角度一：评估指标

一、目前我使用的基础评估指标有六个：

1、Hit@k

2、Precision@k

$$\text{Precision} = \frac{\sum_{u\in U}|R(u)\cap T(u)|}{\sum_{u\in U}|R(u)|}$$

3、Recall@k

$$\text{Recall} = \frac{\sum_{u\in U}|R(u)\cap T(u)|}{\sum_{u\in U}|T(u)|}$$

4、F1Score@k

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

5、NDCG @k：衡量Top k推荐命中的那个点在推荐的列表中有多靠前

$$NDCG \ @ \ K = \frac{DCG}{iDCG}$$

$$\text{DCG}(b,L) = \sum_{i=1}^{b} r_i + \sum_{i=b+1}^{L} \frac{r_i}{\log_b i}$$

6、MAP @k：顺序敏感的recall

$$AP_u = \frac{1}{\Omega_u} \sum_{i\in\Omega_u} \frac{\sum_{j\in\Omega_u} h(p_{uj} < p_{ui}) + 1}{p_{ui}}$$

$$MAP = \frac{\sum_{u\in U} AP_u}{|U|}$$

二、其余具有参考意义的论文中所使用的评估方法：

暂无

角度二：”自定义指标“以证明模型强度

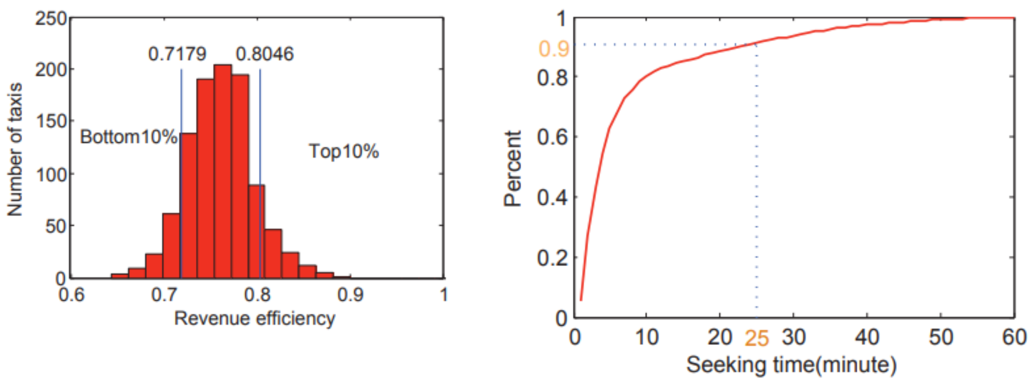
一、具有参考意义的论文中使用的自定义指标：

1、使用“收入提升”来说明模型的强度

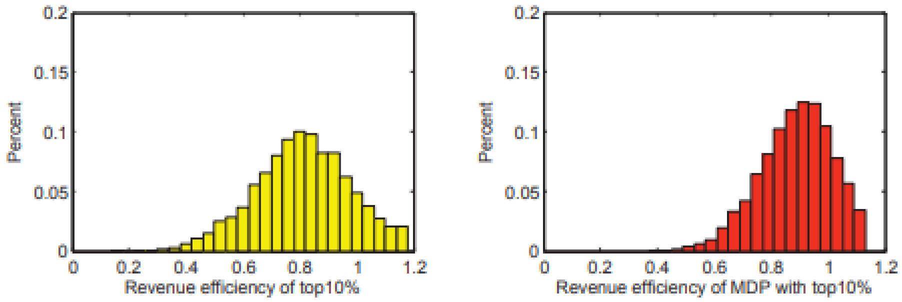
论文一：The Rich and the Poor: A Markov Decision Process Approach to Optimizing Taxi Driver Revenue Efficiency

来源：CIKM ’16

问题一：如何定义“高收入”“低收入”？

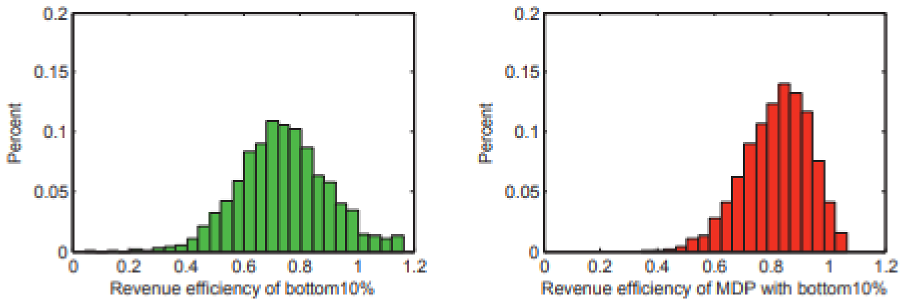
> Revenue Efficiency. An obvious way to increase revenue is by driving taxi for longer time duration. However, given time constraints, it makes sense to maximize revenue per unit of time.
>
> $$E_{rev} = \frac{M}{T_{bus}} = \frac{M}{T_{drive} + T_{seek}}$$
>
> About 80% drivers have a revenue efficiency between 0.72 and 0.80. We find the top 10% and bottom 10% drivers as successful and not-so-successful drivers.
>
> 
>

问题二：针对高低收入如何评估和比较：



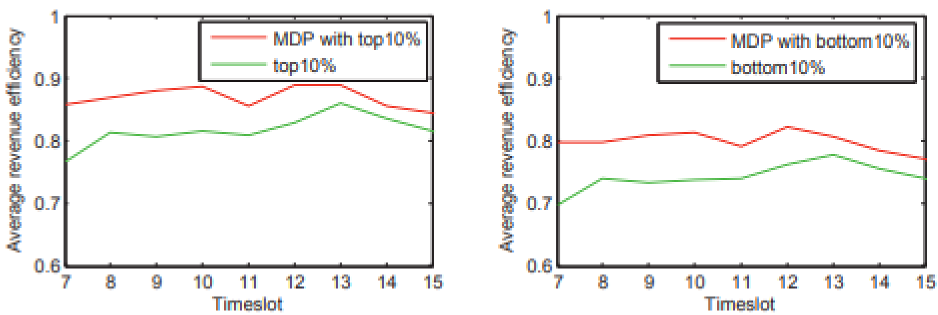
(a) Distribution of revenue efficiency of top10% drivers for 12-13

(b) Distribution of simulated revenue efficiency baed on top 10% drivers



(c) Distribution of revenue efficiency of bottom10% drivers

(d) Distribution of simulated revenue efficiency based on bottom 10% drivers



2、使用“载客率提升”来说明模型的强度

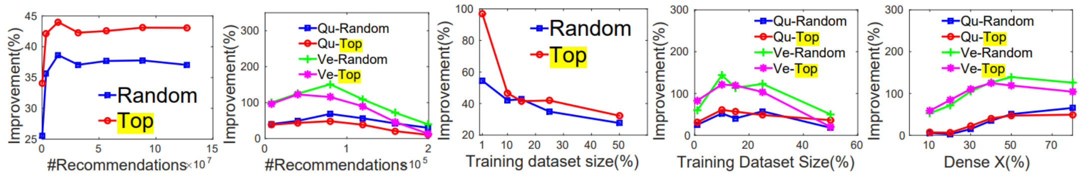
论文二：Route Recommendations for Idle Taxi Drivers: Find Me the Shortest Route to a Customer!

来源：kdd18

问题一：如何选择random/top数据集并评估

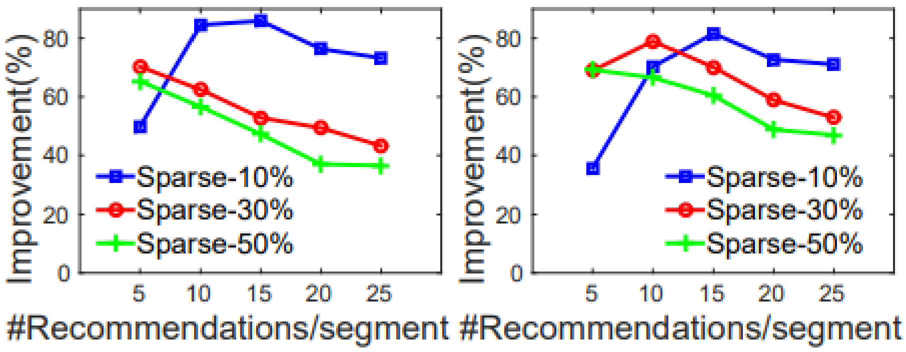
对比random/top两数据集在不同情况下的improvement%，在我的情况下，这个提升率可以换成两种不同收入群体的收入提升率，比如低收入群体/高收入群体原本的pick-up numbers对比应用模型后的pick-up numbers.

> The training set is constructed using two strategies. In the first strategy, we **randomly select X%** of all taxi drivers. All customer pick-ups serviced by these selected taxi drivers and their corresponding driving routes form the training set. X guides the size of the training set. We call this the Random training set. The second strategy is inspired by [13]. Here, each taxi driver is sorted based on the number of customer pick-ups they have on record. The

> top-X drivers with the **highest customer pick-ups** are selected and their customer pick-ups and driving routes form the training set.
>
> 
>
>
> (a) NY (b) SF (c) NY (d) SF (e) SF
>
> **Figure 4: Percentage improvement of MDM against (a-b) the number of recommendation queries and (c-d) training data size. (e) Comparison of MDM, *Qu*, and *Ve* in dense regions.**
>
> $$Improvement(\%) = \begin{cases} \frac{d_M - d_{MDM}}{d_{MDM}} \times 100 \text{ if } d_{MDM} < d_M \\ -1 \times \frac{d_{MDM} - d_M}{d_M} \times 100 \text{ otherwise} \end{cases}$$

问题二：如何对密集/稀疏区域适应（Case Study）

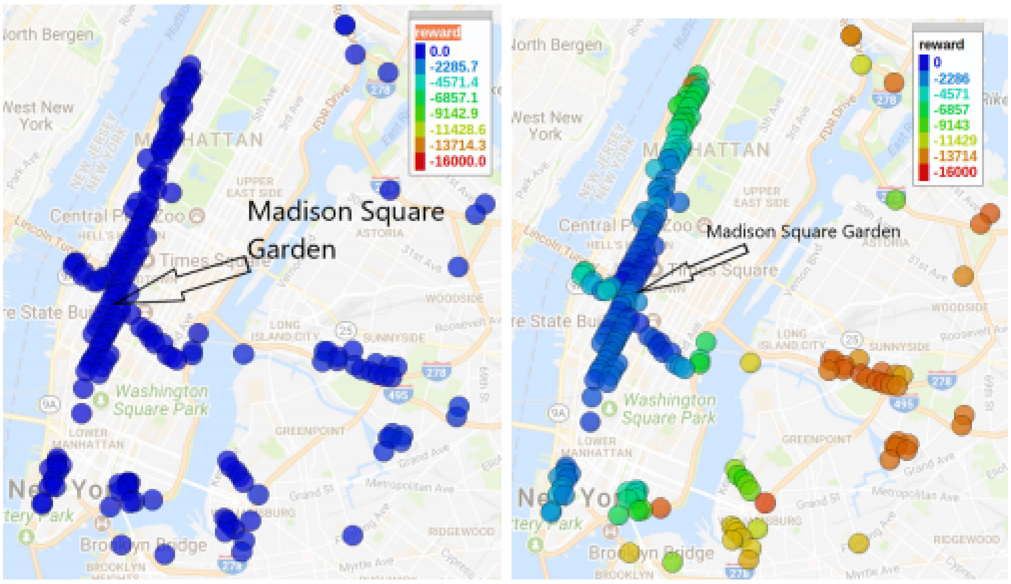
可以作为一个Case Study，用两块人流密集/人流稀疏的区域做对比：

> Impact of customer density: The experiment against number of recommendations indicated that when customer density is low, performance improvement of MDM is more profound. We next verify the impact of customer density directly.
>
> We further substantiate the ability of MDM to do well in difficult situations by computing the improvement in customer-sparse neighborhoods. In these experiments, we choose the query regions from only sparse neighborhoods of various intensities. For example, Sparse-10% contains only top-10% sparsest neighborhoods.
>
> 
>
>
> (a) SF, *Qu*, Top (b) SF, *Qu*, Random
>
> In the x axis we plot the number of query recommendations per segment, instead of total number of recommendations since the number of segments in a region varies. For example, the number of segments in Sparse-10% is much smaller than in Sparse-50%. As clearly visible, the sparser the region, the higher is the improvement provided by MDM.

问题三：对异常的敏感性，特殊拥堵（Case Study）

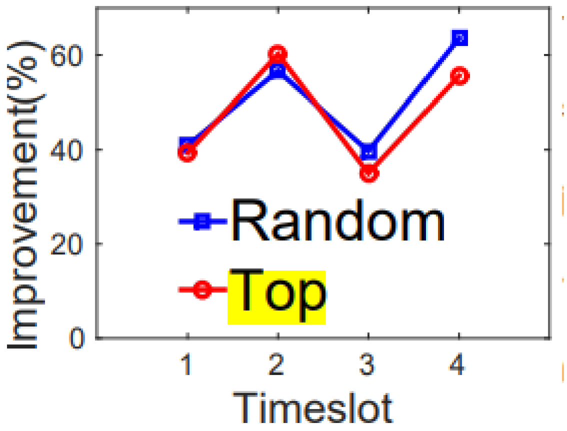
虽然目前没有对拥堵的解决办法，但可以添加一个功能：

在推荐时，每当一个地点出现在推荐累计x次之后，在t分钟内该地点不会再次出现在推荐列表中，也算是动态避免拥堵。



(g) Prior to anomaly (h) Reaction to anomaly

问题四：对时间片的敏感性：

> Adapting to time-slots: How does the performance of MDM vary against various times of the day? We next answer this question. First, we partition the dataset into four different timeslots: 1. 12AM −8AM, 2. 8AM −12PM, 3. 12PM −5PM, 4. 5PM −11:59PM. Generally slot 1 and 3 are non-peak hours, whereas 2 and 4 are peak hours. As shown in Fig. 5(i), regardless of the time of the day, MDM has a performance improvement in the range of 35% to 65%.
>
> 
>
>
> (i) SF

论文三：A Cost-Effective Recommender System for Taxi Drivers

来源：kdd14

Case Study 1：A Case Study on Cost-Effective Route Recommendations

作者用自己模型的两条推荐路径和谷歌地图的推荐路径对比，说明：

自己的推荐路径长度虽然和谷歌地图的推荐路线不一样，但是计算出租车所得收入却高出不少

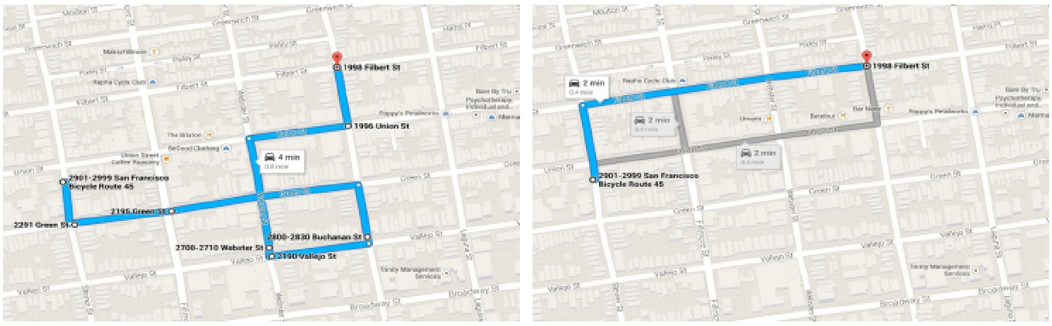


Figure 9: Case Study (a)

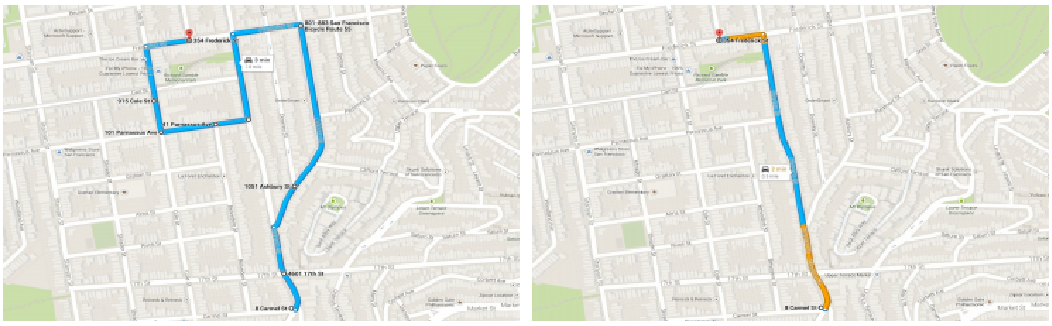
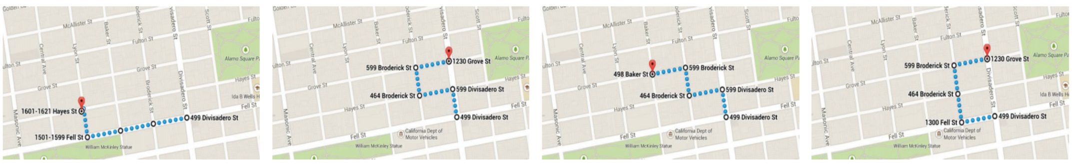


Figure 10: Case Study (b)

> the left figures are the driving route recommended by the MNP recommender system and the right figures are the route suggested by the Google Map based on the shortest driving distance. However, this driving route suggested by the Google map cannot maximize taxi drivers' net profit.

Case Study 2：A Case Study on Top-K Recommendations

这个Case Study，作者想说明，自己的Top-K推荐可以明显地通过减少冗余来提升模型性能

> we introduced a minimum redundant strategy to recommend the Top-K driving routes and solved the overloaded problem. In figure 11, we demonstrate the Top K driving routes starting from the same location, where K equals to 4 in this case. The figure shows that each route has different driving directions and the correlations between those driving distances are very small. Therefore, the minimum redundant strategy can improve the performance of our recommender system.
>
> 
>
> (a) (b) (c) (d)
>
> Figure 11: The Top 4 driving routes starting from the same location (longitude: 37.77407074 and latitude:-122.4376221)

论文四：Augmenting Decisions of Taxi Drivers through Reinforcement Learning for Improving Revenues

问题一：定义指标

指标一：分配概率/载客概率

> As the agent is competing with other drivers present in the state at that time, we compute an assignment probability (pst assign) with which a trip can be assigned to the agent. The probability can be computed as the **number of trips available divided by the number of cruising drivers** present in the state at that time. To compute pst assign, we consider durations of γ minutes. We maintain a list of trips available in every γ minute interval and the number of cruising drivers available in the corresponding interval. This way we have multiple assignment probabilities for a single state, and a trip is assigned based on when (time of day, t) the agent visits the state.

指标二：驾驶员收入

> a driver's revenue in a time interval is computed as the revenue from all the trips of the driver in the time interval minus cost of travelling all trip distances minus cost of all cruising.
>
> Table 3: Weekdays (revenues in SGD)
>
> | Strategy | 0-6 hours | 6-9 hours | 9-12 hours | 12-17 hours | 17-20 hours | 20-24 hours |
> |---|---|---|---|---|---|---|
> | Average of top 1% drivers | 143.26 | 91.86 | 70.81 | 107.75 | 97.09 | 126.97 |
> | Average of top 5% drivers | 112.82 | 77.85 | 60.09 | 91.97 | 82.52 | 110.27 |
> | Average of top 10% drivers | 97.09 | 69.71 | 54.26 | 83.13 | 74.89 | 101.55 |
> | Average of top 20% drivers | 78.66 | 59.71 | 47.24 | 72.20 | 65.59 | 90.97 |
> | Heuristic | 147.4 | 72.42 | 52.7 | 85.13 | 66.3 | 93.59 |
> | Static zone | 164.77 | 84.98 | 57.71 | 93.32 | 74.76 | 100.99 |
> | Dynamic zone | 186.52 | 88.37 | 58.3 | 91.43 | 75.25 | 110.95 |
>
> Table 4: Weekends (revenues in SGD)
>
> | Strategy | 0-6 hours | 6-9 hours | 9-12 hours | 12-17 hours | 17-20 hours | 20-24 hours |
> |---|---|---|---|---|---|---|
> | Average of top 1% drivers | 188.21 | 76.35 | 76.68 | 117.69 | 102.29 | 136.93 |
> | Average of top 5% drivers | 161.27 | 63.74 | 66.22 | 102.44 | 89.31 | 121.35 |
> | Average of top 10% drivers | 145.95 | 56.79 | 60.35 | 93.41 | 82.07 | 112.67 |
> | Average of top 20% drivers | 126.69 | 48.15 | 52.94 | 81.86 | 72.89 | 102.01 |
> | Heuristic | 175.92 | 60.48 | 55.43 | 95.23 | 70.31 | 99.94 |
> | Static zone | 189.35 | 69.67 | 59.91 | 108.13 | 79.24 | 104.81 |
> | Dynamic zone | 195.35 | 74.37 | 69.77 | 111.54 | 79.75 | 114.96 |

指标三：设置50%出租车不听从推荐系统的建议

> Another benchmark we employ to evaluate performance of our learning approach is a heuristic strategy that is typically employed by drivers. When a cruising driver is in a locality, he generally takes one of two options - stay in the current locality or move to a nearby locality. When the agent follows the heuristic strategy, it stays in the current zone with a probability equal to pstay and with the remaining probability moves to a nearby zone. Since pstay = 0.5 worked the best, we employ this strategy

分析四：比较不同时段推荐模型与正常出租车的利润

论文中说：该模型在高峰时期带来的利润和普通出租车持平，但是在凌晨和其余时刻，会带来更大的收入差距

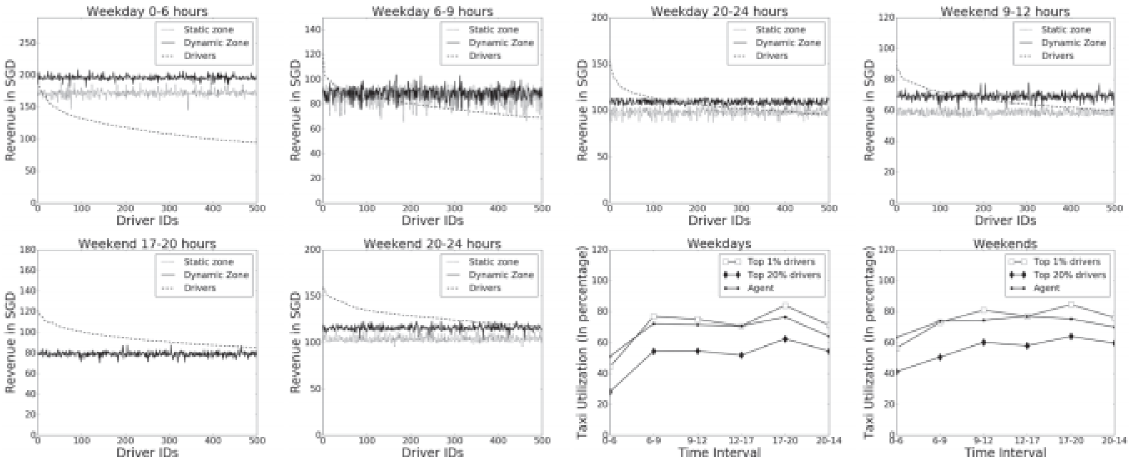


Figure 3: Revenue comparison for various time-intervals and taxi utilization comparison for weekdays and weekends

总结：我该如何做实验

1、评估指标

先使用目前的这些指标

2、自定义指标

1、对比"应用推荐系统的**<u>低/高/随机</u>**收入群体"和"**<u>低/高/随机</u>**收入群体"的**<u>利润</u>**、**<u>出租车时间利用率</u>**、**<u>pick-up number</u>**、

3、Case Study

1、分析人烟稀少。人流密集地区的模型性能（对不同区域的适应性）

具体来说，在这两种情况下，对比同一群体的**<u>利润</u>**、**<u>出租车时间利用率</u>**、**<u>pick-up number</u>**

2、分析对不同时间片的敏感性

在几个有代表性的时间片内（凌晨，早晚高峰，午休），对比本模型对不同司机群体的三个指标变化程度

3、我们使用了topK推荐，如何使用case Study说明topK推荐是好的

4、我们的POI是核心，如何使用Case Study说明POI的加入对推荐的准确度提升非常大

5、可以分析构图时的kdtree对构图速度的影响

6、分析出现交通堵塞的模型性能（很牵强）

举例某个网格出现300次载客记录，但如果2000辆出租车在一个时间片内都前往，会产生总体**<u>利润</u>**、**<u>出租车时间利用率</u>**、**<u>pick-up number</u>**的急剧下降，而应用某个本系统中的机制，将避免这种情况