## **Action3**

****如果你是某P2P租车的技术负责人，你会如何设计个性化推荐和搜索排序；阐述相似车型，搜索排序的设计方法、可能的embedding策略****

P2P租车使用场景：

双边平台（顾客，车主），双边的推荐，既需要考虑用户预订，也需要考虑车主是否接受预订（Host Actions：Reject, Accept, No Response）

顾客通过搜索或者系统推荐找到车源：一个车辆在某时间段内只能被一个顾客租用；单一用户对于车型的匹配较为固定、数据存在严重的稀疏性

相似车源推荐进行的实时个性化：点击率（CTR），网页的观看时长，兴趣度（attention），提高商品购买的转化率等等

针对搜索排序，对于query（带有位置和旅行的时间），同时为host和guest优化搜索结果：

顾客角度：需要根据位置，价格，类型，评论等因素排序来获得客户喜欢的listing

车主角度：需要过滤掉那些有坏的评论，使用时间，使用范围（磨损度预测）、人数，是否携带特殊物品（如宠物、吸烟等行为）以及其他因素而拒绝guest的listing，将这些listing排列的低一点

采用Learnig to rank来做，将问题转换为pairwise regression问题，将预定的listing作为正样本，拒绝的作为负样本。

List Embedding（车源嵌入）：

将每个车源 => 车源embedding

数据集由 N 个用户的点击会话 (Session) 组成，其中每个会话定义为一个由用户点击的 M个车源 id 组成的的不间断序列

只要用户连续两次点击时间间隔超过30分钟，就认为是一个新的Session

目标是通过集合S，学习出每个车源listing的d维（ 32 维）embedding表示，让相似listing在embedding空间中距离更近

借鉴了word2vec中的skip-gram算法

车源embedding，把每个用户连续点击过的车源Session看做一个句子，每个车源当做word，训练出车源的embedding

Word2Vec：

通过Embedding，把原先词所在空间映射到一个新的空间中去，使得语义上相似的单词在该空间内距离相近。

Word Embedding => 学习隐藏层的权重矩阵

word2vec可以让单词之间具有相似性和类比关系

我们想要的结果并不是模型本身，而是隐藏层的参数，即让输入向量转换为新的embedding

List Embedding的冷启动：

新的车源产生时触发冷启动

车主上传新车源时需要上传多个特征，位置，价格，车源类型（包括车型、用户筛选指标（残疾、吸烟、宠物等））

在和新上传车源具有相同类型和相同价格区间的车源中，找到3个综合维度最接近的车源，用这3个车源的embedding求平均作为新车源的embedding

User Type Embedding 和Listing Type Embedding：

一些长期的行为，比如一个人在很久之前在另外一个城市预定过车间，那么在当前城市很有可能还是会喜欢预定同类型的车源

进一步从预定的Listing中来捕获这些信息

构造数据集：由前面的点击序列变成预定序列，数据集为N个用户预定的Listing 组成的session集合

User Type Embedding 和Listing Type Embedding：

为了处理存在的问题，提出了学习type级别的embedding，而不是学习id级别的embedding

对于原始的Listing数据，比如位置，价格，车型等，基于一些硬性的规则生成每个Listing的类别

目的是将原本稀疏的数据（基于id），变得稠密（基于类型），组合出很多session

User Type Embedding 和Listing Type Embedding：

将User Type Embedding方式也采样相同的方法，即拥有一些相同的基础属性和相同行为的用户进行分桶

可以理解为，分别对用户和车源做了聚类，按聚类后的数据进行学习embedding，在更高一个维度（粗粒度）上获取长期兴趣

用type来表示原有的user和listing，可以解决：

实时个性化的体现，即使对于同一个listing或user，对应的type也可以发生变化（比如用户行为发生变化）

冷启动问题，用户层面上的前五个特征是通用的画像特征，对于新用户可以直接通过这五个特征完成映射