東京大学　関本研究室 / Sekimoto Lab. IIS, the University of Tokyo.

Towards Pseudo People Flow:Developing a Deep Generative Model based on PT Data to Reproduce Large-Scale Daily People Activity Profiles

Yurong Zhang, Kunyi Zhang, Yanbo Pang, Yoshihide Sekimoto

Background

Human mobility data plays a crucial role in urban and transportation planning, such as simulating human behavior changes in the presence of new infrastructure, unprecedented disaster, or large-scale activities. However, due to privacy concerns and stringent regulations, acquiring real people flow data require a significant amount of time, resources, and funds. To overcome the accessibility issues concerning privacy-sensitive cell phone datasets, the reproduction of synthetic datasets depicting the large-scale movement of people is a promising approach.

Object

Develop a deep generative model to reproduce large scale daily people activity profile towards Pseudo People Flow

- Enhance Contextual Relevance (demographics) and Realism of Generated Activities.
- Improve Representation of Underrepresented Activities (shoppings, hospital visits).

Dataset

- Based on the 'purpose' code mentioned in the PT survey, we derive activities for each travel segment and assign them accordingly, with each segment lasting 15 minutes.

- Data from more than 10 metropolitan areas in Japan, primarily centered around the Tokyo metropolitan area, were selected as the training data for the model.

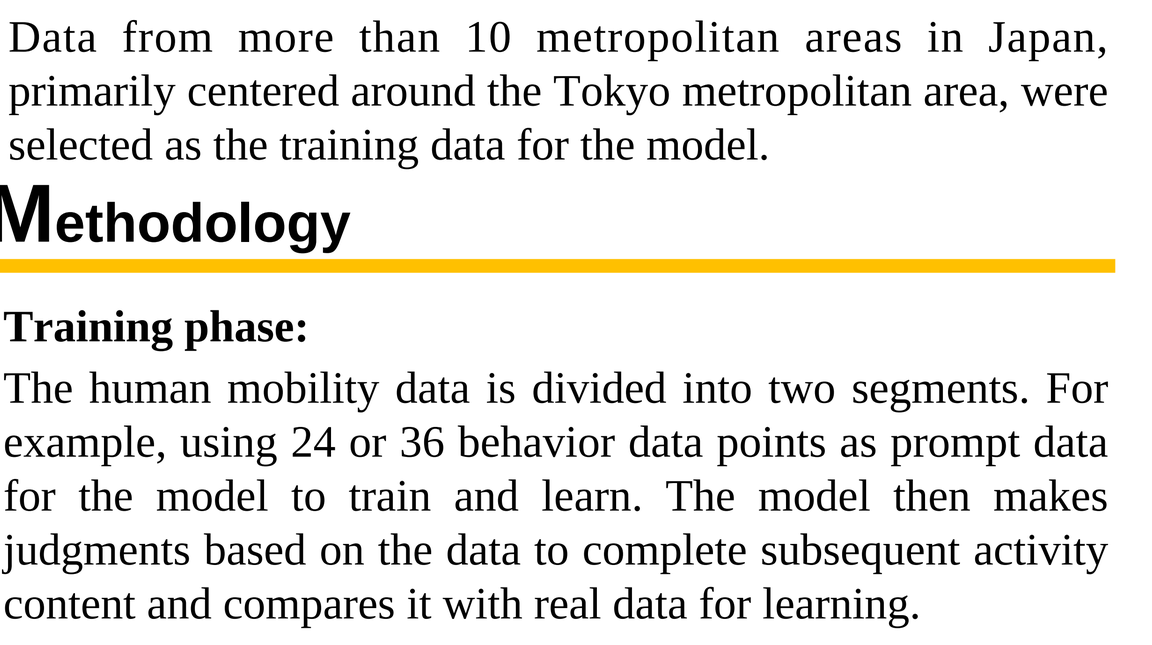
Methodology

- **Training phase:**

  The human mobility data is divided into two segments. For example, using 24 or 36 behavior data points as prompt data for the model to train and learn. The model then makes judgments based on the data to complete subsequent activity content and compares it with real data for learning.

- **Adjustment phase:**

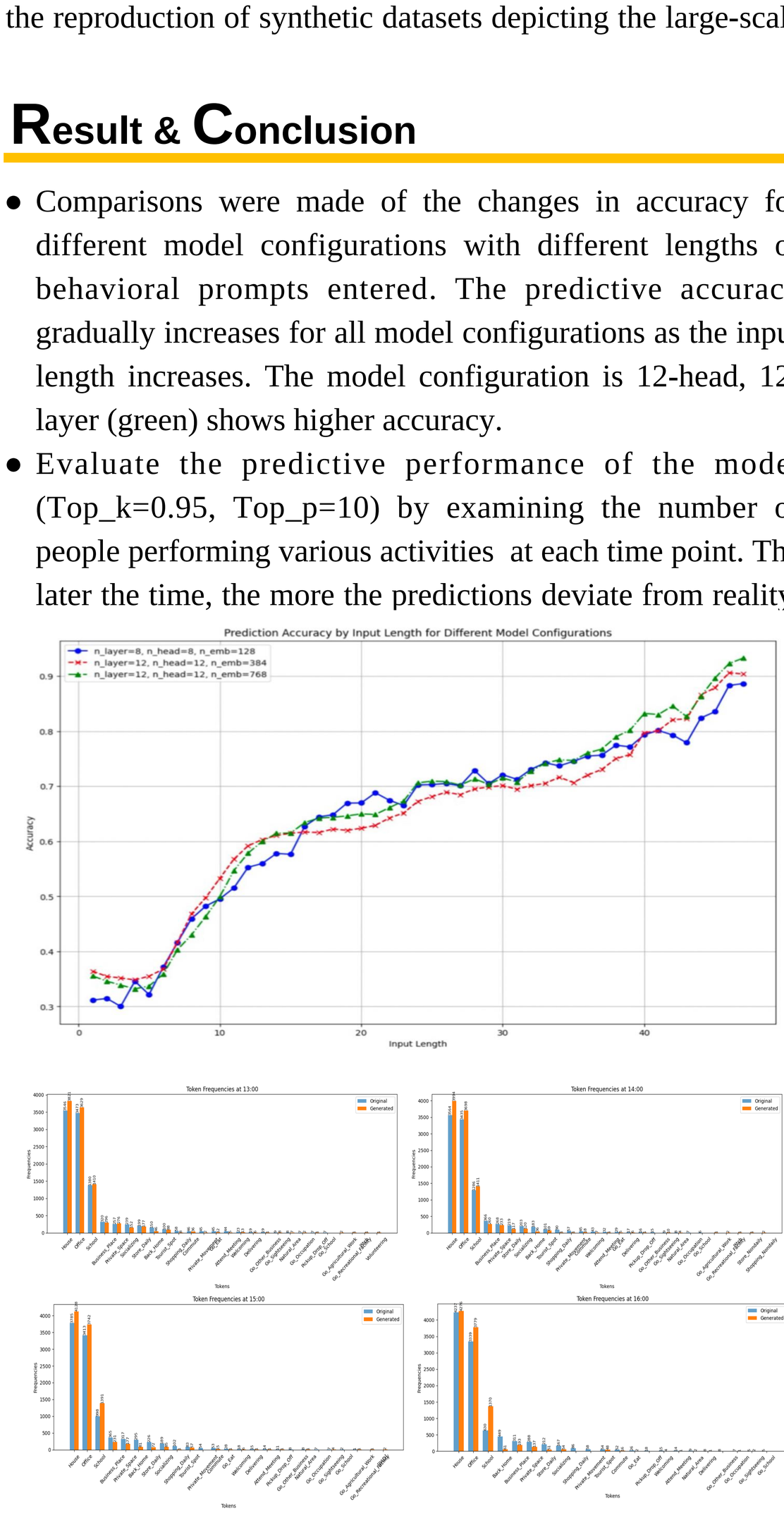
  Optimize the model based on its predictive evaluation capabilities to generate more realistic and vivid travel activity data.



Framework of the proposed method

Result & Conclusion

- Comparisons were made of the changes in accuracy for different model configurations with different lengths of behavioral prompts entered. The predictive accuracy gradually increases for all model configurations as the input length increases. The model configuration is 12-head, 12-layer (green) shows higher accuracy.
- Evaluate the predictive performance of the model (Top_k=0.95, Top_p=10) by examining the number of people performing various activities at each time point. The later the time, the more the predictions deviate from reality,



Comparison Chart of Predicted Results and Actual Data Activities at Various Time Points (Tokyo, prompt token=24)

In summary, we have developed a human mobility data generation model based on the GPT2 model that focuses on daily activity activities. However, compared to the research objectives, more effort is needed to generate vivid and realistic people flow data.