

Report

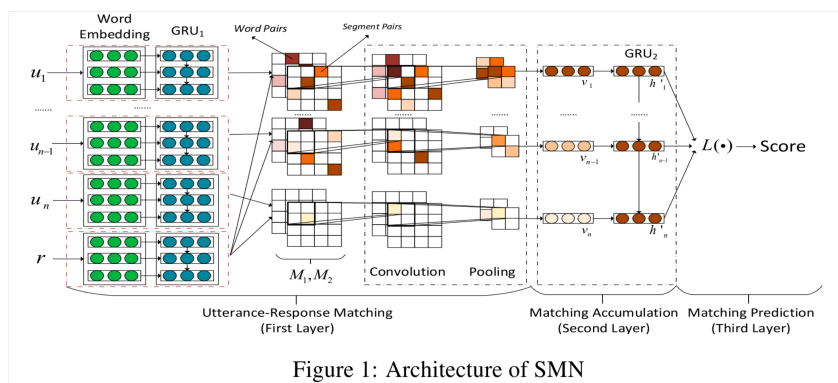
Han Peng, JianFei Chen, Han Li

2019 年 4 月 18 日

1 Model Description

本此实验基于 SMN model(Sequential Matching Network A New Architecture for Multi-turn Response Selection in Retrieval-based chatbots)，针对模型中部分具体实现细节进行了考量，并尝试修改或添加部分组件，以期获得更好的效果。

Sequential matching network (SMN)



ACL 2016

Figure 1: Sequential Matching Network

由上图所示，在 First Layer 中模型构建出了 M1 和 M2 两个矩阵，并将其堆叠以进行卷积操作。M1 和 M2 矩阵的构造方式如下图所示。

由上可知，M1 Matrix 与 M2 Matrix 不同，在构造过程中其并未使用

Specifically, $\forall i, j$, the (i, j) -th element of \mathbf{M}_1 is defined by

$$e_{1,i,j} = e_{u,i}^\top \cdot e_{r,j}. \quad (1)$$

Figure 2: M1 Matrix

hidden vectors of \mathbf{R} . Then, $\forall i, j$, the (i, j) -th element of \mathbf{M}_2 is defined by

$$e_{2,i,j} = h_{u,i}^\top \mathbf{A} h_{r,j}, \quad (3)$$

Figure 3: M2 Matrix

Bilinear 的方式并加入 A Matrix。所以在本次任务中，尝试仿照 M2 为 M1 添加了相同的矩阵组件，并观察效果。

此外，本次实验针对 Embedding 模块，采取了两种不同的处理方式：其一是较为简单的随机初始化，其二为统计出本次任务出现的单词，并依照 Tencent Embedding Matrix(Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings) 制作了符合本次任务的 Embedding Matrix(218563*200 维)，并允许其能够在模型训练中依照梯度更新参数。

最后指出，在本次任务中，模型的 Utterance 中最多包含 10 句话 (如果 utterance 句子数量少于 10，则在距离 response 最远处添加只包含 0 的句子；如果句子数量大于 10，则删除距离 response 最远的几个句子)；在每句话中最多包含 20 个单词 (单词不足则向后补零，单词过多则删除句尾)；embedding size 为 200 维；在 First Layer 中，GRU 的 hidden size 为 200 维；在 Second Layer 中，GRU 的 hidden size 为 50 维。

2 Experiment Result

2.1 Experiment 1

在本次实验中，模型使用了 attention matrices 并且未使用 Tencent Embedding。如 Figure 4 所示，模型在到达第五个 epoch 时效果最佳，此

时 $P@1$ 达到最大, 为 0.4062, 其他指标有: $MAP = MRR = 0.5697$, $P@1 = R10@1 = 0.4062$, $R10@2 = 0.5412$, $R10@5 = 0.7894$ 。

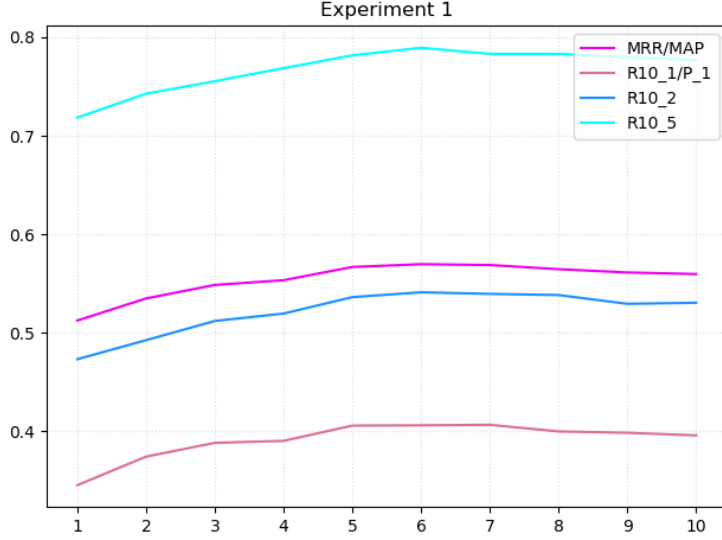


Figure 4: Experiment 1

2.2 Experiment 2

在本次实验中, 模型未使用 attention matrices 并且未使用 Tencent Embedding。如 Figure 5 所示, 模型在到达第五个 epoch 时效果最佳, 此时 $P@1$ 达到最大, 为 0.4144, 其他指标有: $MRR = MAP = 0.5740$, $P@1 = R10@1 = 0.4144$, $R10@2 = 0.5411$, $R10@5 = 0.7921$ 。

可以得出结论, 在不使用 attention matrices 时, 模型的效果会更加优秀。

2.3 Experiment 3

在本次实验中, 模型未使用 attention matrices 并且使用了 Tencent Embedding。如 Figure 6 所示, 模型在到达第二个 epoch 时效果最佳, 此时 $P@1$ 达到最大, 为 0.5489, 其他指标有: $MAP = MRR = 0.6942$, $R10@1 = P@1 = 0.5489$, $R10@2 = 0.7013$, $R10@5 = 0.8986$ 。

可以得出结论, 使用预先训练好的 embedding matrix 不仅会加快模型的收敛速度, 还能够大幅度提升模型的效果。

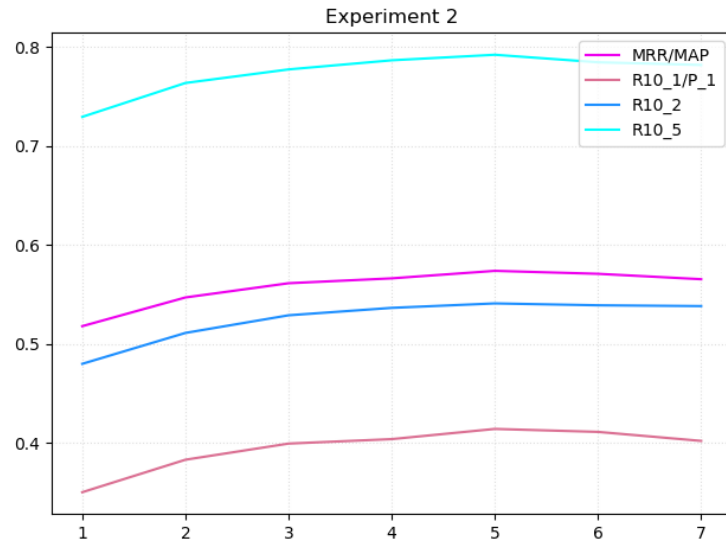


Figure 5: Experiment 2

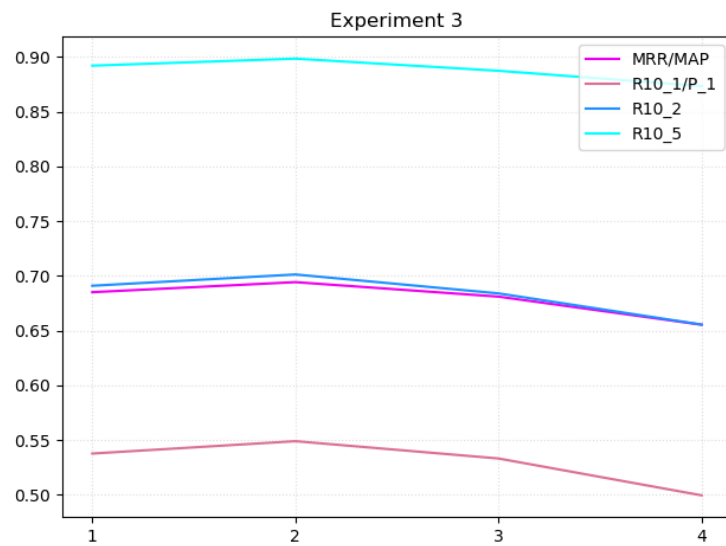


Figure 6: Experiment 3