

口语对话系统

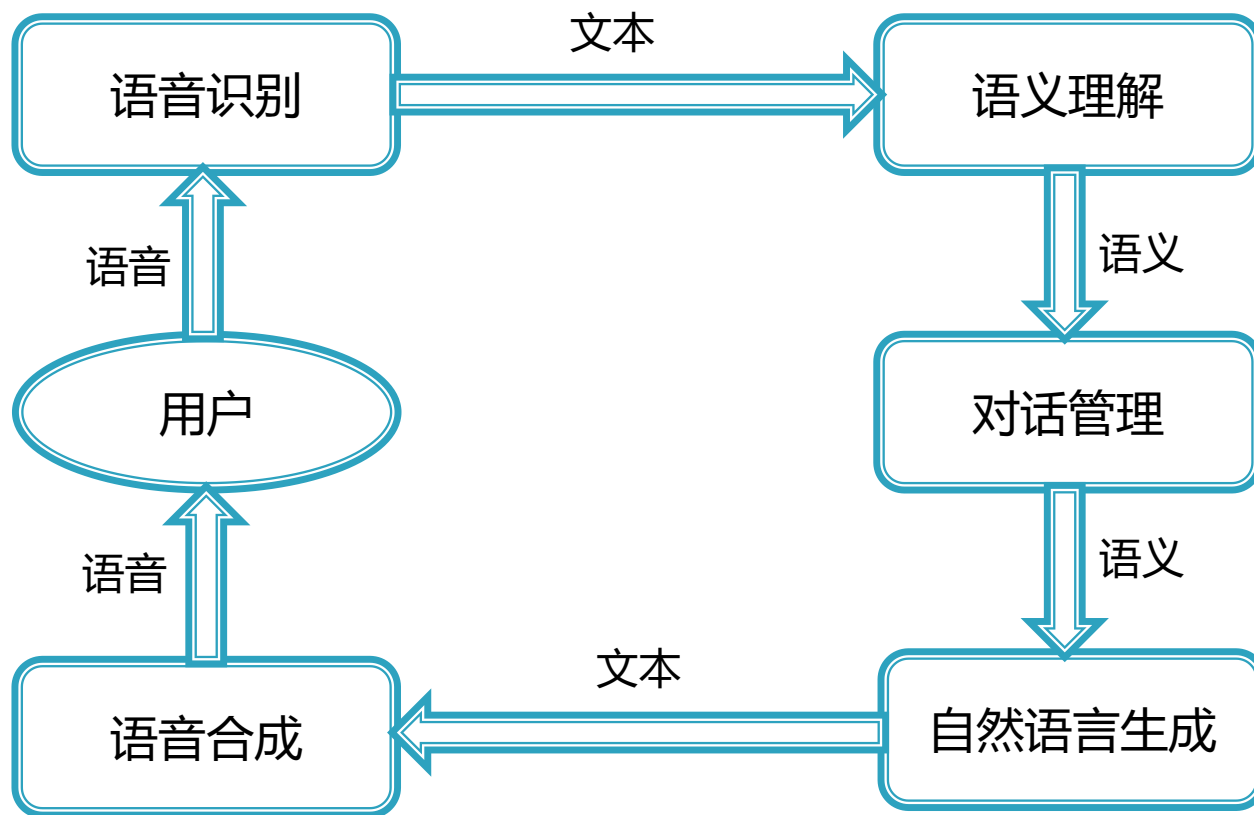
洪青阳 副教授

厦门大学信息科学与技术学院
qyhong@xmu.edu.cn

纲要

- ▶ 口语对话系统主要模块
- ▶ 语义理解模块
- ▶ 对话管理模块
- ▶ 面向服务机器人的口语对话系统
- ▶ 人工智能与机器人

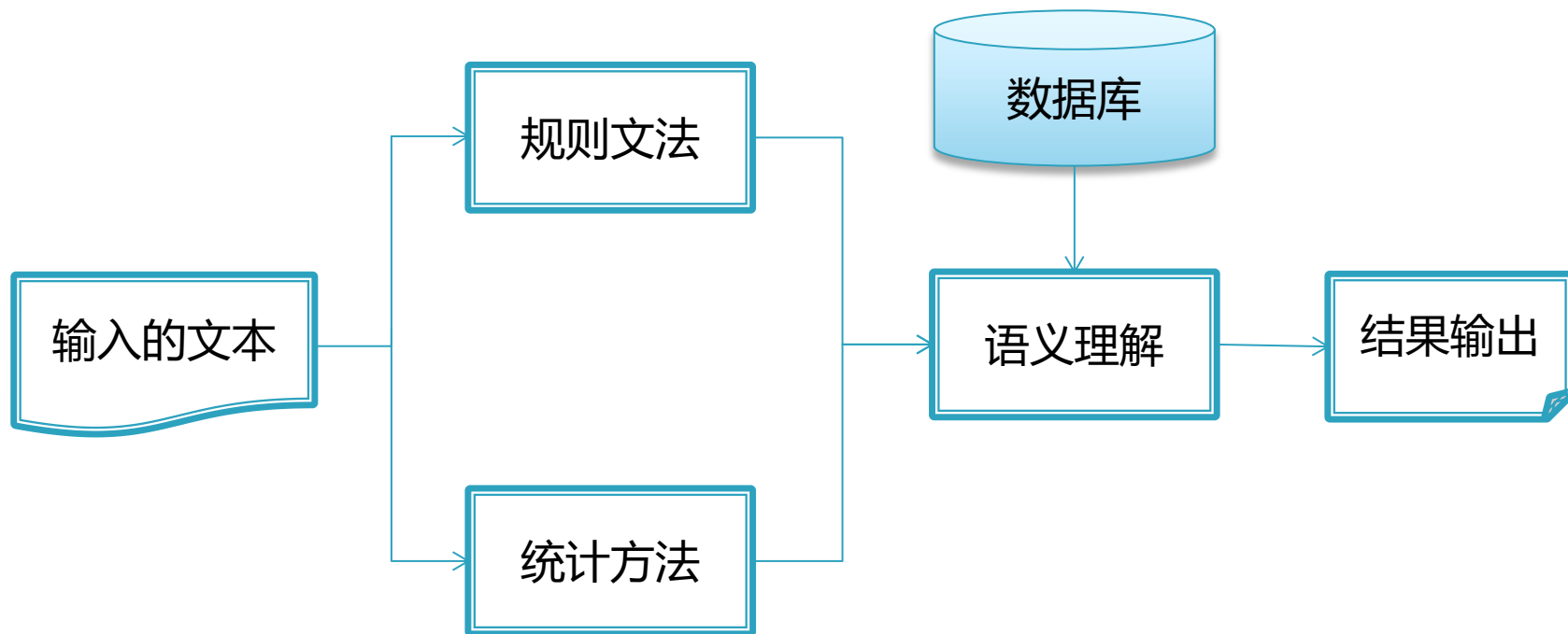
口语对话系统



主要模块

- ◆ 语音识别(ASR)
- ◆ 语义理解(NLU)
- ◆ 语音合成(TTS)
- ◆ 对话管理

语义理解模块



语义理解模块

▶ 规则语法模块

- 正则语法匹配
- 有限状态词表匹配
- 基于Levenshtein编辑距离的模糊匹配

▶ 统计方法模块

- 用户意图分类模块
 - 最大熵分类器：速度快，要求数据多
 - 支持向量机分类器：训练慢，适用于数据稀疏
- 命名实体识别模块
 - 词表匹配
 - 条件随机场(CRF)

对话管理策略

- ▶ 基于有限状态机的对话策略
- ▶ 基于帧（槽）的对话策略
- ▶ 基于任务的对话策略

基于有限状态机的对话策略

- ▶ 用户和系统的交互过程是依照状态图进行的。
- ▶ 工作原理：

$$s(t + 1) = \lambda(s(t), x(t))$$

$$y(t + 1) = \mu(s(t), x(t))$$

给定状态集合 S 和输入集合 X ，如果当前状态是 $s(t)$ 并且输入是 $x(t)$ ，那么通过映射，有限状态机的下一个状态为 $s(t+1)$ ，输出是 $y(t+1)$ 。

- ▶ 这种方式适合流程确定的对话，在对话系统中系统占主导地位。

基于帧（槽）的对话策略

- ▶ 无需定义状态的跳转关系，系统在对话过程中，遵循的原则是填写一张信息表格，但不必要严格按照表格内容填写。
- ▶ 对话的流程可能因用户的输入和当前的上下文不同而不同。

电话订车票例子

S: 您好，这里是火车票订票系统，请问您需要什么帮助？

U: 我要订一张去往上海的火车票。

S: 终点是上海，请问您从哪里出发呢？

U: 我明天下午2:00从厦门出发？

S: 您要预订明天下午两点从厦门出发到上海的火车票，请问您的身份证号码？

U: 123456789

S: 您的身份证号是123456789吗？

U: 是的。

S: 您的订单已生成，请您在明天下午1:00前持身份证到火车站售票大厅领取火车票。

基于帧（槽）的对话策略

▶ 车票信息

◦ 起点：

- 城市：厦门
- 日期：2016年5月27日
- 时间：2:00 pm

◦ 终点：

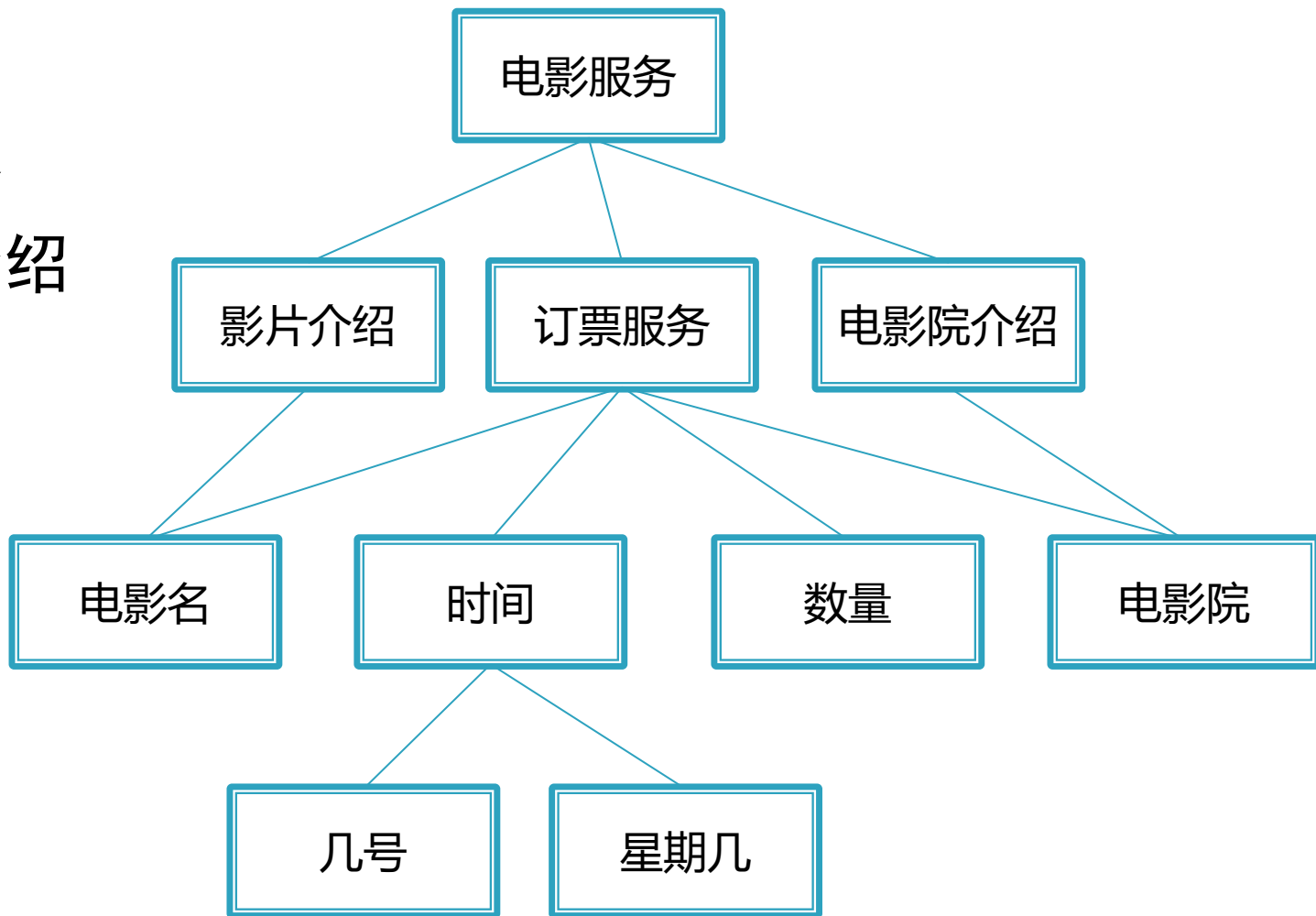
- 城市：上海

基于任务的对话策略

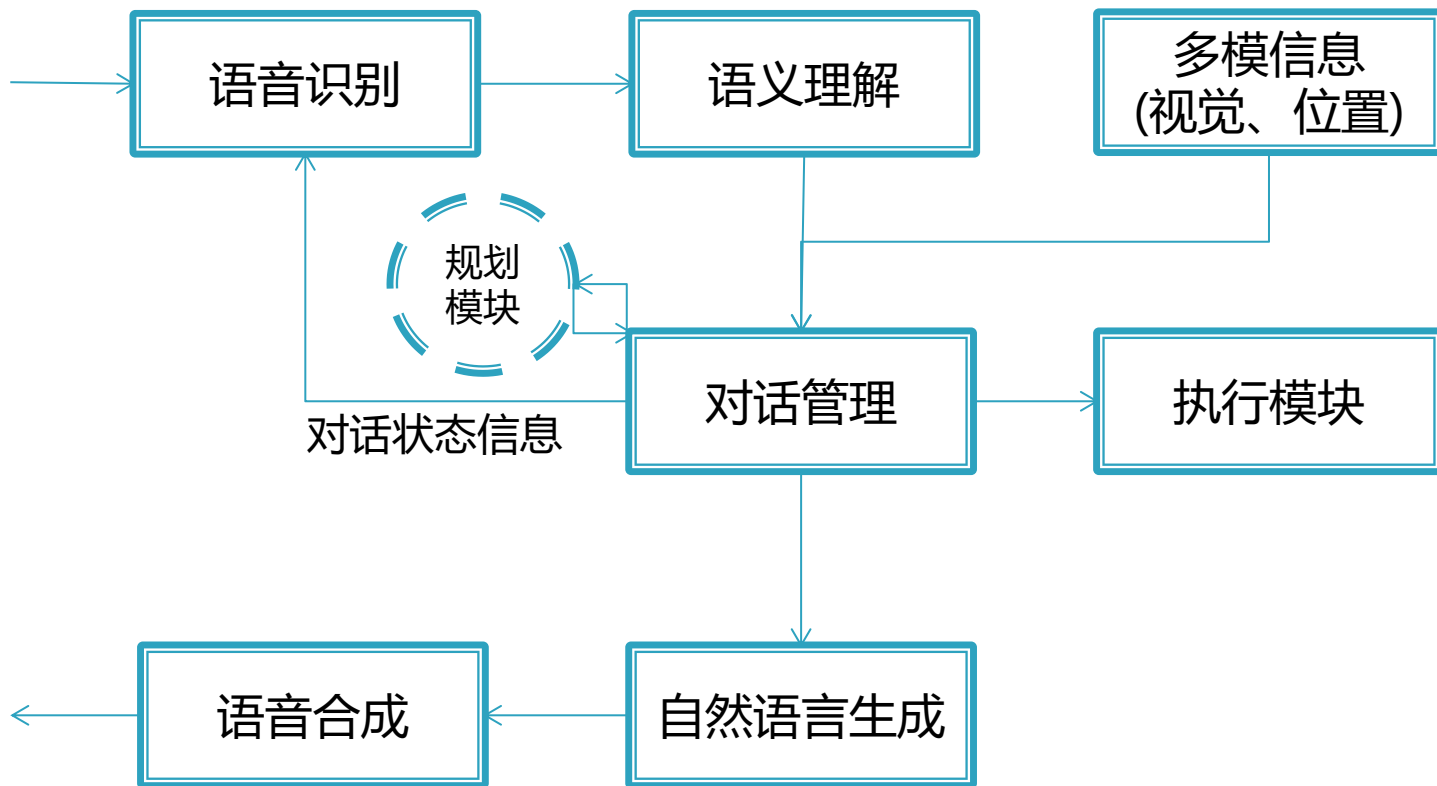
- ▶ 基于帧（槽）的对话可以看作是简单的基于任务的对话管理。
- ▶ 基于任务的对话策略一般采用树状的任务结构图来描述系统所能提供的服务。
- ▶ 根节点用来表示系统的服务领域，在根节点的下一层，可能会有多个节点，表示具体的任务节点。

订电影票服务

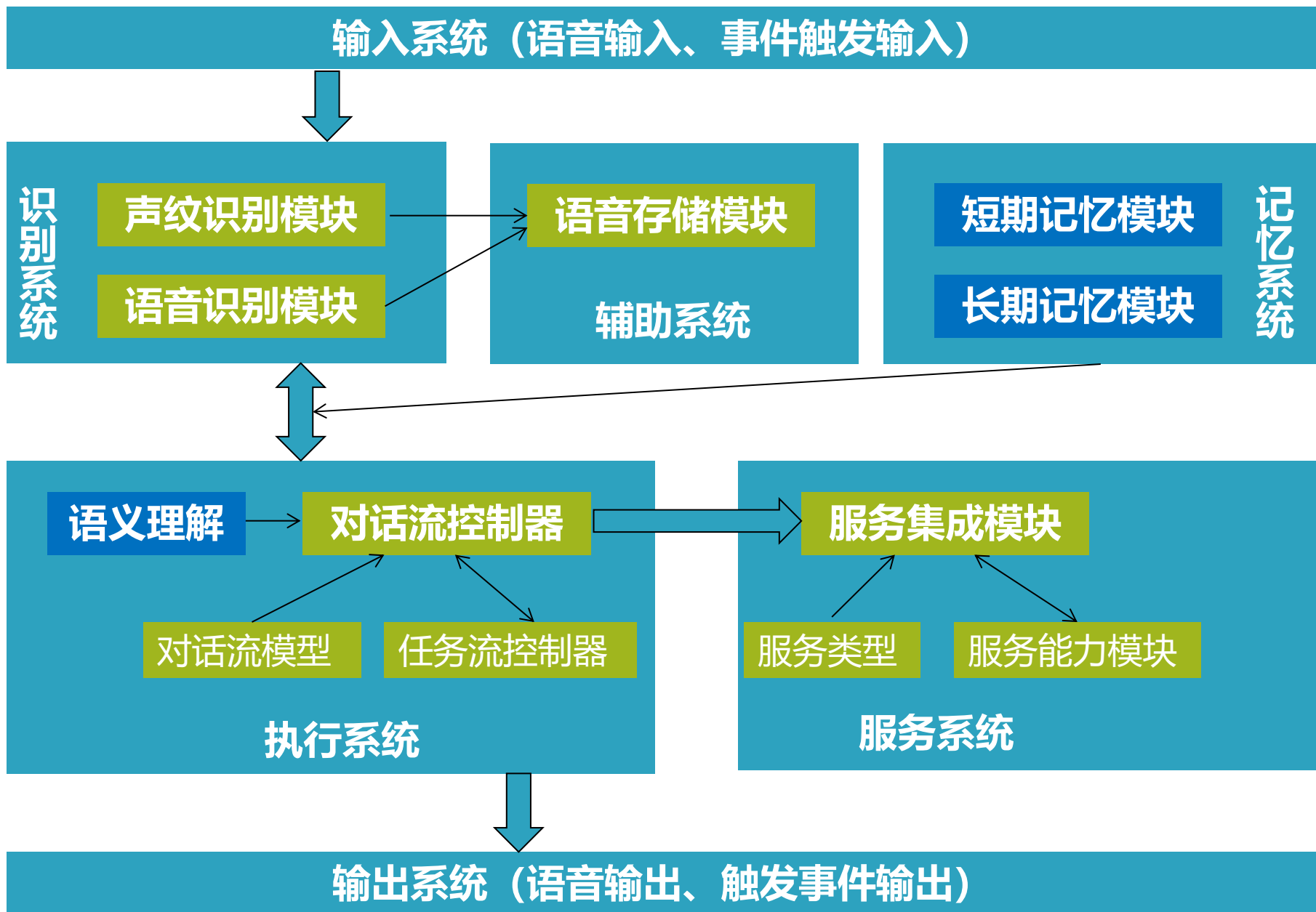
- ▶ 影片介绍
- ▶ 订票服务
- ▶ 电影院介绍



面向服务机器人的口语对话系统



语音系统整体架构图



语音系统功能模块

- **输入系统**

输入系统支持多模态输入方式：

- 1、语音输入：机器人自动检索语音，当检索到有效语音后，触发执行命令
- 2、事件触发输入：通过触摸机器人某一部位，触发执行命令
- 3、移动端输入：通过移动端APP发出的指令来，触发执行命令

- **识别系统**

识别系统包括声纹识别和语音识别

语音识别：识别用户所说话中的内容

语音系统功能模块

声纹识别：识别用户角色（孩子、家长、老人），根据不同的角色进行有意识的引导，将角色带入系统设定好的服务上来。

声纹模型建立方式：

- 1) 产品使用初期，引导用户配合进行角色声纹模型注册
- 2) 通过对输入的语音进行分析，辨别出相对应的角色，进行建模

● 记忆系统

记忆系统分为：短期记忆模块和长期记忆模块

短期记忆模块：记录用户最近一段时间内的对话内容等

长期记忆模块：记录用户相关的个人信息及历史偏好信息等

语音系统功能模块

- **辅助系统**

语音存储模块：存储用户过程中所有的语音数据，用于分析用户行为习惯（在wifi条件下传输语音数据到服务器）

- **执行系统**

执行系统将用户原始输入的信息解析为内部的语义表示，并且决定整个系统应该给用户反馈什么内容。

语义理解、会话流控制器和任务流控制器之间分工有异同又密切合作

第一阶段:没有语义理解模块，通过简单的协议与会话流控制器和任务流控制器协同合作。

语音系统功能模块

- **服务系统**

服务功能有三个子部分：服务类型模块，服务能力模型和多服务集成模块

- 1、服务类型模块：记录了可供使用的各种服务的详细信息
- 2、服务能力模块：存储了哪些服务可以提供什么类型的服务等映射关系
- 3、服务集成模块：调用其他两个模块提供给用户最终的服务内容

- **输出系统**

输出系统包括语音输出和动作指令输出

指令输出：将结果指令给机器人，由机器人来执行最后的任务

语音输出：结合TTS输出个性化语音

语音识别噪声干扰控制

- ▶ 噪声干扰：
 - 周围人群的喧哗声
 - 机器人马达声和回答声
- ▶ 干扰控制：
 - 回答时关闭识别模块
 - 行走时关闭识别模块
 - 麦克风阵列

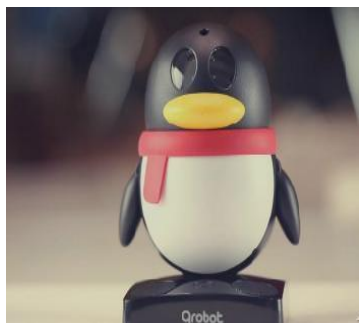
多模信息输入

- ▶ 视觉信息
 - 人脸识别
 - 视频检测
- ▶ 位置信息
 - 地图导航
 - 路径规划
- ▶ 触摸信息
- ▶ APP控制

人工智能和机器人

- ▶ 比尔盖茨早在2007年2月就在《科学人》杂志预言：机器人学将成为下个热门领域，对工作、通信、学习与娱乐带来冲击，就像个人电脑过去30年来的影响一样。
- ▶ 人工智能和机器人产业是最有可能实现1万亿元产值的新兴行业。
- ▶ 智能机器人将走进千家万户。2035年，每5人拥有1台机器人。

智能机器人



Thank you!

Any questions?