# 任务型对话系统中的口语理解

## 评测任务概述

在任务型对话系统中，口语理解（Spoken Language Understanding，SLU）作为一个重要模块，主要完成将用户的自然语言表达文本解析为结构化信息。结构化信息的解析主要包括意图识别和槽填充两个步骤。用户的自然语言表达文本通常是自动语音识别的结果，所以其中会存在一定的识别错误（错误率主要取决于用户对话的场景）。

本次评测的数据采集自一个商用的任务型对话系统的线上数据，我们选取了三个领域music、navigation和phone\_call（其余标注为OTHERS）进行评测。为了简化任务，对领域设计进行了简化，仅挑选了其中的常见意图和语义槽。通过用户连续两次对话的时间差，我们将用户对话流拆分为sessions。在进行用户对话的意图和语义槽标注时，我们会考虑对话在session中的上文。下面给出了两个sessions的示例：

1 打电话 phone\_call.make\_a\_phone\_call 打电话

1 我想听美观 music.play 我想听<song>美观</song>

1 我想听什话 music.play 我想听<song>什话||神话</song>

1 神话 music.play <song>神话</song>

2 播放调频广播 OTHERS 播放调频广播

2 给我唱一首一晃就老了 music.play 给我唱一首<song>一晃就老了</song>

格式说明：

* 每行4个字段，以“\t”分割，字段分别是session ID、用户query、意图和语义槽标注；
* Session中仅给出用户queries，没有给出对话系统响应；
* Session内对话按先后顺序给出；
* 当语义槽值包含语音识别错误时，同时给出了纠错后的正确语义槽值，例如“我想听<song>什话||神话</song>”其中“神话”表示纠错后的正确语义槽值。

为了进行语义槽值的错误纠正，我们同时会给出每一个语义槽的取值词典。词典使用评测集中具体数据进行了裁剪，仅用于评测目的。

## 评测任务描述

本次评测任务分为4个子任务，参赛者可以根据兴趣选择参加其中的一个或多个任务。

## 子任务1：意图识别封闭式评测

参赛者只允许使用主办方提供的评测数据进行意图识别的训练和开发，评测时，主办方给定测试集，在测试集上得出评测结果。

评价指标：所有意图（不包括OTHERS）的值，计算方法如下：

，

，

。

## 子任务2：意图识别开放式评测

参赛者允许使用任何数据进行意图识别的训练和开发，评测时，主办方给定测试集，在测试集上得出评测结果。

评价指标：同子任务1。

## 子任务3：意图识别和槽填充封闭式评测

参赛者只允许使用主办方提供的评测数据进行意图识别和槽填充（包括槽值的错误纠正）的训练和开发，评测时，主办方给定测试集，在测试集上得出评测结果。

评价指标：意图识别和槽填充（包括槽值纠错）的准确率值计算方法如下，

，

其中分母表示测试集中的query条数（包括标注为OTHERS的queries），分子表示被正确解析的query条数。“正确解析”定义为意图识别和语义槽填充（包括槽值的错误纠正）完全正确。

## 子任务4：意图识别和槽填充开放式评测

参赛者允许使用任何数据进行意图识别和槽填充（包括槽值的错误纠正）的训练和开发，评测时，主办方给定测试集，在测试集上得出评测结果。

评价指标：同子任务3。