Each NEXUS project within TULIP Lab is part of an ongoing research initiative. This document provides the specification for the following project:

Name: Tourism Demand Forecasting

Coordinator: Dr *Yishuo Zhang*

1 Project Background

Tourism demand forecasting is an important and challenging task in the tourism industry, whose goal is to forecast the number of tourists arriving at a destination in a certain period. Accurate tourism demand forecasting can help policymakers and practitioners to make appropriate strategies and plans, which is beneficial for promoting local development. Many studies focus on developing diverse models for tourism demand to improve the prediction ability, such as time-series models, econometric models, and artificial intelligence (AI) models [1, 2]. In general, tourism demand forecasting is defined as the task of predicting or estimating the number of tourists arriving at the target destination in a certain period of time.

Despite numerous advancements in the three modeling categories of tourism forecasting literature, several crucial methodological problems still require further attention. For example, the common key challenges on tourism demand forecasting is on how to improve accuracy and omit overfitting of the developed model. Another important aspect focused by researchers recently is to explore the pattern between data characteristics and theoretical forecasting performance. With all above research challenges ahead, this project aims to explore and implement the most recent studies regarding to the abovementioned challenges, benchmark, discuss and review those solutions to find promising research directions on tourism demand forecasting.

2 Research Problems and SOTA Methods

In the tourism demand forecasting area, currently there are several interesting research problems in this literature.

Decomposition based tourism demand forecasting Decomposition-based forecasting methods, such as the classical decomposition method (e.g., seasonal decomposition of time series), offer a powerful approach for analyzing and forecasting tourism demand by decomposing the time series data into its underlying components, including trend, seasonal, and residual components. Under this line of research, several directions could be explored such as choosing the most suitable decomposition model, developing advanced decomposition techniques capable of handling non-linear trends and seasonality, model interpretability and Transparency in decomposition-based forecasting. The most common used library for decomposition is the `statsmodels`.

Data augmentation based tourism demand forecasting Availability of reliable and comprehensive data poses a significant challenge in tourism demand forecasting. Data may be fragmented, inconsistent, or outdated, making accurate predictions difficult. In this research problem, research is needed to explore innovative data sources and methodologies for collecting and aggregating tourism-related data, including social media, mobile devices, and sensors, to enhance forecasting accuracy. Upon completing the data augmentation, the consumer behavior and preferences in shaping tourism demand could be explored as well. Also, the data augmentation is one of the efficient solution to resolve the overfitting problems in tourism demand forecasting. The one of classical data augmentation work [4] is to use the `Google Trend` as the external factors to enhance the tourism demand forecasting.

Spatial Heterogeneity based tourism demand forecasting Tourism demand varies spatially across destinations, regions due to factors such as accessibility, infrastructure, and marketing efforts. The tourism demand could be similar cross different destinations according above-mentioned factors as

well. This research topic requires the development of spatially explicit forecasting models that consider the unique characteristics and dynamics of different locations. The explored aspect could be but not limited to `seasonal and temporal Variability Across Regions`, `Destination Spatial Auto-Correlations`, `Spatial Data Integration In Tourism Demand Forecasting` and `Tourism Demand Interpretability Cross Destinations`. One common approach to measure the region similarity could be found in the work [3], which is open sourced via `https://github.com/tulip-lab/open-code/blob/develop/GP-DLM/Dynamic_time_warping_distance.py`

3 Benchmark Data and Evaluation Metrics

3.1 Benchmark Data

In this area, the following benchmark datasets and evaluation metrics are commonly used:

Multi-variate Tourism Demand Data The common used benchmark data is `Hong Kong Tourism Demand Data`, which is one of the Multi-variate benchmark data has been used in many tourism demand forecasting research works. The factors are collected in `Google trend` and acted as the external factors in the tourism demand forecasting.

Uni-variate Tourism Demand Arrival Data The new tourism demand arrival information could be found via `Hong Kong Tourism Board Website`, which is funded by `Hong Kong Government` and regularly provides the new trend and statistics on the arrival volume information on tourism demand for `Hong Kong` region.

3.2 Evaluation Metrics

The following evaluation metrics are commonly used:

Mean Squared Error (MSE) The `MSE` is one of the most common used metrics on tourism demand forecasting. It measures the average squared difference between the actual value y_i and the predicted values $\hat{y}_i$ generated by the model

$$MSE = \frac{1}{D}\sum_{i=1}^{D}(y_i - \hat{y}_i)^2 \tag{1}$$

Mean Absolute Percentage Error (MAPE) MAPE measures the average absolute percentage difference between the actual value y_i and the predicted value $\hat{y}_i$ generated by the model. The MAPE is calculated as follows:

$$MAPE = \frac{1}{D}\sum_{i=1}^{D}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{2}$$

Mean Absolute Error (MAE) MAE used to evaluate the accuracy of a forecasting model. It measures the average absolute difference between the actual values y_i and the predicted value $\hat{y}_i$.

$$MAE = \frac{1}{D}\sum_{i=1}^{D}|Y_i - \hat{Y}_i| \tag{3}$$

4 Project Tasks

In this project, you are expected to select no more than two research topics, update the latest literature review, collect/up the benchmark datasets, explore the method and implement with the code to prepare an extensive empirical results report.

This report should detail the performance of SOTA methods based on the benchmark datasets, using common performance metrics derived from the references. The report should offer insightful analysis and critical evaluation of the methods in context.

References

[1] Law, Rob and Li, Gang and Fong, Davis Ka Chio and Han, Xin, *Tourism demand forecasting: A deep learning approach*, Annals of Tourism Research,75, 410–423, 2019.

[2] Zhang, Yishuo and Li, Gang and Muskat, Birgit and Vu, Huy Quan and Law, Rob, *Predictivity of tourism demand data*, Annals of Tourism Research,89, 103234, 2021.

[3] Zhang, Yishuo, Gang Li, Birgit Muskat, Rob Law, and Yating Yang, *Group pooling for deep tourism demand forecasting*, Annals of Tourism Research,82, 102899, 2020.

[4] Bing Pan and Doris Chenguang Wu and Haiyan Song, *Predicting Hotel Demand Using Destination Marketing Organizations' Web Traffic Data*, Journal of Tourism Research, 2013.