山有木兮


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

Contents

目录

团队介绍

团队成员简介

团队介绍

林有夕
算法工程师

孙泽勇
广州工业大学 计算机硕士

唐 静
同济大学 计算机硕士

一台2080Ti

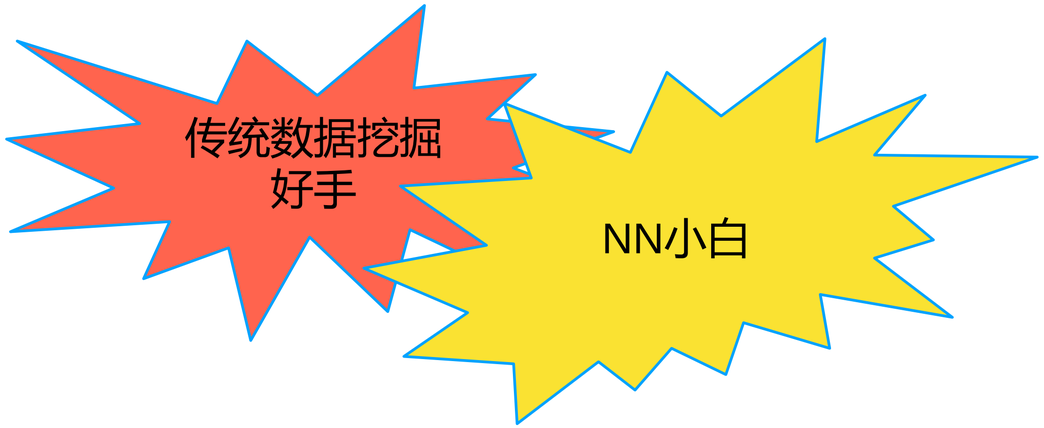
2019 & 2020 DCIC 数字中国创新大赛 卫冕冠军

2020 PAKDD 天池阿里巴巴智能运维大赛 冠军

芒果TV 国际音视频算法大赛 冠军

DC厦门国际银行“数创金融杯”数据建模大赛 冠军

KDD CUP、CCF、天池等十余次亚军

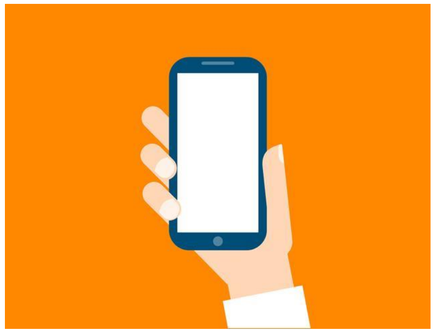


02

赛题理解

序列预测

赛题理解 - 序列 多/二分类



用户浏览广告，产生用户的广告序列

产生数据

Day1 点击序列 … …

Day2 点击序列 … …

⋮ ⋮

Day n 点击序列 … …

根据用户点击的广告序列，预测用户的性别、年龄信息

预测

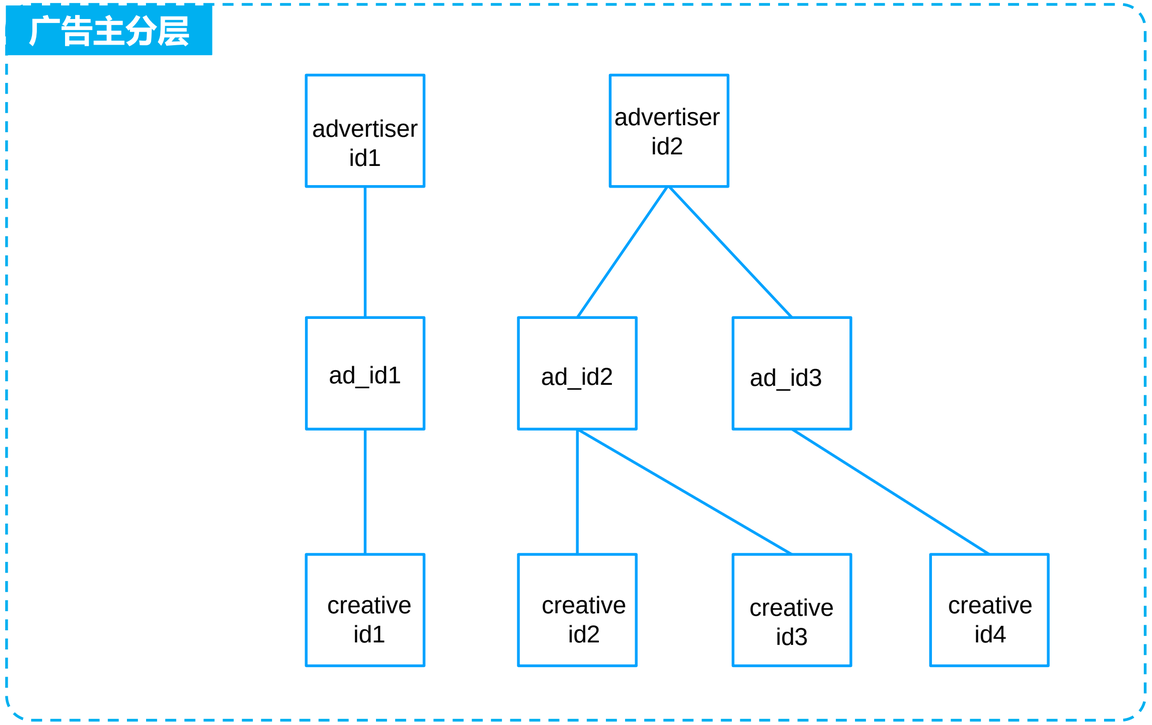
性别?

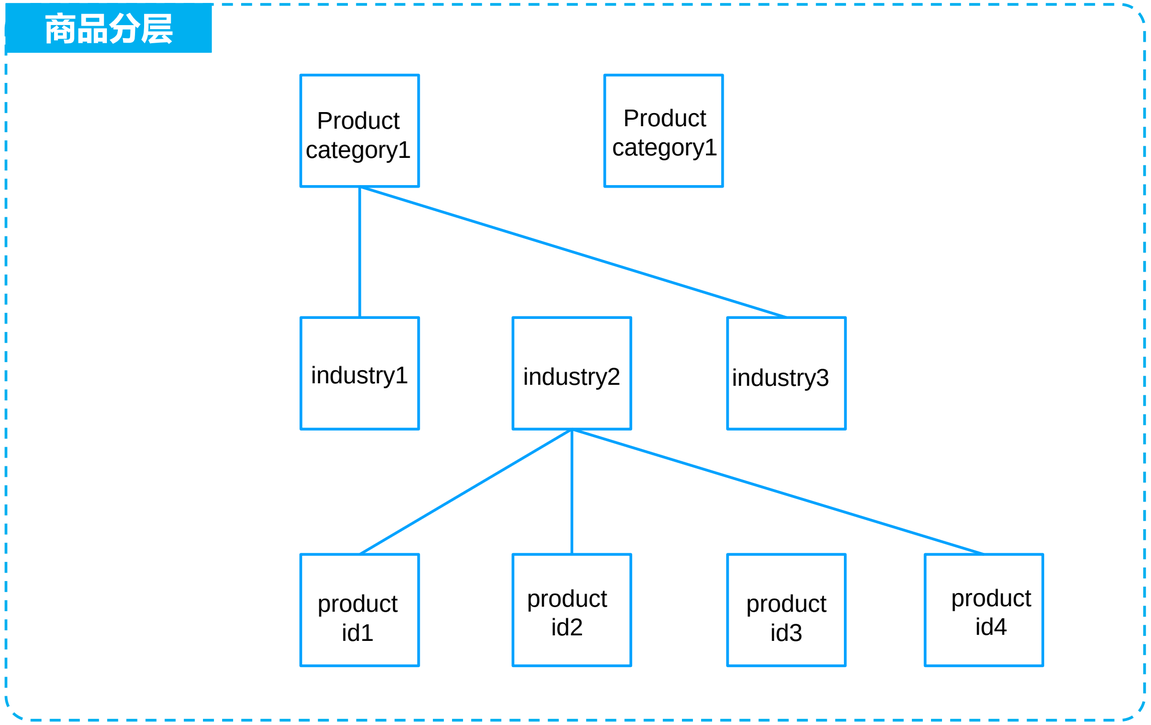
年龄?

特征工程

分层结构/稀疏特性/目标编码

赛题理解- 分层结构





广告稀疏属性具有层级结构，为一对多的关系。

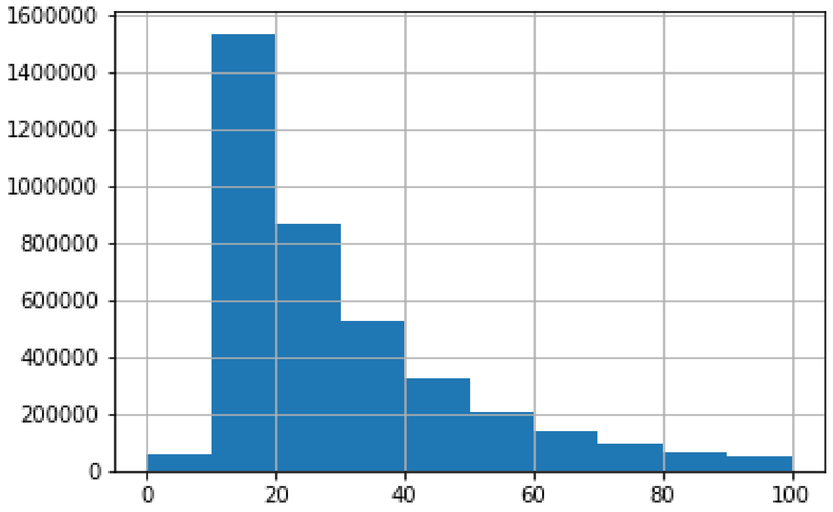
分层编码：将一对一的词使用同一个id进行编码。可以无损的降低一半词表规模。

低频hash：词表大小大幅度降低。

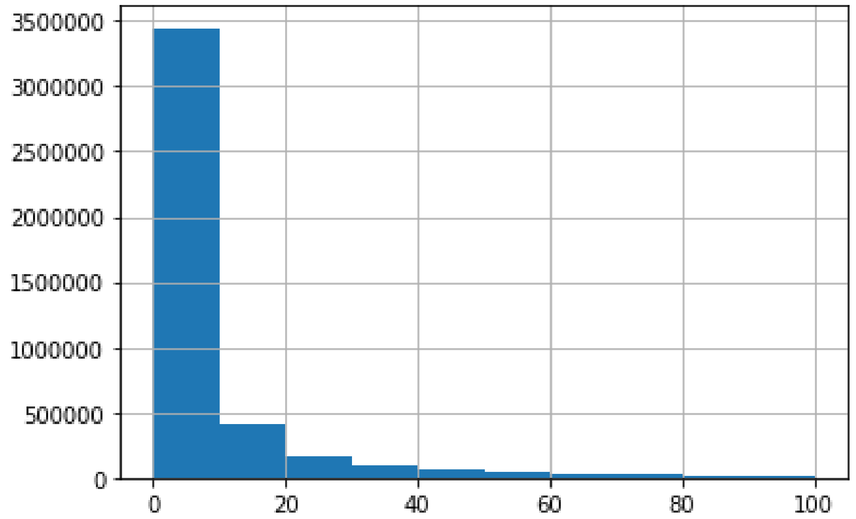
fea_name	advertiser_id	ad_id	creative_id
num	62965	3812202	4445720
fea_name	product_category	industry	product_id
num	18	335	44314

赛题理解- 稀疏特性

id长尾特性+为现实意义明确的实体的场景下：相似度流派完胜统计流派



用户行为稀疏



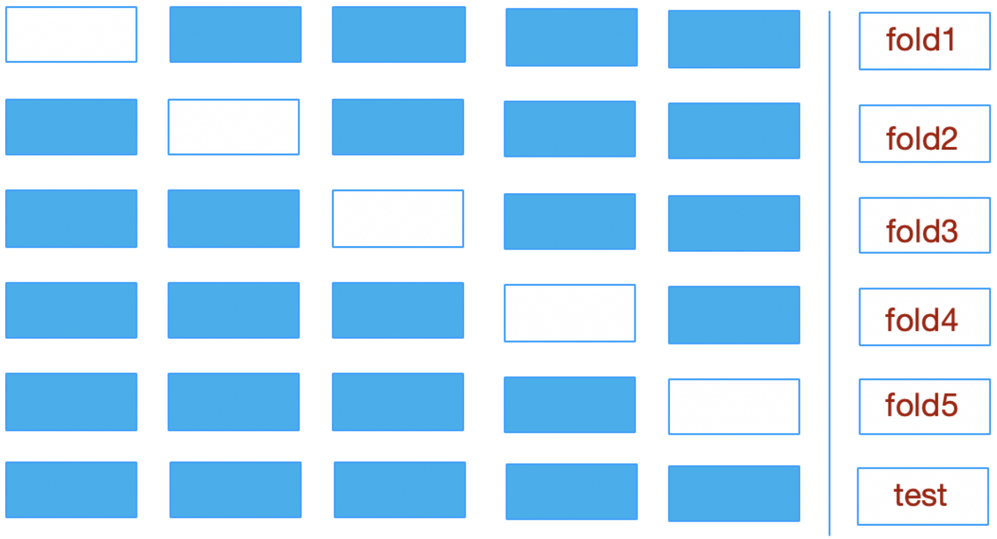
广告投放稀疏

id为现实意义明确的实体时，往往具有很丰富的信息，在分布较为稀疏时，
往往基于低频特征无法很好的学习到id的具体信息。所以需要稠密化转化。

赛题理解- 目标编码（Kfold VS 去一法）

最直接的做法：构建标签预测解，由模型实现平滑，并结合特征矫正

5折法



将样本划分为k份，对于其中每一份数据，我们都用另外k-1份数据提取标签分布特征，复杂度K * On

去一法

	click_times	creative_id	time	user_id	ad_id	advertiser_id	inc
0	1	821396	20	1	724607	7293	
1	1	209778	20	1	188507	9702	
2	1	877468	20	1	773445	29455	
3	1	1683713	39	1	1458878	14668	
4	1	122032	40	1	109959	11411	
...	...	...	...	...	...	...	
133878440	1	3596158	75	4000000	3096233	36668	
133878441	1	3642395	75	4000000	3135640	18422	
133878442	1	366858	76	4000000	331268	36890	
133878443	1	3333680	76	4000000	2868147	32830	
133878444	1	3697105	77	4000000	3181227	52421	

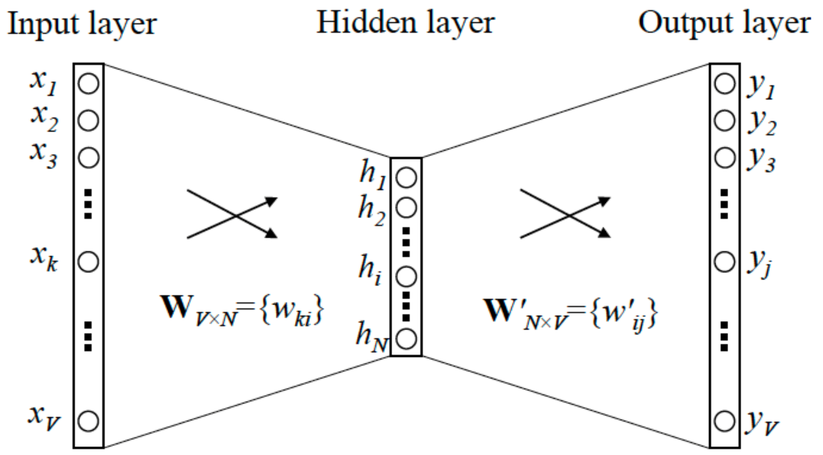
统计全局概率分布，去除当前行样本
复杂度On

模型介绍

w2v&层级结构/ bert or not bert / LINet

模型介绍- w2v&层级结构

W2V编码



不分日期训练——小窗口加速

同一个用户的点击商品序列作为word2vec模型的一个sentence

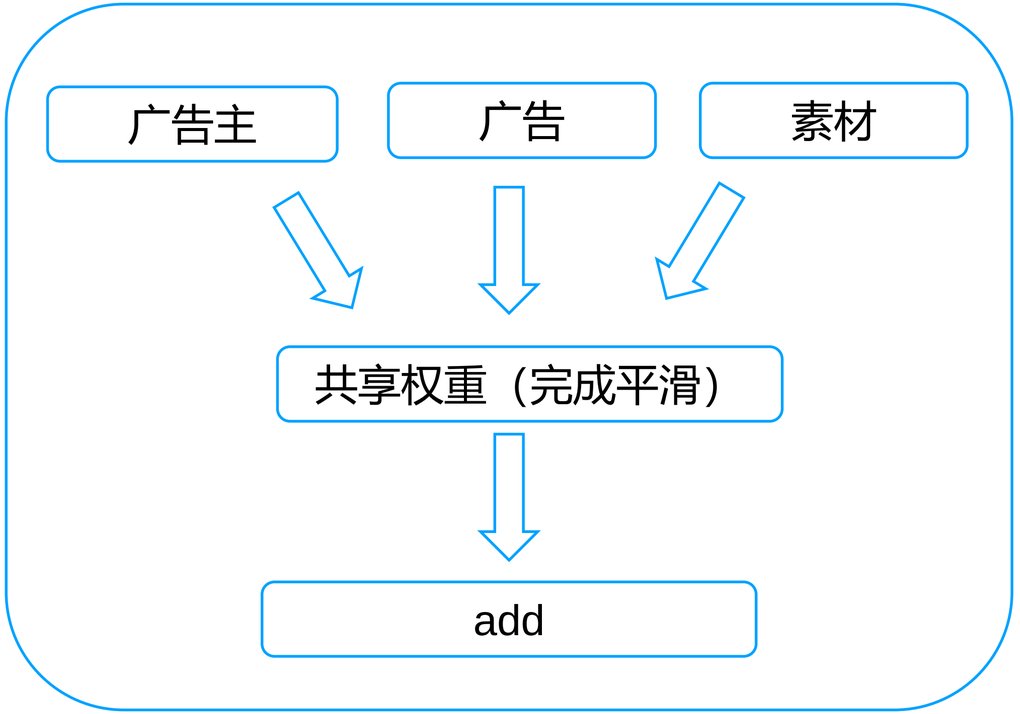
分日期训练——大窗口增加覆盖面

一天内，同一个用户的点击商品序列作为word2vec模型的一个sentence

Level Layer

层级模型，降低模型规模，提高提取器覆盖样本规模

Level Target encoding ＆ Level w2v encoding



层级结构数据处理模块

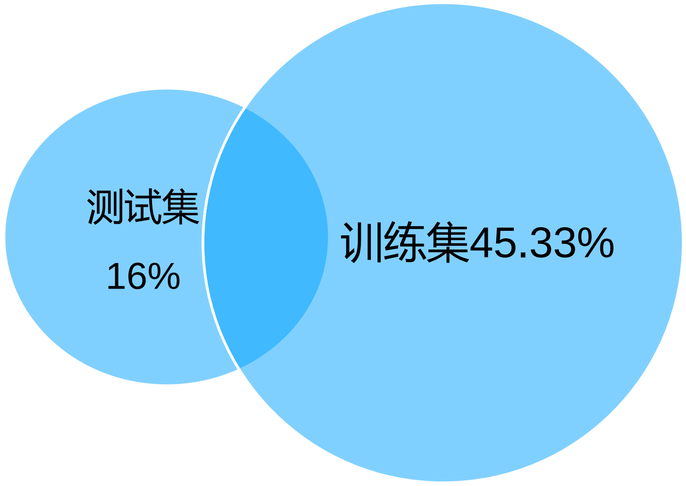
模型介绍- 泛化的启示

泛化问题

随机采样下，由于稀疏性带来的影响我们会得出以下结论：

1、16%的测试集广告，无法获得主体相关泛化信息。

2、训练集中45%的广告主体信息，无法泛化至测试集



素材ID 分布

典型场景

Q：用户点击广告c1 至 c9，其中c1、c9具有强标签相关信息，c2-c8弱相关或无关。如何建模?

A：一般结构Maxpool 提取强相关信息预测

1、好的模型需要在最坏的情况下提供预测的能力

2、无法保证同时学到c1、 c9与结果的关系，c1或c9满足预测能力之后，没有足够的loss供c2-c8学到与结果的关系。

3、 c1或c9和标签的关系。只能泛化到c1或c9所在样本

4、如果c2-c8存在稀疏问题。则无法从主体上，挖掘和目标的关系。

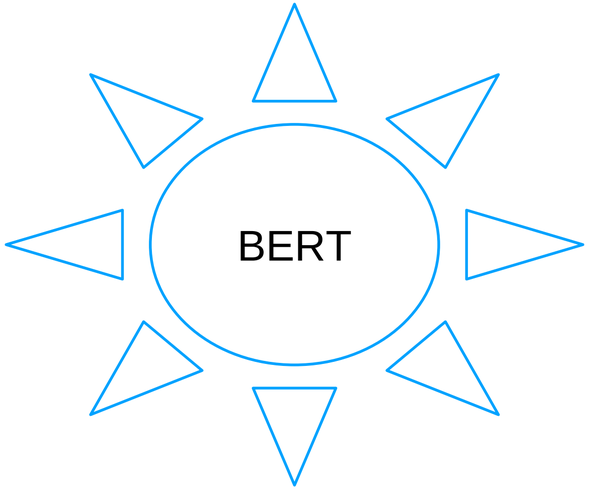
5、如何从训练集分布上，实现信息转移。实现泛化

模型介绍

why bert

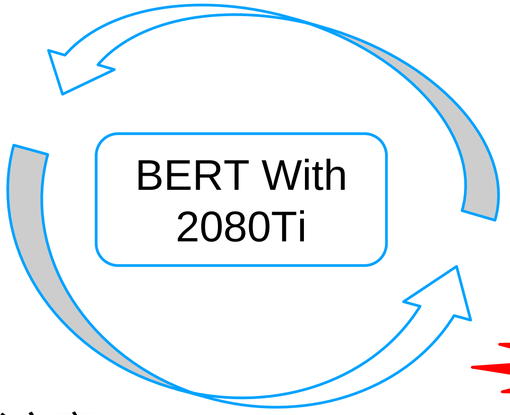
使用bert的几个明显信号

- 时间弱相关假设成立
- 样本信息过载
- 主体富信息



- w2v效果优异
- 低频词依赖环境注入
- 当下最牛预训练模型

why not bert



测试方案

等待反馈….
long times later….

out of memory

缺点：

训练时间长，调试成本高
需要内存较大,显存OOM

PS：尝试过embedding size 16 的ELMO需要运行约24hours

我想试试BERT

模型介绍- Level Inject Net

BERT 可以实现将词级别的**完整**信息注入，理想情况下可获得单个词的丰富的多维度信息，而针对当前场景，是否可以实现一种只将**target** 紧密相关的信息注入的方法？从而大幅度降低模型规模。

2020腾讯广告算法大赛官方交流群

预训练是不是普通机器跑不了啊?

有个四卡 应该可以跑的

我大概算了一下，跑bert-base一星期就差不多了

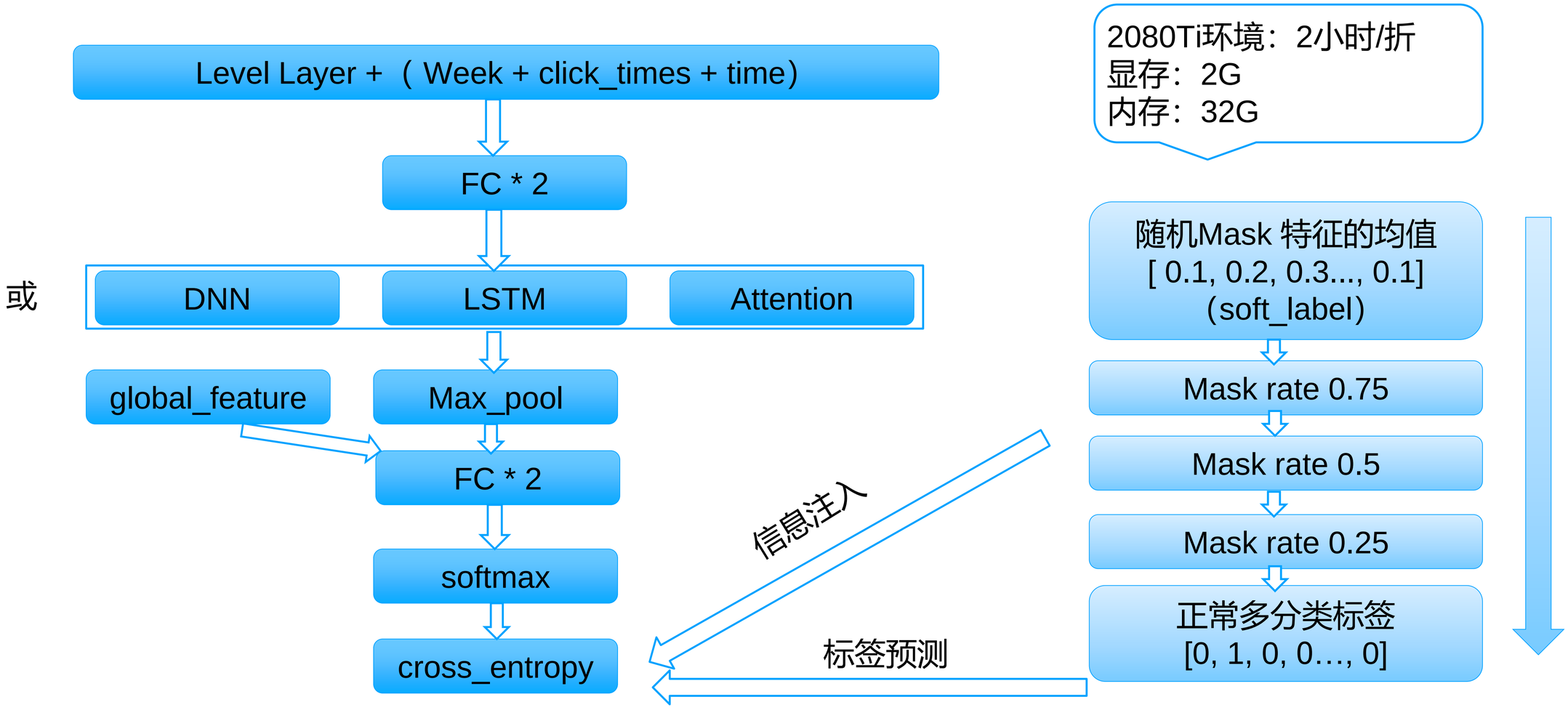
把需要的信息注入就可以了。全空间不见得好。

目标就预测个性别、年龄

看来自己造个针对这个数据集的玩法了

我想给他取名为，Focus\target bert

模型介绍- Level Inject Net



模型可以不切换状态连续训练，经测试发现，在标签预测阶段，学习率降低50倍，
效果明显，具有fine-turning的特性。

Mask rate 随epoch逐步降为0

模型介绍- Level Inject Net

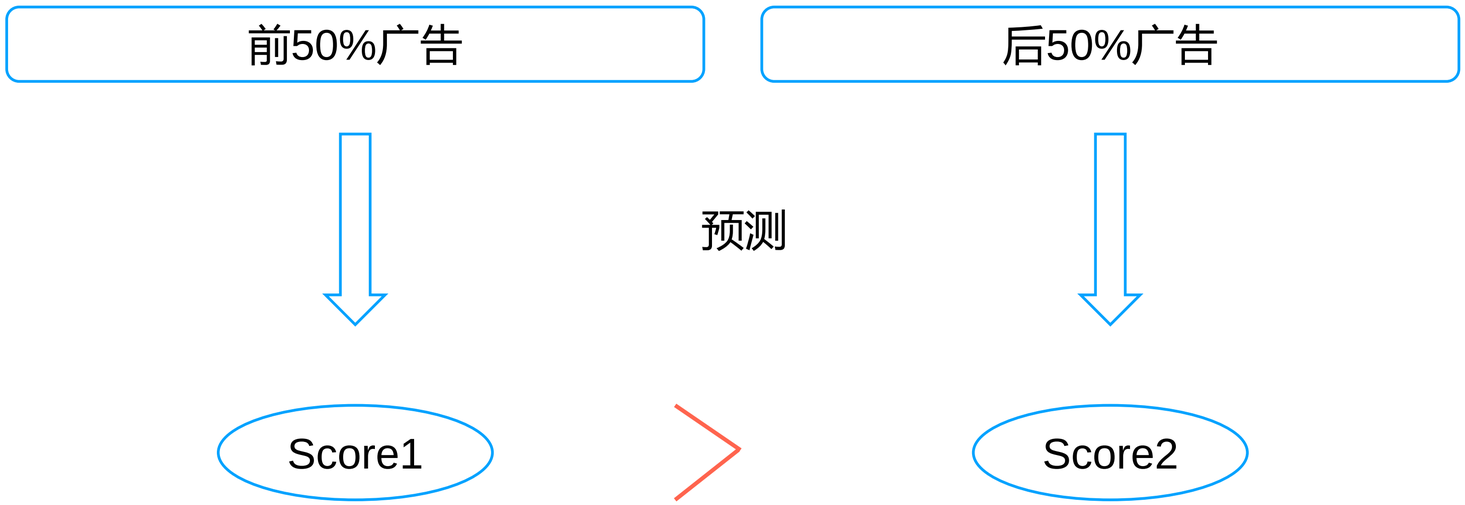
BERT 与 Target Inject对比

	BERT	Target Inject
目标	序列 mask的词 完成词级别的信息注入	序列 mask 的词的target encoding特征 完成target信息的注入
目标类型	序列多分类 （词粒度）	soft_label 多分类 / 回归 （标签粒度）
模型结构	需要特定结构容纳富信息	使用目标模型结构即可，所需参数少
计算效率	低	高（几乎与目标模型一致的效率）
适用性	一次训练、广泛运用	不同目标需要单独训练（可与目标模型同结构训练）
效果	未测试	好

总结与思考

冷启动问题/模型有效性/其它的可能性/成效表现

思考与总结- 冷启动问题

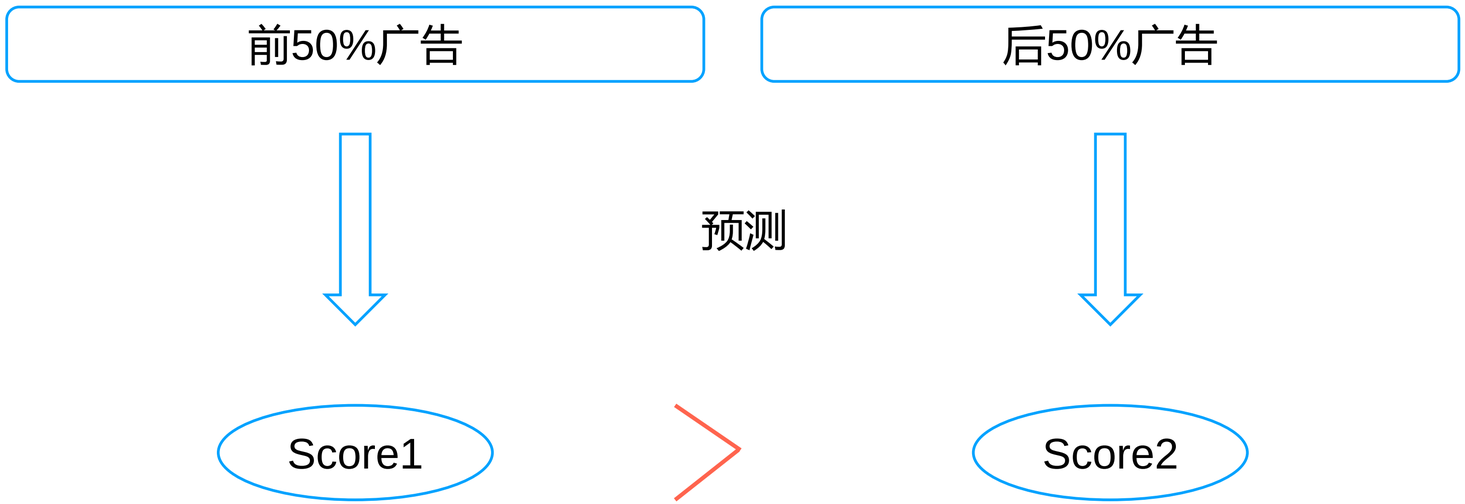


处理方法：

1、对序列进行采样或计算加权，越往前的广告具有越大的权重。

2、序列翻转后入LSTM模块，输出并使用last_output特征。

思考与总结- 模型有效性



启示一：前期用户点击行为较少，所以按照用户基础属性进行推荐。后期用户具有点击行为后，按照行为进行推荐，如果基础属性未包含标签，则说明基础属性之间具有更高的关联性

启示二：如果前期广告依赖标签进行推荐，则形成了信息穿越。该样本不能用来建模。

启示三：可设计标签修正模型校验此类现象。

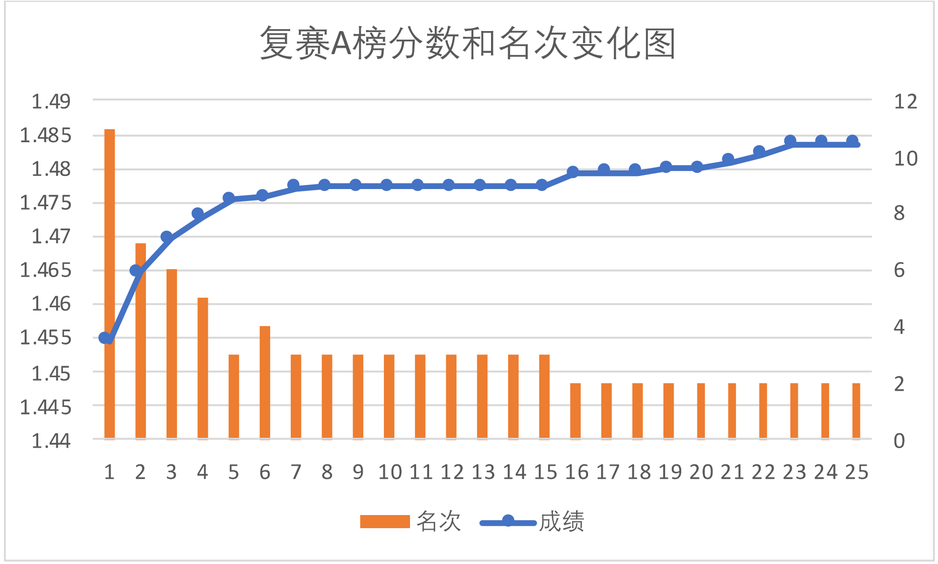
思考与总结- 其它可能性

多维度信息注入　　目标信息注入　　Fine-turning 目标模型

样本即特征，特征即标签：
1、在信息损失等较坏情况下具有良好预测能力。**每条样本**都应为此付出贡献。
2、特征或特征之间应该具有**相互备份容灾**的能力，具有丢失情况下的恢复能力。这个过程中形成的相互记忆的中间态，具有**桥梁**的作用，具有更强的**泛化**能力。

谨防信息孤岛\空岛：
1、稀疏实体富信息现象，容易形成信息孤岛、空岛。 如何将此类信息拆解分发或者注入。是接下来研究的重点。从而实现由**记忆到泛化**的转变。

思考与总结- 成效表现



1、复赛正式参与比赛，从160名左右，一周内进入前三
2、分数持续上涨，方案潜力大
3、模型效率高是迭代的前提、最终方案所需模型少，单模成绩可达第二。

方法	原理
分层标签注入	仿BERT，实现分层次的标签信息注入
Target encoding	基于标签主体信息的传递
Level layer	层级拼接，同质信息共享提取器参数
模型参数 宽度大： hidden_size 2080 深度适中	越深的模型，适宜越复杂的逻辑 越宽的模型，适合记忆越多的信息 （词表巨大，且交叉信息不明显）
冷启动处理	给予前期数据更高的权重
Week & holiday	不同年龄的人，具有明显的行为差异 广告主投放策略也会有所不同
动态学习率	提高收敛效果
性别单独建模	提高性别预测效果
多模型融合	提高模型稳定性


昨日已去，未来可期！

THANKS