



中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

2019—2020学年(春)第二学期
中国科学院大学课程

语音交互技术 ——语音对话系统



中国科学院自动化研究所
模式识别国家重点实验室

陶建华

jhtao@nlpr.ia.ac.cn

章节内容

- 对话系统概述
- 自然语言理解
- 对话管理
- 自然语言生成
- 人机对话评测方法
- 未来发展趋势
- 多模态交互技术
- 本章小结

章节内容

■ 对话系统概述

➤ 对话系统发展历程

➤ 语音对话系统

■ 自然语言理解

■ 对话管理

■ 自然语言生成

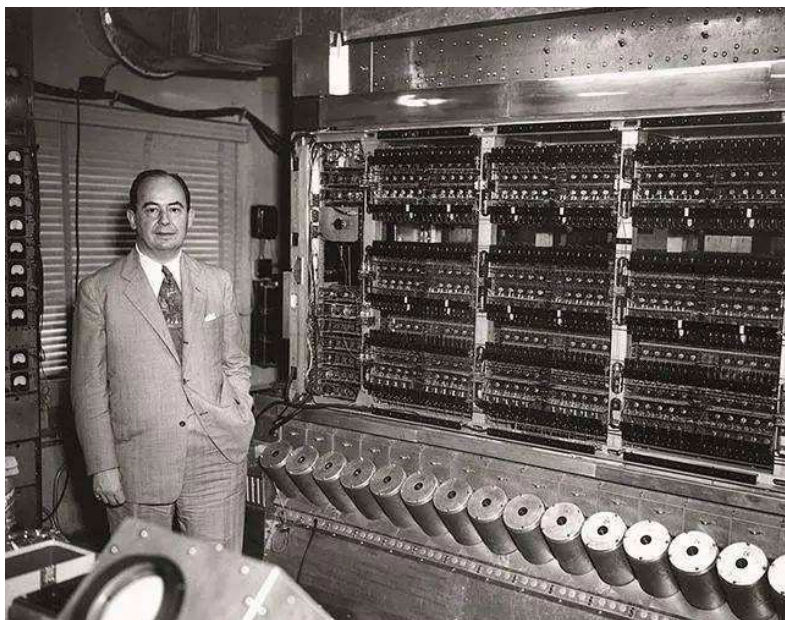
■ 人机对话评测方法

■ 未来发展趋势

■ 多模态交互技术

■ 本章小结

对话系统发展历程



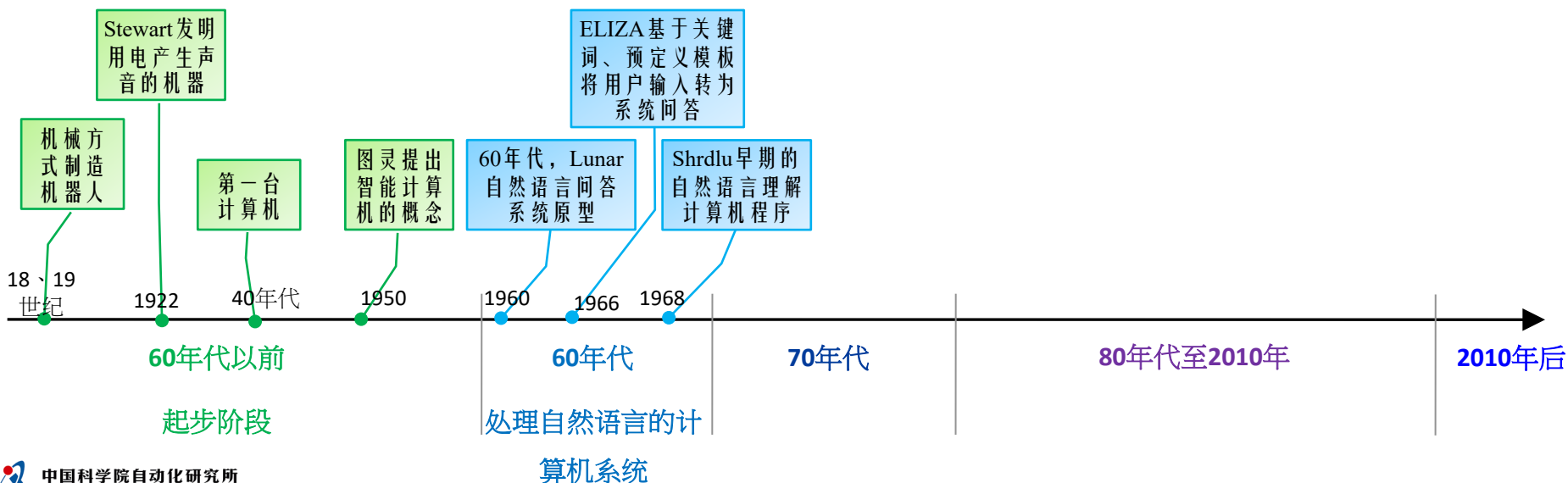
1946年，第一台现代电子计算机ENIAC在美国宾夕法尼亚大学诞生



对话系统发展历程

```
=====
EEEEEEEE L      IIIIII ZZZZZZZ AAA
E         L      I      Z      A  A
E         L      I      Z      A  A
EEEE     L      I      Z      A  A
E         L      I      Z      AAAAAA
E         L      I      Z      A  A
EEEEEEEE LLLLLLLL IIIIII ZZZZZZ  A  A
=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE...!
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====
```

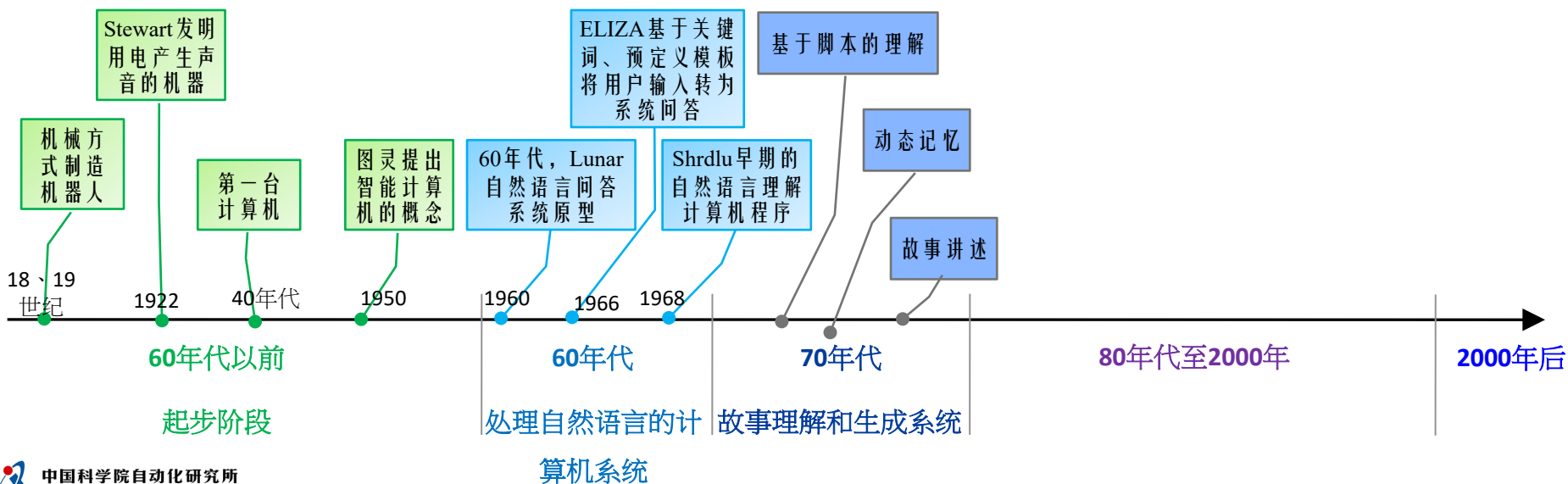
1966年，ELIZA基于关键词提取以及预定义模板将用户输入转换为系统回答，主要用于心理治疗。



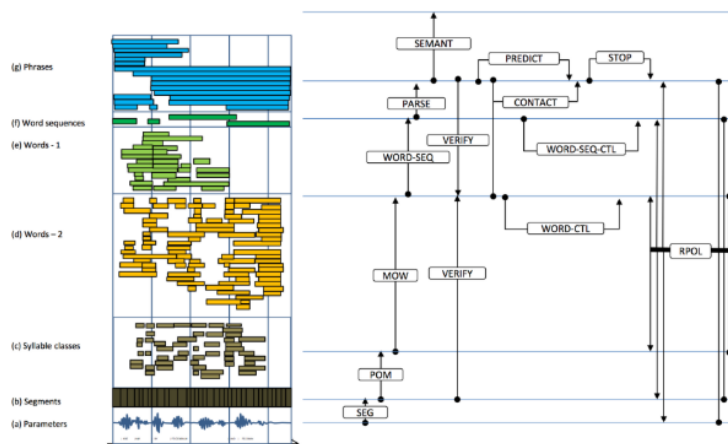
对话系统发展历程

70年代，出现故事理解和生成系统，这些系统可以根据人类话语推断动作、动作执行者和对象。这些系统可以分为不同的类型：基于脚本的理解、动态记忆和故事讲述。其典型系统有：

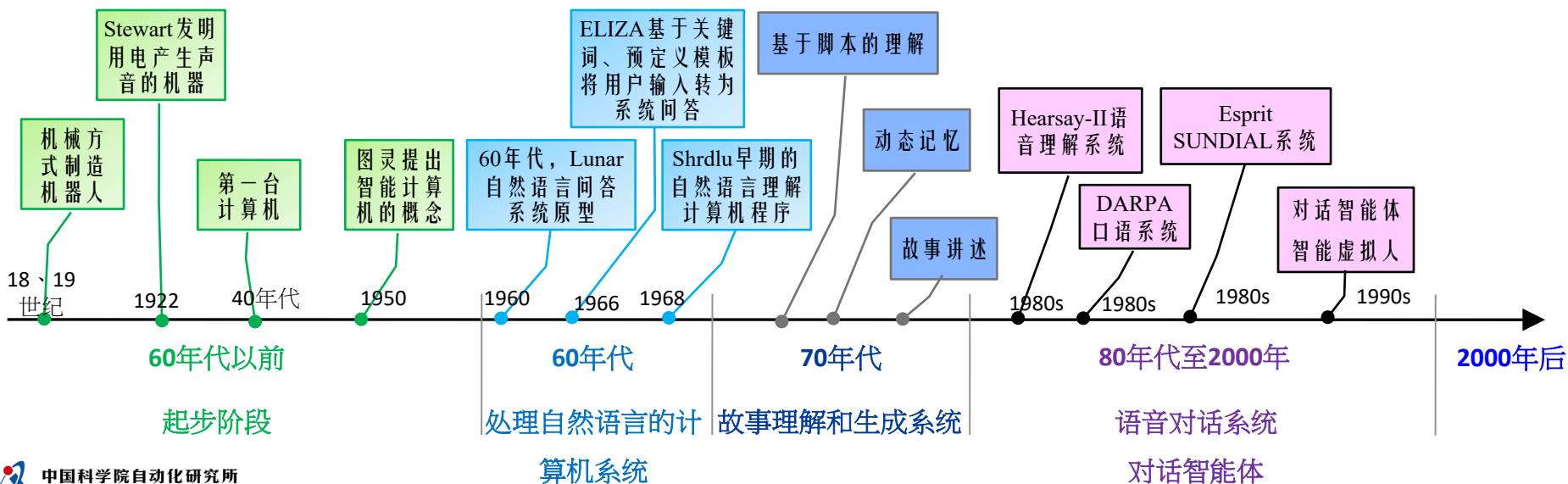
- 基于脚本的理解：SAM、FRUPM
- 动态记忆：IPP、BORIS、CYRUS
- 故事讲述：TALE-SPIN



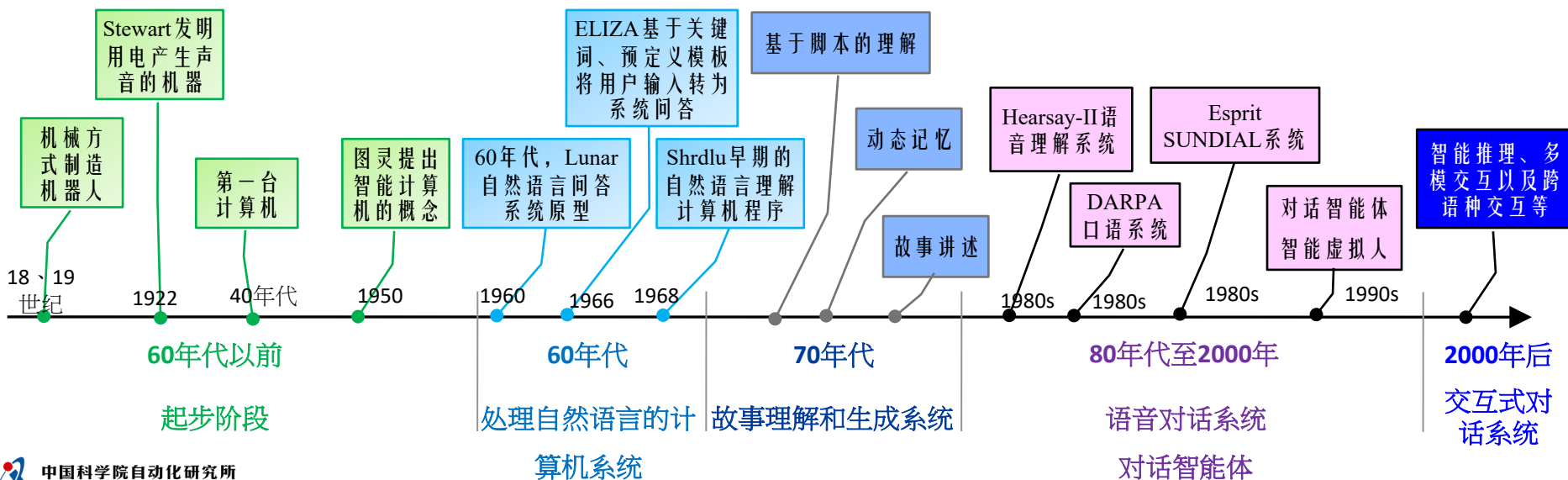
对话系统发展历程



早期的语音对话系统，研究重心在于孤立词语语音识别上，大都结合了线性预测技术，其中较具代表性的是由卡耐基梅隆大学研发的Hearsay-II语音理解系统。



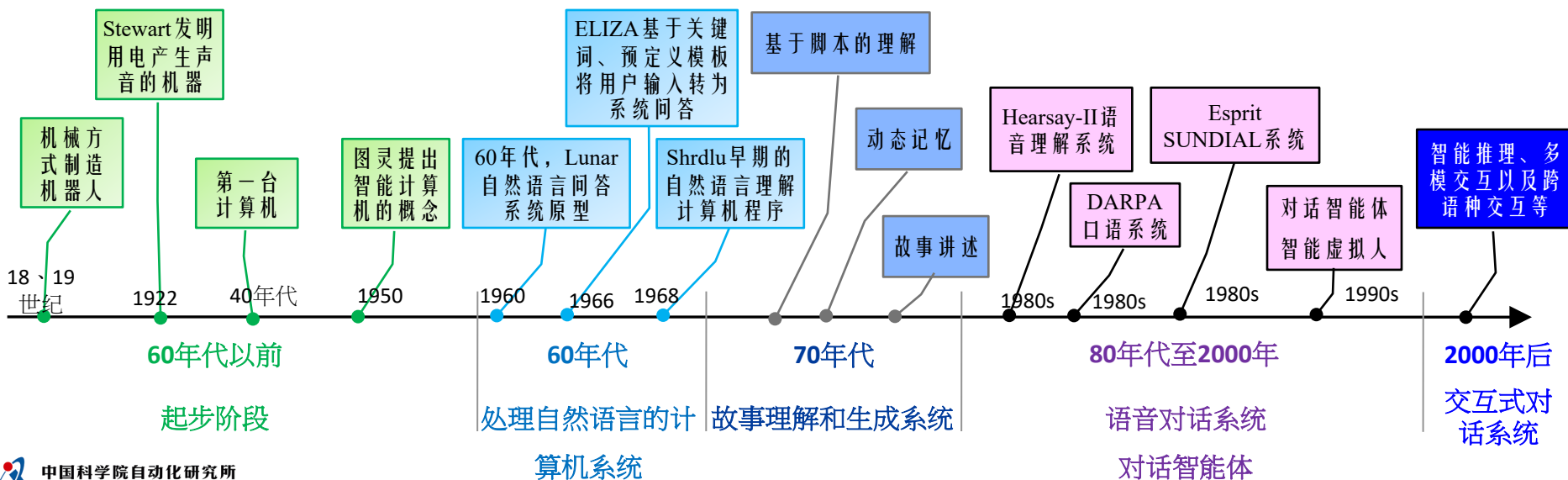
对话系统发展历程



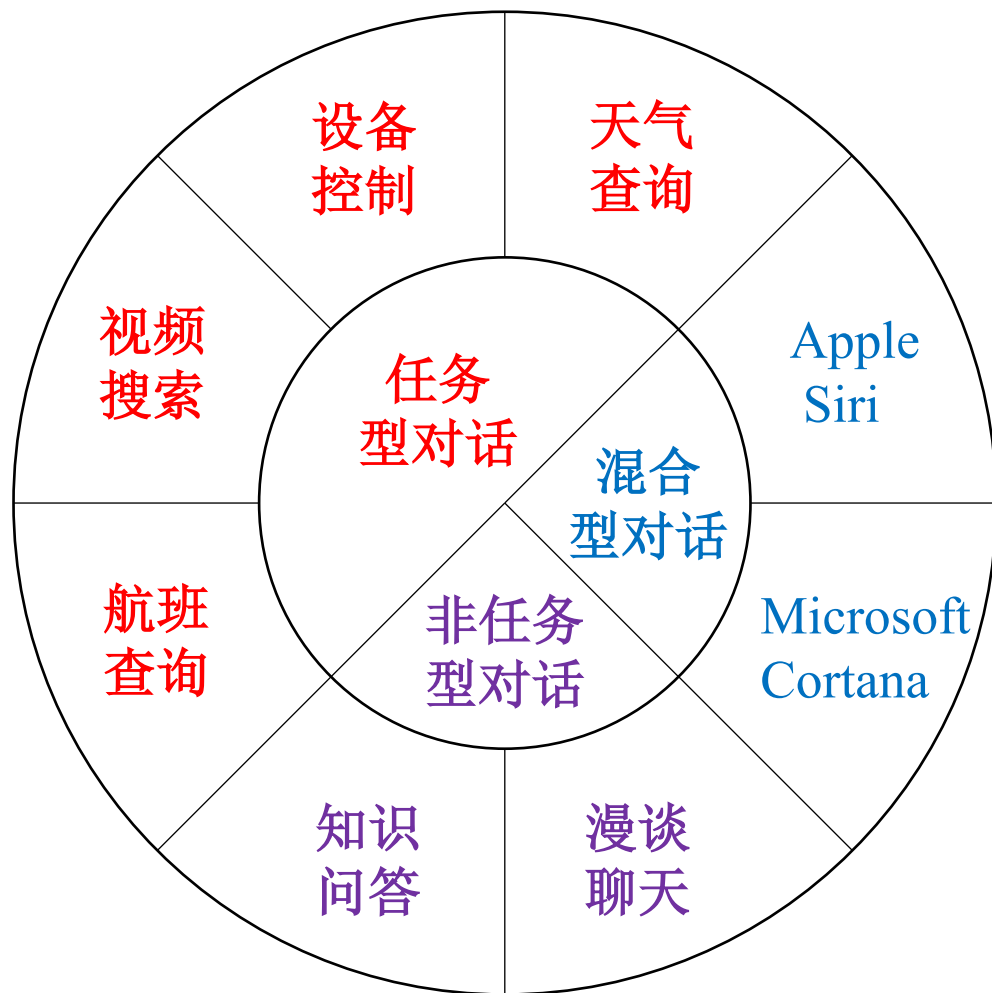
对话系统发展历程

对话系统有着广泛的应用实例：

- 自动旅游旅行信息系统
- 天气预报系统
- 银行系统
- 会议协作系统
- 医疗诊断系统
- 车载系统
- 智能家居
- 个人助手



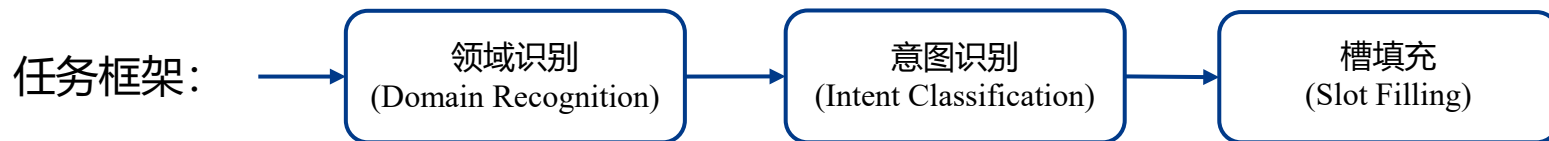
对话系统



章节内容

- 对话系统概述
- 自然语言理解
 - 自然语言理解概述
 - 领域识别
 - 意图识别
 - 槽填充
 - 联合模型
- 对话管理
- 自然语言生成
- 人机对话评测方法
- 未来发展趋势
- 多模态交互技术
- 本章小结

自然语言理解概述



例子： 用户输入“我下周三要在北京开会”，NLU处理后的结果为：

Domain：会议室预定

Intent：提供信息

Slots：{Time：下周三, Location：北京}

自然语言理解概述

任务描述： 给定用户输入，需要从输入中提取出任务相关的信息，包括对话领域、用户意图及部分语义。其中，语义表示由槽及对应的槽值构成。

实现方法： 领域识别与意图识别一般建模成句子分类模型。
槽填充可通过有监督学习和无监督学习实现。其中有监督学习可由分类或序列标注模型完成。目前大部分有监督模型都是以序列标注的形式实现槽填充。
同时，意图识别与槽填充任务可以联合建模，使用一个模型同时完成两个任务。

接下来的部分将从以下几个方面介绍NLU部分：

- 领域识别
- 意图识别
- 槽填充
- 联合建模

领域识别

任务描述： 领域识别是自然语言理解的前导任务，指不同的信息服务任务，如餐馆推荐和会议室预订就属于两个不同的领域。在不同领域下，系统根据任务需求可能会设计不同的意图种类和不同的槽。

实现方法： 领域识别一般被建模成句子分类模型。

- 基于规则的方法（CFG、JSGF）
- 基于统计学的方法（SVM、ME）
- 基于深度学习的方法（CNN, RNN, LSTM, RCNN, C-LSTM, FastText）
- 基于预训练模型的方法（BERT、GPT）

意图识别

任务描述： 在识别出领域信息后，机器需要对用户的意图进行判断。比如在会议室预定领域中，用户输入了一句话，他是想预定会议室，还是在仅仅咨询相关信息？他确认了预定结果，还是取消了预定结果？等等。

实现方法： 意图识别一般也被建模成句子分类模型。用户的输入可以有多个意图，多意图识别可抽象成多标签分类任务，而单意图识别则可抽象成多分类任务。

- 多标签分类

Methods for Multi-label Classification

- 单标签多分类

传统的基于句子的分类方法包括SVM、MaxEnt等。随着神经网络的发展与应用，分类任务中逐渐用神经网络来提取数据特征，例如DBN、RNN、LSTM等。

槽填充

任务描述： 槽填充的目的是在用户输入中提取与任务相关的重要信息，即槽信息，其实质是与领域任务相关的语义标签值。

实现方法：

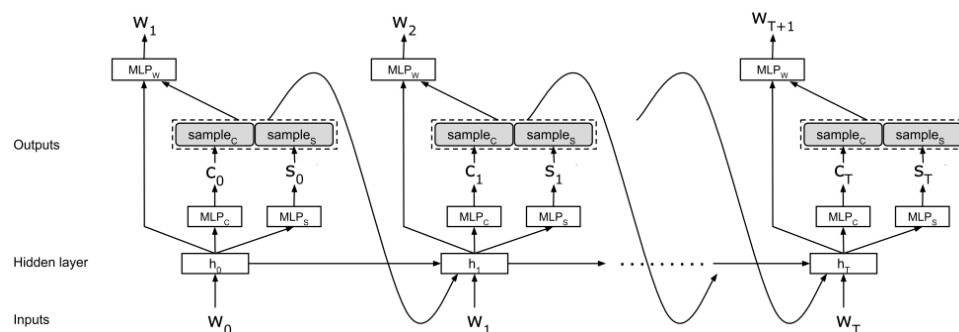
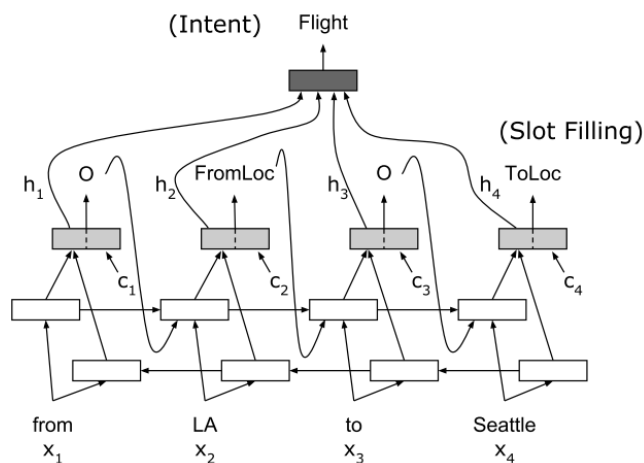
- 基于规则的方法
- 分类模型
- 一般通过序列标注模型完成。传统的序标模型包括CRF、HMM等。随着深度神经网络的兴起，各式各样的基于神经网络的模型涌现了出来。

联合模型

联合原因： 一方面，后面子任务的性能会因为前面子任务的误差逐步累积放大而受到严重影响，比如，错误的意图识别会造成对槽的误判；另一方面，管道式方法只能反映子任务间的单向约束，而无法建模子任务间的交互约束信息，如：后续的槽填充任务会受到先导的意图识别任务的影响，但后续任务对先导任务的影响却无法反传。

实现方法： 联合模型将意图识别与槽填充任务联合起来，用同一个模型完成两个任务：

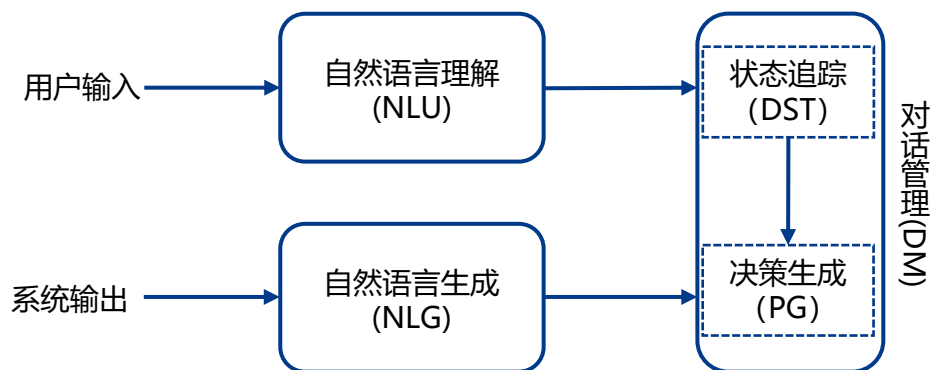
- 将意图接在输入序列后，整体作为一个序列标注任务，
- 将意图和序列标注分开，在每一个时刻输出槽标注结果的同时，利用句子编码对输入进行意图分类。



章节内容

- 对话系统概述
- 自然语言理解
- 对话管理
 - 对话管理概述
 - 基于有限状态自动机的方法
 - 基于槽填充的方法
 - 基于信息状态更新的方法
 - 基于马尔科夫决策过程的方法
 - 知识库
- 自然语言生成
- 人机对话评测方法
- 未来发展趋势
- 多模态交互技术
- 本章小结

对话管理概述

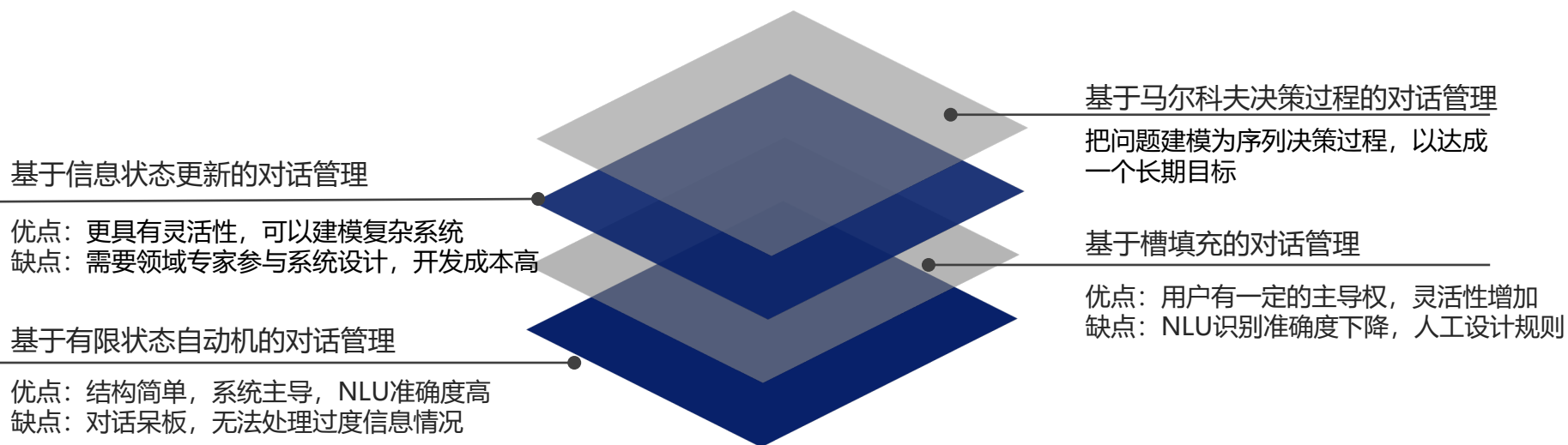


对话管理是对话系统的核心，负责引导人机对话过程，帮助用户高效自然地完成对话，其设计的优异性直接关系到整个对话系统的性能。对话管理可分为两个子部分：状态追踪（State Tracking, ST）和决策生成（Policy Generation, PG）。

- 对话状态追踪：对话状态追踪定义了对话状态的表示形式，随着对话的进行，更新每个时刻对话状态的信念分布；任务型人机对话的对话状态由若干个槽及其槽值组成。
- 决策生成：基于对话所处的状态，决策生成在用户与对话系统的交互中采取合适合理的动作，促使交互有序进行，最终为用户提供有价值的信息。

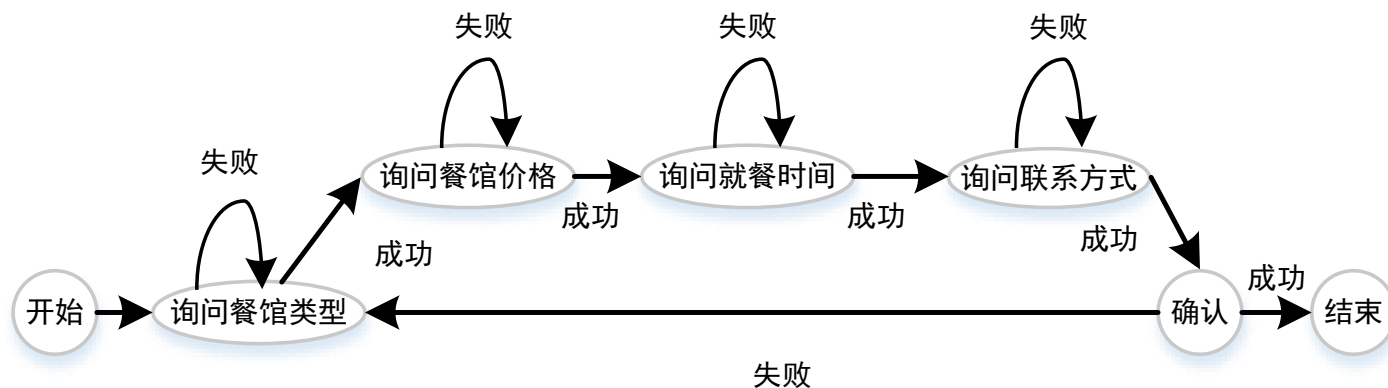
对话管理概述

用于建模对话管理的方法：基于有限状态自动机的对话管理，基于槽填充的对话管理，基于信息状态更新的对话管理，基于马尔科夫决策过程的对话管理.....



基于有限状态自动机的方法

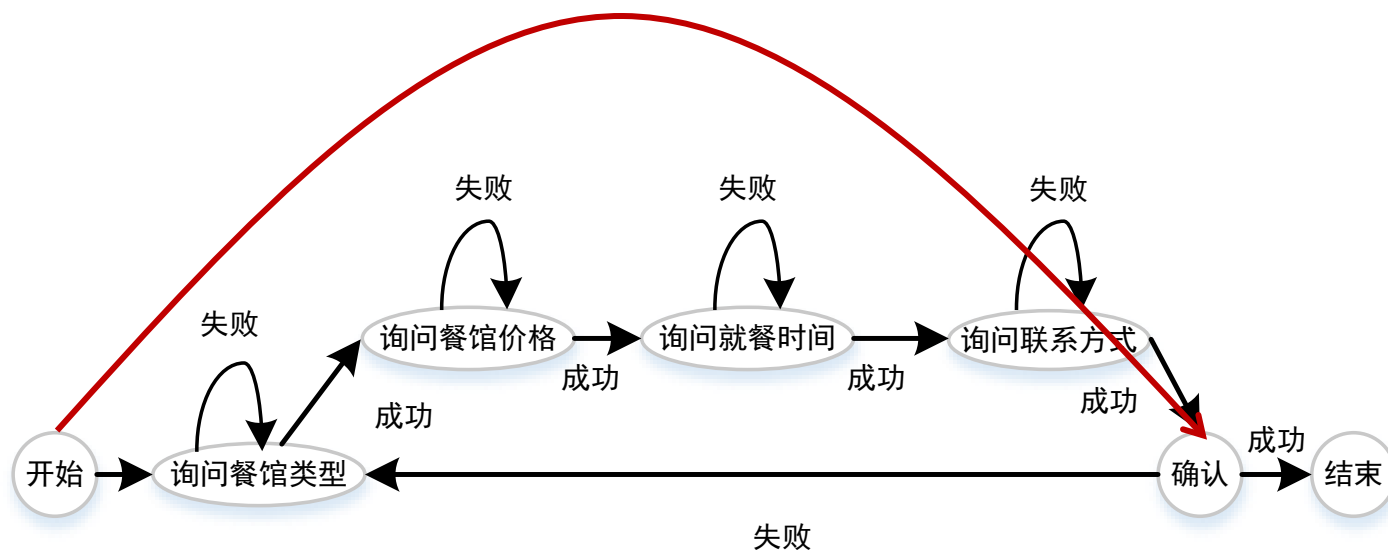
有限状态自动机是建模在有限个状态之间的数学模型，用于描述状态间的转移，是一种最基础的对话管理技术。



订购餐馆状态转移图

基于有限状态自动机的方法

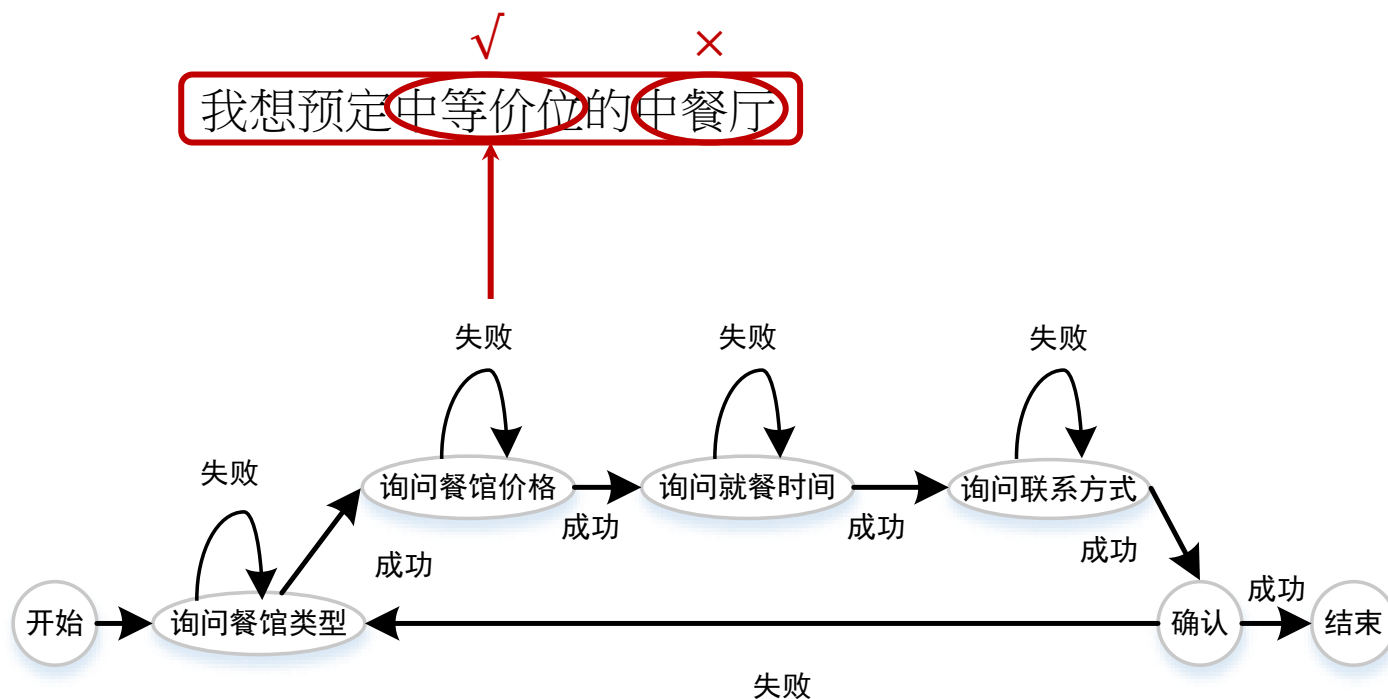
有限状态自动机是建模在有限个状态之间的数学模型，用于描述状态间的转移，是一种最基础的对话管理技术。



订购餐馆状态转移图

基于有限状态自动机的方法

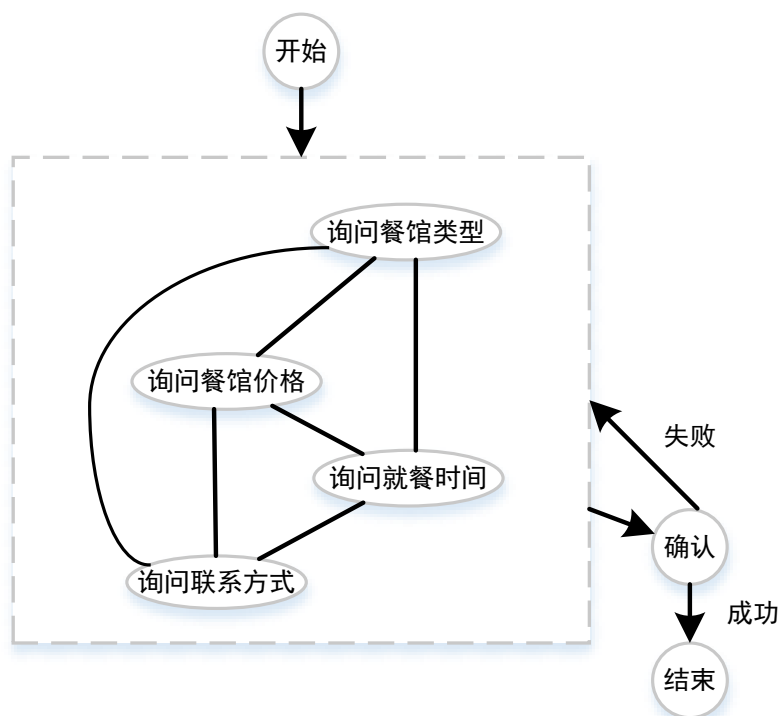
有限状态自动机是建模在有限个状态之间的数学模型，用于描述状态间的转移，是一种最基础的对话管理技术。



订购餐馆状态转移图

基于槽填充的方法

基于槽填充的对话系统的对话情况是由一个多维特征向量来记录的，该向量的值会在对话的过程中进行不断填充和修改。特征向量的元素表示的是对话系统对用户的意图进行分析后得到的信息和一些状态标志，对话管理根据特征向量的填充情况对对话的过程进行控制。



订购餐馆的槽结构

初始状态:

类 型	价 格	时 间	电 话
Null	Null	Null	Null

基于槽填充的方法

- 系统：您好，我们提供餐馆预定服务，您需要什么样的餐馆呢？
- 用户：我要订一个中等价位的中餐厅。
- 系统：好的，请问要订哪一天餐馆呢？
- 用户：后天中午。
- 系统：后天中午中等价位的中餐厅。您的联系方式是？
- 用户：XXXX。
- 系统：您订的是后天中午中等价位的中餐厅，电话是XXXX，对吗？
- 用户：对的。
- 系统：好的，您的订单已生成。谢谢您的惠顾，再见。

初始状态：

类 型	价 格	时 间	电 话
Null	Null	Null	Null

填充：

类 型	价 格	时 间	电 话
中 餐 厅	中 等	Null	Null

基于信息状态更新的方法

基于信息状态更新的对话系统在进行对话决策时用到更丰富的上下文信息。基于信息状态更新的对话系统包括信息部件、信息部件的形式化表示、对话动作、对话规则和对话更新策略。

- **信息部件**：信息组件包含了与用户交互过程中需要用到的组件。组件可以分成两类，第一类是私有（PRIVATE）组件表示系统私有的信息的，包括系统在对话过程中执行动作的计划（PLAN）；第二类是共享（SHARED）组件表示系统与用户对话过程中获得的信息，包括共享的槽信息（SLOT）、当前讨论的问题（Question Under Discussion, QUD）和最近一个执行对话动作（Latest Dialogue Move, LM）。
- **形式化表示**：形式化表示信息组件。

PRIVATE	:	PLAN	:	[Q1, Q2, ..., QN]
		SLOT	:	[Slot1=val, ..., SlotN=val]
SHARED	:	QUD	:	Question
		LM	:	Move

基于信息状态更新的方法

- **对话动作**：对话动作指对话系统可以发送和接受的消息类型，用于触发更新信息状态。这里只考虑提问（ask）和回答（answer）两种类型的动作。
- **更新规则**：更新规则形式化描述了信息状态在对话过程中的变化方式，每个规则都由规则执行条件和规则执行效果两部分组成。这里考虑三种规则：

```
U-RULE: integrateSysAsk  
  
PRE {  
    value(SHARE.LM, ask(usr,Q))  
    first(PRIVATE.PLAN, raise(Q))  
}  
  
EFF {  
    set(SHARE.QUD, Q)  
    pop(PRIVATE.PLAN)  
}
```

规则一：最近执行的动作是ask并且PLAN的第一个动作是提问问题，那么设置QUD为PLAN的第一个问题，并且从PLAN中删除第一个问题。

基于信息状态更新的方法

U-RULE: selectAsk

PRE { first(PRIVATE.PLAN, raise(Q))
EFF { set(SHARE.LM, ask(usr,Q))

规则二：选择PLAN的第一个问题进行提问。

U-RULE: integrateUserAnswer

PRE { value(SHARE.LM, answer(usr,A))
value(SHARE.QUD, Q)
DOMAIN::resolve(A, Q)
EFF { add(SHARE.SLOT, Q(A))
unset(SHARE.QUD)

规则三：判断用户的回答是否解决了QUD的问题。如果解决了问题，那么就把这个问题的答案记录到SLOT中，然后清空QUD。

基于信息状态更新的方法

- **更新策略：**更新策略描述的是系统按什么样的策略来应用更新规则，例如规则执行的顺序。

```
PRIVATE : [ PLAN : [raise(room-type(x)), raise(num(x)),  
                    raise(time(x)), raise(phone(x))] ]  
SHARED  : [ SLOT : []  
            QUD  :  
            LM   : ] ]
```

U-RULE : selectAsk

EFF : { set(SHARE.LM, ask(room-type(x)))

R: 你好, 请问你需要什么类型的房间

U-RULE : integrateSysAsk

EFF { set(SHARE.QUD, room-type(x))
 pop(PRIVATE.PLAN)

```
PRIVATE : [ PLAN : [raise(num(x)), raise(time(x)),  
                    raise(phone(x))] ]  
SHARED  : [ SLOT : []  
            QUD  : room-type(x)  
            LM   : ask(room-type(x)) ] ]
```

U: 需要双人房(这句话会使得LM变成answer(room-type(双人房)))

U-RULE : integrateUserAnswer

EFF { add(SHARE.SLOT, room-type(双人房))
 unset(SHARE.QUD)

```
PRIVATE : [ PLAN : [raise(num(x)), raise(time(x)),  
                    raise(phone(x))] ]  
SHARED  : [ SLOT : [room-type(双人间)]  
            QUD  :  
            LM   : answer(room-type(双人间)) ] ]
```

U-RULE : selectAsk

EFF : { set(SHARE.LM, ask(time(x)))

R: 请问您要哪天入住呢?

基于马尔科夫决策过程的方法

马尔科夫决策过程 (Markov Decision Process, MDP) 用来建模连续决策性问题, MDP是从马尔科夫链演变而来, 具有以下几个特性:

- 马氏性 (Markov Property, 也称无后效性): 当前时刻的状态只和上一个状态有关, 和其它的历史状态无关。

$$p(s_{t+1} = s' | s_t, a_t, \dots, s_0, a_0) = p(s_{t+1} = s' | s_t, a_t)$$

- 状态转移不确定性: 在MDP中, 状态之间的转移概率可以具有不确定性, 可以有概率的转移到不同的新状态, 这些转移的概率组成了状态转移矩阵。
- 状态完全可观: MDP应用的具体问题, 其状态的组成元素是完全可观测的, 也就是说, 模型的输入是整个系统的内部情况表示, 状态是完全可观的。

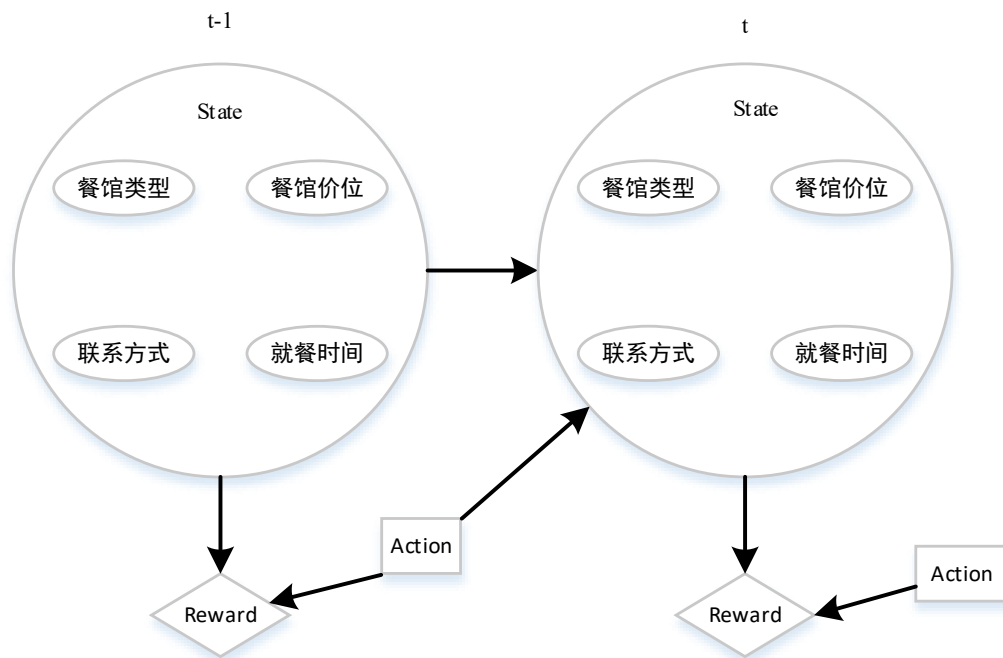
基于马尔科夫决策过程的方法

一般情况下, (S, A, P, R, γ) 使用五元组来表示MDP, 其中:

- 状态集合 S (state), 指目标系统的所有可能的状态的集合, 表示在 t 时刻系统所处的状态;
- 动作集合 A (action), 指的是系统可以执行的所有动作的集合, 也是状态之间的转换行为, 表示在 t 时刻系统执行的动作;
- 状态转移概率矩阵 P , $P_a(s, s') = P(s_{t+1} = s' | s_t = s, a_t = a)$ 表示在 t 时刻下, 系统处于状态 s_t , 执行动作 a_t 后转移到新状态 s_{t+1} 的概率;
- 回报函数 R , 通常使用 $R(s, a, s')$ 来表示执行动作 a 后, 从状态 s 转移到状态 s' 获得的即时回报, 使用 r_t 来表示在 t 时刻得到的即时回报;
- 折扣因子 γ , $\gamma \in [0, 1]$, 通过折扣因子来衡量即时回报和长期回报的重要性, 同时也是用来避免计算结果趋于无穷大。

基于马尔科夫决策过程的方法

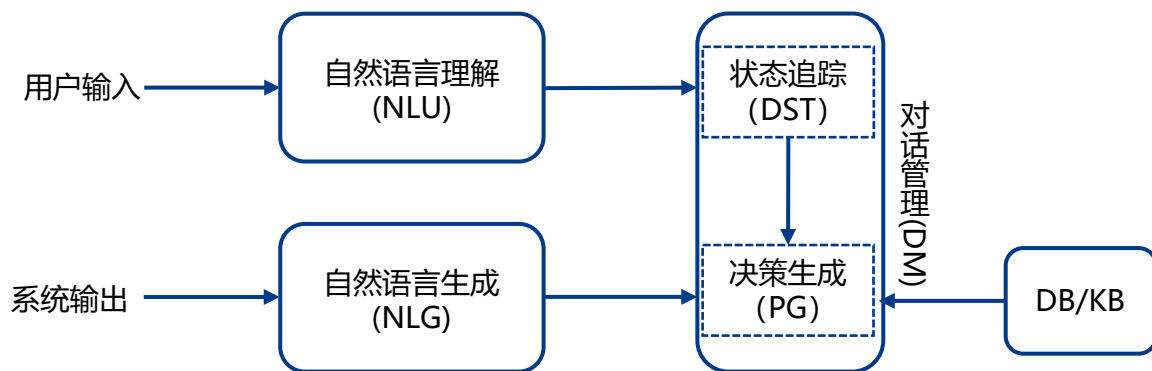
马尔可夫决策过程是一种可以应用在对话管理系统上的概率模型，它学习决策规则的方式是通过分析对话数据，在一定程度上克服人工定义规则与覆盖率不够的问题。使用MDP模型建模对话系统，然后通过强化学习技术求出最优策略。



订购餐馆的MDP模型

知识库

为了进行准确、自然、智能、友好的人机交互，在获得对话的状态之后可以查询知识库来进行智能决策，以完成更复杂的任务。



知识库

名称	开始时间	依赖资源	规模# (实体/概念/关系/事实)
Cyc/OpenCyc	1984	专家知识	239,261/116,822/18,014/2,093,000
WordNet	1985	专家知识	155,287/117,659/18/-
ConceptNet	1999	群里智能 (多语言)	-/8,000,000/36/21,000,000
YAGO	2007	WordNet+Wikipedia	4,595,906/488,469/77/≈40M
DBpedia	2007	Wikipedia+ 专家知识	17,315,785/754/2843/79,030,098
Freebase	2008	Wikipedia+领域知识 + 群体智能	58,726,427/2,209/39,151/3,197,653,841
NELL	2010	机器学习	-/287/327/2,309,095
BabelNet	2012	WordNet+Wikipedia (多语言)	9,671,518/6,117,108/1,307,706,673/-
Wikidata	2012	Freebase 群体智能	45,766,755/-/-/-
Google Knowledge Graph	2012	基于Freebase	570M/1500/35000/18000M
Knowledge Vault	2014	机器学习	45M/1100/4469/271M

常见的部分大规模知识库及其规模

章节内容

- 10.1 对话系统概述
- 10.2 自然语言理解
- 10.3 对话管理
- 10.4 自然语言生成
 - 自然语言生成概述
 - 模板式方法
 - 基于神经网络的方法
- 10.5 人机对话评测方法
- 10.6 未来发展趋势
- 10.7 多模态交互技术
- 10.8 本章小结

自然语言生成概述

- 自然语言生成任务可以视为自然语言理解任务的逆过程，即以要表述的语义项作为输入，以自然语言描述作为输出。
- 人机对话系统中的自然语言生成任务的目标包括了：语言流畅、语义准确、可扩展性、表达多样性。
- 根据系统最后产生回复的机制，自然语言生成主要有模板式、基于数据驱动的方法。

模板式方法

模板生成方式是最早采用的一种生成技术，要根据不同任务的系统设计出不同的模板来处理对应的可能出现的情况。当用户输入结构化的语义表示后，可以通过后台的模板将用户输入的结构化语义表示作为字符串嵌入到模板中。

例： 输入：confirm（route=10号线）
 模板：你是想了解value的信息吗？
 输出：你是想了解10号线的信息吗？

 模板：是关于value的信息吗？
 输出：是关于10号线的信息吗？

模板式方法

优点：

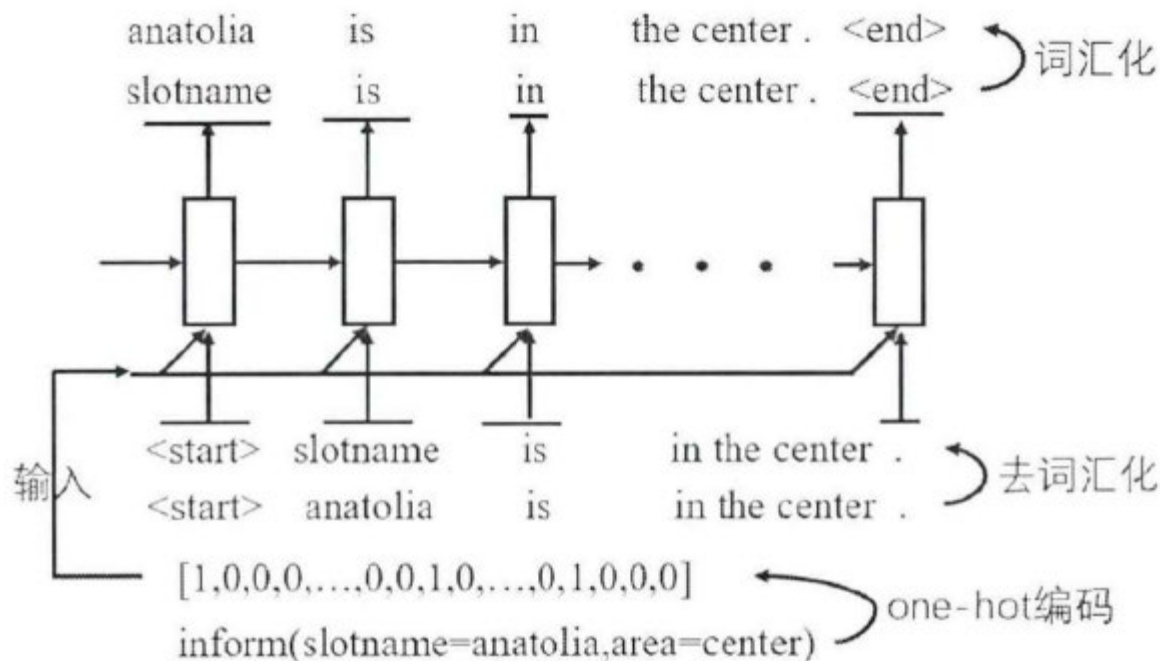
- 思路简单，较容易实现；
- 模板匹配的速度较快，工作效率高；
- 生成的回复不会产生语义层面问题，能够在一些轻度应用中发挥预期的效果。

缺点：

- 无法做到全覆盖，随着模板增多，后期维护的工作较为困难；
- 生成语句表达单一、不够灵活，难以满足人们各种各样的要求；
- 需要领域专家编制文档结构，或是需要语言学家编制语法规则，极大提高了基于规则的语言生成系统的构建成本；
- 在一个领域编制的规则，往往极难适应于另一个领域，极大限制了基于规则的语言生成系统的通用性和可移植性。

基于神经网络的方法

随着可获取的训练数据越来越丰富，以及计算机计算能力的不断增长，近些年完全数据驱动的基于神经网络的生成方法受到了关注。在多种神经网络中，循环神经网络因为其具有解决上下文依赖的能力被广泛的应用在自然语言生成领域。



基于LSTM的自然语言生成

章节内容

- 10.1 对话系统概述
- 10.2 自然语言理解
- 10.3 对话管理
- 10.4 自然语言生成
- 10.5 人机对话评测方法
 - 任务型对话系统评价方法
 - 开发域对话系统评价方法
- 10.6 未来发展趋势
- 10.7 多模态交互技术
- 10.8 本章小结

人机对话评测方法

基于任务型和开放式两种对话系统类型，研究者们针对性地提出了各种不同的评测方法，主要分为客观指标和主观指标。



任务型对话系统评价方法

- 通过构造用户模拟系统进行评价

用户模拟是一种有效且简单的评价策略，并且是最有可能覆盖最大对话空间的评价策略。通过模拟不同情境下的对话，可以有效的在大范围内进行测试和评价。

- 人工评价

通过雇佣测试人员对对话系统生成的结果进行人工评价，这样做的好处是能够产生更多真实的评价数据。

- 在公开部署的系统中进行评价

在商业推广中植入对话系统或构建一个能够让公众主动去使用对话系统的服务设施。

非任务型对话系统评价方法

- 基于词重叠率的评价矩阵

根据已有的NLP任务经验，当提到如何评价模型生成结果的质量时，可以根据生成的回复与标准答案之间的词重叠率来进行评价。常见指标有: BLEU、METEOR、ROUGE等。

- 基于词向量的评价矩阵

通过词向量矩阵来判断回答的相关性。依据语义分布，采用Word2Vec等方法给每个词分配一个向量，所有的词向量矩阵通过向量连接就可以近似为句子级的句向量，通过这种方法可以分别得到候选回复句与目标回复句的句向量，再通过余弦距离进行比较，就可以得到二者的相似度。

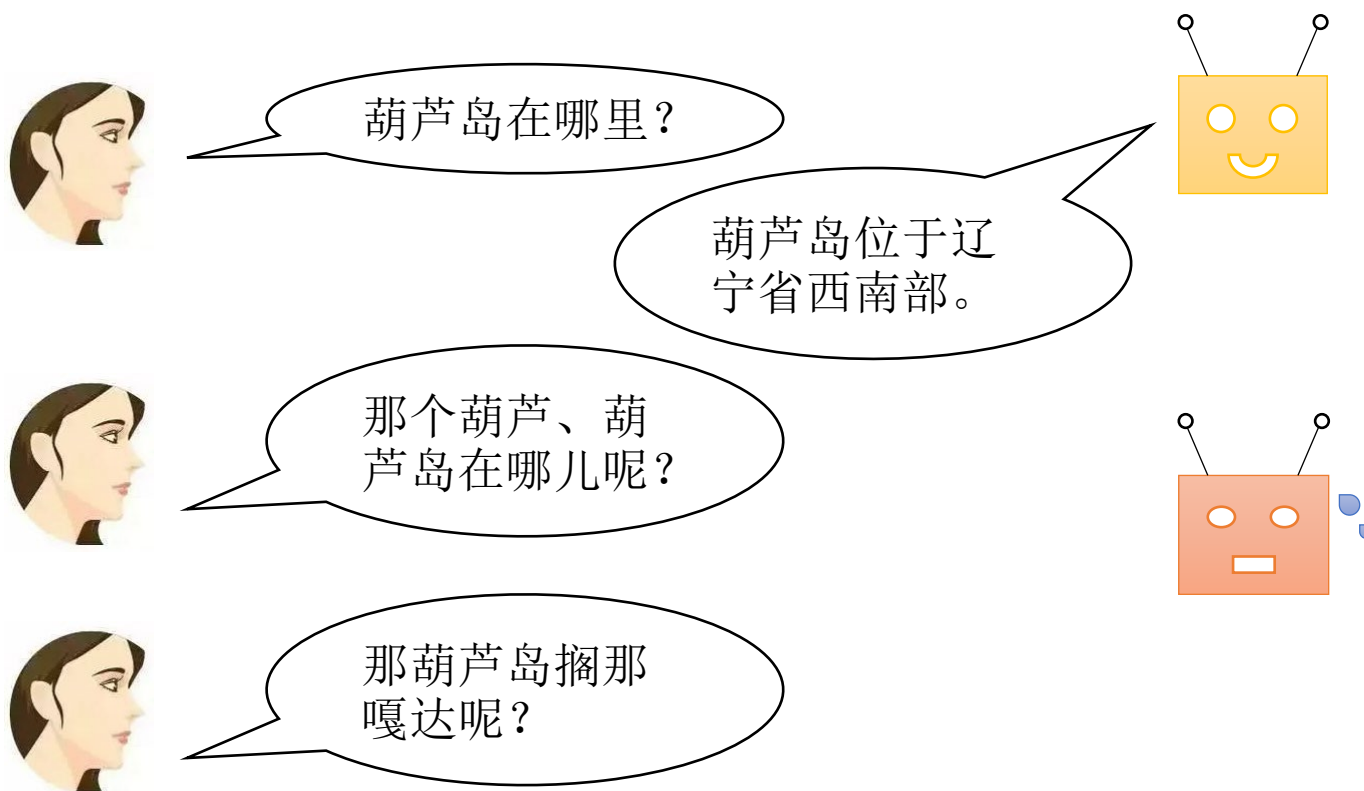
- 基于评分模拟的评价方法

研究人员提出了一种类 GAN 结构的对抗评价模型，直观评价生成器产生的回复结果与人类回复的相似程度；使用RNN的方法进行自动评分模型的训练，进而提出了一个对话系统自动评价模型用于预测回复的人工评价结果等。

章节内容

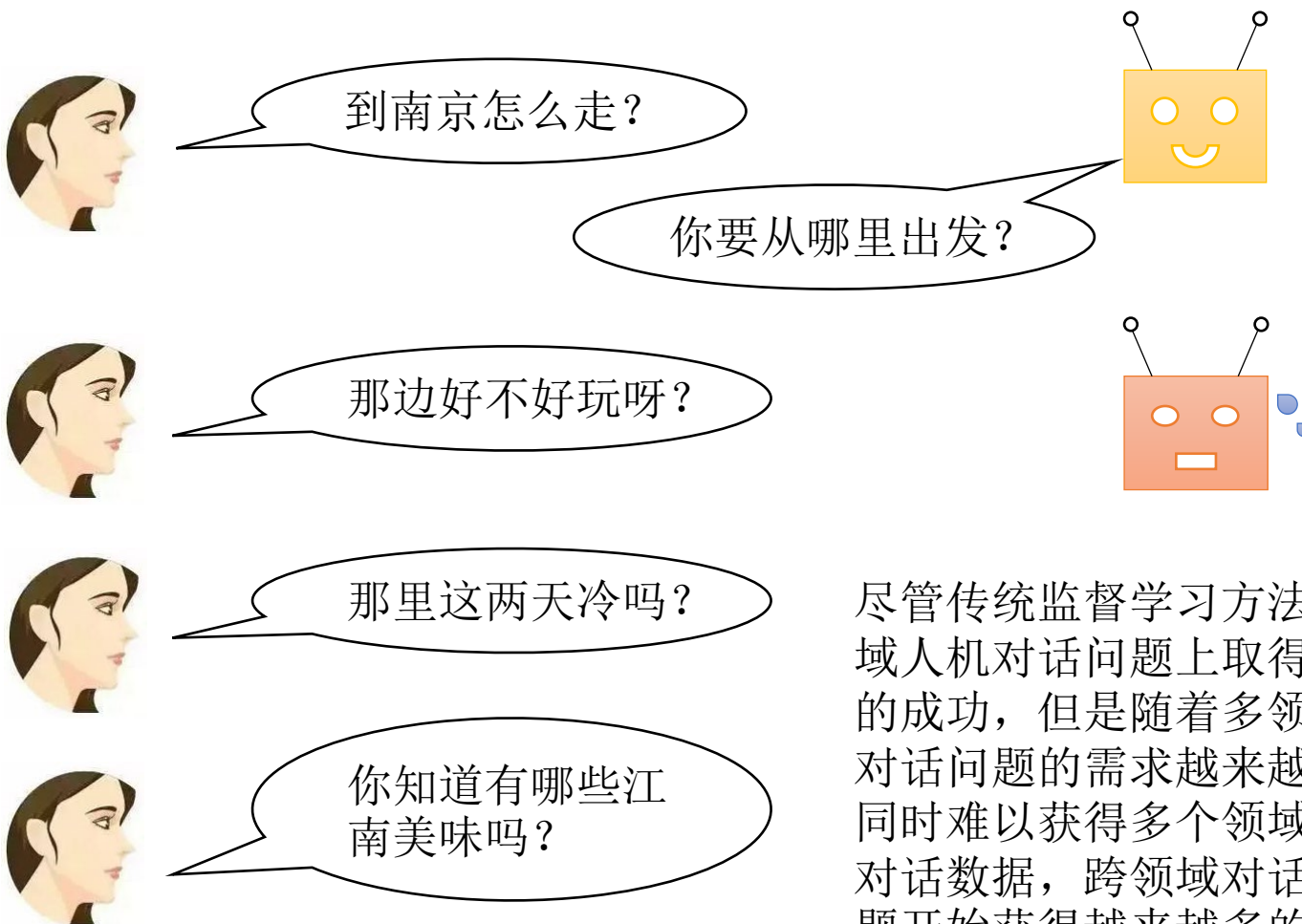
- 10.1 对话系统概述
- 10.2 自然语言理解
- 10.3 对话管理
- 10.4 自然语言生成
- 10.5 人机对话评测方法
- 10.6 未来发展趋势
 - 复杂口语下的对话系统
 - 跨领域人机对话
 - 自主学习与知识融合
 - 多模态人机交互
- 10.7 多模态交互技术
- 10.8 本章小结

复杂口语下的对话系统



现有的语音识别技术还无法完全准确识别自然口语，特别是对于带有口音方言和远场条件下语音识别的低准确率极大降低了用户对问答系统的体验度。另外，对话过程中副语现象也给对话系统带来了很大的挑战，需要从语音识别和语义理解两方面入手改进。

跨领域人机对话



尽管传统监督学习方法在单领域人机对话问题上取得了较大的成功，但是随着多领域人机对话问题的需求越来越显著，同时难以获得多个领域的有标对话数据，跨领域对话理解问题开始获得越来越多的关注。

自主学习与知识融合

小白养成计划

嘿,小白,你知道我是谁吗

不知道你的名字呀,要不你告诉我好吗

我是西门吹雪

嗯,原来你的名字是西门吹雪呀

你知道吗,西门吹雪身高两米四哦

我靠,居然比姚明还高!

他拿过几次NBA总冠军呀

姚明从未拿过NBA总冠军呀

那姚明的膝盖受过多少次伤,伤的有多严重呢?

这种问题.....你去问百度和知乎不是更方便吗?

那你觉得我可能擅长做什么

可能擅长打篮球,因为主人身高非常高

不对,你应该说,主人明明可以靠脸吃饭,却偏偏要靠才华

哦,好吧,记住了

我可能擅长做什么呀

主人明明可以靠脸吃饭,却偏偏要靠才华

属性询问

反问

记忆

反馈

记忆

评论,跨库属性对比—表情:惊讶

上下文

事实类知识图谱查询

对复杂问题,进行CQA导流—表情:鄙视

推理、联想

调教

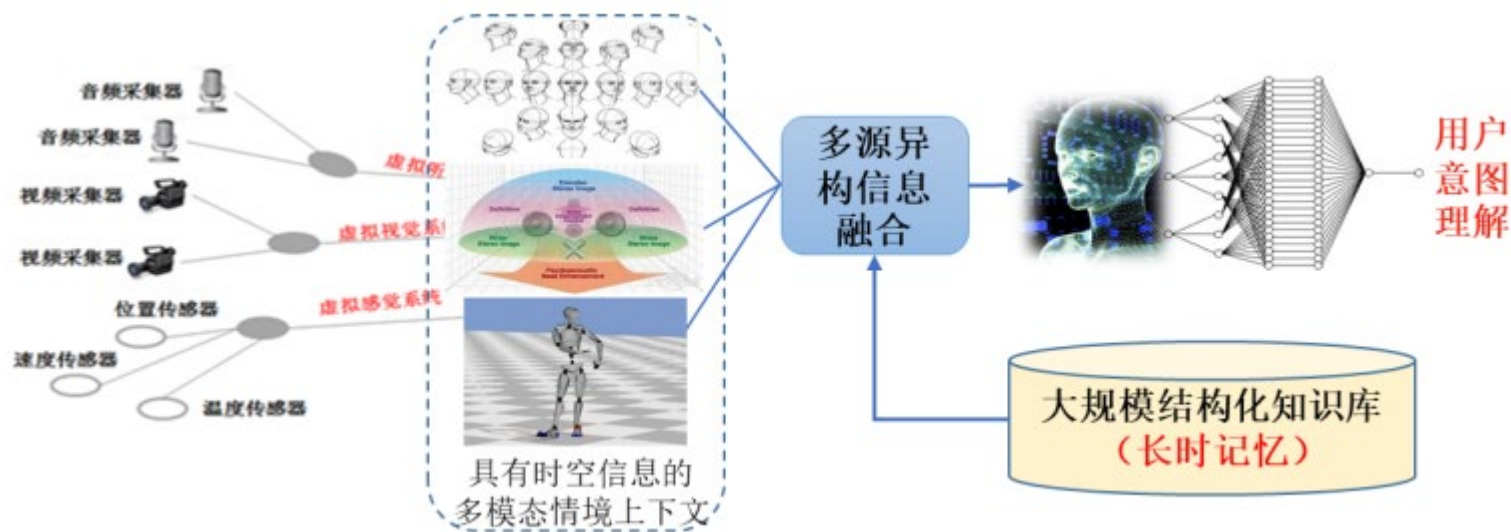
调教后,原句类似句子进行询问

用户个人知识图谱查询

目前基于知识抽取和知识图谱构建,对话系统可实现简单的自主学习和知识推理,但在实际应用中尚难以形成真正的理解和常识记忆。

多模态人机交互

人机是一个高度协同的系统，为了降低对人的要求和限制，提升人机对话的自然度，必须通过技术手段加强机器对用户的音视图文等多模态信息识别与综合理解能力，即多模态融合意图理解能力。



多模态人机交互

多模态智能人机对话系统：

系统基于多模态输入输出融合技术

实现了人机自然交流与互动

支持对语音、情感、姿态等多通道协同感知

采用混合主导式对话策略

以及话题和对话意图相关性分析

支持与多个虚拟人同时聊天



章节内容

- 10.1 对话系统概述
- 10.2 自然语言理解
- 10.3 对话管理
- 10.4 自然语言生成
- 10.5 人机对话评测方法
- 10.6 未来发展趋势
- 10.7 多模态交互技术
 - 多模态人机交互优势
 - 多模态人机交互面临的挑战
- 10.8 本章小结

多模态人机交互



多模态人机交互优势

- 交互自然

多模态的智能人机交互最明显的优点是交互的自然度和自由度，用户可以根据场合以及对不同模态的熟悉度来选择一种或多种方式与设备交互。

- 信息融合

通过多模态信息的融合，比如同时基于手势、口型或面部表情我们可以获得更精确的语义信息。

- 互为监督

不同模态之间可以互相提供关联和监督信息。

- 多维感觉

多维感觉对用户尤其重要，因为用户可以在多模态交互系统中从视觉、听觉、触觉等多个维度来感知和体会智能设备想要表达的情感和语义。

多模态人机交互面临的挑战

- 复杂场景下的语音唤醒和识别

在多模态交互系统中，语音作为对人来说最自然和最常用的交互手段是不可或缺的。但语音识别技术在很多真实远场场景下性能还远远不够，这里主要的挑战来自于重口音和鸡尾酒会这样的场景。

- 对话、场景和情感理解

多模态交互的理解能力，不仅包含自然语言的理解还包含场景和情感层面的理解。对自然语言、场景和情感的理解是目前系统里比较薄弱的环节。

- 智能对话管理

对话需要逻辑上一致、语义上连贯，并且多种目的的交互方式往往会在同一个对话段里出现，对话管理需要能灵活处理不同类型的对话而且需要因人而异。

- 高逼真度多模态对话回复

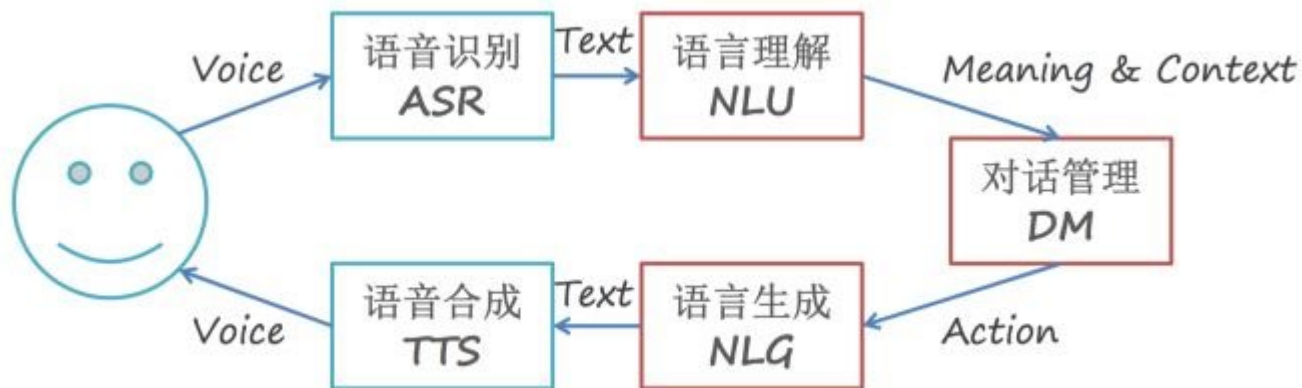
为了使用户能在与机器交互的过程中得到自然的信息反馈，需要不断突破视频展示方面、对话回答生成方面的关键技术。

章节内容

- 10.1 对话系统概述
- 10.2 自然语言理解
- 10.3 对话管理
- 10.4 自然语言生成
- 10.5 人机对话评测方法
- 10.6 未来发展趋势
- 10.7 多模态交互技术
- 10.8 本章小结

本章小结

本章围绕语音人机对话系统研究领域，特别是任务型人机对话系统，从自然语言理解、对话管理、自然语言生成及评测方法等方面做了详细介绍，并针对未来发展趋势展开讨论。



谢谢！