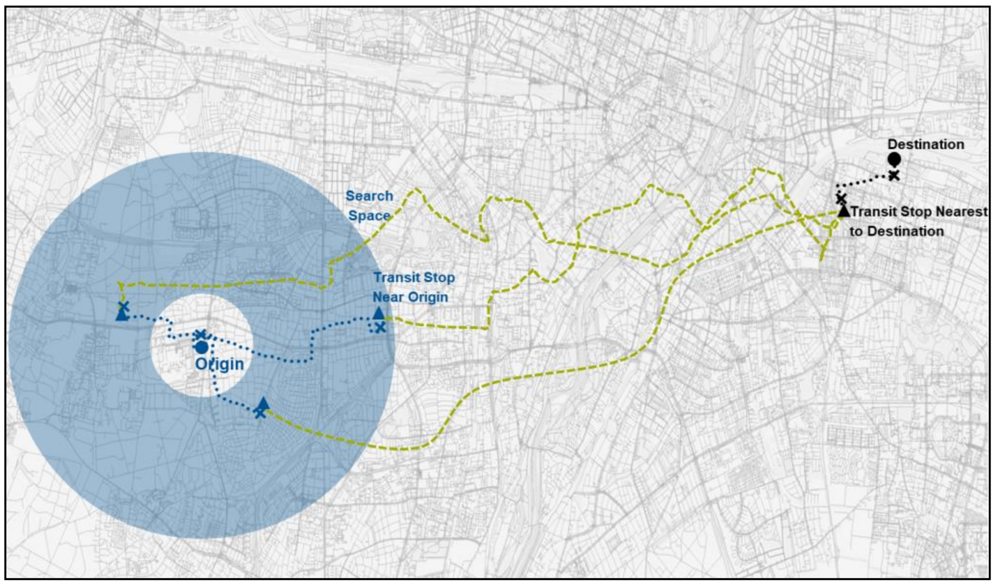
Development of Heuristics and Machine-learning Approaches to Find Transfer Points in Intermodal Travel

Master‘s Thesis of Anirudh Ray

Mentoring:
Dr.-Ing. Florian Dandl
Chenhao Ding, M. Sc.
Santiago Álvarez-Ossorio Martínez, M. Sc.

External Mentoring:
Josef Jaud (Stadtwerke München GmbH)

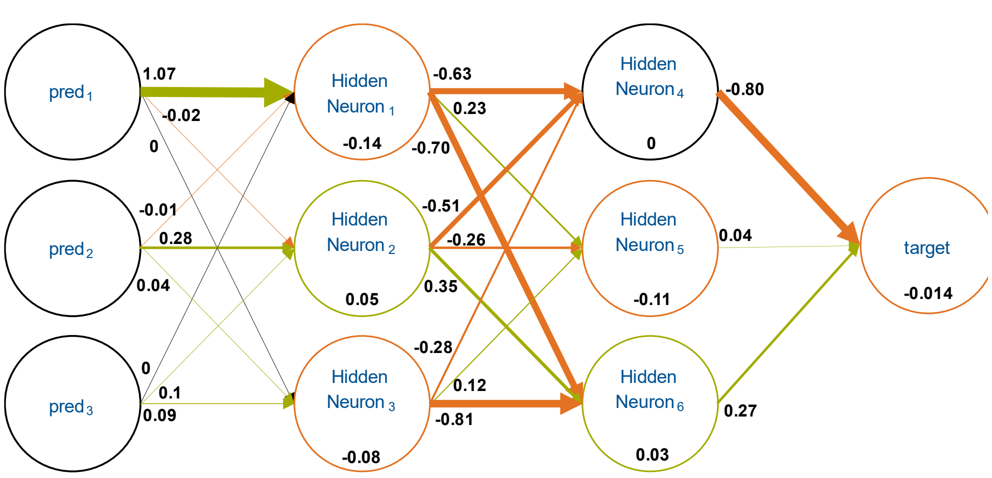


This research explores heuristics and machine learning models to speed up multi-modal journey-finding. The study focuses on the challenge of selecting the most suitable transit stops from a large set near the origin of a journey, wherein using the exhaustive set is computationally expensive for large-scale and real-time applications. The two heuristics developed – stop type and daily service count – are designed to reduce the number of transit stops considered during journey planning. The stop type heuristic filters out bus-only stops, while the daily service count heuristic selects stops based on the number of daily services. These heuristics achieve substantial computational gains, reducing the stops considered by up to 87% and 69%, respectively, while maintaining high solution accuracy with less than a 6% and 3% difference (respectively) from exact solutions. Parallelization of program code and utilization of custom-built data structures leads to further speedups in associated earliest-arriving journey computations.

A machine learning model was developed and trained using data generated from the daily service count heuristic, designed to predict the suitability of transit stops in forming part of optimal multi-modal journeys. The model uses key attributes such as the daily service count of transit stops, the type of stop, and the average transfer duration at each stop to make its predictions. These attributes are widely available in most geographies due to the proliferation of General Transit Feed Specification (GTFS) data, making the model highly transferable and adaptable to various transit networks.

The model was tested on a dataset of 17,000 records and demonstrated a mean squared error (MSE) of approximately 6%, with a mean absolute error (MAE) of 3% and a median absolute error (MedAE) of 0.9%. This level of performance highlights the model's ability to predict travel times and stop suitability with a high degree of accuracy, even in complex, multi-modal journey environments. The model is especially adept at identifying small percentage differences between exact and heuristic-based solutions, which is crucial for minimizing total journey time while reducing computational costs. Notably, it can generalize well to new data, with balanced training and testing errors indicating its potential applicability to real-time journey planning tasks across various network conditions.

Performance Parameter (Unit)	Value
GTFS data processing time (min)	15.34
OpenStreetMap-OPL data processing time (sec)	3.11
KD-Tree building time (sec)	0.21
Multi-modal queries’ parsing time (sec)	0.33
Time taken per exact solution (sec)	3.76
Transit stops considered near origin for exact solution	399
Time taken per stop type heuristic-based solution (sec)	0.57
Transit stops considered for a stop type heuristic-based solution	49
Share of accurate stop type heuristic-based solutions (%)	83.33
Average percentage difference between exact and stop type heuristic-based solutions (%)	5.94
Time taken per daily service count heuristic-based solution (sec)	1.67
Transit stops considered for a daily service count heuristic-based solution	180
Share of accurate daily service count heuristic-based solutions (%)	93.58
Average percentage difference between exact and daily service count heuristic-based solutions (%)	2.75



The study highlights that the daily service count heuristic performs better in terms of both flexibility and accuracy, though at the cost of increased computation time. However, both heuristics and the machine learning model present valuable tools for balancing efficiency and accuracy in multi-modal journey planning.

The machine learning model further showcases its ability to enhance multi-modal journey planning by avoiding the need for computationally expensive brute-force evaluations. Instead, it leverages a stop’s service frequency and transfer potential to rapidly identify optimal or near-optimal transit stops. This capability could significantly speed up the decision-making process in real-time applications, especially when integrated into existing journey-finding algorithms. Future enhancements could incorporate additional dynamic factors, such as real-time traffic conditions or user preferences, further improving the accuracy and efficiency of multi-modal transportation systems.