

# 实验说明及安排

## 数据说明：

[train.csv](#)中包括了用户1-用户53975的过往行为，字段说明如下：

文件字段说明：

名称	标签
event_time	When event is was happened
event_type	Event type: one of [view, cart, remove_from_cart, purchase]
product_id	Product ID
category_id	Product category ID
category_code	Category meaningful name (if present)
brand	Brand name in lower case (if present)
price	Product price
user_id	Permanent user ID
user_session	User session ID

[test.csv](#)中包括了用户53978-用户67469的过往行为，字段和上图一样

[submit\\_example.csv](#)中包括了我们要提交的文件格式，字段如下：

字段名	类型	取值范围	字段解释
user_id	Int	-	用户ID
product_id	Int	-	商品ID

即对测试集的每一个用户，给出他最有可能购买的下一个商品

## 实验目标：

对测试集的每一个用户，给出他最有可能购买的下一个商品

# 思路：

用Transformer预测。

1. 将每个商品和起始符、终止符、padding符（共32734+3个）embedding 到一个长为508维的向量中（使用nn.embedding）。
2. 提取出 TrainSet 中同一个用户的所有行为（最多不超过20行，超过的截断），记作 PartSet，在 PartSet 起始处添加起始符，结束处添加终止符，添加过后不足22行的在末端补齐padding符。
3. 对于 PartSet 中的每一个条目（每一行），将其对应的商品的embedding编码与用户对该商品的行为（one-hot编码）一起组成一个长为512的向量。这部分 PartSet 数据总共组成一个22\*512的tensor，将其交给Transformer训练。
4. 同时对每一个512维的向量训练两个 MLP，vec2prod 用来输出对应的商品id，vec2act 用来输出对应的用户行为
5. 对于测试集的每一个人的数据，将其用同样的方式 embedding，然后交给训练好的Transformer，对输出结果同样通过 vec2prod 和 vec2act，若 vec2act 的结果是 purchase，则输出 vec2prod 的结果，否则将这个向量加入该用户的行为，重复5的步骤。

# 分工：

邬靖宇：1.2.3

郑鹏飞：4

李泽林：5