travelers' predictability-CEUS

▼ Research Gap:

1. Entropy-based method is not adequate since the observation for travellers are usually temporary. **How to measure the predictability** on this context?
2. How different factors(length of stay, movement frequency,...) affect their predictability? What's more, considering space, what's the heterogeneity over different cities?

▼ Dataset

Area: South Korea

Dataset: national-wide travellers mobie phone data(signal data)

scale: 192302 individual, Aug. 1 - Aug. 15, 2018

▼ Findings

1. **For global model performance level**: Generally speaking, LSTM outperforms Markov, which is intuitive. Yet with the K of ACC@K increase, the difference would be smaller.(Might indicates that LSTM can more sensitivly capture the unique destination of K, compare to Markov. **LSTM is more exactly.**) Global Performance of Acc@5 is more than 60%, indicating that **travelers' movement is still predictable.**
2. **For individual level:** Markov would fail to predict many individual at Acc@1 **since the mechanism is based on 'mainstream'**. While LSTM can still capture individual level well at acc@1. What's more, the variation of LSTM(upper bound and lower bound of boxplot) is smaller, indicating the long term dependency is significant.
3. **For Length of Stay level:** Grouping Accuracy for different length of stay visitors, **For Markov:** the predictability tends to peak for those who stays 2-4 days. But decrease from more than 4 days. **For LSTM**: All time stable. **Indication**: LSTM helps recognize personal travel style.
4. **For Length of Sequence:** length of sequence represents the activeness of a traveler, especially when controling the length of stay. General speaking, **the more active the traveler, the higher predictability.** (A little bit different with the statment with the paper, but from the same graph, both are ok for me, not signifcant, i think.)
5. **For *Journey Ratio***: Journey ratio describes the relative stages. For instance, User A with Length of Traj 10 while B with 20. the journey ratio of 0.5 is their 5th step and 10th step, respectively. **The results shows that the journey ratio do not change significantly excepts the start and end stage**, which is mainly the airpo

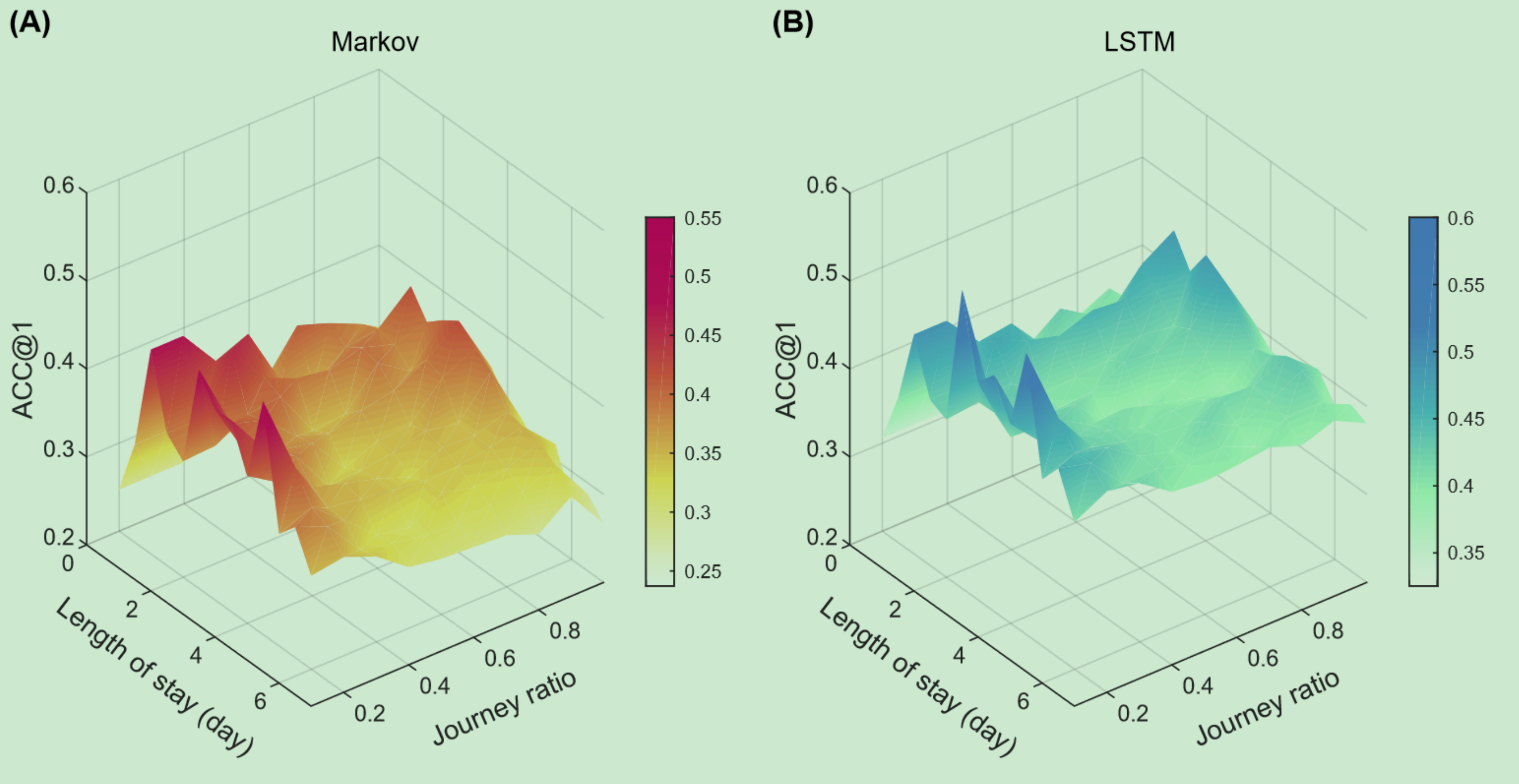


Figure 8: Movement predictability of travelers at different stages of their journeys.

6. **For movement within or across cities: The movement predictability is much higher for intra-citiey, compare with inter-cities.** Author says that movement across cities is a lot difficult to predict which city do user going to go.

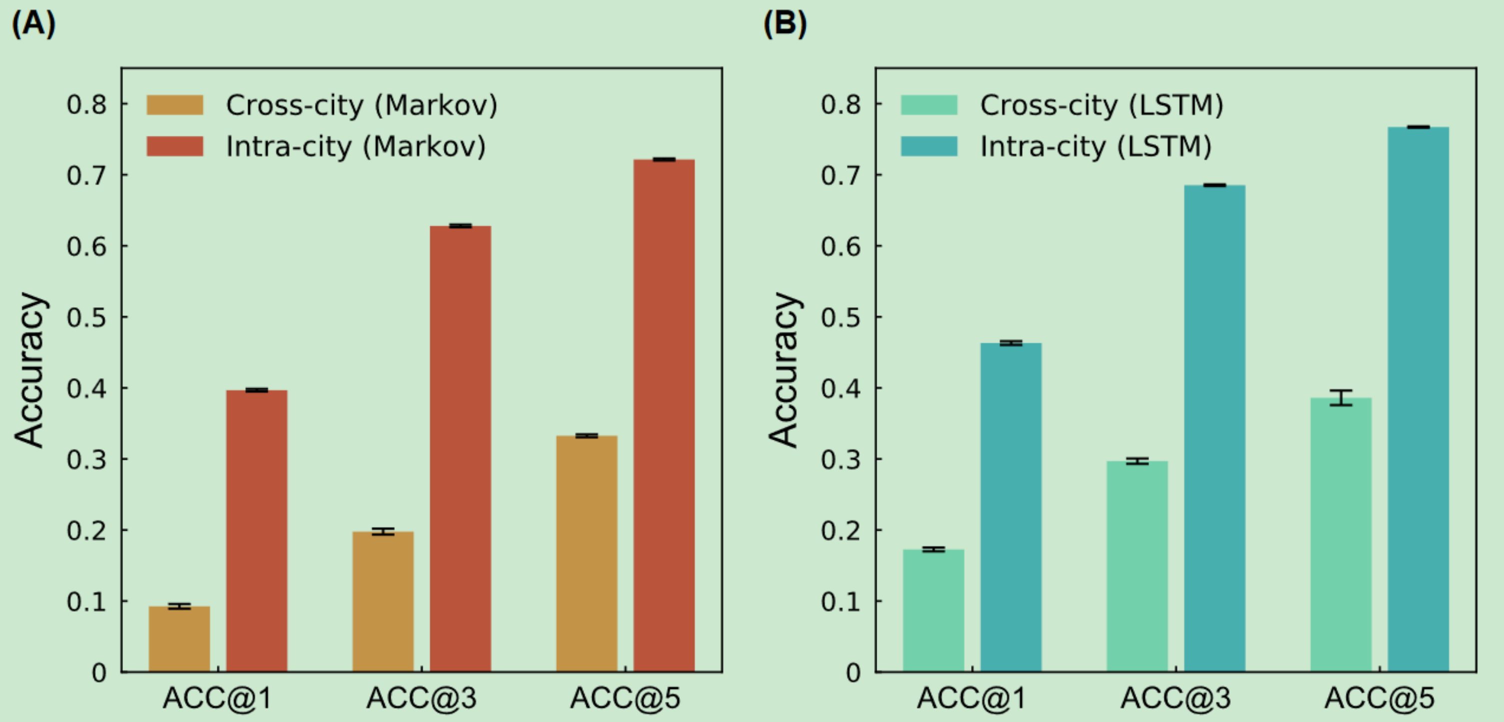


Figure 9: Prediction accuracy over intra-city and cross-city trips: (A) Markov chain model; (B) LSTM Model.

7. For Intra-City level, the spatial heterogeneity: Predictability for intra-city movement have strong heterogeneity from spatial perspective.

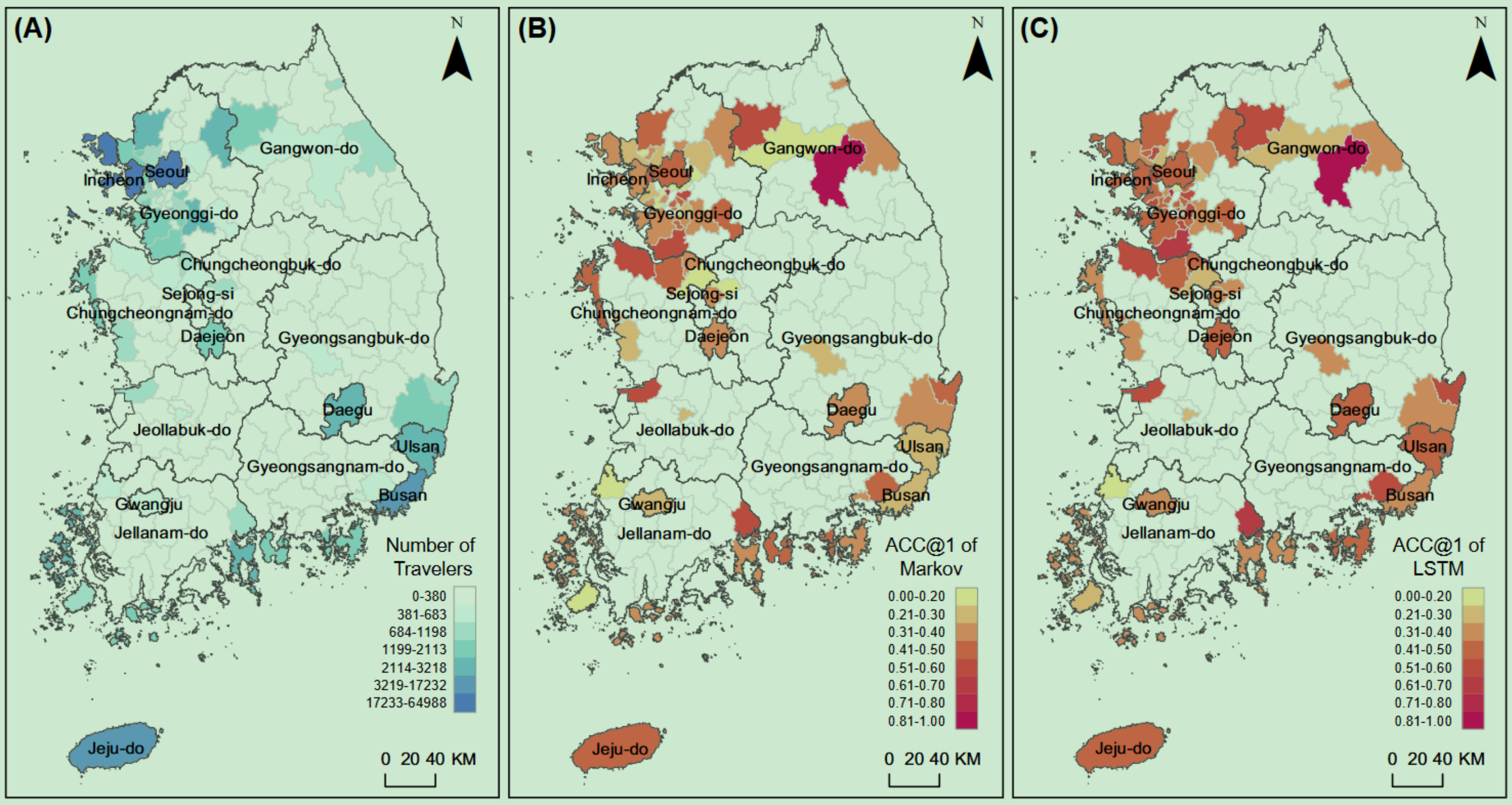
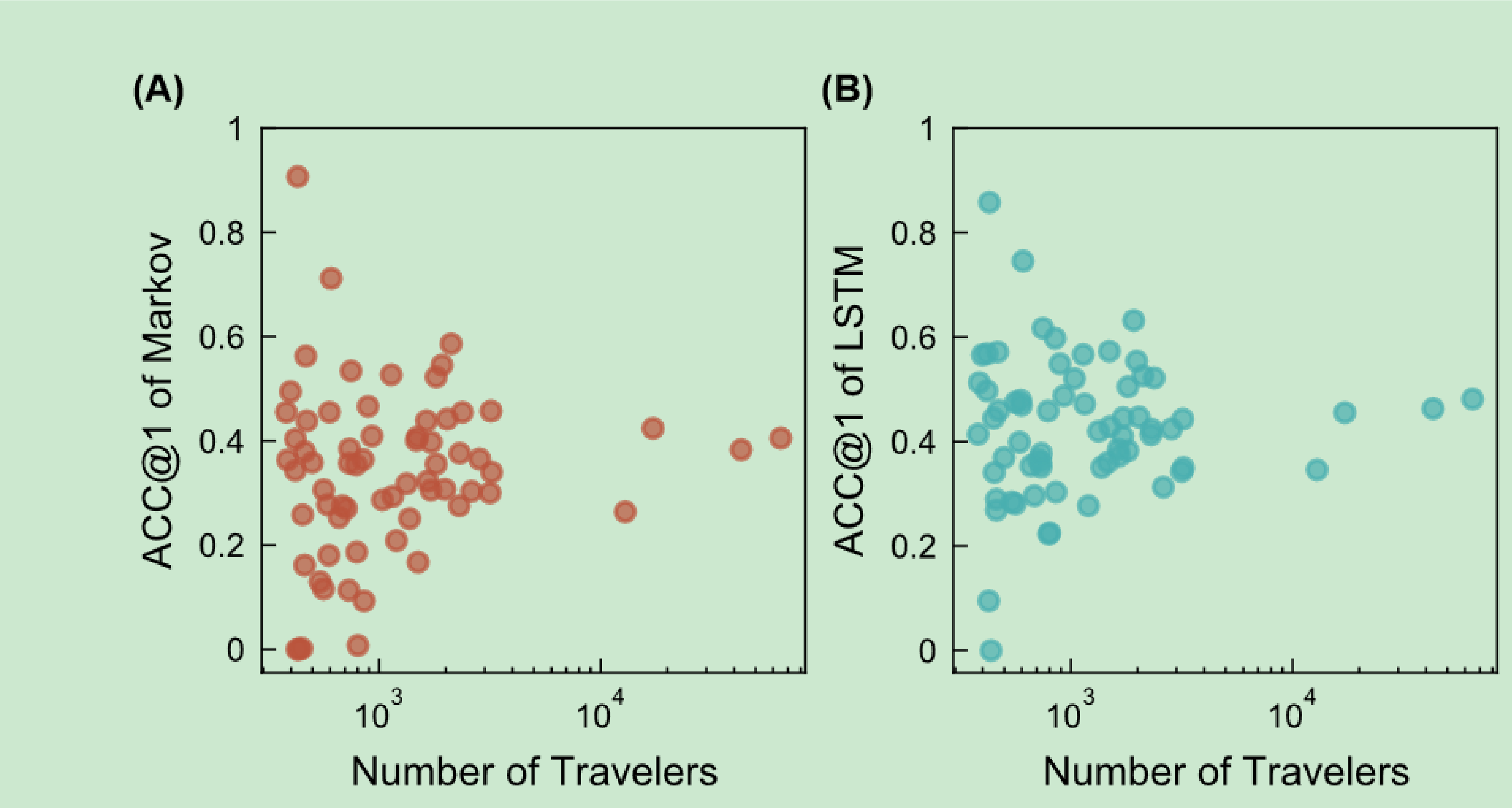
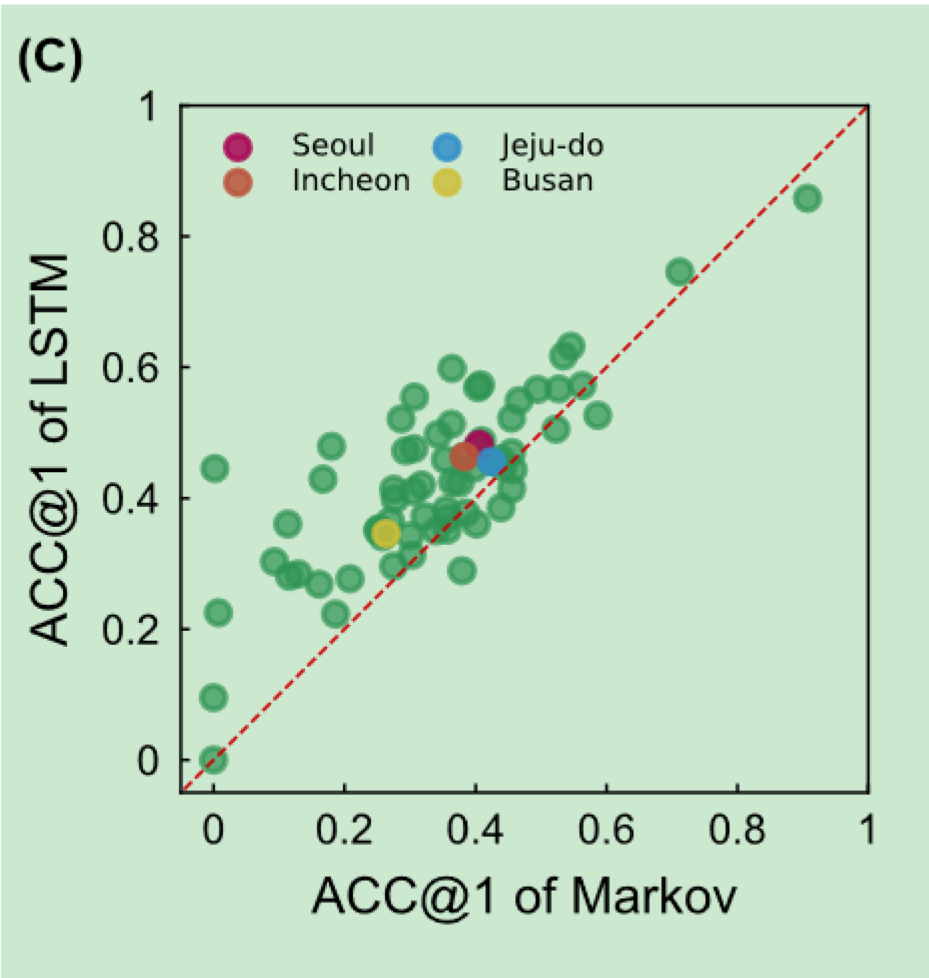


Figure 10: (A) Total number of travelers to each city derived from the mobile phone dataset; (B) Acc@1 metric of intra-city trips for each city based on the Markov chain model; (C) Acc@1 metric of intra-city trips for each city based on the LSTM model;

8. **The relationship of visitation frequency and city predictability: visitation frequency is not the decisive factor of the predictability of the city**. But converge. This is mainly due to the mix of those regular pattern and unique style visitors. (Not influenced by the number of travelers but may be influenced by the characteristics of cities that shape individual's travel eg. spatial organization of attractiveness or some transportation deployment or social backgrounds of visitors, and so on).



9. Also, The correlationship between two model indicates that LSTM perform better for most of cities. But For some most popular cities, they shows a similar performance, indicating that for some specific city, we can use some collective data(OD matrix) for some commercial recommendation when we can not access personal data.



10. **Although travel is regared as 'an escape of daily routine', the results shows that travelers' movement still exhibits a high degree of predictability.**

▼ Useful Views

1. Aggregated data(OD matrix) can still perform well in many cities, which is helpful to do decision.
2. Differential strategies should be deployed over different charateristic of people(age, planned staying length...) by adopting different sub models.
3. The results show that there is no "one-model-fits-all" solution, indicating that building local model for each city can help improve tourist experiences

4. According to the low performance of predicting cross-city movement, we can build next-location model as well as next-city model to serve some recommendation.

▼ Personal Inspiration

1. How to measure movement predictability?
    a. **Ideal and personal theoritical quantification: Entropy**. Entropy can theoritcally explain individual's predictability based on historical information. This requires a relatively long-term dataset to include individual's entire possibile visiting locations. (Entropy for regular patterns measurement)
    b. **Prediciting Model:** Model is a compression on data. If we can not get a regular pattern dataset, we can use predicting model to get a collective information and use it for individual prediction. The accuracy of the model would be the 'predictability'. **Yet model can not be as personalized or perfectly describe individuals' unique pattern, it is a relatively acceptable way to explain for movement predictability.**
2. For spatial units using cell tower, uncertainties should be discussed and avoid using raw thiessen units directly. Some units clustering should be introduced if the scale is large enough.
3. **Sampling Strategy:**
    a. What is the appropriate samples (From spatial unit aspect and population aspect)? Defining your problem and sample those who behaves differently on your topics.
    b. uneven distribution: if the raw data presents a uneven**(number uneven or spatial uneven)**, ramdon sampling would **under-represent some relatively low-frequency samples**. Hence, uniformly sampling with repeat tries might be a strategy.
4. **Why also adopting 1-order-Markov here?**
    a. Inspiration: if the markov presents well, it means that collective data is ok(OD matrix) for travellers prediction, which orders a more general method for tourism management when personal data is not accessbile
    b. to be a comparison with lstm, indicating that **the effects of individual diversity in travel**. **As well as long-term temporal** (more than 1) **context.**