20190804

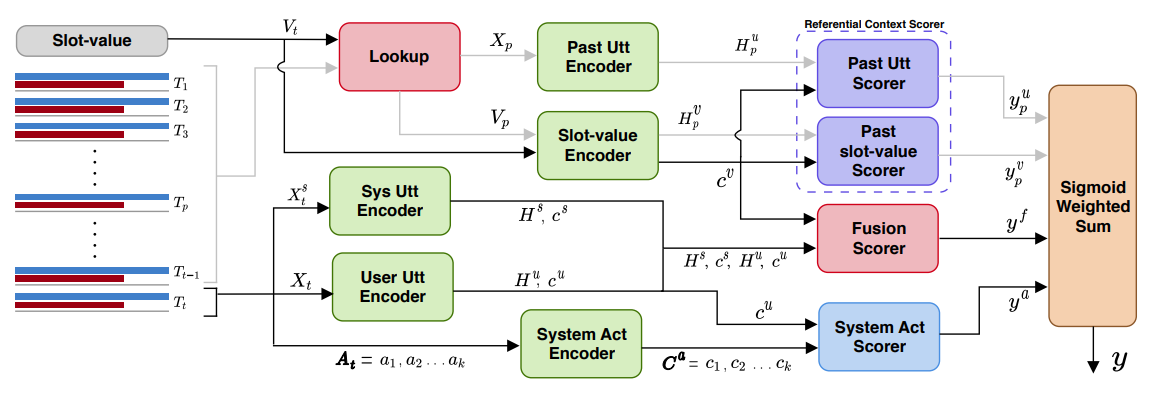
基于上下文辨识的对话状态追踪（NAACL2019）

Sharma S, Choubey P K, Huang R. Improving Dialogue State Tracking by Discerning the Relevant Context[J]. arXiv preprint arXiv:1904.02800, 2019.

数据集：WoZ2.0，MultiWoZ2.0(Restaurant)

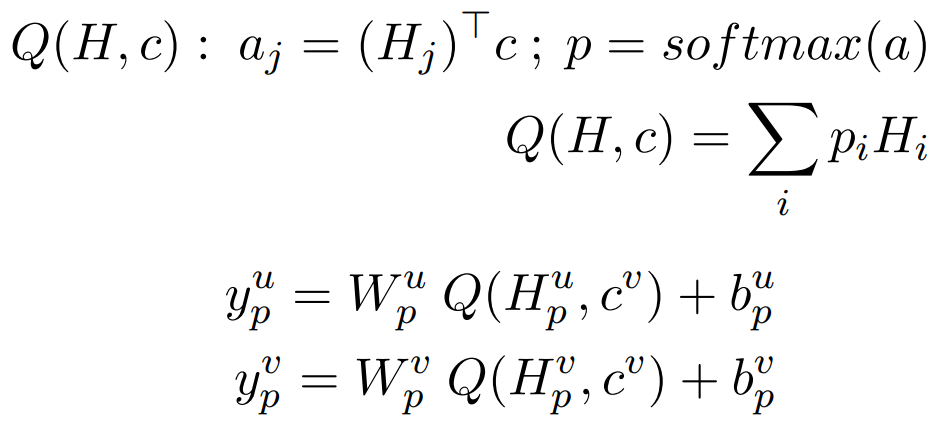
在对话状态追踪任务中，对于slot-value转换，即话题转换时的对话状态追踪一直是一个难点，由于话题转换过程中会存在引入多余噪声的问题，使得这个问题一直没有被妥善解决。对于解决这种任务的模型的评估表明，确定相关背景对于追踪对话状态极为重要。该模型致力于识别非相邻对话中的重要细节，参考历史轮对话的记录，达到了比之前的SOTA（GLAD）更好的效果。

文章大体的思路和之前的许多工作类似，先将DST的多标签分类问题分解为二分类问题，对于每轮对话，对每个value-slot分别进行评分，取高于阈值的分数加入当前对话状态中。为了保证选择的正确性，使用当前论用户话语和之前轮的系统话语组成的融合话语进行预测。模型的结构如下：

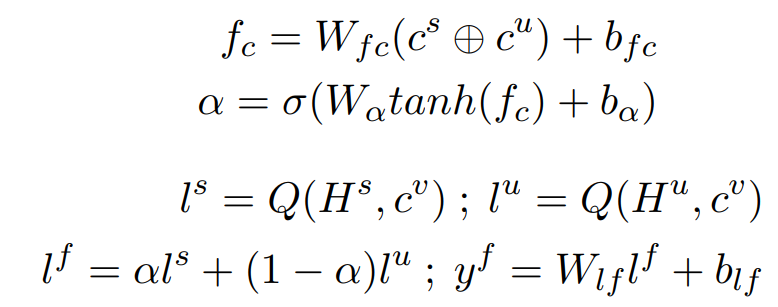


该模型分为以下几个部分：

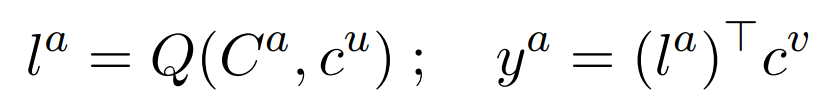
1. Lookup Module：存放用户的历史论对话。
2. GLE Module：图中绿色的部分都是GLE（Global local self-attentive Encoder） Module的部分，即之前说的GLAD模型的全局-局部注意力机制的内容。该模型输入用户对话文本，输出隐状态向量和上下文表示向量。
3. Referential context scorer：根据候选slot-value对，用户历史对话和先前的slot-value对确定候选的slot-value对是否在历史轮对话中有引用。然后对各个slot-value对进行打分。



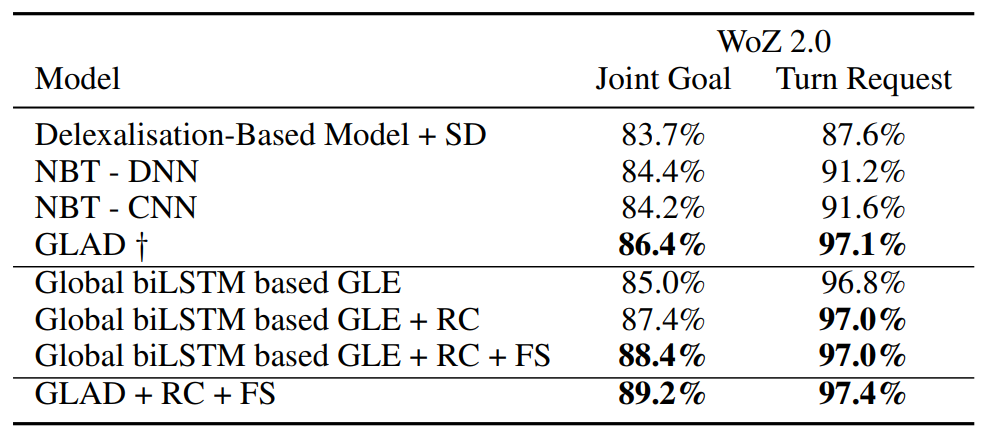
1. Fusion Scorer：该融合打分模块用于计算上下文摘要的表示，最终计算基于当前的用户表示和前轮用户表示的候选slot-value对的打分：



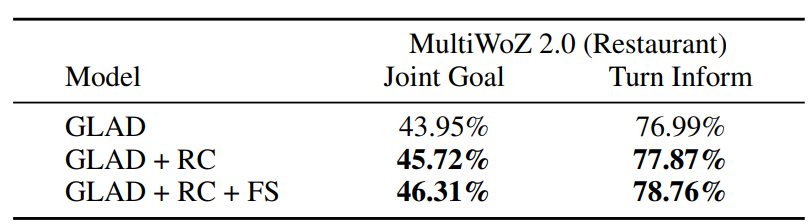
1. System Act Scorer：计算系统的动作摘要(Action Summary)：



该实验的结果分析如下：



该模型与ＧＡＬＤ相比达到了更好的效果。



同时研究者对比了GALD模型和该模型在multiWoZ 2.0上的表现，也表现了很好的提升。但是在8月1日论文（Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems）的结果中的模型结果更好。但两者跑的GLAD结果不同，无法比较。