

Capstone 成果汇报

基于RASA的任务型机器人构建

OPPO 技术平台中心

实习生 王晶冰

2019/6/18

内容



RASA介绍



机器人DEMO



总结展望



/01

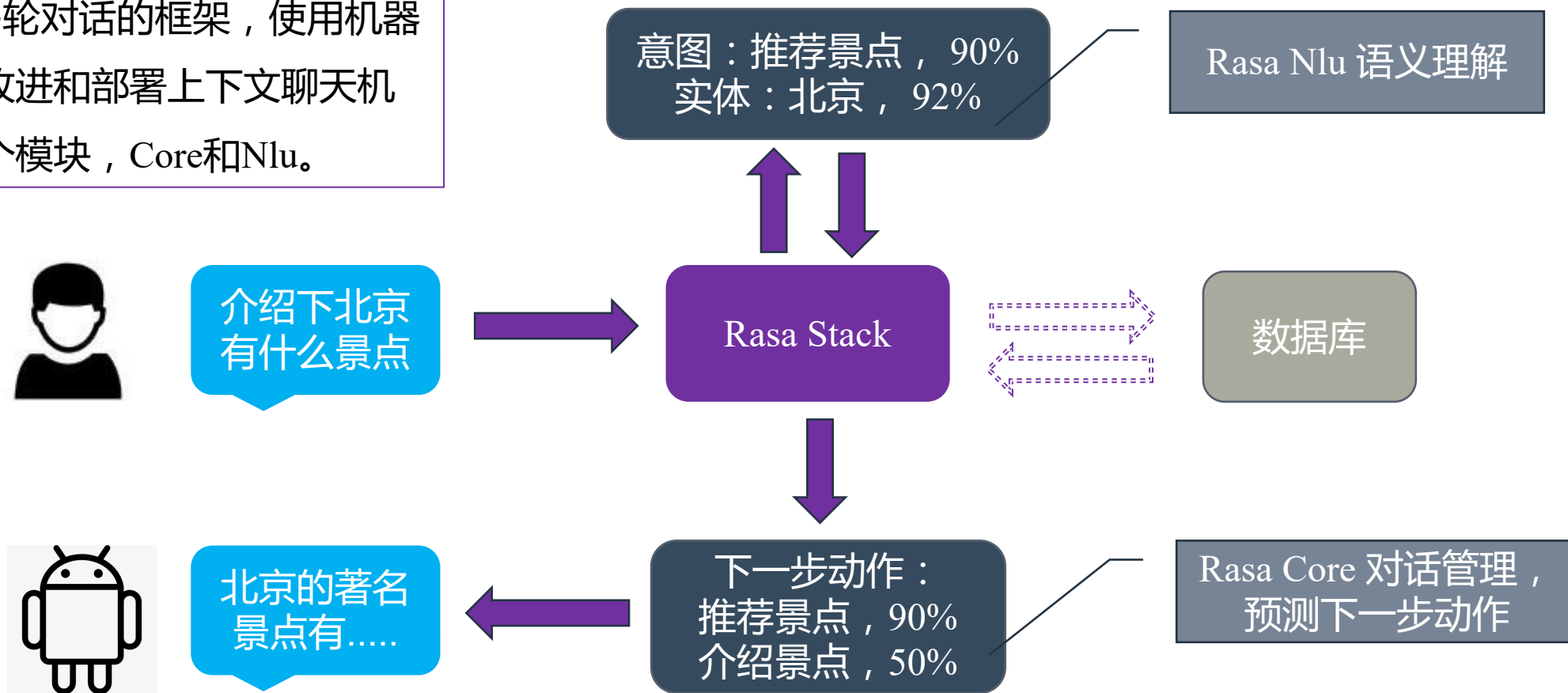
RASA介绍

RASA CORE

RASA NLU

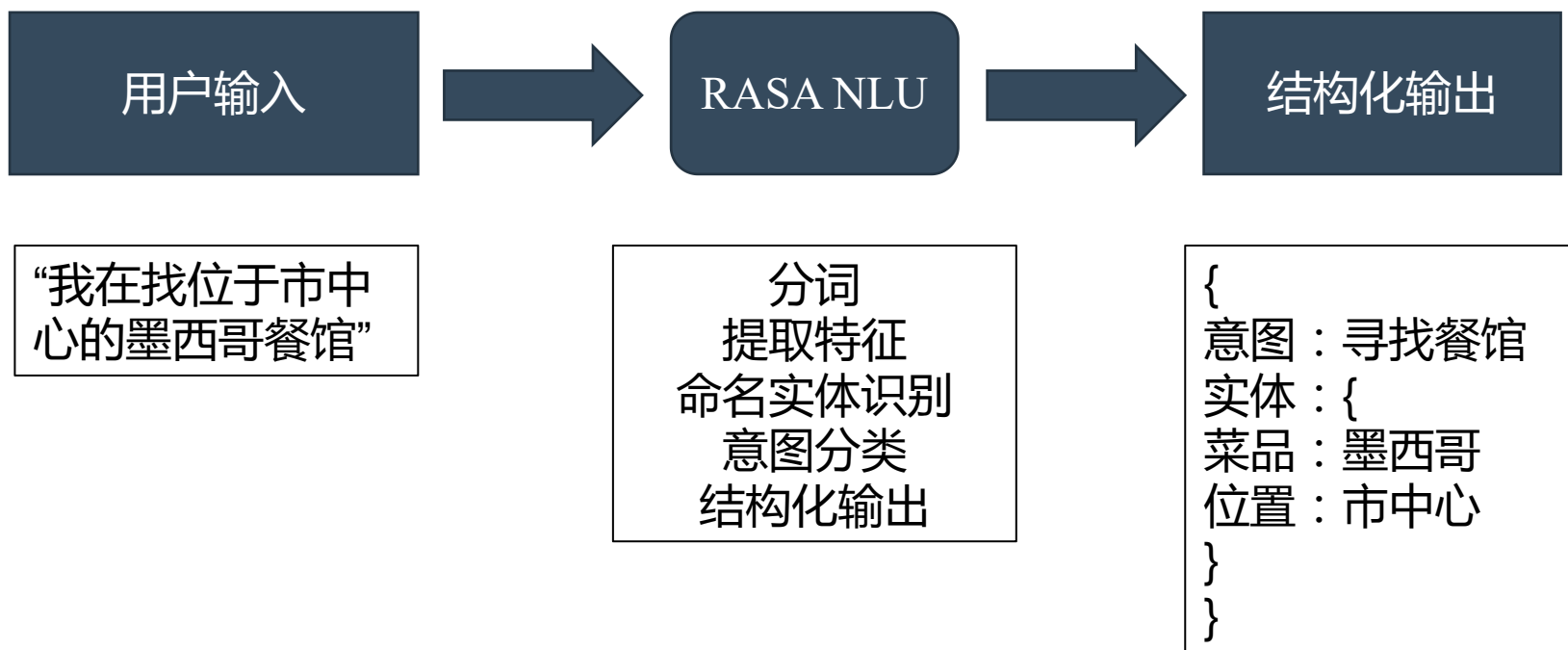
RASA介绍

RASA是一个基于多轮对话的框架，使用机器学习工具，构建、改进和部署上下文聊天机器人，主要包括两个模块，Core和Nlu。



RASA介绍

Rasa Nlu：用于理解语义，包括意图识别，实体识别，将用户的输入转换为结构化数据。



RASA介绍

Rasa Nlu：用于理解语义，包括意图识别，实体识别，将用户的输入转换为结构化数据。

Step 1 准备数据：支持Markdown和Json格式

Markdown示例

```
## 意图：检查余额
- 我的账户余额是多少？
- 我的[招商银行卡](账户) 有多少钱？
```

```
## 正则表达式：邮编
- [0-9]{5}
```

```
## 表查询：货币
- 元
-美元
```

```
## 表查询：外部资源
- path/to/currencies.txt
```

Json示例

```
{
  "rasa_nlu_data": {
    "common_examples": [
      {
        "text": "找个吃拉面的店",
        "intent": "寻找餐馆",
        "entities": [
          {
            "start": 3,
            "end": 5,
            "value": "拉面",
            "entities": "食物"
          }
        ]
      }
    ]
  }
}
```


RASA介绍

Rasa Nlu：用于理解语义，包括意图识别，实体识别，将用户的输入转换为结构化数据。

Step 1 准备数据：包括4个模块（example最为重要）

例子 examples

意图的示例直接列举，用-或*或+标记。实体用“[]”标记，实体类别用“()”标记，
例如：
[实体]（实体类别）

同义词 synonyms

定义同义词将不同的实体映射到同一实体。两种实现方式，1) 在examples中提取实体时直接将value替换；2) 在entity_synonyms定义同义词列表。

正则表达式特征 regex features

对于有特定结构的实体，可以通过定义其正则表达式来提取所有符合该结构特征的实体。

表查询 Lookup table

list格式列举或者提供外部文件路径导入，外部文件每一个选项单独成一行），查询时精确匹配。

RASA介绍

Rasa Nlu：用于理解语义，包括意图识别，实体识别，将用户的输入转换为结构化数据。

Step 2 设定模型框架：配置模型Pipeline

数据量 < 1000 选用pretrained_embedding
(使用预训练的词向量)

```
language: "en"

pipeline:
- name: "SpacyNLP"
- name: "SpacyTokenizer"
- name: "SpacyFeaturizer"
- name: "RegexFeaturizer"
- name: "CRFEntityExtractor"
- name: "EntitySynonymMapper"
- name: "SklearnIntentClassifier"
```

根据需要，设置
组件即可，训练
过程按照组件顺
序进行。支持自
定义组件。

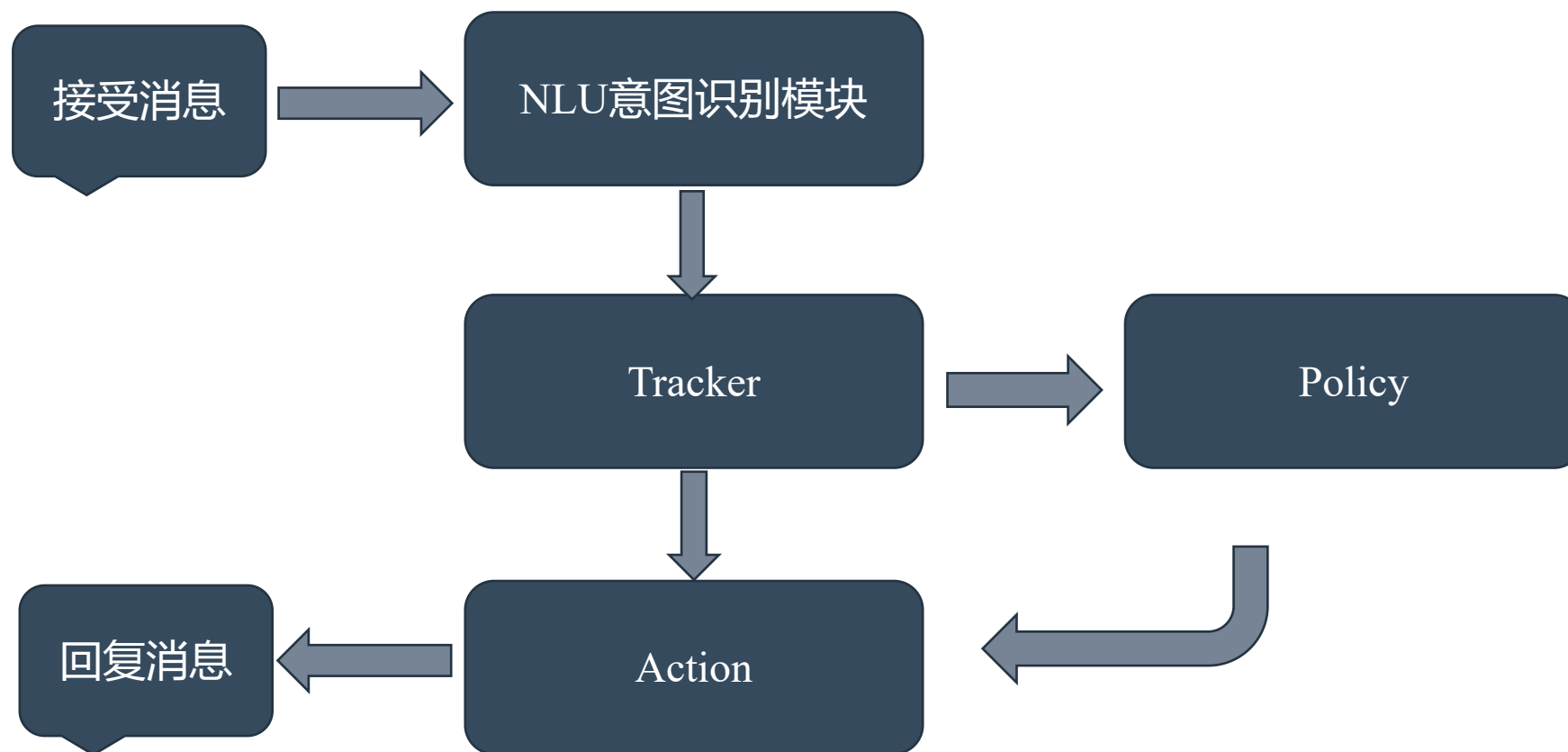
数据量 > 1000 选用supervised embedding
(直接根据特定任务数据训练词向量)

```
language: "en"

pipeline:
- name: "WhitespaceTokenizer"
- name: "CRFEntityExtractor"
- name: "EntitySynonymMapper"
- name: "CountVectorsFeaturizer"
- name: "EmbeddingIntentClassifier"
```


RASA介绍

Rasa Core：对话管理平台，决定接下来机器人该返回给用户的内容。



RASA介绍

Rasa Core：对话管理平台，决定接下来机器人该返回给用户的内容。

Step 1 训练数据：Stories

编写可能的对话路径，
用于训练模型

```
## happy path 1
* greet
  - utter_greet
* inform
  - utter_ask
* recommend_spot
  - action_provide_spots
* introduce_spot
  - action_introduce_spots
  - slot{"spot": null}
* introduce_food
  - action_introduce_food
* thankyou
  - action_noworries
* goodbye
  - utter_goodbye
```

story的名字，便于debug
* 用户意图
- 机器人动作

```
## sad path 1
* greet
  - utter_greet
* inform
  - utter_ask
* recommend_spot
  - action_provide_spots
* introduce_spot
  - action_introduce_spots
  - slot{"spot": null}
* introduce_food
  - action_introduce_food
* insult
  - action_sad
```

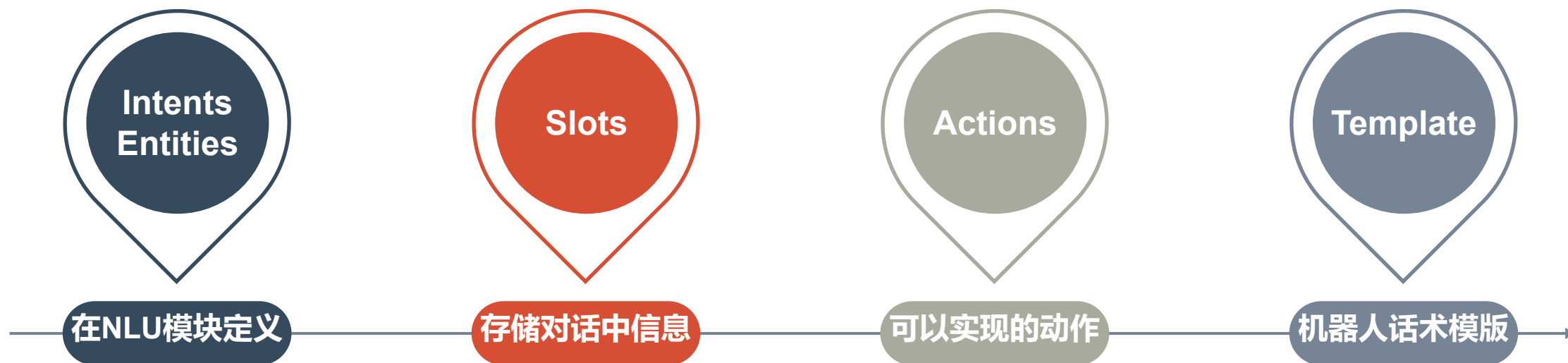
交互学习自动保存增加
训练数据

```
## Generated Story 8535577808949347764
* greet
  - utter_greet
* inform{"province": "广东"}
  - slot{"province": "广东"}
  - utter_ask
* recommend_spot
  - action_provide_spots
* introduce_spot{"spot": "罗浮山"}
  - slot{"spot": "罗浮山"}
  - action_introduce_spots
  - slot{"spot": null}
* introduce_food
  - action_introduce_food
* thankyou
  - action_noworries
```

RASA介绍

Rasa Core：对话管理平台，决定接下来机器人该返回给用户的内容。

Step 2 定义整体操作环境：Domain



```
intents:  
- greet  
- goodbye  
- inform
```

```
entities:  
- province  
- spot  
- name
```

```
slots:  
  province:  
    type: text  
  spot:  
    type: text
```

```
actions:  
- utter_greet  
- action_noworries  
- action_provide_spots
```

```
templates:  
  utter_greet:  
    - text: "主人好"  
    - text: "小主好，今天想去哪个省旅游呢"
```

RASA介绍

Rasa Core：对话管理平台，决定接下来机器人该返回给用户的内容。

Step 3 模型设定：Policy

Keras Policy

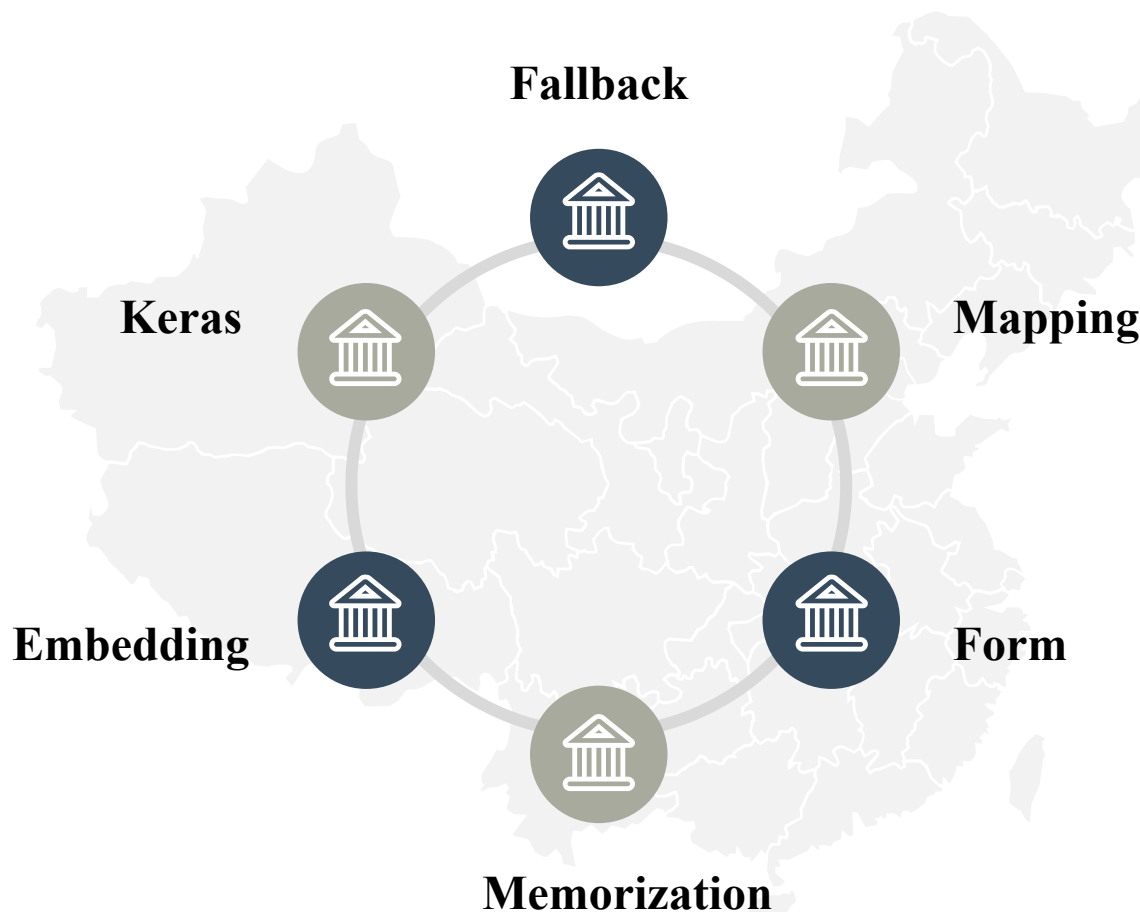
使用神经网络模型进行下一步动作预测，可以重写。

Embedding Policy

使用论文提出的Recurrent Embedding Dialogue Policy。

Memorization Policy

完全基于训练数据记忆，若训练数据没有出现过该对话，则置信度为0。



Fallback

Fallback Policy

如果所有意图的置信度或者所有动作的置信度都低于阈值，则该动作会要求用户进行澄清话术。

Mapping Policy

直接建立意图和动作的映射，一个意图最多对应一个动作。

Form Policy

执行一系列某动作所需要的槽位填写，槽位全部填充完毕，执行动作。

RASA介绍

Rasa Interactive Learning：提供交互学习界面，可以在每一步给机器人提供反馈

```
-----
Chat History

#      Bot      You
-----
1      action_listen
-----
2                                     你好
                                     intent: greet 0.74

Current slots:
      name: None, province: None, spot: None
-----
? The bot wants to run 'utter_greet', correct? (Y/n)_
微软拼音 半 ;
```

RASA介绍

Rasa Interactive Learning：若机器人判断错误，可以直接给予纠正，帮助改善模型

```
? The bot wants to run 'action_listen', correct? Yes
? Your input -> 我想喝水
? Your NLU model classified '我想喝水' with intent 'insult' and there are no entities, is this correct? No
? What intent is it? (Use arrow keys)_
» <create_new_intent>
  0.29 insult
  0.15 greet
  0.11 goodbye
  0.10 unknown
  0.09 introduce_spot
  0.08 thankyou
  0.07 inform
  0.05 introduce_food
  0.03 whatcanyoudo
  0.02 recommend_spot
  0.00 chat
  0.00 oppo
微软拼音 半：
```

RASA介绍

Rasa X：提供交互界面，可以在对话过程中对数据进行标注，可视化对话过程

The screenshot displays the Rasa X interactive learning interface. The main window is titled "Talk to your bot" and shows a conversation history. The conversation starts with a user saying "hi", followed by the bot's response "greet". The user then says "Hi. What are you looking for?", and the bot responds "hospital". The user then says "Please provide your city name.", and the bot responds "Bunnell". The user then says "Sorry, we could not find a hospital in Bunnell.", and the bot responds "Sitka". The conversation history is shown in a list of messages, with the bot's messages in blue and the user's messages in light blue. The interface also includes a "Story till now" section on the right, which shows the current story state. The story state is a list of actions and slots, including "greet", "find_facility_types", "search_provider", "facility_form", "slot", "inform", "find_healthcare_address", and "utter_address". The "Slots" section on the right shows the current slot values: "facility_type" is "rbry-mqwu", "requested_slot" is "null", "location" is "bunnell", and "facility_address" is "not found".

Talk to your bot

Mode: Interactive Learning

action_listen

hi

greet

find_facility_types

Hi. What are you looking for?

action_listen

hospital

search_provider{"facility_type":"rbry-mqwu"}

facility_form

Please provide your city name.

slot{"facility_type":"rbry-mqwu"}

slot{"requested_slot":"location"}

action_listen

Bunnell

inform{"location":"bunnell"}

facility_form

Sorry, we could not find a hospital in Bunnell.

slot{"location":"bunnell"}

slot{"requested_slot":null}

action_listen

Sitka

inform{"location":"sitka"}

find_healthcare_address

Start typing a message...

Story till now

```
## New Story
* greet
- find_facility_types
* search_provider{"facility_type":"rbry-mqwu"}
- facility_form
- slot{"facility_type":"rbry-mqwu"}
- slot{"requested_slot":"location"}
* inform{"location":"bunnell"}
- facility_form
- slot{"location":"bunnell"}
- slot{"requested_slot":null}
* inform{"location":"sitka"}
- find_healthcare_address
- slot{"facility_address":"not found"}
- utter_address
```

Clear Save

Slots

facility_type
rbry-mqwu

requested_slot
null

location
bunnell

facility_address
not found

A wide-angle landscape photograph showing a dark asphalt road that curves through a valley. On the left, a steep, grassy hill rises. On the right, a metal guardrail follows the edge of the road, with more hills and a small town visible in the distance. The sky is filled with soft, golden clouds, suggesting a sunset or sunrise. The overall color palette is warm, with shades of orange, yellow, and brown.

/02

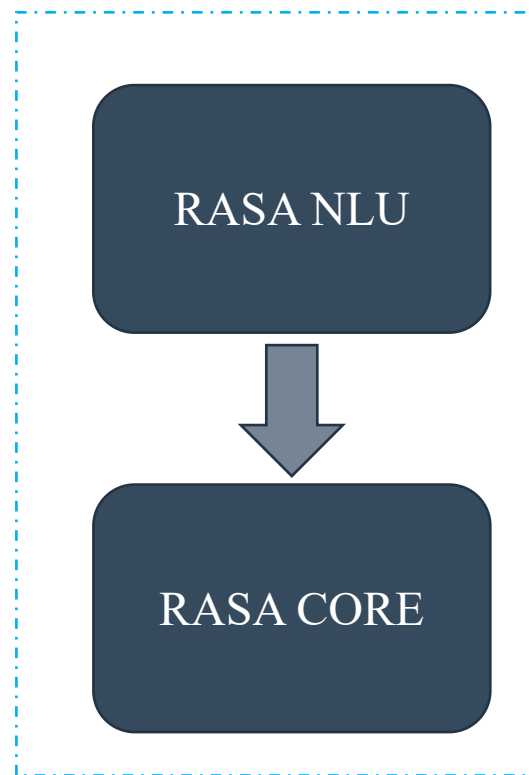
机器人DEMO

机器人DEMO

1. 自定义demo数据
2. 整理各个省份景点名称，美食信息
3. 景点介绍采用实时在线爬虫
4. 预训练词向量
5. 在交互学习中获取更多训练数据

数据准备

交互学习



1. 结巴分词
2. 预训练词向量
3. 命名实体识别模型
4. 意图识别模型

1. MemorizationPolicy
2. KerasPolicy (LSTM)
3. MappingPolicy
4. FallbackPolicy

旅游信息咨询机器人

1. 查询给定省份的5A级旅游景区
2. 介绍旅游景点
3. 推荐各个省份的特色美食

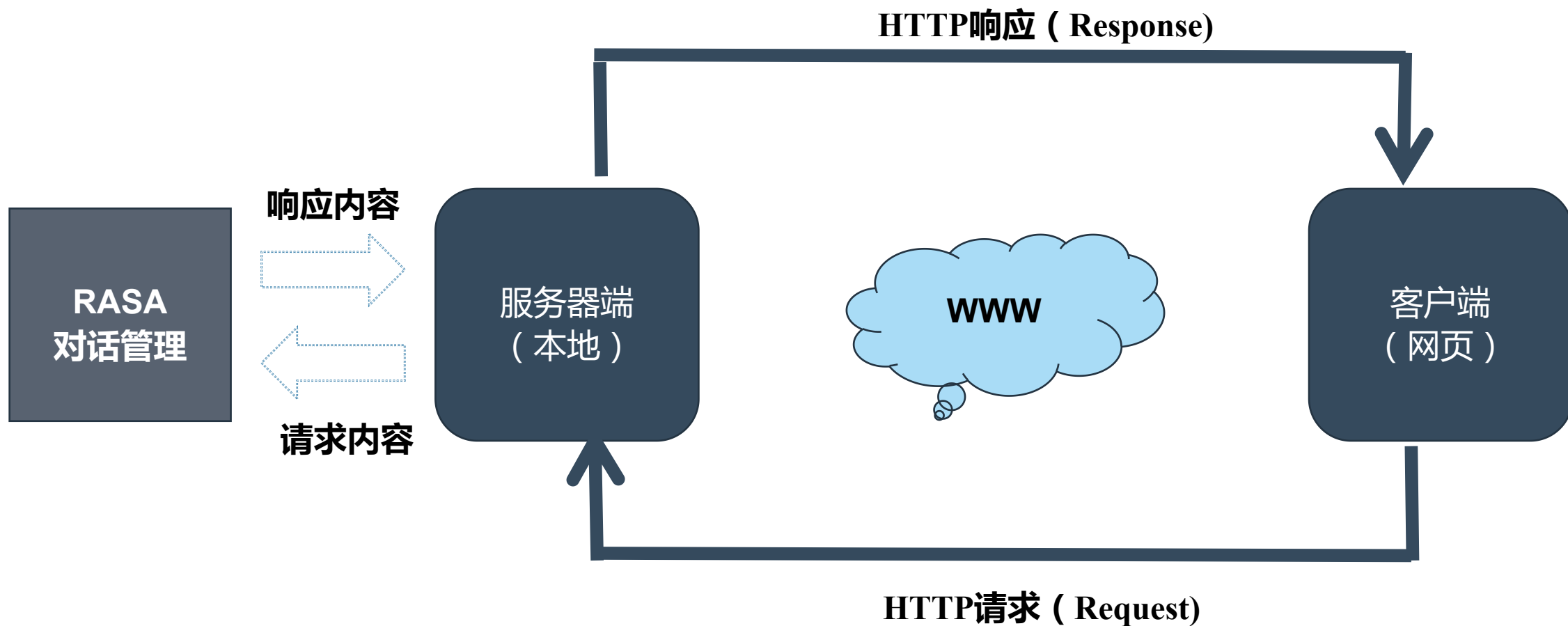


Flask

UI对话界面

机器人DEMO

Flask + Rasa 工作流程



机器人DEMO

笨笨在么



小主好，今天想去哪个省旅游呢

福建



好棒！主人是想知道福建有哪些著名景点还是想查询好吃的呢

有什么好玩的



福建的5A级旅游景区主要有
厦门市鼓浪屿风景名胜区
武夷山
泰宁风景名胜区
福建土楼
白水洋鸳鸯溪
泉州清源山

Type your messages here...

发送

介绍一下太姥山



太姥山，位于福建省东北部。狭义的太姥山系指福鼎市南部秦屿镇以覆鼎峰为中心的山地，也是太姥山风景名胜区的核心地带，在福鼎市正南距市区四十五公里，约在东经120度与北纬27度的附近。挺立于东海之滨，三面临海，一面背山。主峰海拔917.3米。广义的太姥山，则指展布于宁德市东北部交溪以东 - 东海之间的一系列山地，或称太姥山脉。相传尧时老母种蓝（蓝草，其汁色蓝，榨之以染布帛）于山中，逢道士而羽化仙去，故名“太母”，后又改称“太姥”。传说东海诸仙常年聚会于此，故有“海上仙都”的美誉。武夷、太姥、雁荡，构成闽越三大名山。太姥山为交溪与福鼎、霞浦两县所有独流入海的溪流的分水岭。该山脉山体主要由中生代火山岩、花岗岩构成。山体蕴藏的矿产主要有铅、锌、银、镉、明矾石、石英岩、高岭土、玄武岩等。森林覆盖率达46%以上。太姥山风景名胜区，位于太姥山区东北端福鼎市境内，是国家级风景名胜区，同时，太姥山是国内唯一的花岗岩丘陵的地形上发育的峰林地貌的地区，也是国内晶洞花岗岩带上唯一的峰林地貌，因此，也为国家地质公园。

Type your messages here...

发送

有什么好吃的呀



福建的特色美食主要有
泉州四果汤
佛跳墙
漳州鸡仔胎
泉港浮粿
红糟片鸭

你会查询天气么



笨笨没有听懂，麻烦主人重新表述一遍


天气查询



抱歉主人，这些超出我的能力范围了

Type your messages here...

发送

A wide-angle landscape photograph showing a paved road that curves through a valley. On the left, a steep, grassy hill rises. On the right, a metal guardrail follows the edge of the road, with more hills and a small town visible in the distance. The sky is filled with soft, golden clouds, suggesting a sunset or sunrise. The overall tone is warm and contemplative.

/03

总结展望

总结展望



A scenic landscape featuring a winding asphalt road that curves through a valley. To the left, a steep, grassy hill rises, with a few small figures visible on its crest. The sky is a mix of deep blue and warm orange, suggesting a sunset or sunrise, with scattered clouds and a flock of birds in flight in the upper right. The overall mood is peaceful and expansive.

Thanks !

OPPO 技术平台中心

实习生 王晶冰

2019/6/18