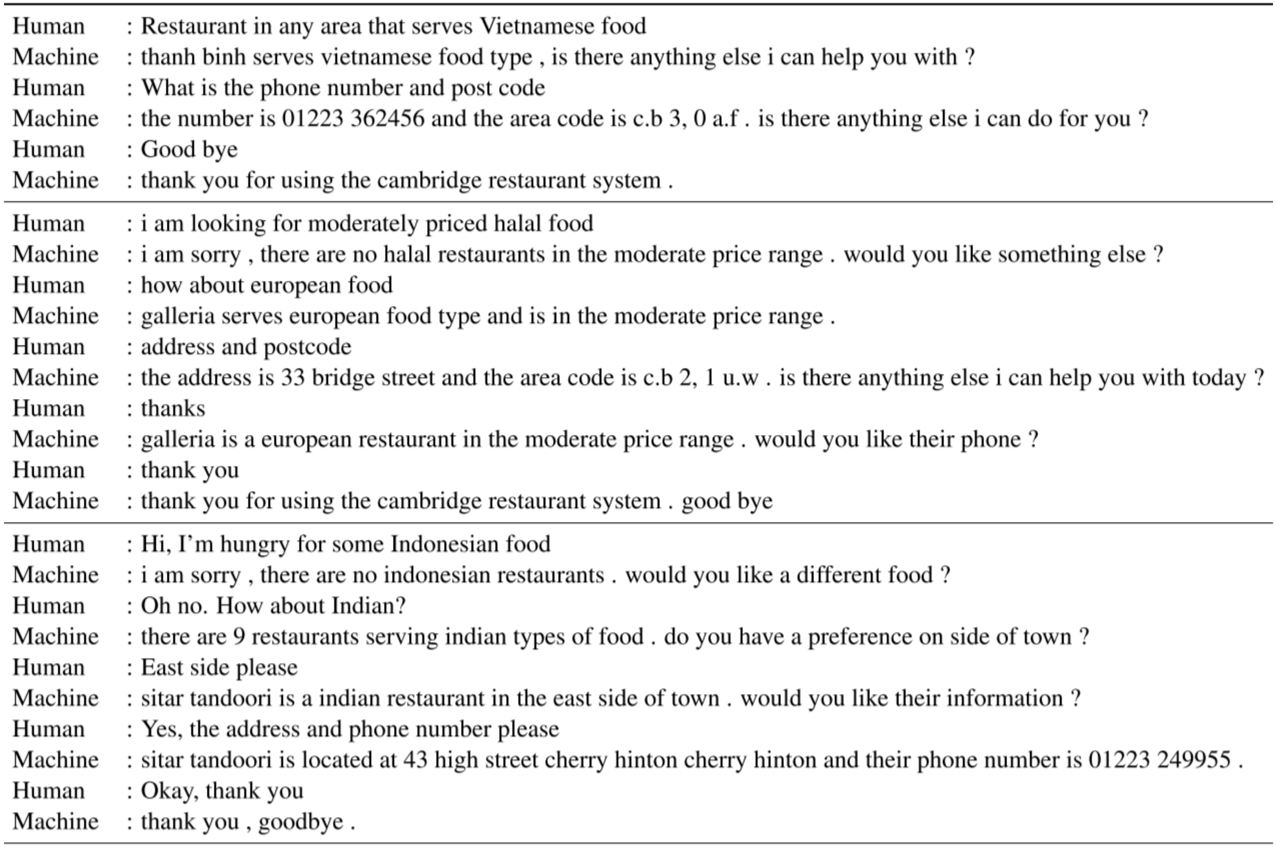
I would recommend backstreet bistro, a great gastropub restaurant in the centre. do you want their phone number ?

☐ end-of-dialogue?

Name	Food	Area	Price Range	Phone	Address	Postcode
backstreet bistro	gastropub	centre	expensive	01223 306306	2 Sturton Street City Centre	C.B 1, 2 Q.A
royal standard	gastropub	east	expensive	01223 247877	290 Mill Road City Centre	C.B 1, 3 N.L
the cow pizza kitchen and bar	gastropub	centre	moderate	01223 308871	Corn Exchange Street	C.B 2, 3 Q.F
the slug and lettuce	gastropub	centre	expensive	--	34 - 35 Green Street	C.B 2, 3 J.U
nil	gastropub	nil	nil	nil	nil	nil

Showing 1 to 4 of 4 entries (filtered from 110 total entries)

模型效果



Human : Restaurant in any area that serves Vietnamese food
Machine : thanh binh serves vietnamese food type , is there anything else i can help you with ?
Human : What is the phone number and post code
Machine : the number is 01223 362456 and the area code is c.b 3, 0 a.f . is there anything else i can do for you ?
Human : Good bye
Machine : thank you for using the cambridge restaurant system .

Human : i am looking for moderately priced halal food
Machine : i am sorry , there are no halal restaurants in the moderate price range . would you like something else ?
Human : how about european food
Machine : galleria serves european food type and is in the moderate price range .
Human : address and postcode
Machine : the address is 33 bridge street and the area code is c.b 2, 1 u.w . is there anything else i can help you with today ?
Human : thanks
Machine : galleria is a european restaurant in the moderate price range . would you like their phone ?
Human : thank you
Machine : thank you for using the cambridge restaurant system . good bye

Human : Hi, I'm hungry for some Indonesian food
Machine : i am sorry , there are no indonesian restaurants . would you like a different food ?
Human : Oh no. How about Indian?
Machine : there are 9 restaurants serving indian types of food . do you have a preference on side of town ?
Human : East side please
Machine : sitar tandoori is a indian restaurant in the east side of town . would you like their information ?
Human : Yes, the address and phone number please
Machine : sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955 .
Human : Okay, thank you
Machine : thank you , goodbye .

目标导向的模型训练方法

强化学习

强化学习概述

- Agent interacting with a dynamic environment with the objective of maximizing a long term reward
    - 强化学习(Reinforcement Learning，RL) 关注目标导向 (goal-directed learning) 的学习和决策问题.
    - 强化学习中，agent 通过(states, actions 和 rewards) 和环境交流从中学习知识.
        - agent 通过 policy 从 state 出发寻找 actions
    - 强化学习中，agent 试图优化长期目标(long-term goal)，而 rewards 的反馈通常有延迟.

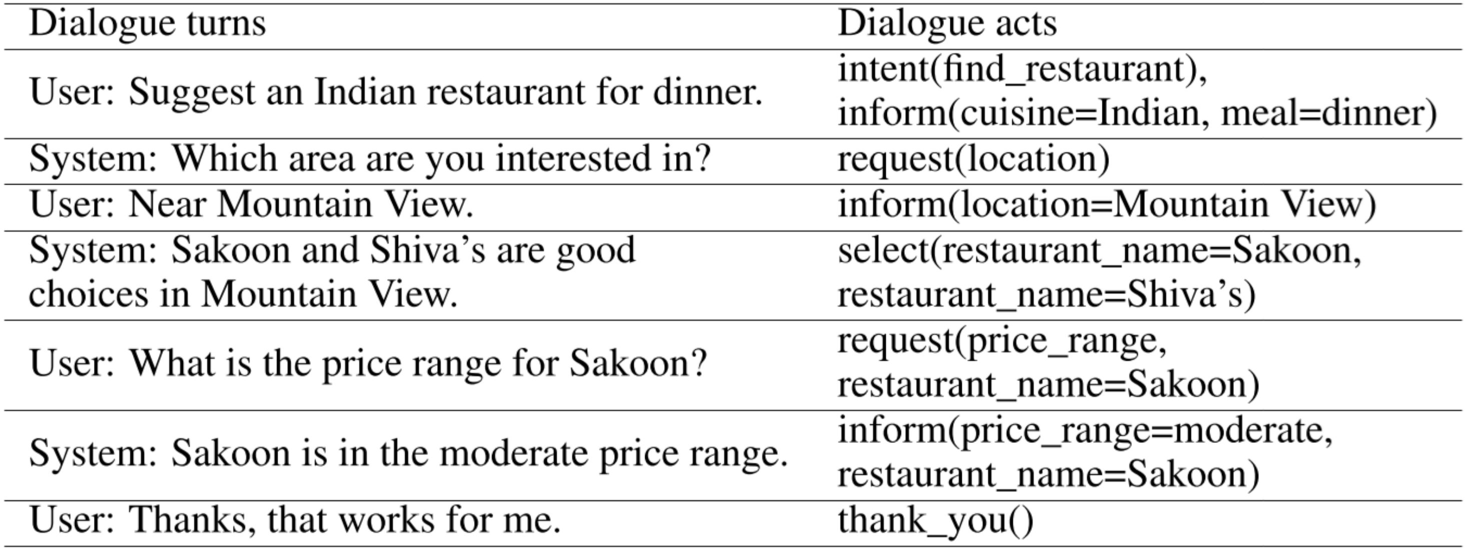
强化学习概述

- 强化学习使用目标导向的(agents)，获取环境的状态，作出决定(action)，影响环境
- 强化学习中的 探索-开发 (exploration-exploitation dilemma) 过程
  - 在学习初期，agent 对环境有很多不确定，侧重使用探索(exploration) 方式作出不确定性比较大的决策
  - 在学习后期，agent 侧重使用开发(exploitation)作出认为当前最好的决策

强化学习概述

- 例子
  - 围棋软件：了解棋盘局面(state)，作出符合围棋规则的决策(action)，期望最终赢得比赛(reward)；每一步的决策选项相对较小，轮数相对较大
  - 聊天系统：了解用户意图和总结当下语境(state)，给出有质量的回复(action)，期望尽快给出用户想要的信息，避免无用或者错误回复，使得用户满意(reward)；每一步的决策选项相对较大，轮数相对较小

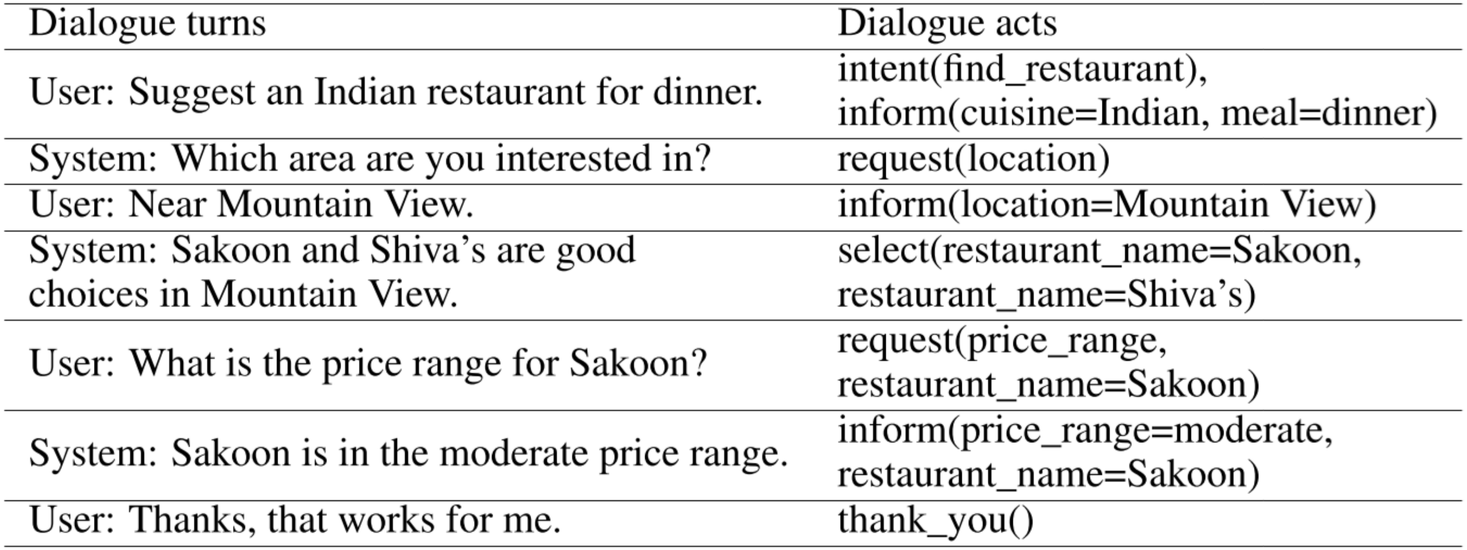
使用MDP描述一个对话问题



Dialogue turns	Dialogue acts
User: Suggest an Indian restaurant for dinner.	intent(find_restaurant), inform(cuisine=Indian, meal=dinner)
System: Which area are you interested in?	request(location)
User: Near Mountain View.	inform(location=Mountain View)
System: Sakoon and Shiva's are good choices in Mountain View.	select(restaurant_name=Sakoon, restaurant_name=Shiva's)
User: What is the price range for Sakoon?	request(price_range, restaurant_name=Sakoon)
System: Sakoon is in the moderate price range.	inform(price_range=moderate, restaurant_name=Sakoon)
User: Thanks, that works for me.	thank_you()

- Markov Decision Process: $\{S, A, P, R, \gamma\}$
  - {state, action, policy, reward, discount factor}

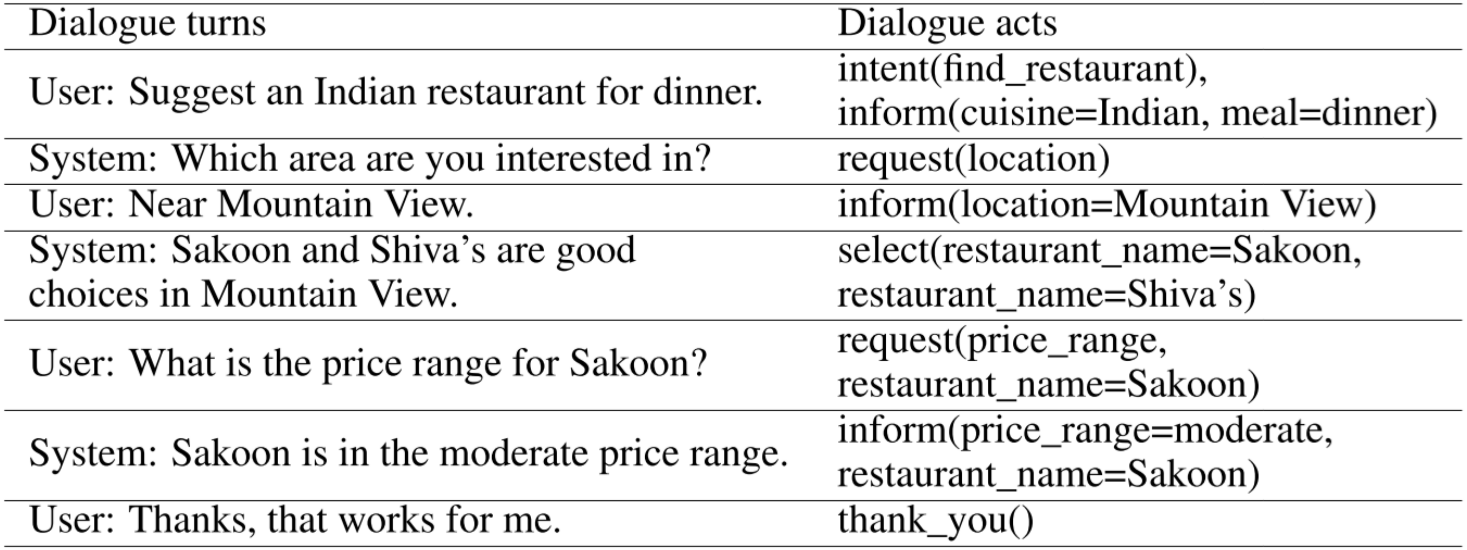
使用MDP描述一个对话问题



Dialogue turns	Dialogue acts
User: Suggest an Indian restaurant for dinner.	intent(find_restaurant), inform(cuisine=Indian, meal=dinner)
System: Which area are you interested in?	request(location)
User: Near Mountain View.	inform(location=Mountain View)
System: Sakoon and Shiva's are good choices in Mountain View.	select(restaurant_name=Sakoon, restaurant_name=Shiva's)
User: What is the price range for Sakoon?	request(price_range, restaurant_name=Sakoon)
System: Sakoon is in the moderate price range.	inform(price_range=moderate, restaurant_name=Sakoon)
User: Thanks, that works for me.	thank_you()

- 环境状态通常使用{用户动机，slot-value pair}等特征表示
  - 例：intent=find_restaurant; location=MV, price_range = moderate

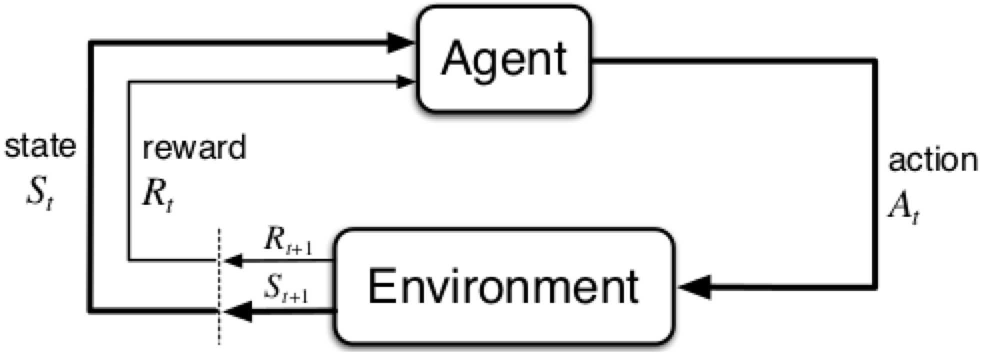
使用MDP描述一个对话问题



Dialogue turns	Dialogue acts
User: Suggest an Indian restaurant for dinner.	intent(find_restaurant), inform(cuisine=Indian, meal=dinner)
System: Which area are you interested in?	request(location)
User: Near Mountain View.	inform(location=Mountain View)
System: Sakoon and Shiva’s are good choices in Mountain View.	select(restaurant_name=Sakoon, restaurant_name=Shiva’s)
User: What is the price range for Sakoon?	request(price_range, restaurant_name=Sakoon)
System: Sakoon is in the moderate price range.	inform(price_range=moderate, restaurant_name=Sakoon)
User: Thanks, that works for me.	thank_you()

- 在找餐馆的环境下，agent的action通常有提供信息，询问slot-value等
  - 例：request(location), select(…)

使用MDP描述一个对话问题



- 目标：寻找一个policy $\pi_\theta: S \rightarrow P(A)$，最大化期望的reward: "**cumulative reward agent receives in the long run**《Richard S. Sutton, Andrew G. Barto. Reinforcement Learning - An Introduction》"
  - $J_t(\theta) = E[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots]$
  - 当前（时间t）的决策会**长期**地影响将来的进程
  - 一段时间以后才便于更准确地估计当下的决策会有什么样的**长期**后果，即，rewards 的反馈通常有**延迟**

强化学习的chatbot上的应用案例

- 拟合数据（观察到的实际回复）训练模型
  - 目标函数和人类对话的目的不符
  - 数据有限，安全回复
- 强化学习
  - 设计+优化 和人类对话目的更相近的长期目标

一个开放领域的强化学习应用

- Deep Reinforcement Learning for Dialogue Generation （2016）
- 通过强化学习手段，鼓励两个互相聊天的agents产生高质量的对话

设计一个强化学习问题

- 强化学习模型: 两个使用seq2seq模型的 agents 对话
  - State: $[p_i, q_i]$，一轮对话
  - **Policy**: $P(p_{i+1}|p_i, q_i)$，使用seq2seq模型，包含需要学习的参数
  - Action: agent产生的回复
  - **Goal**: 两个使用seq2seq模型的agent的对话接近有意义的人类对话

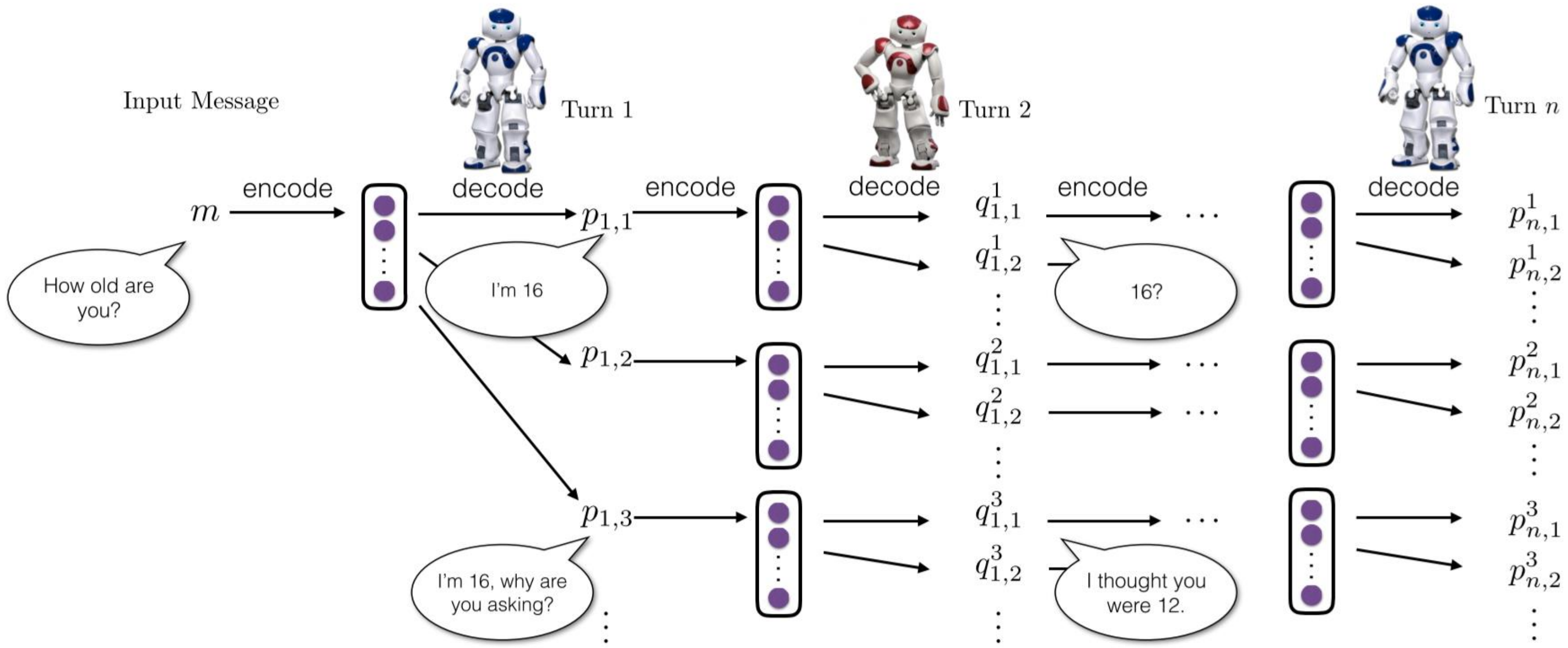
设计一个强化学习问题

- 强化学习模型: 两个使用seq2seq模型的agents 对话
  - State: $[p_i, q_i]$，一轮对话
  - Policy: $P(p_{i+1}|p_i, q_i)$ ，使用seq2seq模型，包含需要学习的参数
  - Action: agent产生的回复
  - **Reward**:
    - □ Ease of answering: 如果本回复导致对方更倾向于回复“不知道”，“不清楚”等表示难以继续交流下去的意思，扣分
    - □ Information flow ：如果这一轮的回复和上一轮本agent的回复非常相似，扣分
    - □ Semantic coherence ：如果（1）给定上一轮对话(state)，本轮回复的概率比较大(传统的SEQ2SEQ里面的cond. prob)，且（2）从本轮回复能比较容易地反推出上一轮对话的信息，加分

强化学习过程

- 模型(policy)初始化：基于优化likelihood的seq2seq模型
- 模型simulation:
  - 两个agent聊天产生聊天样本数据集
  - 每个样本是多轮对话
  - 通过优化样本的reward更新模型(policy)的参数，使模型产生更好的对话效果

强化学习过程


Input Message
Turn 1
Turn 2
Turn n
m
encode
decode
encode
decode
encode
decode
How old are you?
I'm 16
I'm 16, why are you asking?
16?
I thought you were 12.
$p_{1,1}$
$p_{1,2}$
$p_{1,3}$
$q^1_{1,1}$
$q^1_{1,2}$
$q^2_{1,1}$
$q^2_{1,2}$
$q^3_{1,1}$
$q^3_{1,2}$
$p^1_{n,1}$
$p^1_{n,2}$
$p^2_{n,1}$
$p^2_{n,2}$
$p^3_{n,1}$
$p^3_{n,2}$

强化学习效果

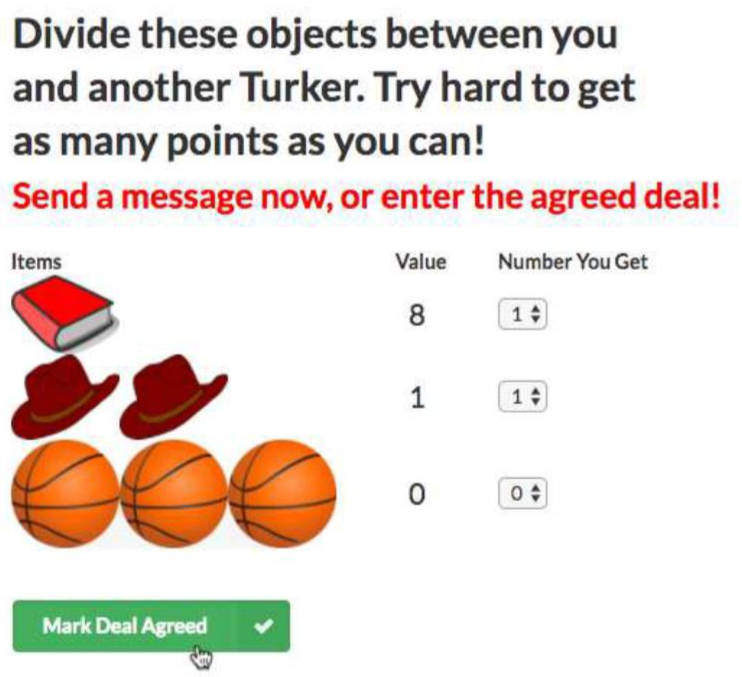
Input Message	Mutual Information Model	RL
How old are you ?	I’m 16.	I’m 16, why are you asking ?
What’s your full name ?	I have no idea .	What’s yours ?
I don’t want to go home tonight.	Really ?	Why ?
Do you have any feelings for me ?	I don’t know what you are talking about.	I don ’t want to hurt your feelings .
How much time do you have here?	Not long enough. Sorry, sir.	Ten seconds .
Shall we get started ?	Of course !	Yes . We ’ve got a lot of work to do here .
Do you play football ?	No, i don’t	Yes. I love football !
We’d have to talk to him.	I mean, he’s a good guy	About what ?
How come you never say it?	Because I don’t want to hurt you .	I don ’t think it ’s a good idea to say it .

Model	# of simulated turns
SEQ2SEQ	2.68
mutual information	3.40
RL	4.48

一个封闭领域的强化学习应用

- Deal or no deal? End-to-end learning for negotiation dialogues (2017-06)
- 在一个人工生成的商议任务数据集上，训练两个chatbot互相聊天谈判
  - 拟合导向的模型(最大化对话数据概率的模型)
  - 目标导向的模型(优化bot收益的模型)

人工产生的 bot-bot 聊天数据



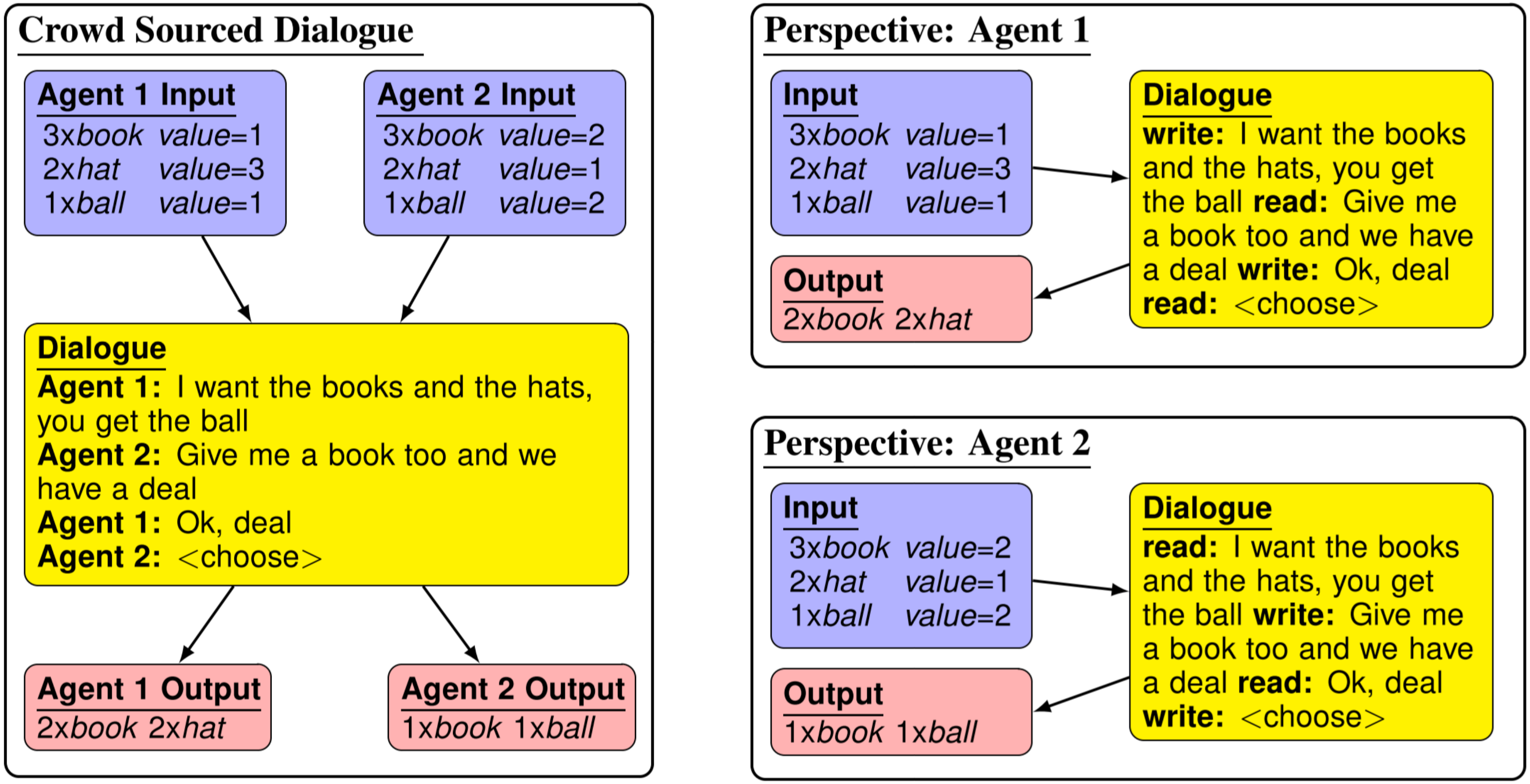
Divide these objects between you and another Turker. Try hard to get as many points as you can!

Send a message now, or enter the agreed deal!

Items	Value	Number You Get
	8	1
	1	1
	0	0

- □ 两个users (bots)协商瓜分一部分物品
  - 对每个user (bot)，所有物品的价值总和是10
  - 每个物品在不同的user(bot)那里可以有不同的价值
  - 有些物品对于两个users(bots)都有价值 (竞争关系)
  - 每个物品至少对于一个user(bot) 是有价值的 (>0)
- □ 目标/收益
  - 最大化自身的收益
  - 需要保证对方也同意

人工产生的 bot-bot 聊天数据



Crowd Sourced Dialogue

Agent 1 Input
3x*book* *value*=1
2x*hat* *value*=3
1x*ball* *value*=1

Agent 2 Input
3x*book* *value*=2
2x*hat* *value*=1
1x*ball* *value*=2

Dialogue
Agent 1: I want the books and the hats, you get the ball
Agent 2: Give me a book too and we have a deal
Agent 1: Ok, deal
Agent 2: <choose>

Agent 1 Output
2x*book* 2x*hat*

Agent 2 Output
1x*book* 1x*ball*

Perspective: Agent 1

Input
3x*book* *value*=1
2x*hat* *value*=3
1x*ball* *value*=1

Dialogue
write: I want the books and the hats, you get the ball **read:** Give me a book too and we have a deal **write:** Ok, deal **read:** <choose>

Output
2x*book* 2x*hat*

Perspective: Agent 2

Input
3x*book* *value*=2
2x*hat* *value*=1
1x*ball* *value*=2

Dialogue
read: I want the books and the hats, you get the ball **write:** Give me a book too and we have a deal **read:** Ok, deal **write:** <choose>

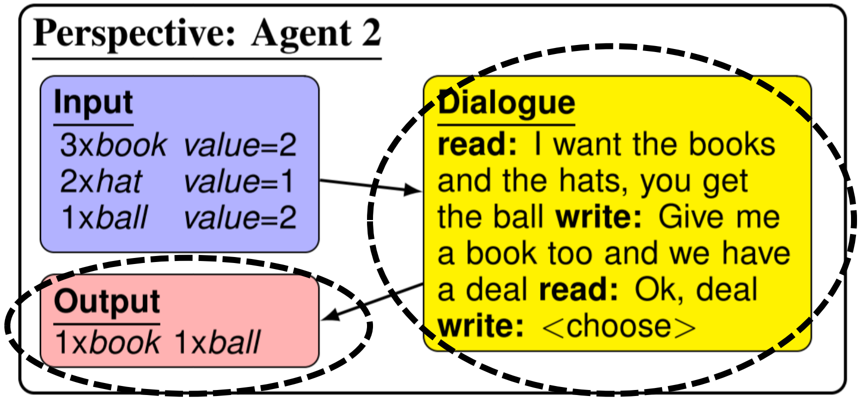
Output
1x*book* 1x*ball*

每段对话分成两个样本，分别对应两个**Bot**视角的对话记录

Baseline: likelihood model

- 输入数据
  - 物品信息g: $[c_1, v_1, c_2, v_2, c_3, v_3]$, 三类物品各自的数量和价值
  - 讨论结果o: $[n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}]$,每个物品每个bot各自挑选多少
  - 对话记录x: $[x_0, \cdots x_T]$, 自己和对方Bot的话
- 目标函数
  - 拟合对话和商议结果

e.g.



Perspective: Agent 2

Input
3x*book* *value*=2
2x*hat* *value*=1
1x*ball* *value*=2

Dialogue
read: I want the books and the hats, you get the ball **write:** Give me a book too and we have a deal **read:** Ok, deal **write:** <choose>

Output
1x*book* 1x*ball*

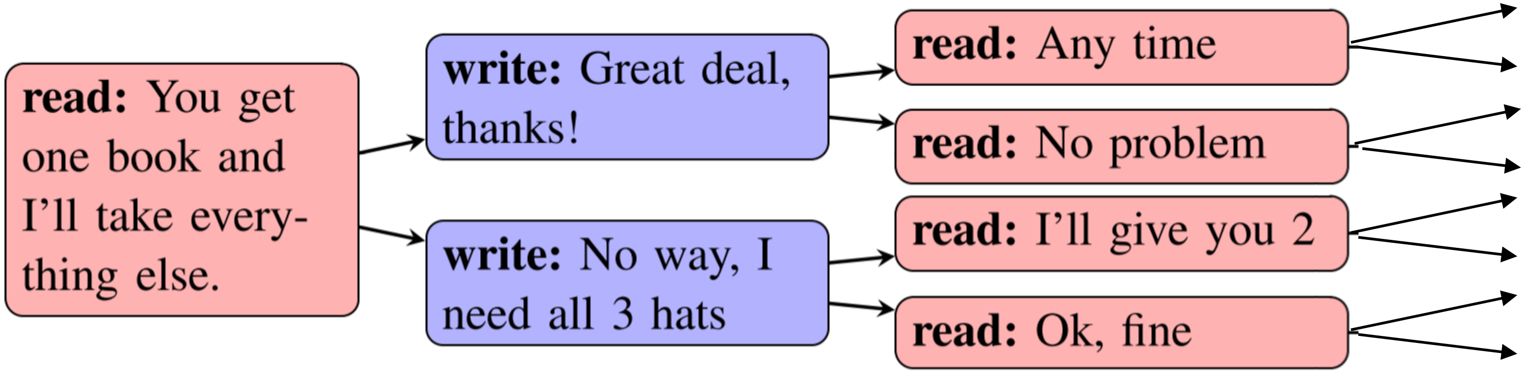
Baseline: likelihood model

- Likehihood 模型的问题在于目标函数和实际目标不符：
  - Likelihood模型：拟合对话数据
  - 实际目标：最大化用户的收益

目标导向的模型

- 首先训练likelihood model初始化模型参数
- 使用强化学习根据衡量标准(收益) finetune模型

目标导向的模型



read: You get one book and I'll take everything else.

write: Great deal, thanks!

write: No way, I need all 3 hats

read: Any time

read: No problem

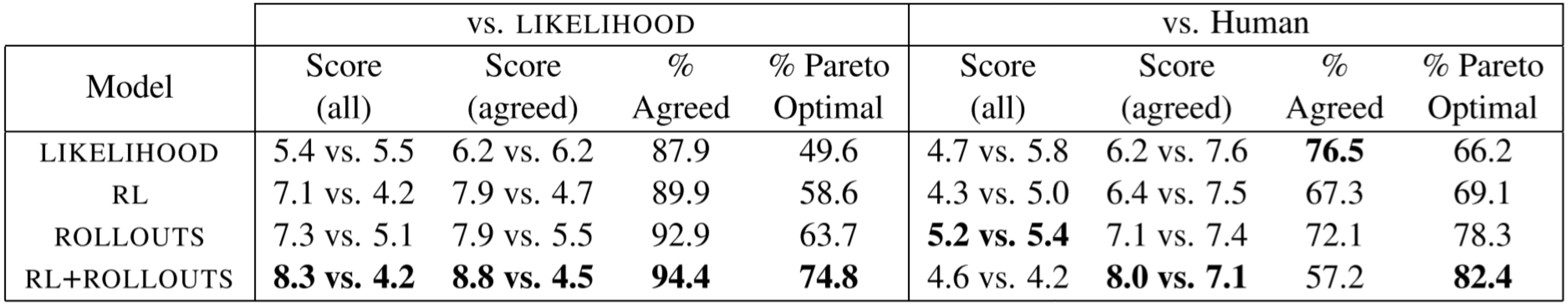
read: I'll give you 2

read: Ok, fine

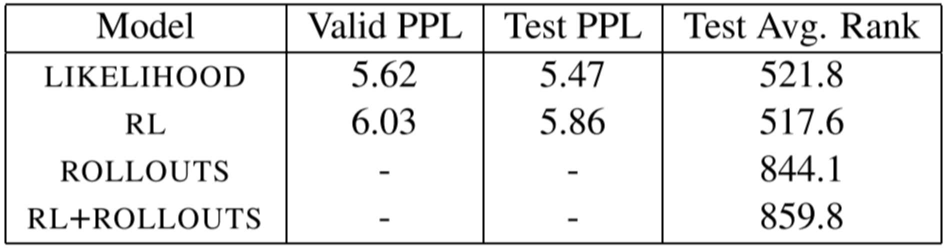
- Reward:
  - r^A：在这句话结束协商的话，收益会是多少
  - 考虑到长期收益：

$$R(x_t) = \sum_{x_t \in X^A} \gamma^{(T-t)}(\gamma^A(o) - \mu)$$

目标导向的模型

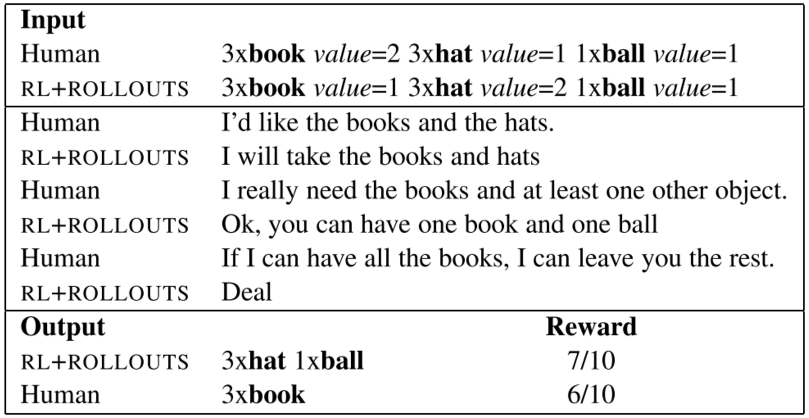


Model	vs. LIKELIHOOD Score (all)	Score (agreed)	% Agreed	% Pareto Optimal	vs. Human Score (all)	Score (agreed)	% Agreed	% Pareto Optimal
LIKELIHOOD	5.4 vs. 5.5	6.2 vs. 6.2	87.9	49.6	4.7 vs. 5.8	6.2 vs. 7.6	**76.5**	66.2
RL	7.1 vs. 4.2	7.9 vs. 4.7	89.9	58.6	4.3 vs. 5.0	6.4 vs. 7.5	67.3	69.1
ROLLOUTS	7.3 vs. 5.1	7.9 vs. 5.5	92.9	63.7	**5.2 vs. 5.4**	7.1 vs. 7.4	72.1	78.3
RL+ROLLOUTS	**8.3 vs. 4.2**	**8.8 vs. 4.5**	**94.4**	**74.8**	4.6 vs. 4.2	**8.0 vs. 7.1**	57.2	**82.4**



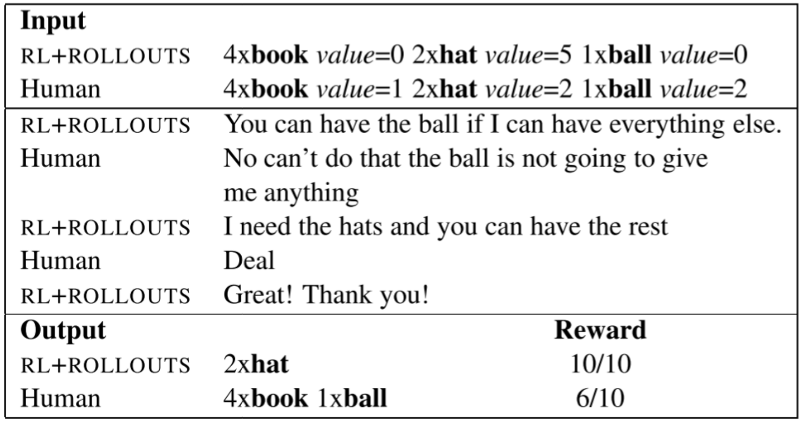
Model	Valid PPL	Test PPL	Test Avg. Rank
LIKELIHOOD	5.62	5.47	521.8
RL	6.03	5.86	517.6
ROLLOUTS	-	-	844.1
RL+ROLLOUTS	-	-	859.8

目标导向的模型



Input		
Human	3x**book** *value*=2 3x**hat** *value*=1 1x**ball** *value*=1	
RL+ROLLOUTS	3x**book** *value*=1 3x**hat** *value*=2 1x**ball** *value*=1	
Human	I'd like the books and the hats.	
RL+ROLLOUTS	I will take the books and hats	
Human	I really need the books and at least one other object.	
RL+ROLLOUTS	Ok, you can have one book and one ball	
Human	If I can have all the books, I can leave you the rest.	
RL+ROLLOUTS	Deal	
Output		**Reward**
RL+ROLLOUTS	3x**hat** 1x**ball**	7/10
Human	3x**book**	6/10

Figure 6: Example of model compromising.



Input		
RL+ROLLOUTS	4x**book** *value*=0 2x**hat** *value*=5 1x**ball** *value*=0	
Human	4x**book** *value*=1 2x**hat** *value*=2 1x**ball** *value*=2	
RL+ROLLOUTS	You can have the ball if I can have everything else.	
Human	No can't do that the ball is not going to give me anything	
RL+ROLLOUTS	I need the hats and you can have the rest	
Human	Deal	
RL+ROLLOUTS	Great! Thank you!	
Output		**Reward**
RL+ROLLOUTS	2x**hat**	10/10
Human	4x**book** 1x**ball**	6/10

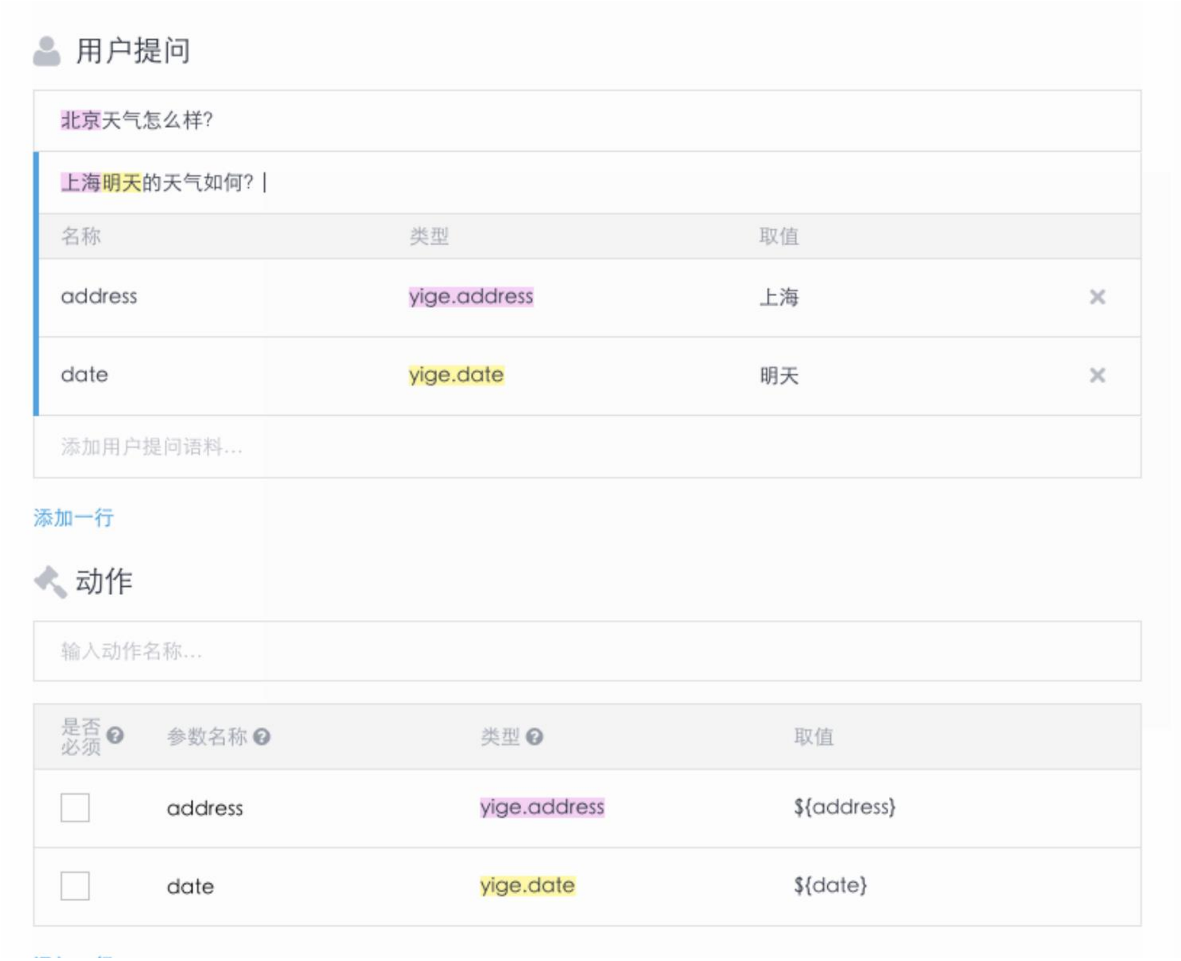
Figure 7: Dialogue in which the model's initial interest in the valueless books allows it to compromise while achieving a maximum score.

工业界的CHATBOT

Api.ai和‘一个’ai的典型工作流程

“一个”AI遵循的流程和Api.ai类似，接收到用户的输入后，流程如下：

- **一个AI首先识别用户输入中的词库和用户场景。**词库和场景的识别并不是独立的，相同的词在不同的场景下可能属于不同的词库类型。在场景识别时也会考虑到场景设定的状态是否存在。如果某场景设定的输入状态不是都存在，则不会把用户输入识别为此场景。
- 查看动作中需要的**必须参数(slog-value)**是否都已获得取值。如果存在必须参数还没有获得取值，就触发设定好的提示语作为机器人回复，要求用户输入对应的参数取值。参数的取值不仅可以来自于此次用户输入中的词库，也可以来自于输入状态中的变量。对于非必须参数，可以为他们设定**默认值**。
- 只有所有**必须参数**都已**收集**到取值，此**场景**才能**完成**，场景设定的AI回复才会作为回复返回给用户。到这里此场景就完成了，用户之后的输入就会触发新的循环。



北京天气怎么样?

上海明天的天气如何? |

名称	类型	取值	
address	yige.address	上海	×
date	yige.date	明天	×

是否必须	参数名称	类型	取值
☐	address	yige.address	${address}
☐	date	yige.date	${date}

API.AI演示

疑问

- 问题答疑：　　http://www.xxwenda.com/
  - 可邀请老师或者其他人回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop