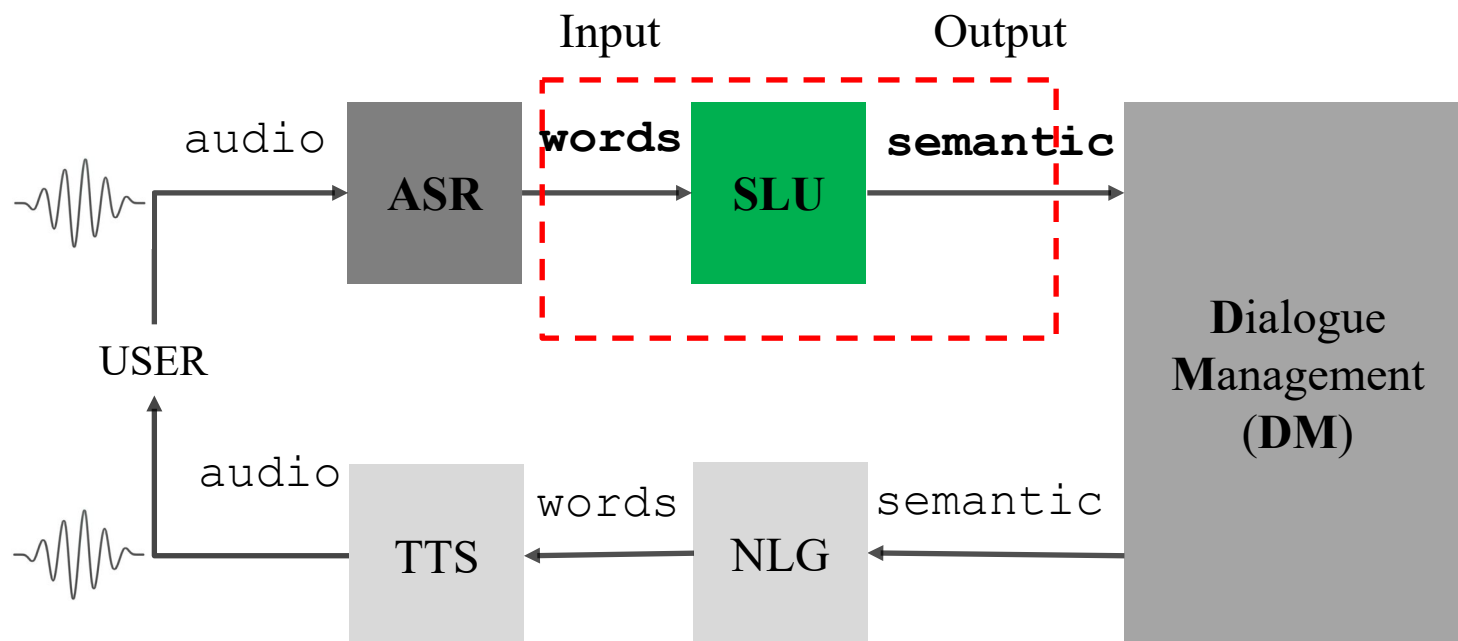


任务简介

▶ 口语语义理解 (Spoken Language Understanding)

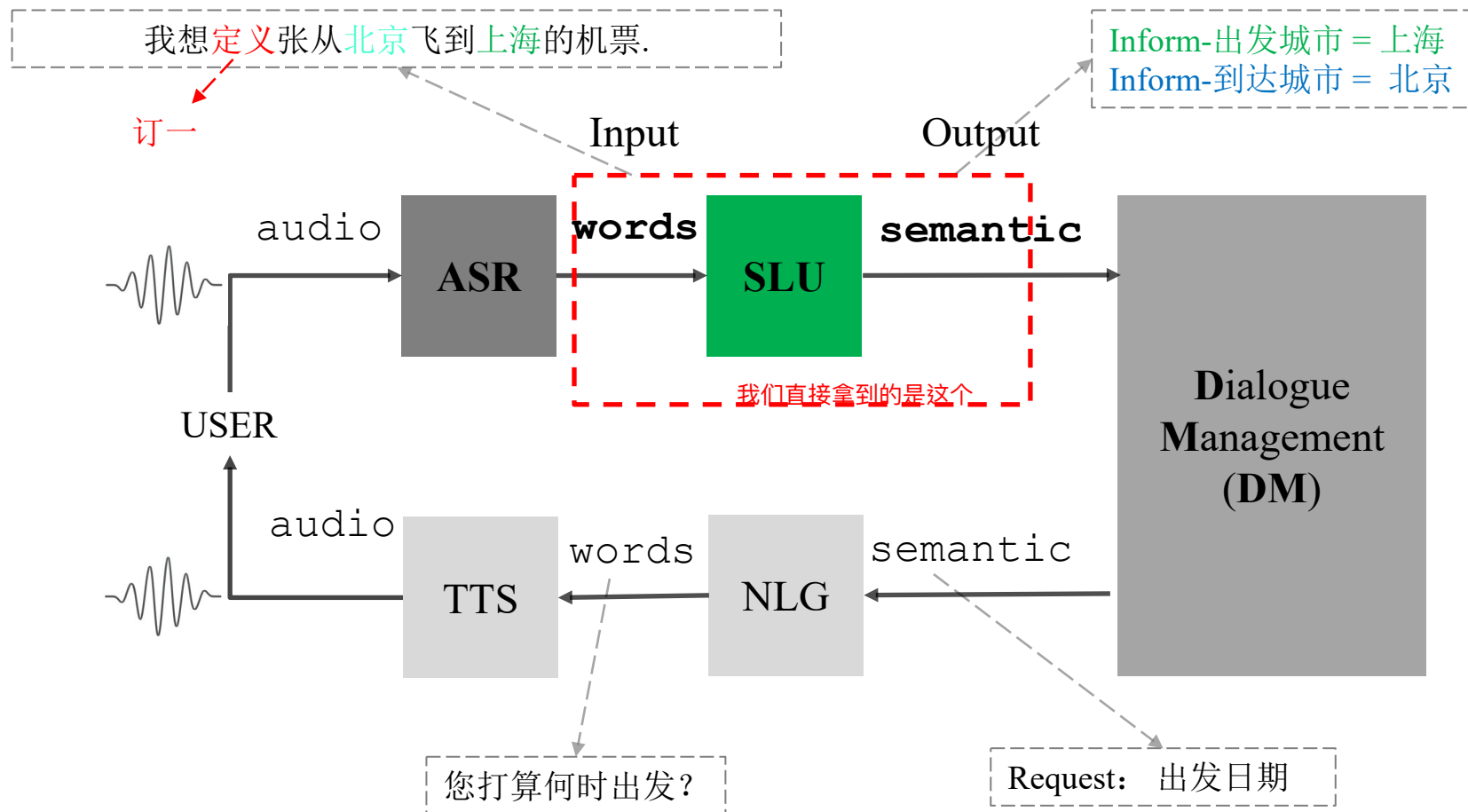


A diagram of Spoken dialogue system

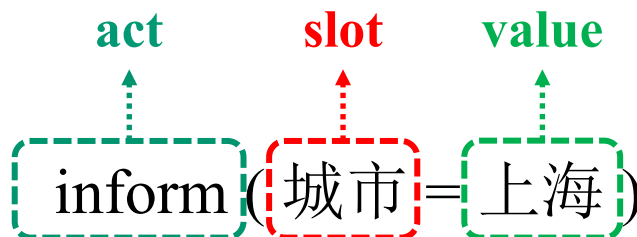
- **ASR : Automatic Speech Recongition**
- **NLG : Natural Language Generation**
- **TTS: Text to Speech**

任务简介

▶ 口语语义理解 (Spoken Language Understanding)



语义三元组

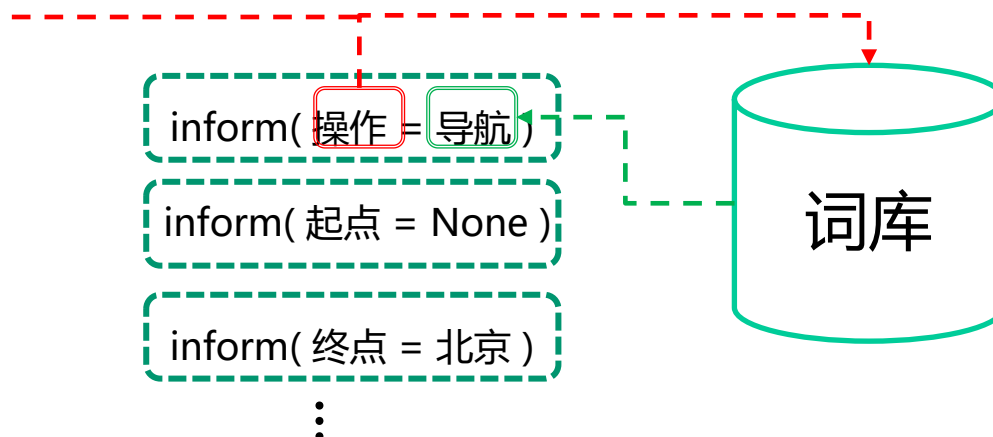


Manual transcripts	Semantic labels
你好。	hello()
上海今天的天气如何？	request(天气预报); inform(城市=上海); inform(日期=今天)
明天是晴天吗？	confirm(天气=晴天); inform(日期=明天)
不是周四，我问的是周五。	deny(日期=周四); inform(日期=周五)

文本分类方法

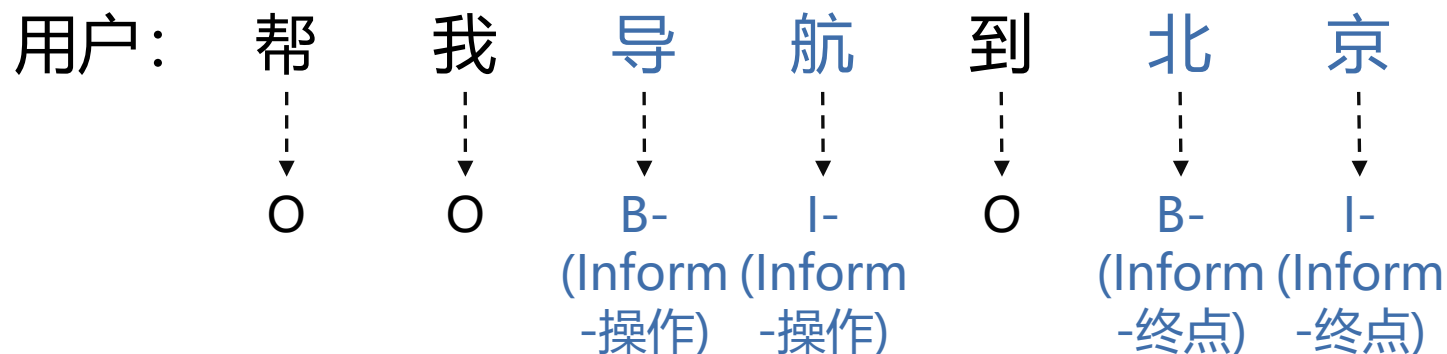
- ▶ 所有的act和slot都是提前设定好的
- ▶ 测试集中出现的value并不一定都在训练集中出现
 - ▶ 可以使用训练集的value构建词库
- ▶ 针对所有的act-slot，在词库中使用分类器选取最有可能的value
 - ▶ 若该act-slot本身是没有value的（如：request（weather_forecast））则可通过二分类方法确定该段对话中是否涉及
 - ▶ 若某一act-slot在该段对话里没有对应的value，可以认为特殊值None是其value

用户：帮我导航
到北京



序列标注方法

- ▶ 类似于命名实体识别任务的序列标注方法
- ▶ 通过预测对话中每个字的标签确定act-slot在文本中的value
 - ▶ 不具备value的act-slot仍然需要使用二分类方法进行预测
- ▶ act-slot可以被合并在一起看作一个标签
- ▶ 使用B- (act-slot) 和I- (act-slot) 标签确定value
 - ▶ B表示词的开头字, I表示非开头字



- ▶ **Accuracy:**

- ▶ 每轮对话所有的 (act, slot, value) 三元组均正确才算该轮对话预测正确

- ▶ **Precision/Recall/F-score**

- ▶ Precision: 输出的所有 (act, slot, value) 三元组中预测正确的比例
- ▶ Recall: 真实标签中所有 (act, slot, value) 三元组被预测正确的比例
- ▶ F-score: Precision和Recall的调和平均

评分标准

▶ 结果报告：

- ▶ 在报告中汇报了符合实验设定的结果，提交的测试集上的预测结果合乎规范
- ▶ 反面教材：没有汇报结果、汇报了训练集上的结果、测试集不符合规范导致无法进行评测、手动预测测试集上结果。

▶ 算法实现：

- ▶ 实现了除给出代码外的其它NLP算法，不局限于课上介绍过的算法，也可以来自于调研的文献
- ▶ 包括并不限于，序列到序列的生成式算法，指针网络（Pointer Network）等，以及BERT、RoBERTa之类的预训练语言模型
- ▶ 每实现一个算法，根据实现的算法本身的复杂程度加分，若实现的算法有误会扣除一定比例的分
- ▶ 在报告中应列出所实现的算法的参考文献

评分标准

► 创新：

- 提出相应的改进策略，包括但不限于数据增强、基于规则的后处理等
- 可以在某个已有的算法基础上进行改进，若是在已有方法上进行的原创性改进，应在报告中汇报其本身算法的参考文献

► 分析与讨论：

- 开放式的讨论，包括但不限于对于性能瓶颈的讨论、对于某种方法为什么会没有效果的讨论等
- 建议提供相关的实验或者错误分析，进行严谨的论证，而非凭空猜测

► 分组完成：

- 每小组人数不超过3人，只需一个人提交
- 需在报告中注明分工和工作量

► 性能问题

- 性能不是决定最终得分的唯一指标，不追求高性能以及大的提升，但要求学术诚信以及研究的规范，希望大家可以抱着探索的精神完成这次作业

大模型推理部分

► 语义三元组解析



You

请对给定的句子进行语义解析，解析成`<act>(<slot>=<value>)`语义三元组的形式，`act`只能为`[inform,deny]`,`slot`包括操作、起点、终点
如果包含多个三元组，请分别输出每个三元组，不要一个`info`里面包含多个`slot`。
句子：帮我导航到北京



ChatGPT

对于句子“帮我导航到北京”，可以进行语义解析成如下的语义三元组形式：

1. `<inform>(<操作=导航>)`
2. `<inform>(<起点=当前位置>)`
3. `<inform>(<终点=北京>)`

► 语义理解



You

你是一个口语语义理解机器人，需要根据用户的提问帮助用户完成操作。
用户：我想定义张从北京飞到上海的机票



ChatGPT

你好！为了帮助您定义张从北京飞到上海的机票，我需要更多的信息。请提供以下信息：

1. 出发日期和时间。
2. 返回日期和时间（如果适用）。
3. 是否有特定的舱位要求，例如经济舱、商务舱或头等舱？
4. 是否有任何特殊的航空公司偏好？
5. 是否需要考虑转机或直飞的选择？
6. 有没有其他特殊需求，比如餐食、座位偏好或其他服务？

提供这些信息后，我将能够更准确地帮助您找到符合张需求的机票。

大模型推理部分

▶ 实现方式:

- ▶ 直接使用生成式大模型进行**语义理解**或/和**语义三元组解析**
- ▶ 大语言模型平台使用不限, 可以使用网页版对话框、Huggingface Demo、或对应调用的API
- ▶ 至少使用三个不同的Setting来测试包含但不限于模型/大小/提示词/Zero/One/Few-shot/中文/英文

▶ 一些探索方向:

- ▶ Prompt优化
- ▶ Few-shot learning 或者 CoT 技术
- ▶ 不同模型生成效果对比
- ▶ 结构化输出(进阶)
- ▶ 工具使用(进阶)
- ▶ 大模型幻觉(进阶)
- ▶ 拒绝回答(进阶)

大模型推理部分

▶ 评分标准:

- ▶ 在报告中汇报了符合实验设定的结果，并对实验结果进行分析
- ▶ 简要讨论大模型时代下口语语义理解现状以及未来可能的研究方向
- ▶ 没有性能要求，要求熟悉大模型提示工程并分析对比结果

▶ 独立完成

- ▶ 该部分需要每位同学独立完成

时间宽限说明

- ▶ 原定7天宽限计入统计，每个同学独立计算，按照以下公式计算总延期时间

$$T_{\text{总}} = T_{HW1} + T_{HW2} + T_{HW3} + \max(T_{\text{Project-LM}}, T_{\text{Project-LLM}})$$

- ▶ 作业二DDL: 12月7日 (本周四) 23:59
- ▶ 大作业DDL: 2024年1月7日23:59