

娱乐播放助手实体抽取 创新与实践

周晓欢

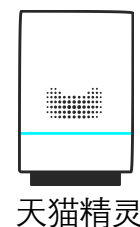
部门：天猫精灵事业部-人工智能部

- 天猫精灵作为一个智能音箱，内容点播是其最重要的功能之一。
- 在天猫精灵点播场景，用户是通过**语音交互**，以**自然语言**请求的方式进行内容点播的。

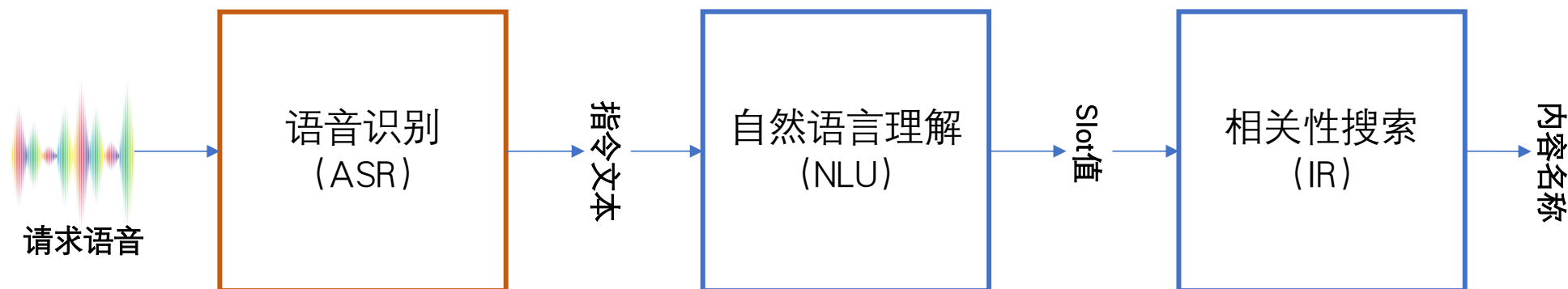


天猫精灵，我想听刘德华的忘情水。

好的，为你送上这首好听的歌：刘德华的忘情水。



- Pipeline的形式
- 实体抽取（Slot Filling）在搜索中起关键作用

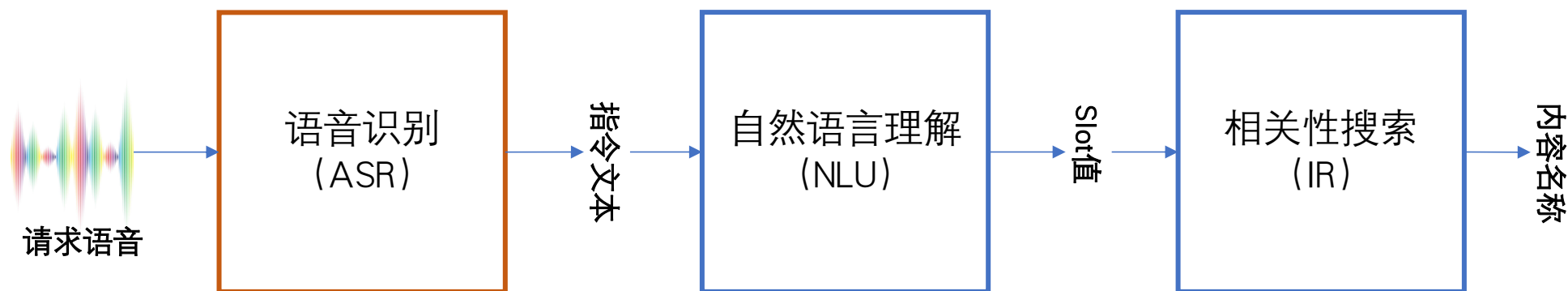


➤ 曝光偏差问题

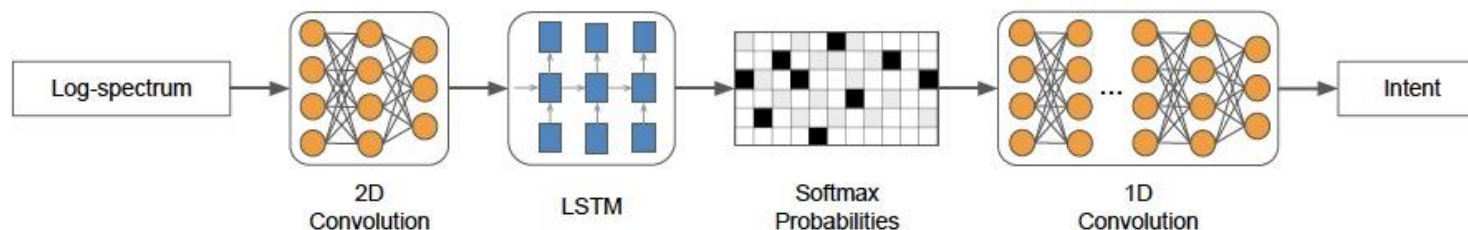
NLU模块的训练是基于准确的文本进行训练的，而在测试阶段是基于ASR模块的输出文本进行，一旦ASR模块出错，就会导致**连环错误**。

➤ 训练目标不一致

ASR训练目标是降低字错误率，然而Slot值的错误率和其他词的错误率对整个系统的影响是不同的，pipeline的训练方式无法对齐这一点。



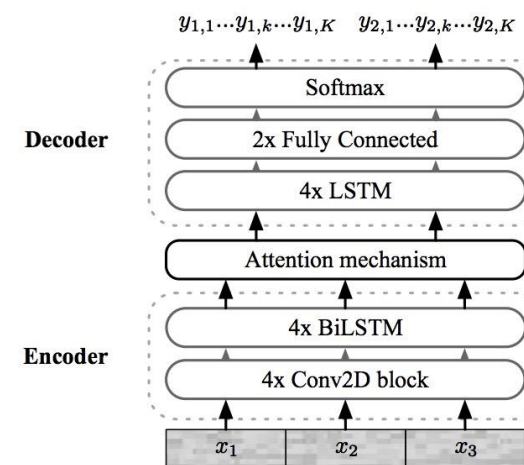
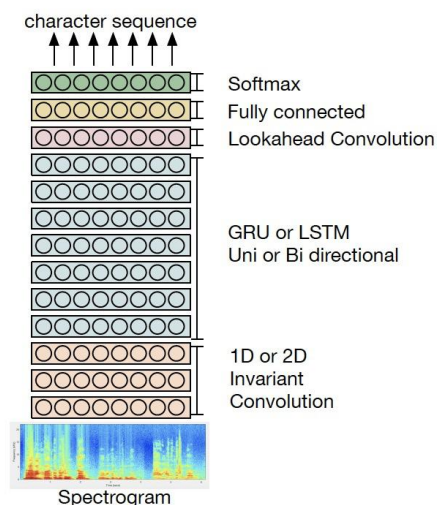
➤ END2END DC/IC (分类任务)



[1] Chen, Yuan-Ping, Ryan Price, and Srinivas Bangalore. "Spoken language understanding without speech recognition." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.

➤ END2END Slot-Filling (生成任务)

我想听\$刘德华\$的#忘情水#

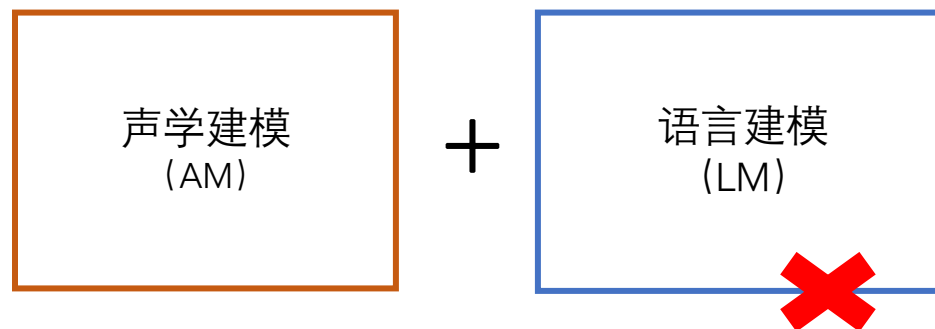


[2] Sahar Ghannay, Antoine Caubrière, Yannick Estève, Nathalie Camelin, Edwin Simonnet, Antoine Laurent, and Emmanuel Morin. 2018. End-to-end named entity and semantic concept extraction from speech. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 692 – 699. IEEE.

[3] Valentin Pelloin, Nathalie Camelin, Antoine Laurent, Renato De Mori, Antoine Caubrière, Yannick Estève, and Sylvain Meignier. 2021. End2end acoustic to semantic transduction. *arXiv preprint arXiv:2102.01013*.

END2END Slot-Filling

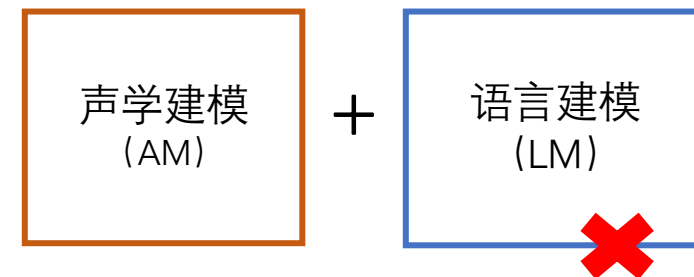
Alibaba



内容实体识别困难

Alibaba

- 一音多字：圣诞节&圣诞结、油精&游京
- 内容名不存在语法约束：巴辛博嘞、加油鸭
- 口音问题：津味鹅&京味儿、澜菲的大雁&南飞的大雁

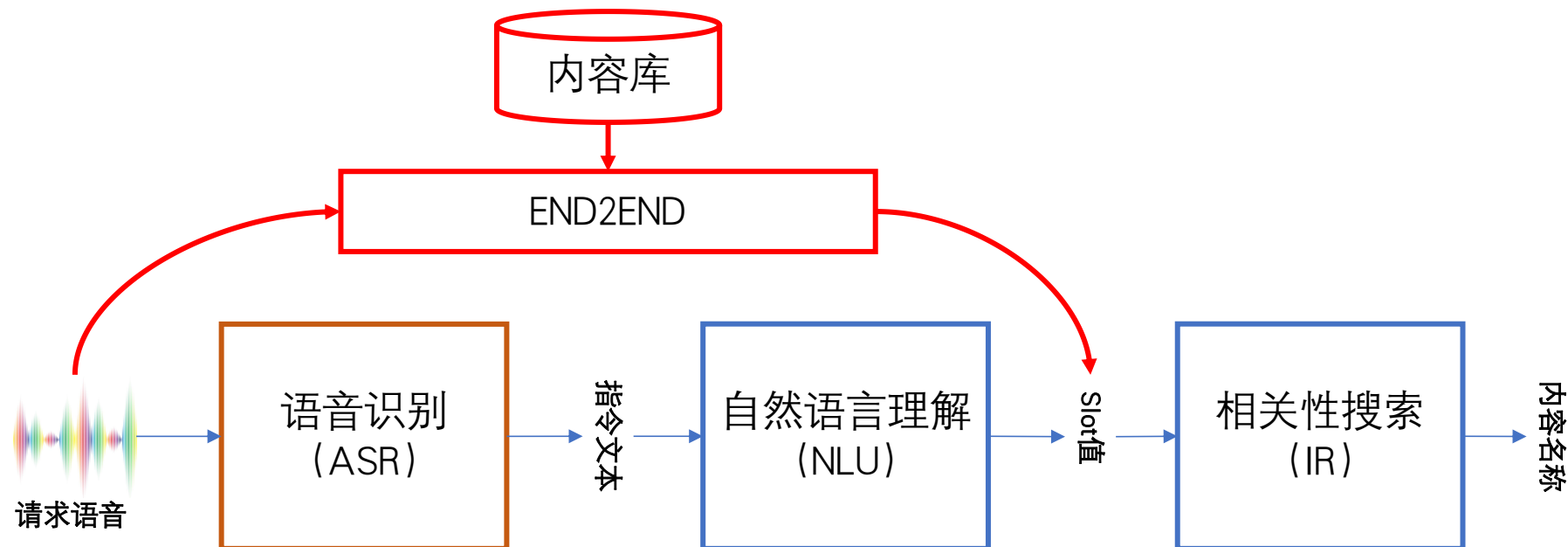


➤ END2END

人在理解一段语音的时候，并不需要经过转化成文字这一环节的思考，不识字的人也能够很好地进行交流。

➤ 引入内容库信息

实体名称是个有限集合，做选择题比做填空题简单。



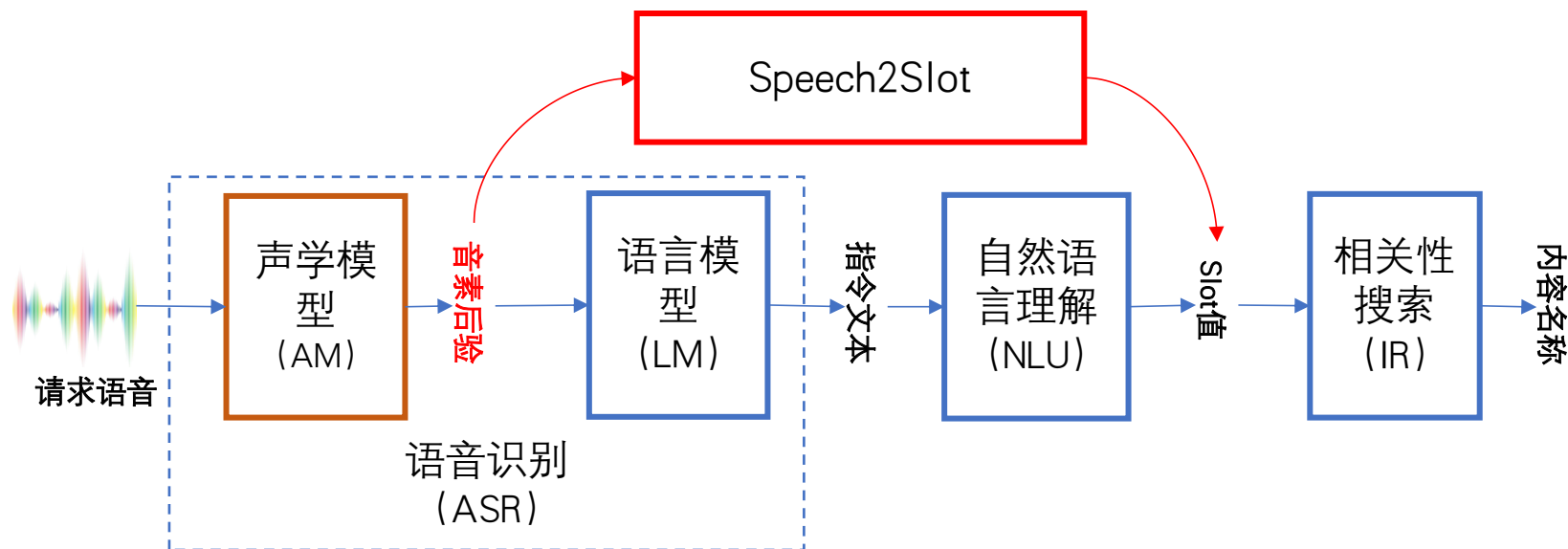
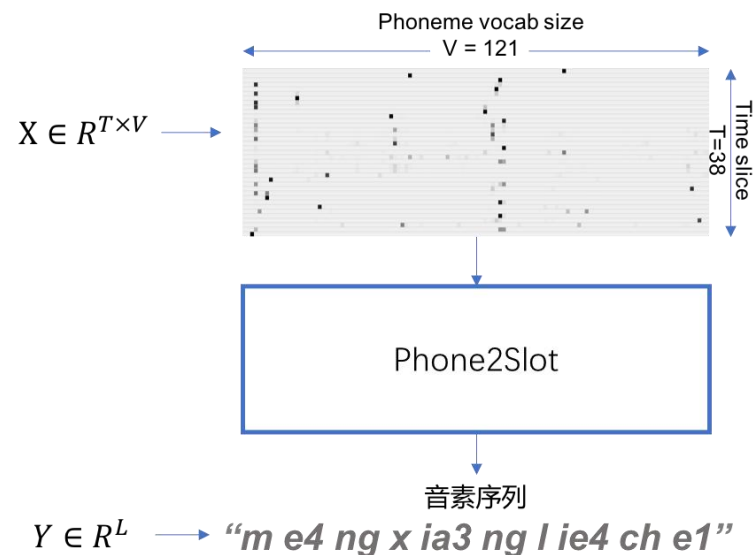
Speech2Slot——Knowledge-based End-to-End

Alibaba

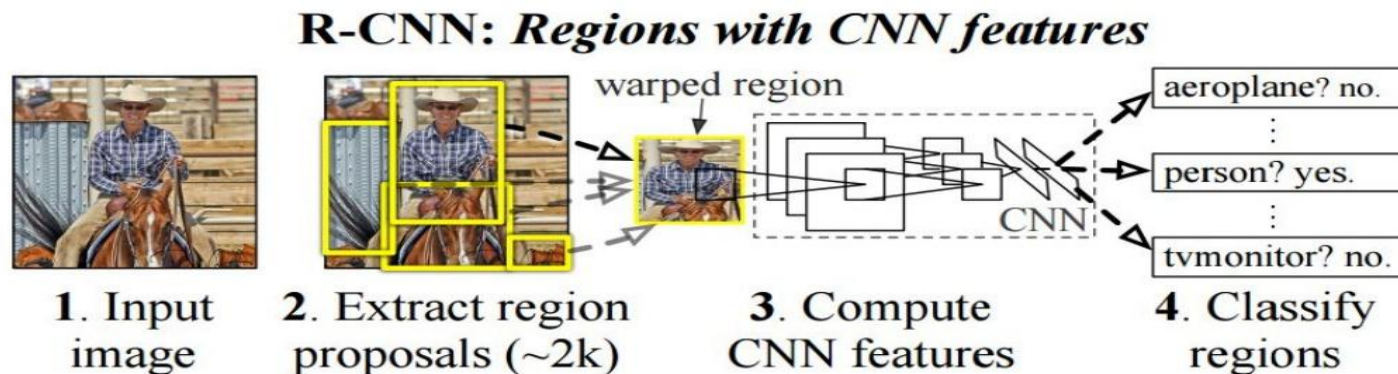
➤ 训练数据不足，难以训练一个绝对END2END的模型

➤ 输入输出：音素后验矩阵 \rightarrow Slot的音素序列

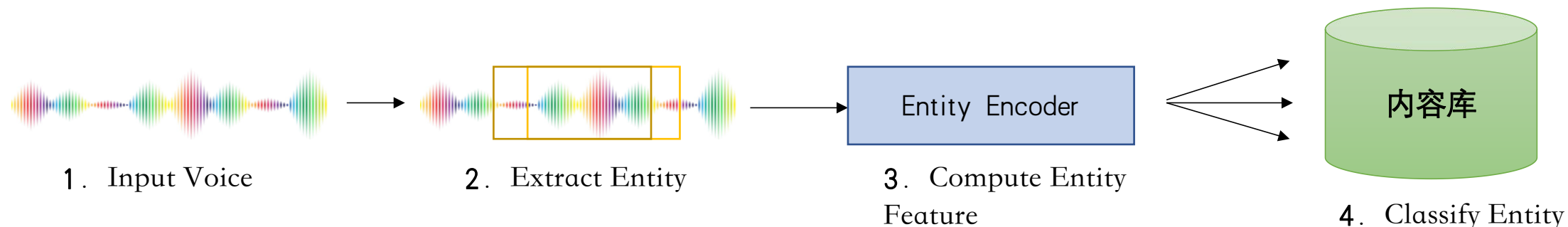
- 声学模型与实体抽取任务相对独立，声学模型是感知层面的学习，而后续的实体抽取是认知层面的学习；
- 音素（phone）恰好为语音信号和文本信息(实体名)提供了一个统一编码方式，降低了模型学习难度。



➤ 图像目标检测任务

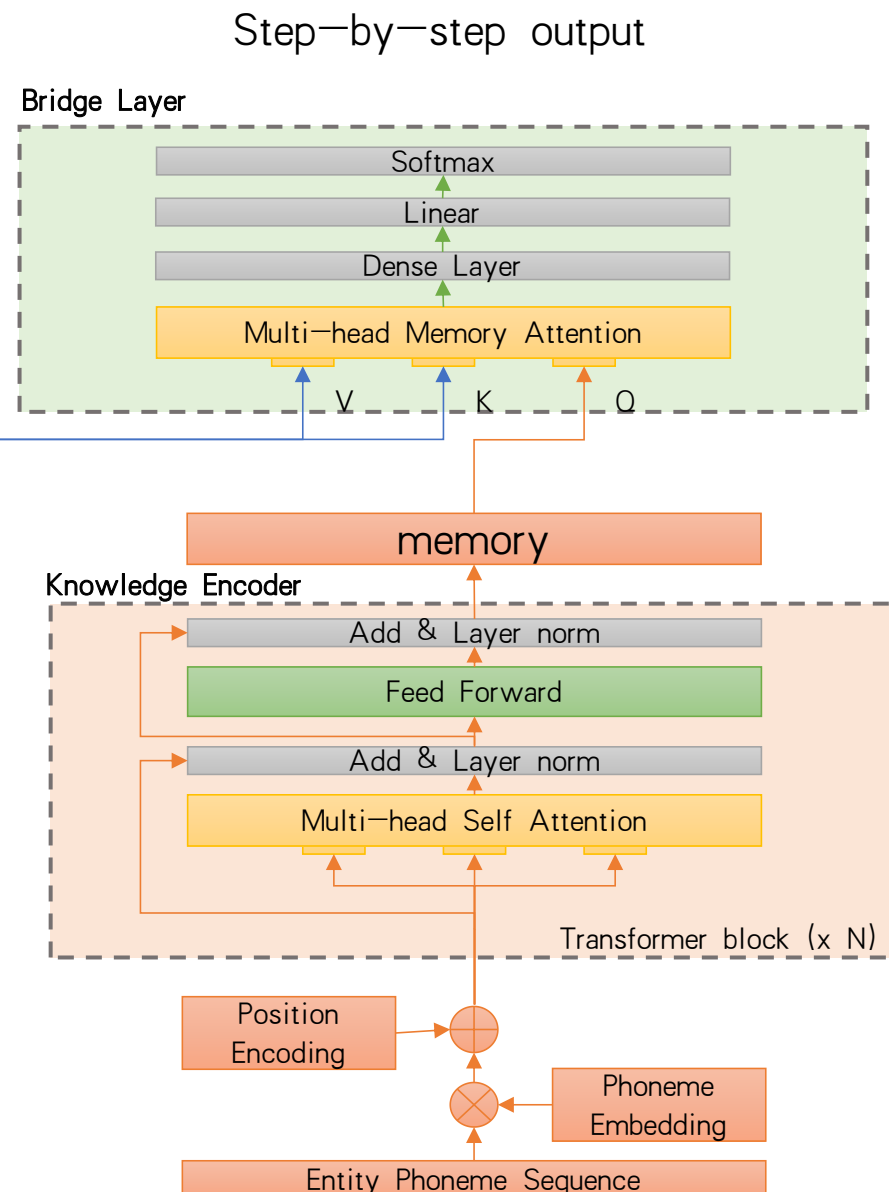
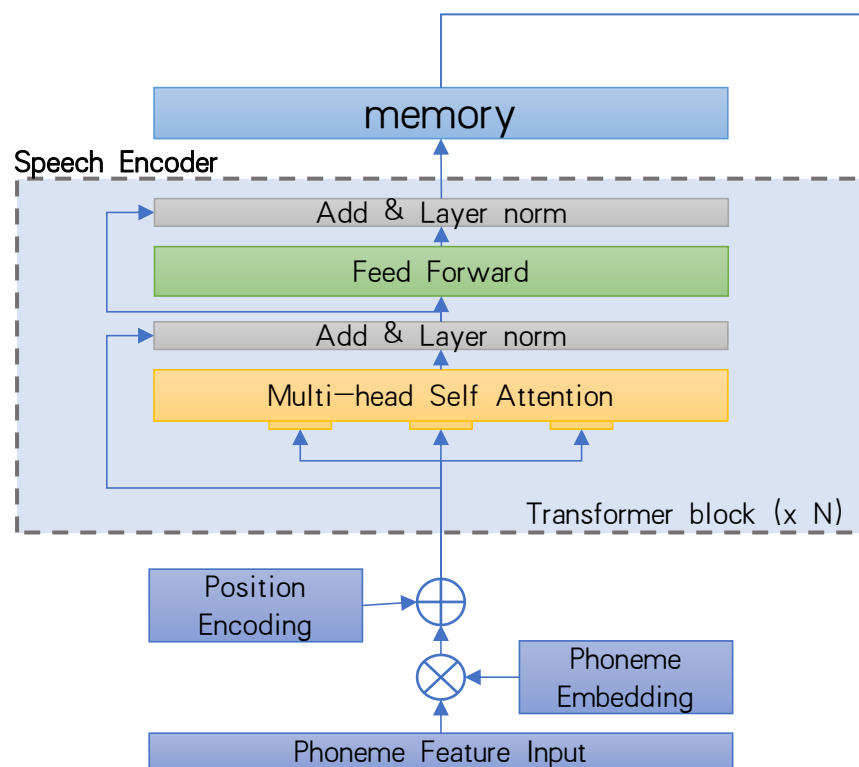


➤ 语音实体识别任务



实体定位+特征抽取

- Speech Encoder: 语音语义理解
- Knowledge Encoder: 内容库知识建模
 - 基于内容库预训练, 并采用Cold-fusion的方式进行Finetune
- Bridge Layer: 实体抽取
 - 用内容库知识对语音片段进行定位, 并做特征分析
 - LM的hidden state不参与结果计算, 应对“反语言模型”问题



Speech2Slot—Knowledge-based End-to-End

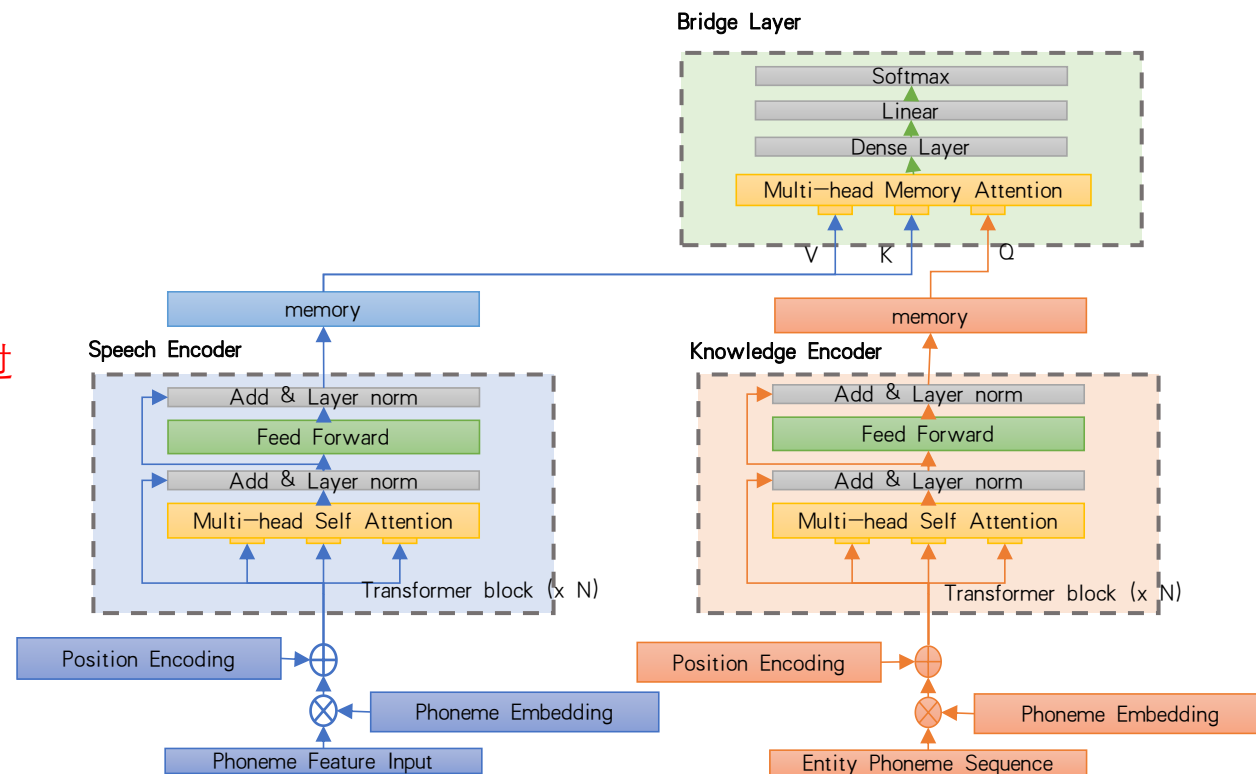
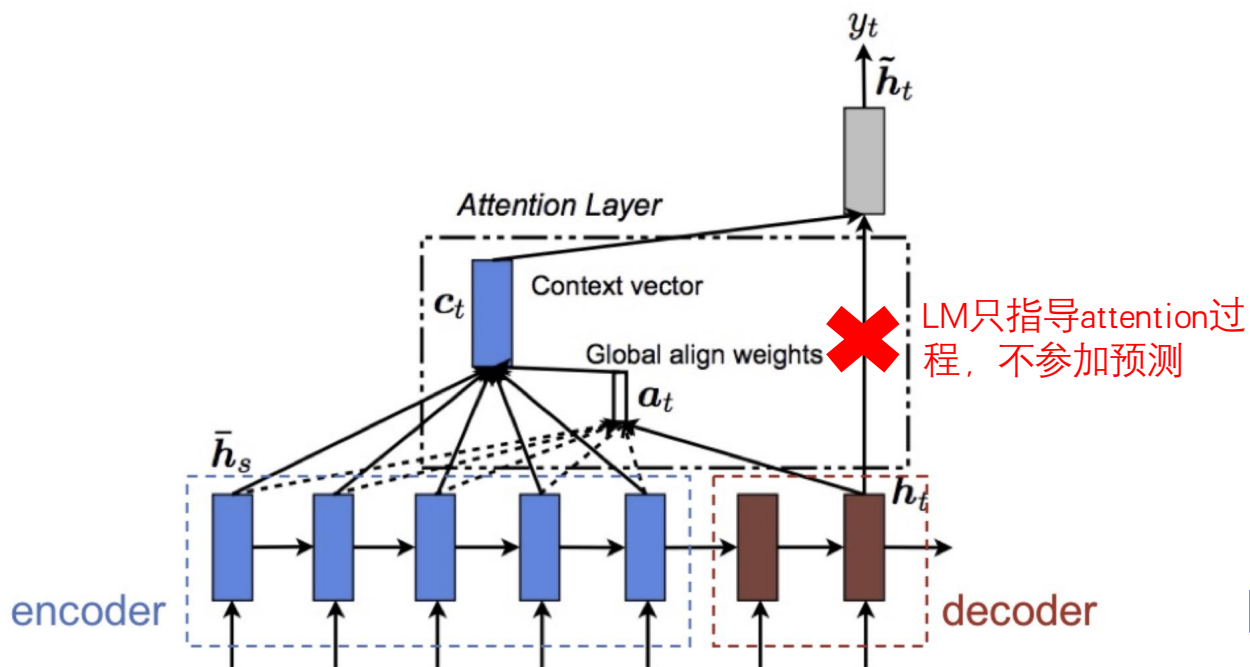
Alibaba

➤ END2END

从AM后验矩阵中直接“翻译”出所包含的实体名称

➤ 引入内容库信息

Knowledge encoder用内容库中的所有内容名称进行预训练，并采用Cold-fusion的方式进行Finetune
推理阶段，Beam-search过程中，构建内容库Trie树来约束模型输出必须指向内容库中的名称



Speech2Slot方案应用在天猫精灵音乐点播歌名识别上，点播第一首歌的切歌率相对下降**6.32%**，完播率相对提升**5.93%**。

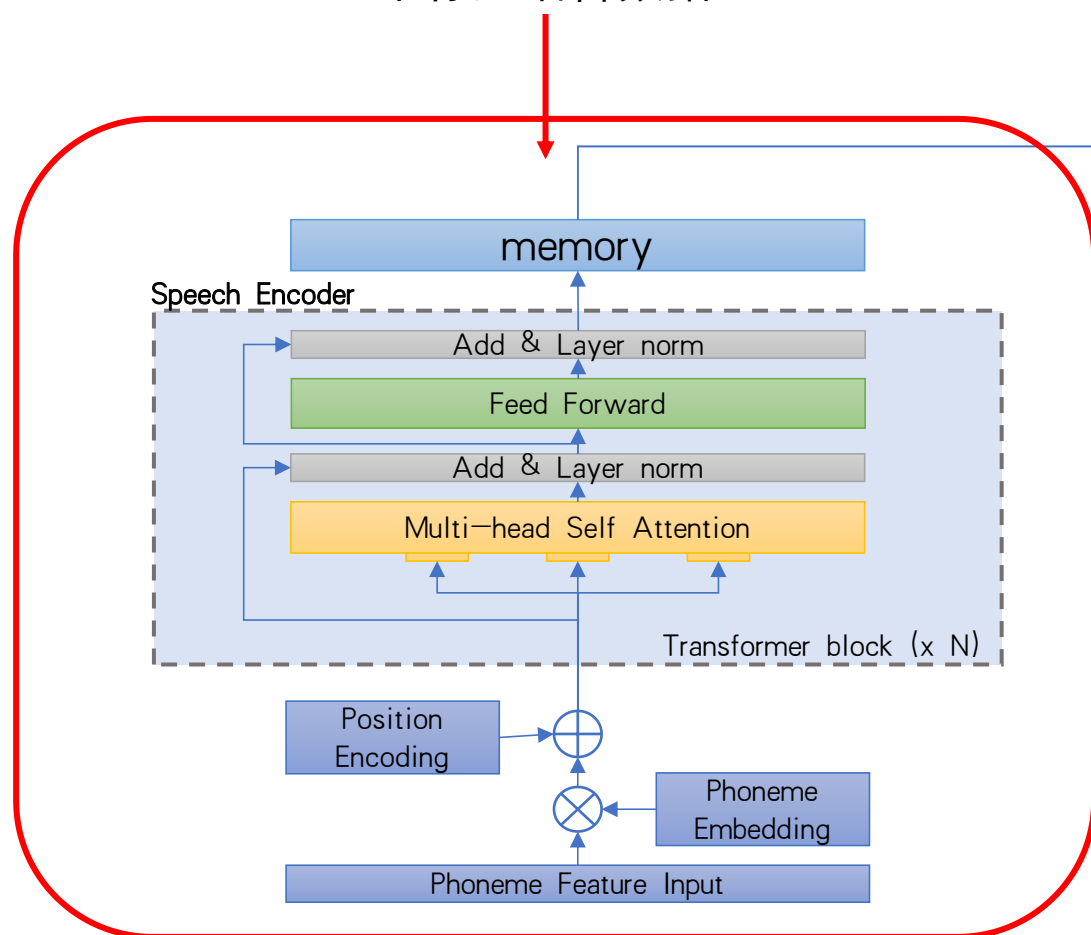
模型	Slot错误率相对降低
baseline(ASR+SF)	—
Speech2Slot	−11.3%

无监督语音语义预训练

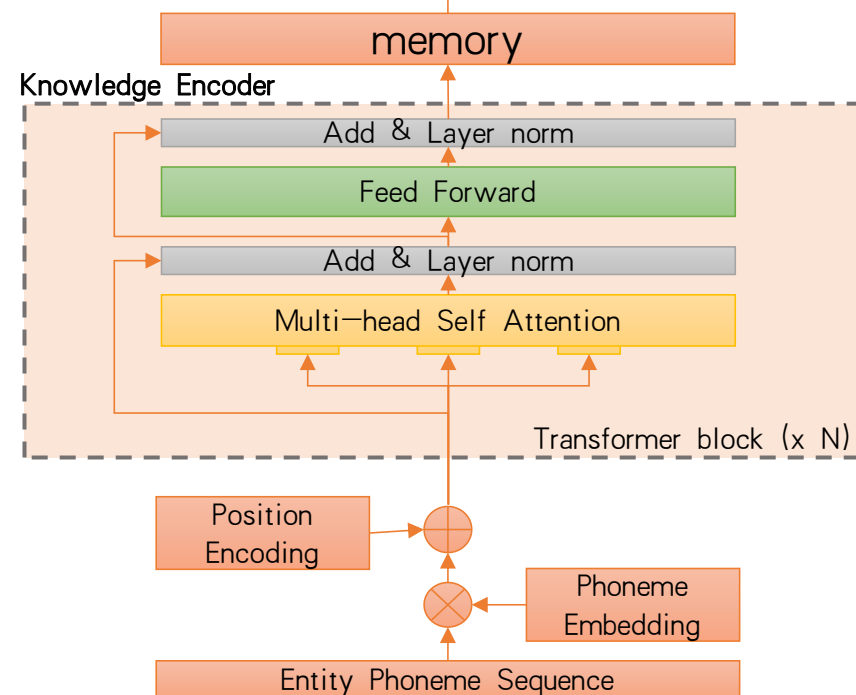
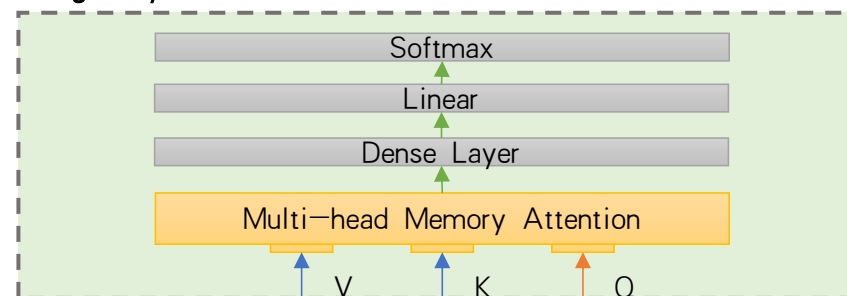
Alibaba

- 端到端训练数据不足
- 语义理解能力很重要

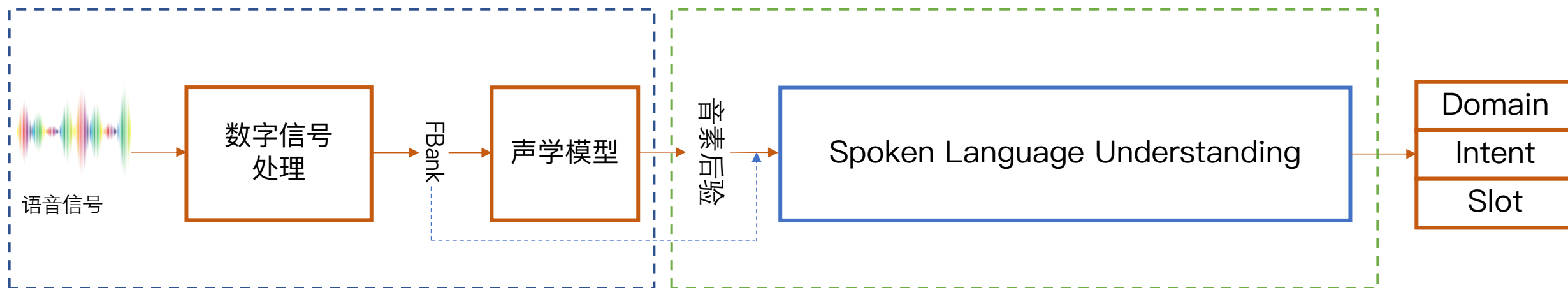
未标注语音数据



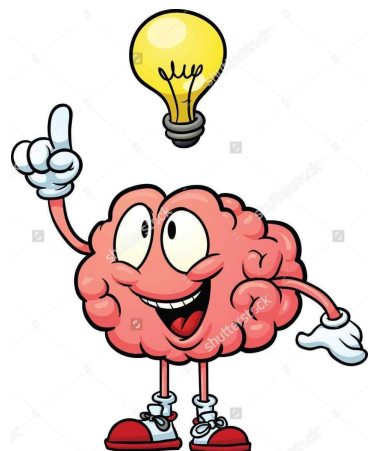
Bridge Layer



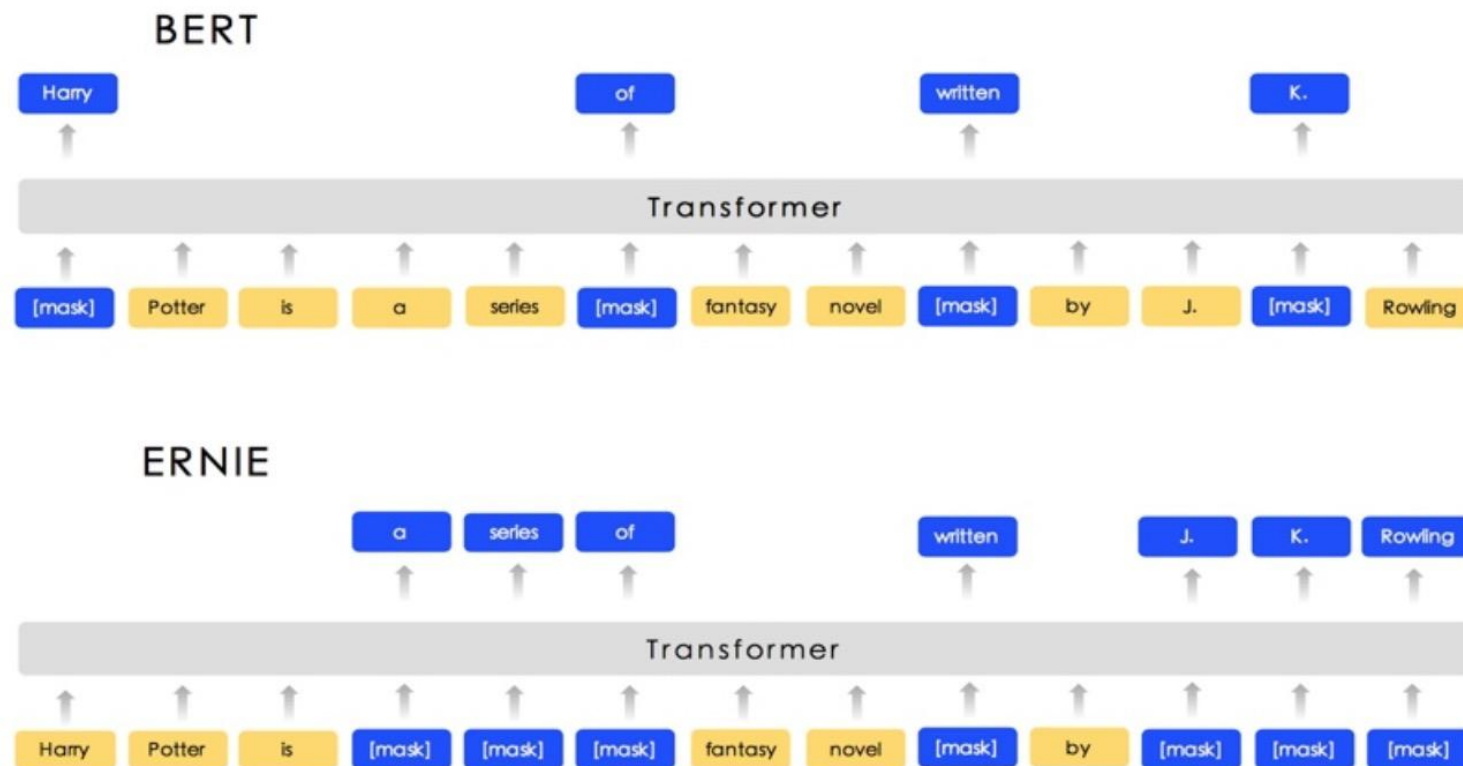
- 能否将语音信息以无监督方式进行编码



Pre-training



无标签
语音数据



[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.

[6] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xi Tian. 2019. Ernie: Enhanced representation through knowledge integration. arXiv.

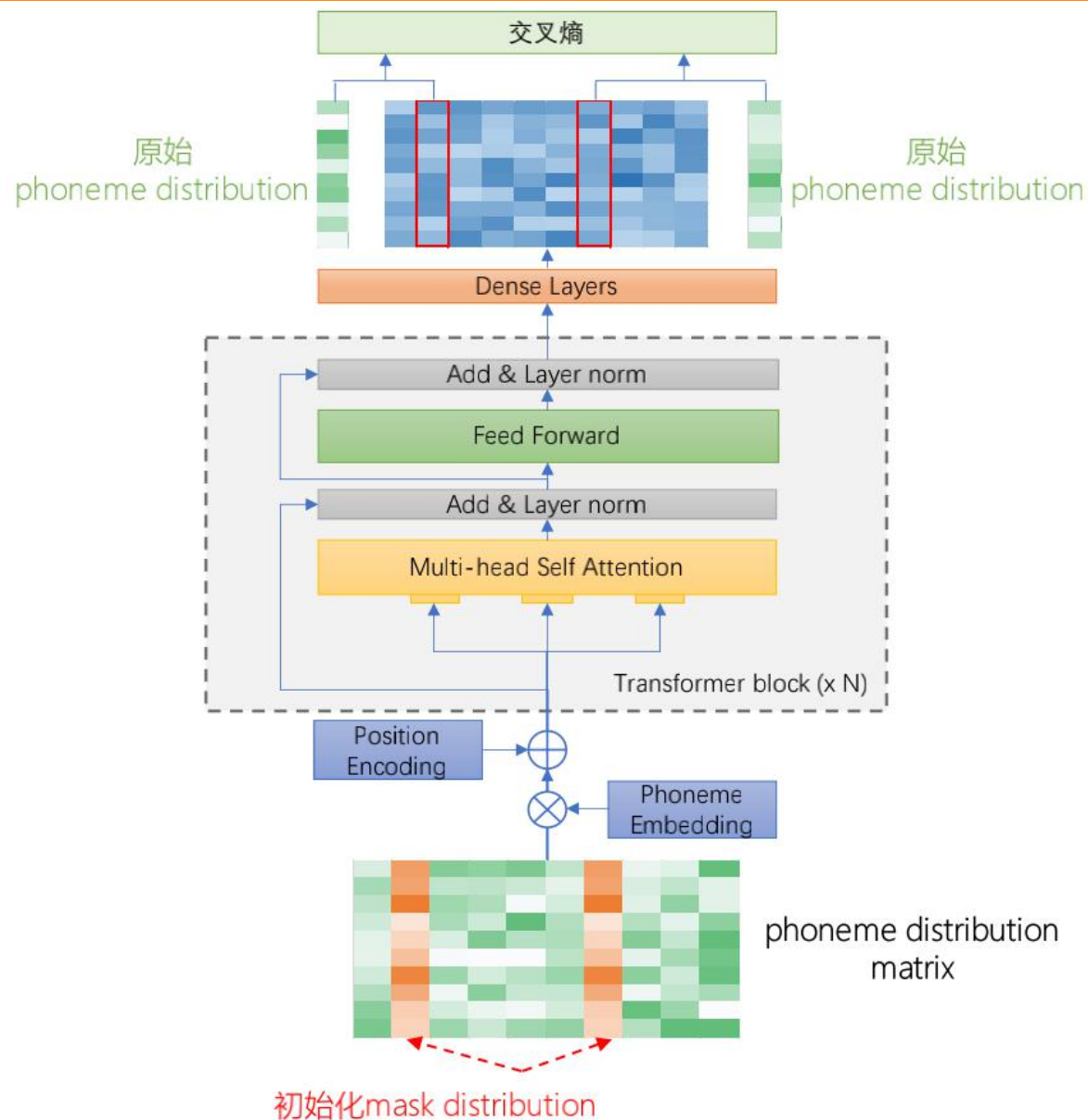
基于音素后验的无监督语义预训练 (Phoneme-BERT) Alibaba

➤ 文本预训练

- 输入输出为token序列
- 序列中每个token都有意义
- 单个token包含信息量多
 - 中文: 2500-7000
 - 英文: 40,000 Tri-letter: 10,000+

➤ 音素后验预训练

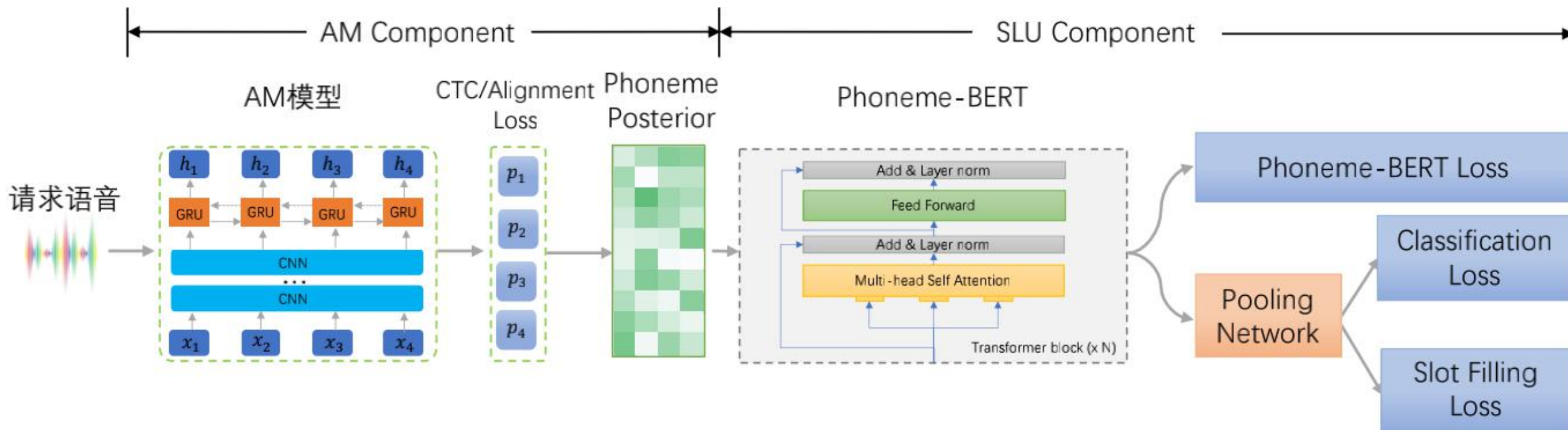
- 输入输出为分布式矩阵 (长度X音素个数)
- 分布中包含大量静音帧
 - Sil音素维度归一大于阈值a
 - 所有音素维度归一最大值小于阈值b
- 单个向量分布包含信息量少
 - 中文音素: 121
 - 英文音素: 42
 - 片段式mask



Phoneme-BERT

Alibaba

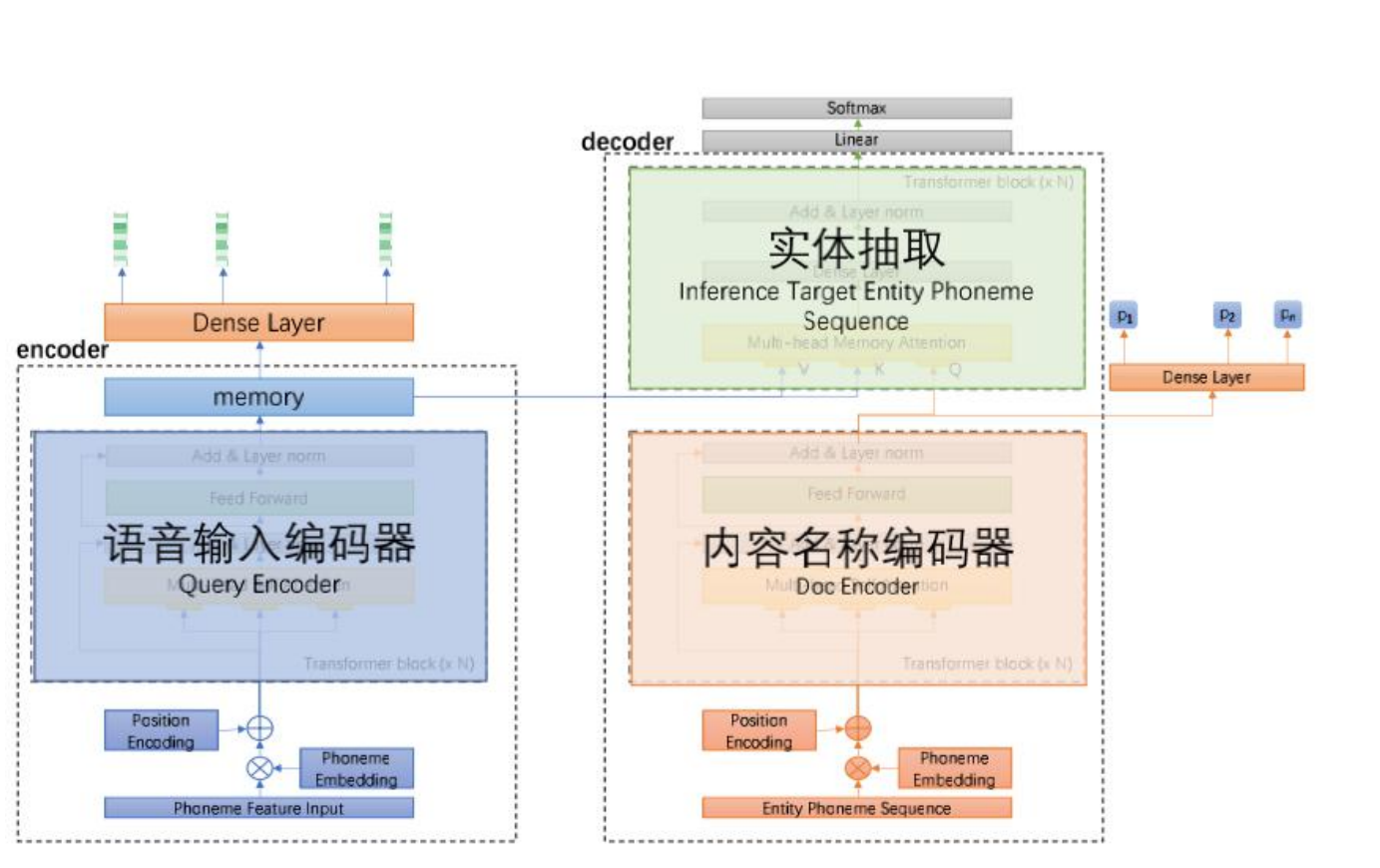
➤ 端到端Phoneme-BERT SLU模型



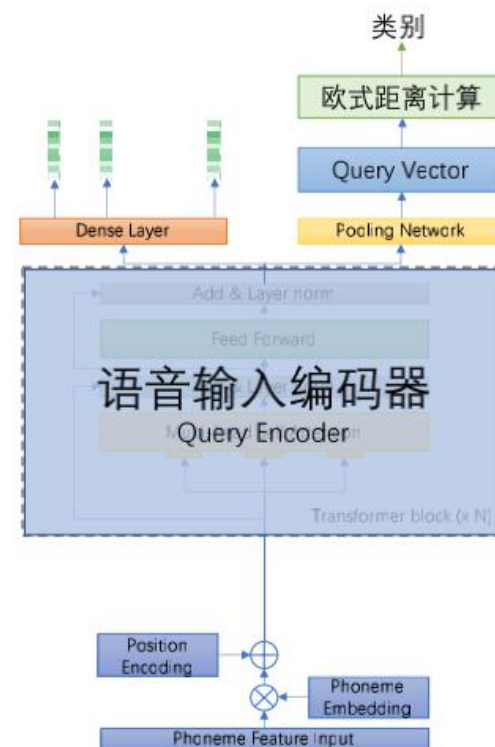
Phoneme-BERT

Alibaba

➤ Phoneme-BERT Fine-Tuning



Speech2Slot翻译模型 Fine-Tuning



分类任务 Fine-Tuning

➤ 实验结果（SF模型）

➤ 预训练数据：

- 类 型：天猫精灵音乐领域
- 数据量：600,000

➤ Fine-Tuning数据

- 类 型：音乐领域标注数据
- 数据量：20,000

Model	Accuracy Delta
Non-Pre-training	-
BERT	+6.41%
ERNIE	+7.42%

Speech2Slot模型

➤ 实验结果（分类模型）

➤ 预训练数据：

- 类 型：互联网开源英文语音
- 数据量：4,087小时

➤ Fine-Tuning数据

- 类 型：Fluent Speech Commands^[1]
- 数据量：23,100

Model	Error Rate	Delta
SOTA	1.2%	--
Non-Pre-training	1.95%	-62.50%
BERT	1.05%	12.5%
ERNIE	0.98%	18.33%

➤ 预训练数据：

- 类 型：天猫精灵语音数据
- 数据量：20,000+小时

➤ Fine-Tuning数据

- 类 型：天猫精灵领域分类数据
- 数据量：20,000

Model	Accuracy Delta
ASR+NLU	-
Non-Pre-training	+4.82%
BERT	+11.14%
ERNIE	+11.81%

欢迎加入天猫精灵人工智能部

Alibaba

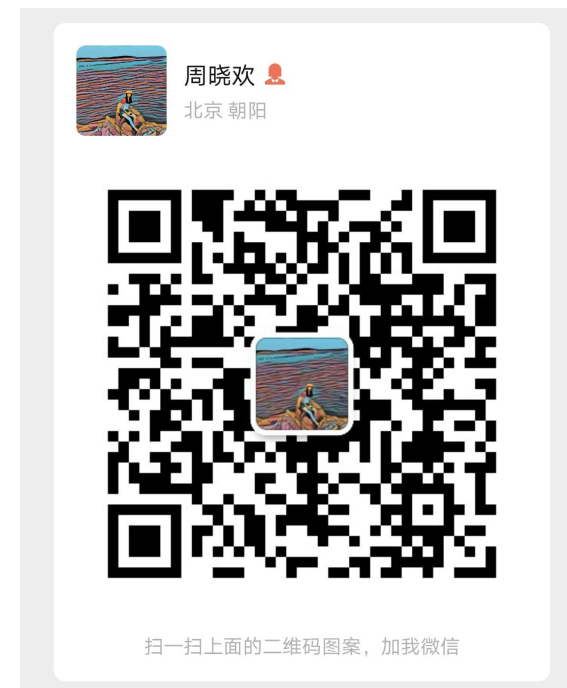
➤ 人工智能部

- 语音识别团队
- 图像识别团队
- 语音合成团队
- 自然语言理解团队
- 内容点播推荐团队

Thanks!



校招群



欢迎交流联系



Thanks!