BBS-KWS

大主干、关键词偏移及混合音节建模

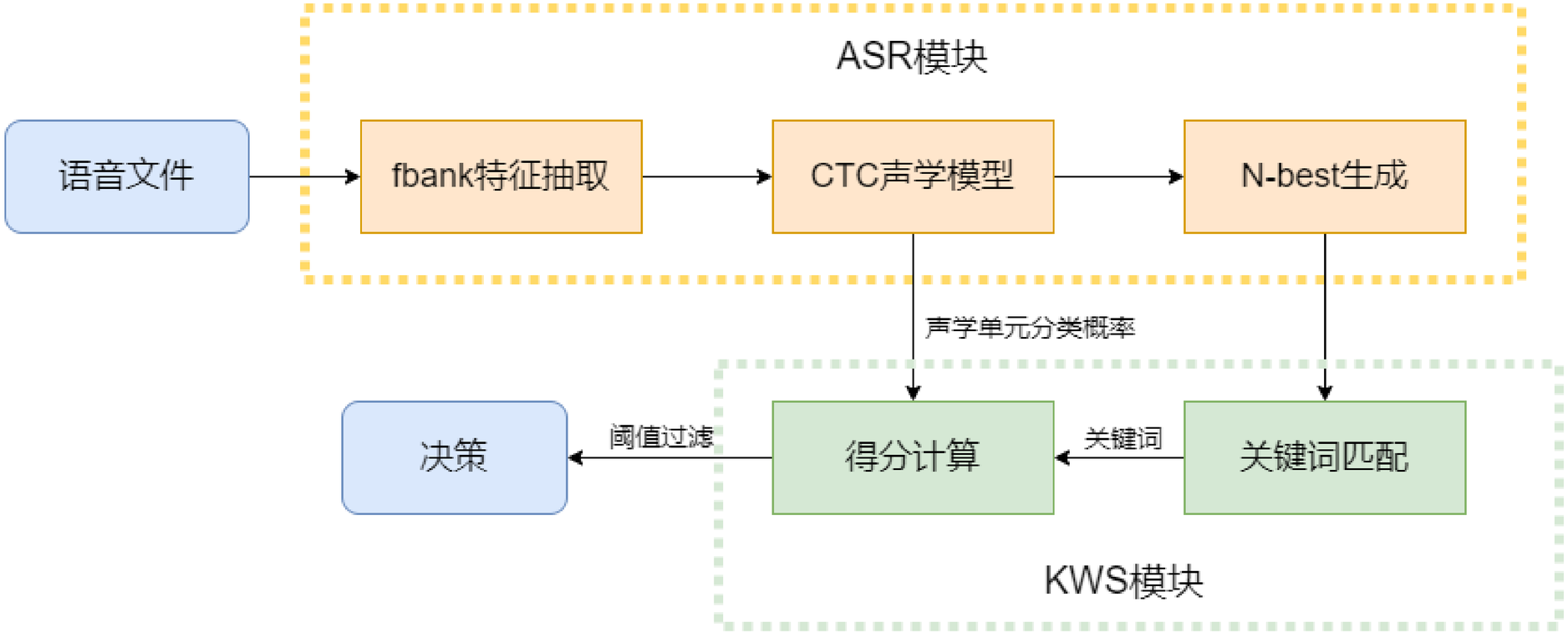
杜彬彬 张潆心 杨玉婷 汪文轩

纲要

- **BBS-KWS结构**
- ASR模块
- KWS模块
- 半监督学习
- 总结与展望

BBS-KWS结构

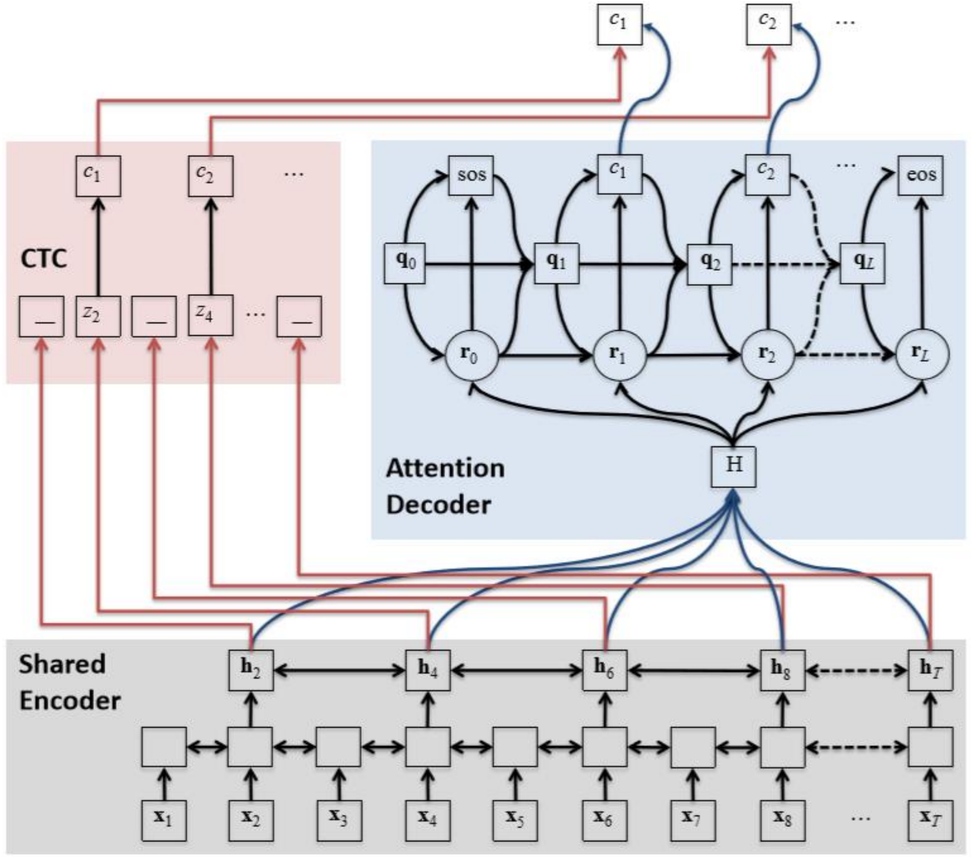
- **Motivation：**使用E2E ASR的技术栈替换传统算法功能
- ASR模块：生成N-best候选，使用端到端算法搭建
- KWS模块：负责关键词匹配与打分决策



纲要

- BBS-KWS结构
- **ASR模块**
- KWS模块
- 半监督学习
- 总结与展望

声学模型



- Hrbrid CTC/Attention训练

$$loss = \lambda loss_{ctc} + (1 - \lambda) loss_{att},$$

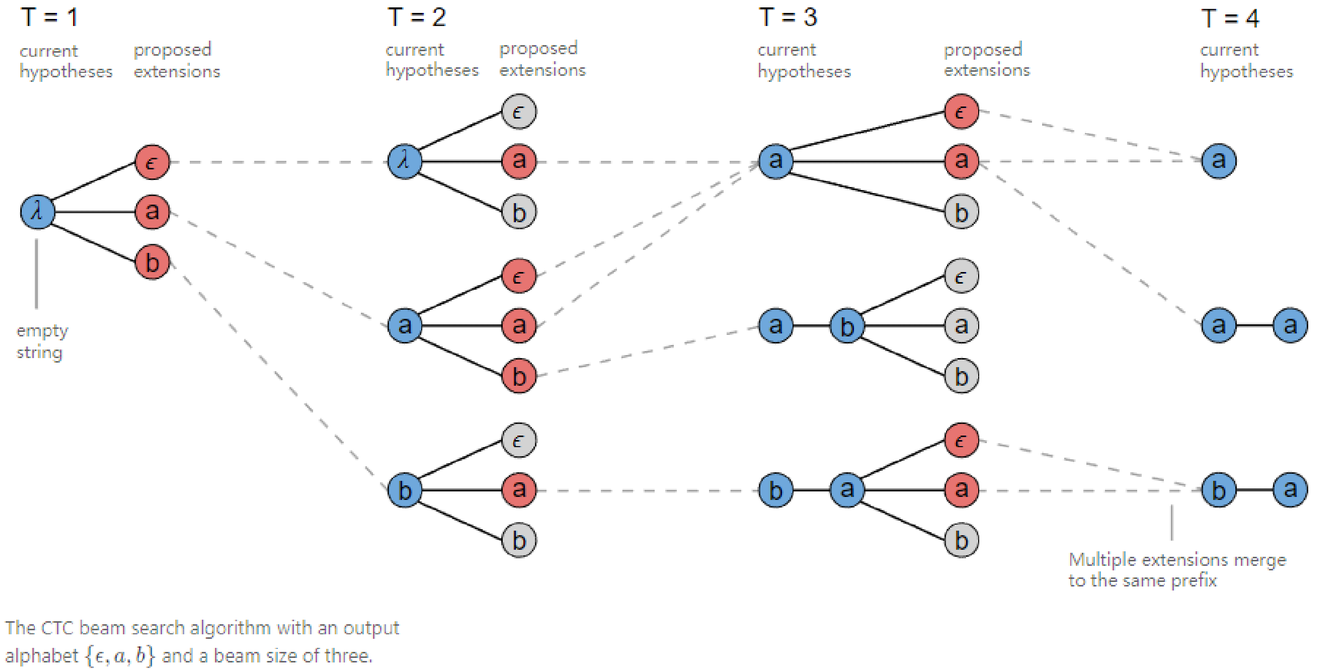
where λ = 0.9

- CTC分支解码

Watanabe S, Hori T, Kim S, et al. Hybrid CTC/attention architecture for end-to-end speech recognition[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(8): 1240-1253.

语言模型与解码

- 解码：ctc prefix beam search
- 语言模型： 4-gram，kenlm训练无剪枝，约8G
- 语料：维基百科、微博、豆瓣、网易新闻

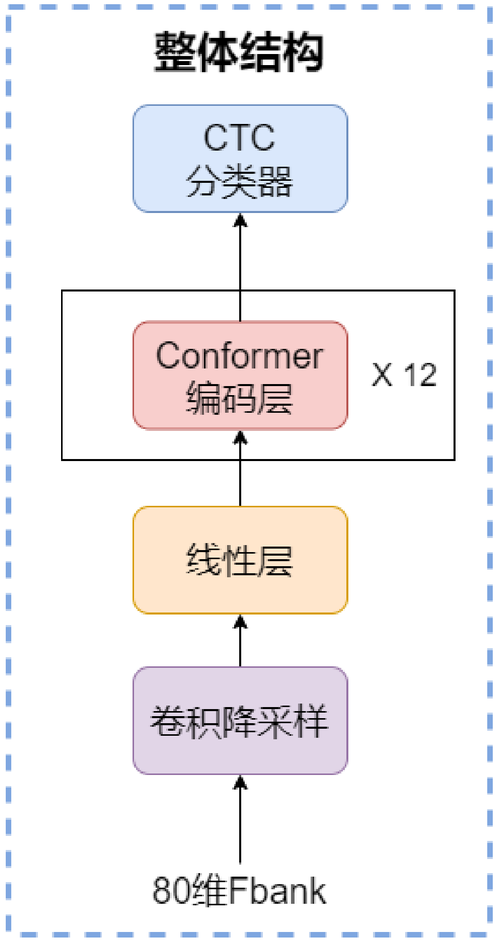


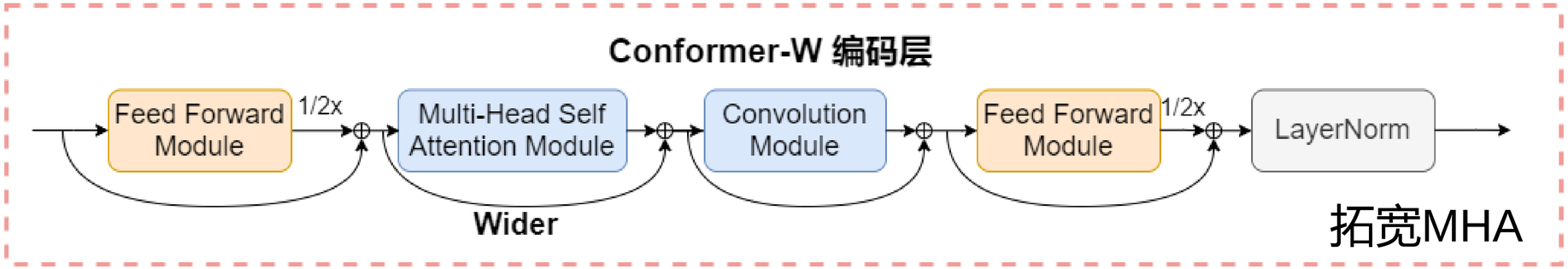
The CTC beam search algorithm with an output alphabet $\{\epsilon, a, b\}$ and a beam size of three.

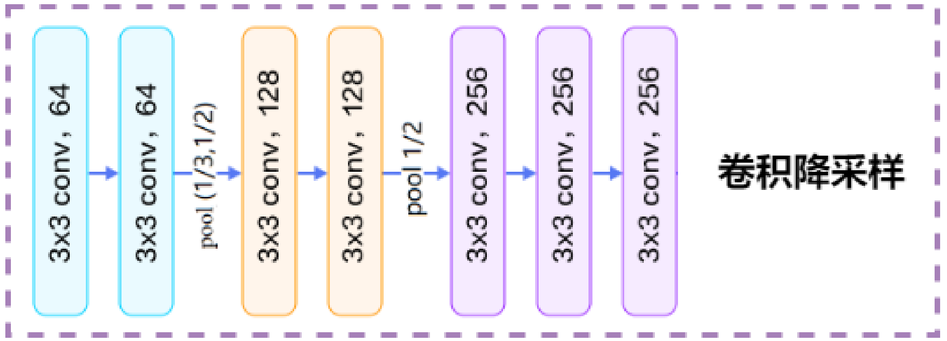
图片：https://distill.pub/2017/ctc/

Big Backbone

- 编码器：Conformer-W结构







时间轴下采样6倍

Gulati A, Qin J, Chiu C C, et al. Conformer: Convolution-augmented transformer for speech recognition[J]. arXiv preprint arXiv:2005.08100, 2020.

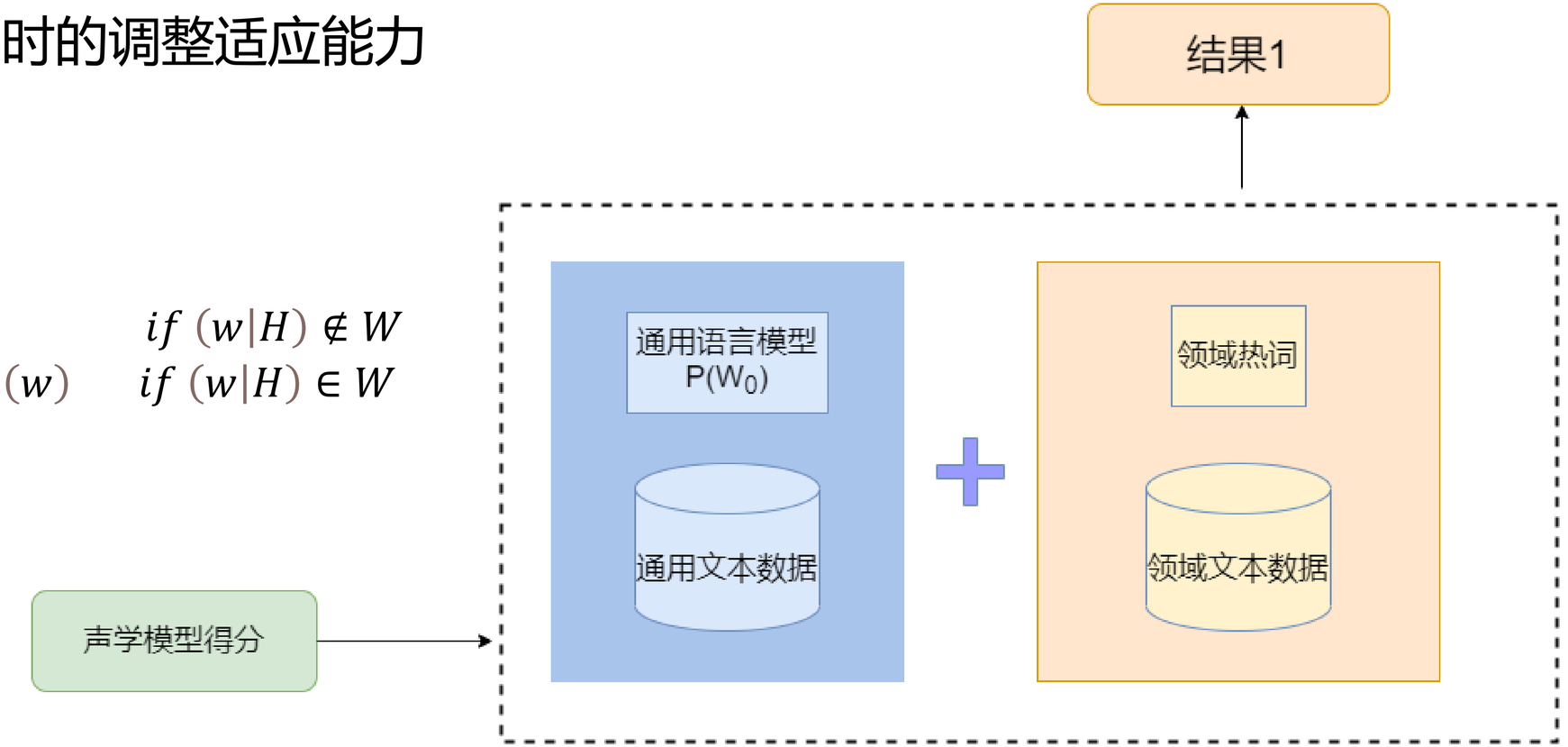
Biasing Keywords

- **Motivation**：增强解码时的调整适应能力
- 自定义关键词增强

$$s(w|H) = \begin{cases} s_G(w|H) & if\ (w|H) \notin W \\ s_G(w|H) + s_B(w) & if\ (w|H) \in W \end{cases}$$

- 自适应关键词权重

$$s_B(w) = -\alpha s_G(w) + \beta$$

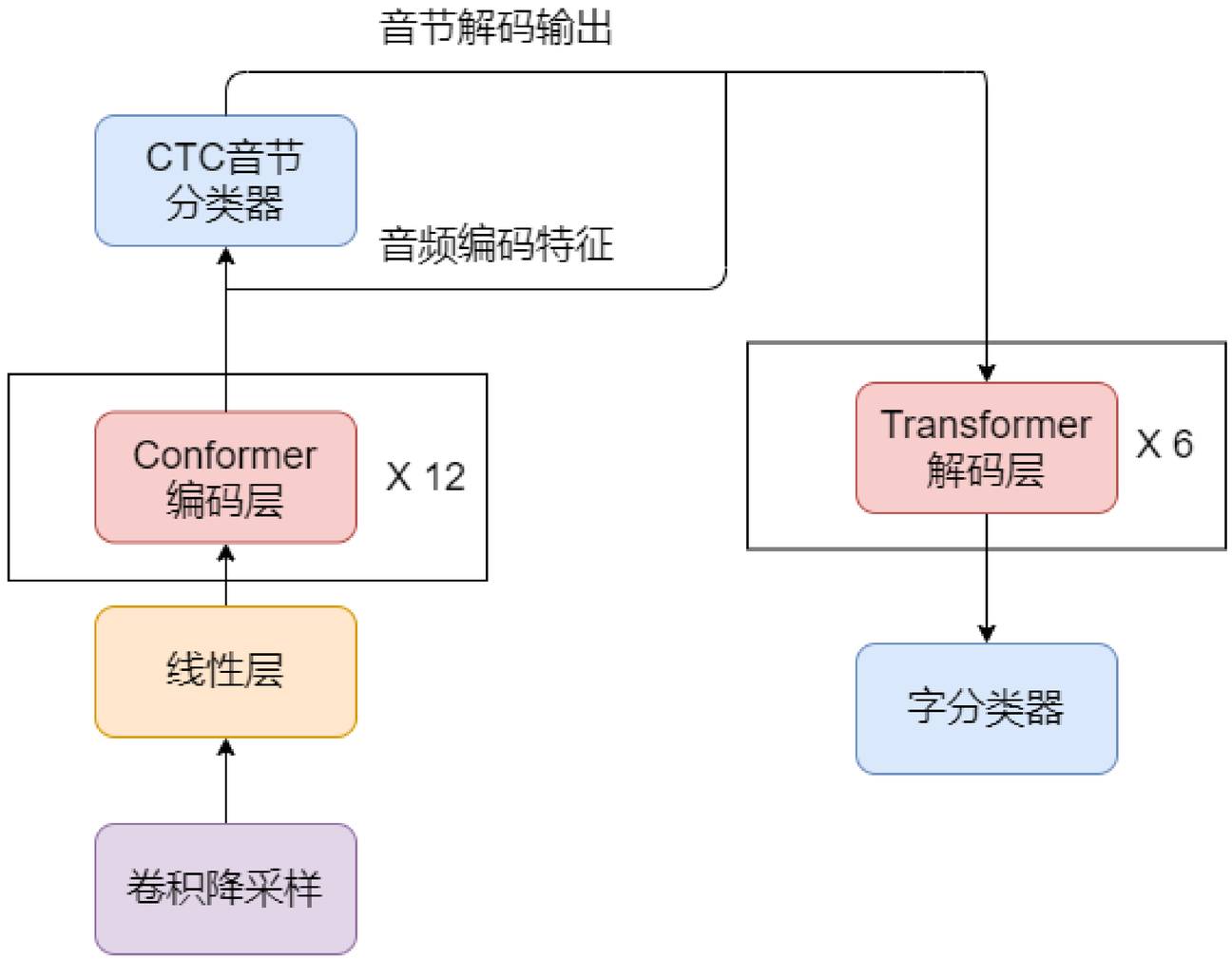


Mixed Syllable Modeling Units

- **Motivation：**更小的建模单元鲁棒性更好
- 混合syllable+character建模

Utterance：互联网 ➡ Syllable：hu4 lian2 wang3

- 模型设计：CTC出音，attention出字

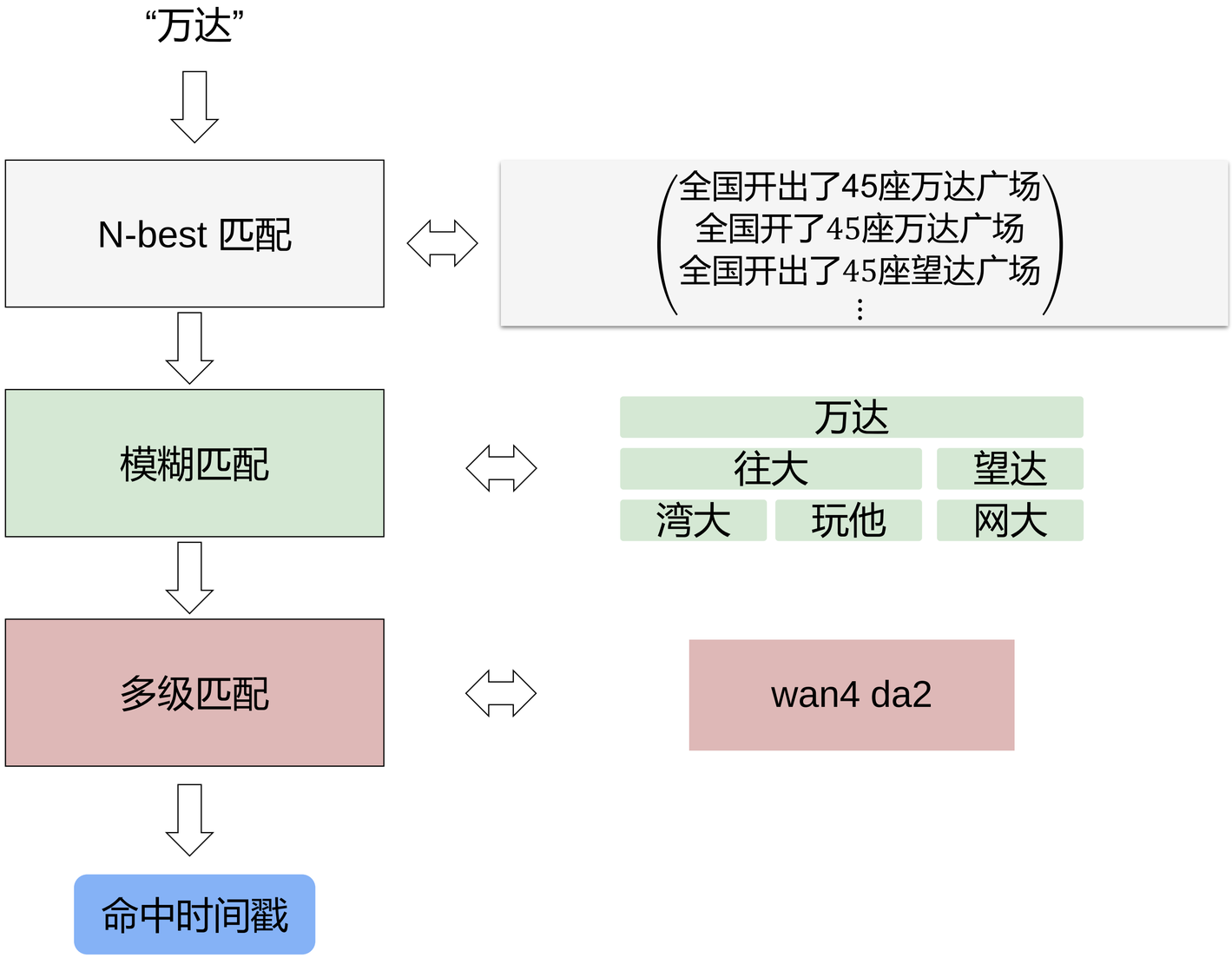


纲要

- BBS-KWS结构
- ASR模块
- **KWS模块**
- 半监督学习
- 总结与展望

匹配策略

- **Motivation：**处理OOV
- 取解码器的N-best结果匹配
- 使用代理词匹配
- 使用音节结果匹配



Li M, Danilevsky M, Noeman S, et al. Dimsim: An accurate chinese phonetic similarity algorithm based on learned high dimensional encoding[C]. Proceedings of the 22nd Conference on Computational Natural Language Learning. 2018: 444-453.

关键词打分

- 计算声学模型中映射到kw的所有路径π的得分之和

$$score(kw) = \sum_{\pi:B(\pi)=kw}^{from\ t_s\ to\ t_e} P(\pi)$$

- 与音节得分取最大值

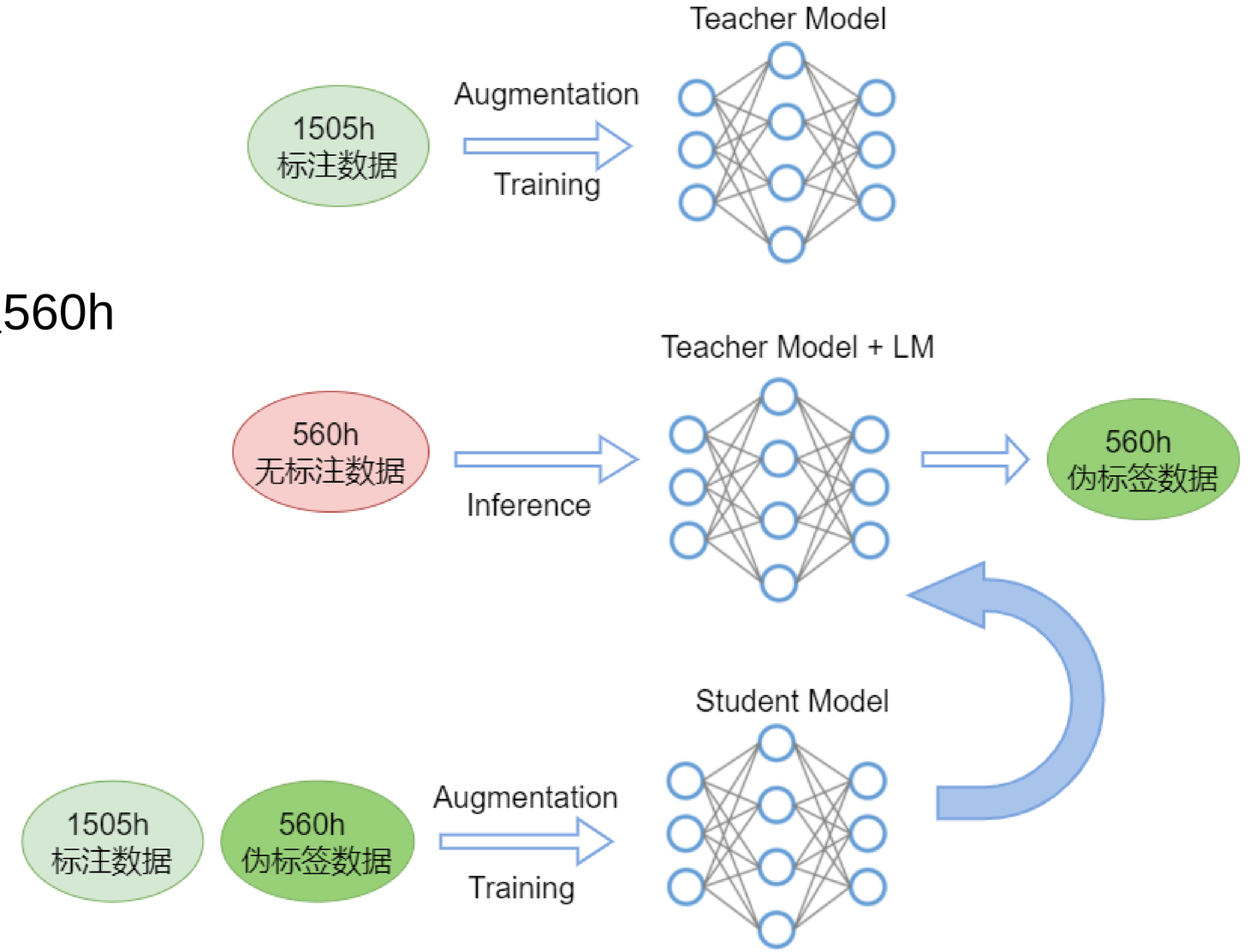
$$score(kw) = \max(score_{char}(kw), score_{syllable}(kw))$$

纲要

- BBS-KWS结构
- ASR模块
- KWS模块
- **半监督学习**
- 总结与展望

半监督学习

- **Motivation：**缓解场景失配问题
- 数据：CN-Celeb数据中随机抽取560h
- 自动打标：模型+LM解码
- 模型结构：全部为Conformer-W
- 循环轮次：4次，每次从M_0开始



纲要

- BBS-KWS结构
- ASR模块
- KWS模块
- 半监督学习
- **总结与展望**

实验结果

BBS-KWS实验对比

方法	Lgv场景（F1+ATWV）/2	提升幅度
官方baseline	59.76	-
Conformer-W 字建模	65.90	+6.14
+ 语言模型	70.89	+4.99
+ 长度归一化	74.70	+3.81
+ beam匹配	76.33	+1.63
+ 一轮SSL	78.68	+2.35
+ 关键词偏移	80.79	+2.11
+ 共四轮SSL	81.61	+0.82
+ 模糊匹配	83.79	+2.18
+ 混合音节建模	85.18	+1.39
+ 模型融合	85.59	+0.41

总结与展望

- **优势：**

  依托E2E ASR技术栈搭建，方便快捷

- **不足：**

  精度模块：BSS-KWS整体设计偏向召回，缺乏一个提升精度的模块

  数据利用：对其他开源有标签数据的利用不足，对TTS利用不足

谢谢！