□□□


TAC
腾讯广告算法大赛
Tencent Advertising Algorithm Competition

Contents

目录

02

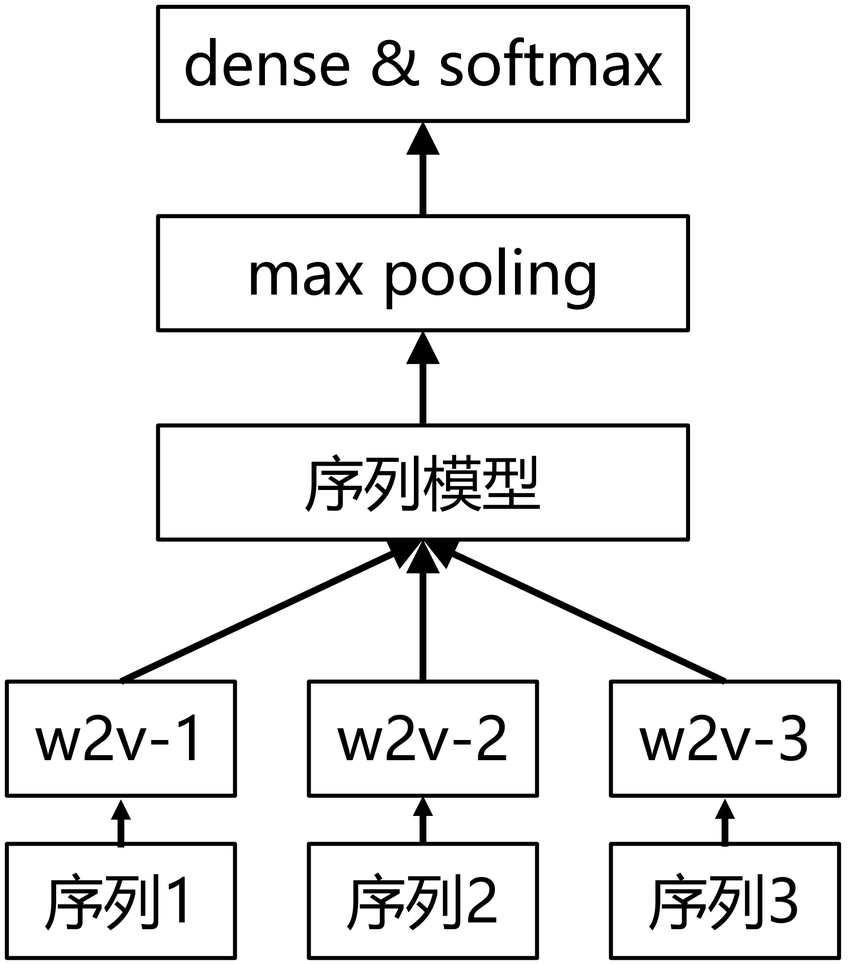
赛题理解

赛题理解

- **任务**：根据用户点击的广告预测其年龄段（10分类）和性别（2分类）
- **数据**：用户点击的广告点击序列，及广告的一些属性信息，如广告主ID、商品ID等
- **指标**：两个分类任务分别计算准确率后相加作为分数

- **基本思路**：从序列分类的角度，利用NLP文本分类技术解决
- **OOV问题**：利用word2vec做预训练，解决数据中存在的OOV问题
- **模型融合**：建立多种具有差异性的模型，借助模型融合进一步提升效果

模型介绍



1. 以用户点击的广告序列为句子，训练素材ID、广告主ID等序列的词向量
2. 选取部分序列的词向量，进行拼接，作为模型输入，词向量在训练过程中冻结不更新
3. 利用lstm或transformer等模型，建立20分类的模型（年龄10分类 * 性别2分类）
4. 预测时对20分类的概率进行聚合再得到最终预测结果

$$P(\text{gender } = 1) = \sum_{i=1}^{10} P(\text{gender} = 1,\ \text{age } = i)$$

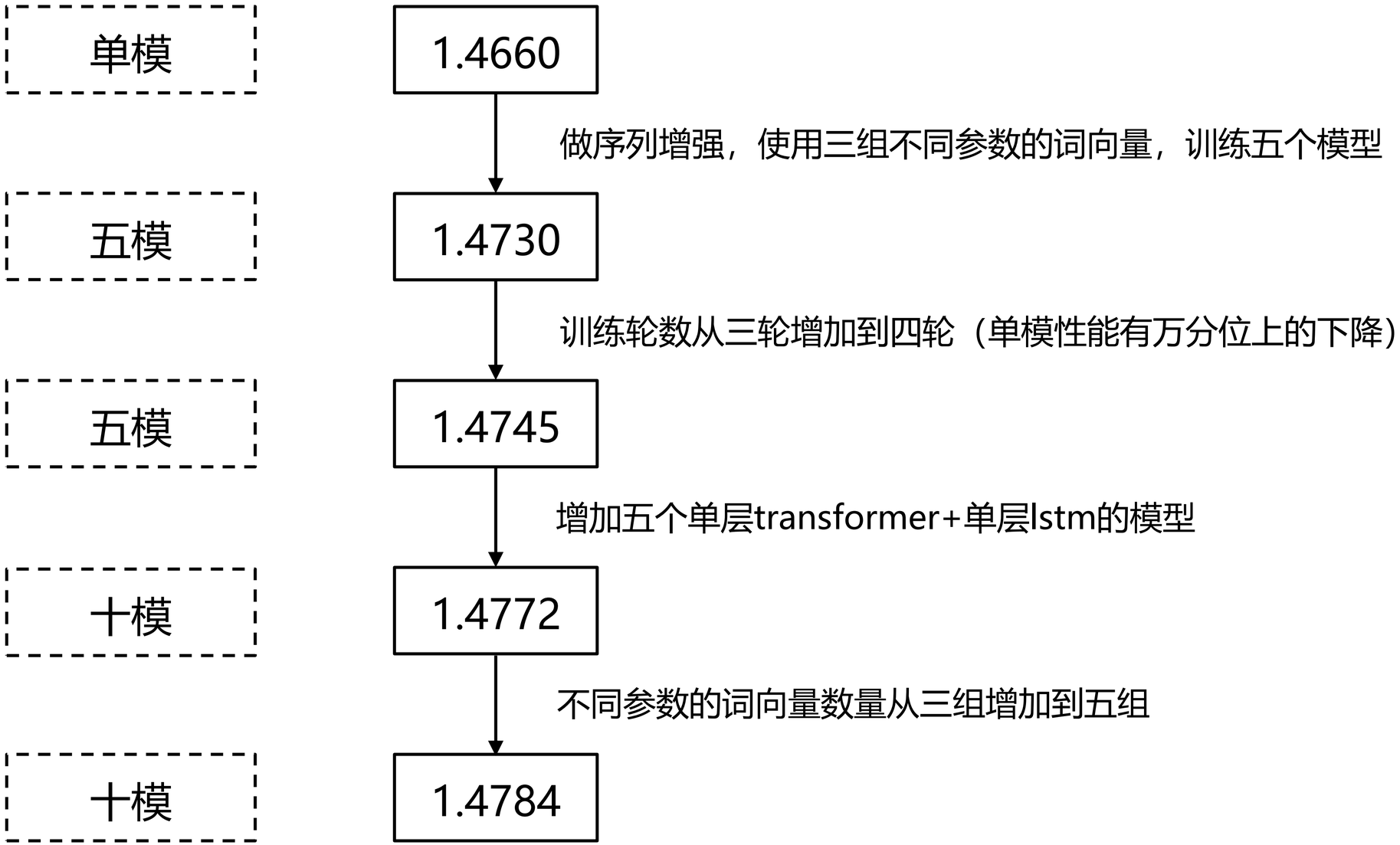
$$P(\text{age } = 1) = \sum_{j=1}^{2} P(\text{gender} = j,\ \text{age } = 1)$$

- 模型融合可以有效提升模型效果
- 如果模型相互之间差异很小，那么融合收益会很快降低为0
- 单模分数越高，并不代表融合也能越高

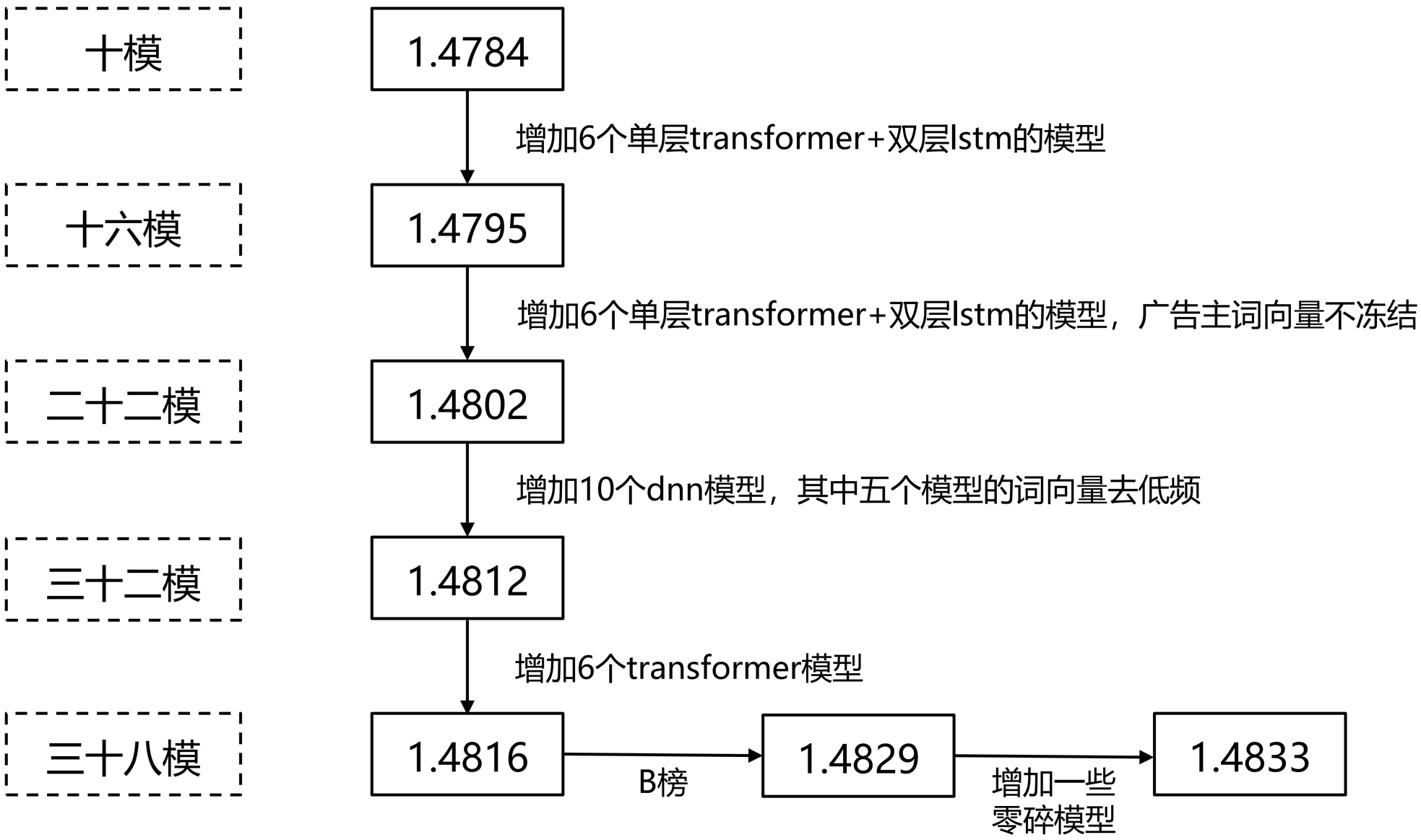
如何有效提高融合效果?

- 模型方面：多种不同的模型，如lstm、transformer、dnn、lstm+transformer等
- 训练策略：多训练一个epoch，牺牲单模性能换取融合效果
- 序列方面：对于过长的序列，使用多种不同的截断类型；对同一用户同一天点击的广告进行shuffle
- 词向量方面：使用多种参数组合训练得到的词向量；部分模型中取消对广告主ID的词向量冻结；部分模型使用去低频的词向量

模型介绍——模型融合


单模
1.4660
做序列增强，使用三组不同参数的词向量，训练五个模型
五模
1.4730
训练轮数从三轮增加到四轮（单模性能有万分位上的下降）
五模
1.4745
增加五个单层transformer+单层lstm的模型
十模
1.4772
不同参数的词向量数量从三组增加到五组
十模
1.4784

模型介绍——模型融合


十模
1.4784
增加6个单层transformer+双层lstm的模型
十六模
1.4795
增加6个单层transformer+双层lstm的模型，广告主词向量不冻结
二十二模
1.4802
增加10个dnn模型，其中五个模型的词向量去低频
三十二模
1.4812
增加6个transformer模型
三十八模
1.4816
B榜
1.4829
增加一些
零碎模型
1.4833

- Scaled Softmax： 不以e为底数的Softmax $\frac{e^{ax_j}}{\sum_{i=1}^{m} e^{ax_i}}$
- 将点击次数取embedding后乘到词向量上
- Transformer的attention矩阵计算时考虑点击次数 $\frac{a_j e^{x_j}}{\sum_{i=1}^{m} a_i e^{x_i}} = \frac{e^{x_j + \log a_j}}{\sum_{i=1}^{m} e^{x_i + \log a_i}}$
- 调整不同序列的词向量长度
- 特征工程
- User Embedding
- BERT
- ……

05

总结与思考

方案总结

- 利用word2vec和多种序列分类模型进行建模
- 探索多种不同的方案增加模型差异性，提高最终的融合效果

思考

- 如何更好地利用多个序列的信息？多种序列的embedding维度和长度比？
- 序列信息在这个任务中的重要性有多大？
- 单模效果与融合效果的一致性？
- ……

THANKS