Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9009

OD Matrix Acquisition Based on Intersection as Origin-Destination

Shujie Wang. Lin Zhang
North China University of Science and Technology, Tangshan, China

How to cite this paper: Wang, S., & Zhang, L. (2025). OD Matrix Acquisition Based on Intersection as Origin-Destination. Advances in Engineering Research: Possibilities and Challenges, 1(1), 117-124. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9009
Published: 2025-03-31

Copyright © 2025 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

OD (Origin-Destination, OD) matrix as a core tool for traffic analysis, its accu-racy and real-time performance directly affect the effectiveness of traffic man-agement strategies. The study explores the division of traffic allocation areas based on intersection nodes to extract the OD matrix of intersections so that it can be used as a comparison of the accuracy of OD results with other models.

Keywords

OD matrix; STL; Origin; Destination

1. Introduction

Dynamic traffic management system, an important research content of the intelligent transportation system, the core of which is the traffic management center according to the dynamic traffic distribution system predicted traffic flow state of the road network, the release of traffic condition information, vehicle path guidance information, to produce instantaneous changes in the traffic optimization and control scheme, to enhance the spatial planning of traffic flow, and to ease the congested traffic flow. And the realization of dynamic traffic distribution system requires the help of OD matrix as the base input. The problem of estimating and predicting the OD matrix becomes fundamental.

2. Overview of the OD Matrix

2.1. Basic Form of OD Matrix

Traffic OD matrix is used to represent the traffic flow from the trip origin to the trip end, it is sorted by traffic neighborhoods by rows and columns, and each element in the matrix represents the amount of trips between the two subareas, reflecting the travel demand of traffic participants on the transportation road network, which is in the form of a two-dimensional n × n table. For a complete OD matrix, the following relationship is satisfied:

$$G_i = \sum_j T_{ij} \tag{1}$$

$$A_j = \sum_i T_{ij} \tag{2}$$

$$T = \sum_i \sum_j T_{ij} = \sum_i G_i = \sum_j A_j \tag{3}$$

G_i - traffic volume occurring in traffic cell i; A_j - the attracted traffic volume of traffic cell i; T_{ij} - the traffic volume where the trip starts at neighborhood i , and the trip ends at neighborhood j ; T - Total traffic for all transit neighborhoods within the study area.

2.2. The Significance of Studying OD

In transportation planning and management, OD matrix plays a crucial role, which provides a scientific basis for the optimization of the transportation system by quantitatively analyzing the traffic flow between Origin and Destination, the starting point of the trip. The significance of studying the generation method of OD is mainly reflected in the following aspects:

(1) Traffic Demand Analysis: OD matrix can reflect the distribution of travel demand between different regions and help transportation planners to identify high traffic areas and potential bottlenecks, so as to develop a more reasonable transportation network layout;

(2) Infrastructure planning: Based on the OD matrix, roads, public transportation routes and hub facilities can be accurately designed to ensure efficient allocation of resources to meet actual travel demand;

(3) Traffic flow simulation and prediction: OD Matrix is the basic data for traffic flow simulation, which can support the traffic management department to predict future traffic flow changes, formulate response policies in advance, and alleviate traffic pressure[1];

(4) Policy evaluation and optimization: OD information can evaluate the impact of traffic policies (such as traffic restrictions, fees, etc.) on travel behavior, provide data support for policy adjustment, and enhance management effectiveness;

(5) Multi-modal Transportation Integration: OD matrix helps to analyze the travel characteristics of different transportation modes (e.g., buses, subways, private cars, etc.), promote the synergistic development of multi-modal transportation systems, and improve the overall transportation efficiency.

3. Feasibility Analysis Using Intersections as Origin-Destination

In traffic planning, OD matrix is an important tool for describing the distribution of traffic flow at the starting and finishing points, but OD estimation methods often rely on models and algorithms, and it is difficult to obtain the real OD matrix, which leads to the deviation between the estimation results and the real situation due to the simplification of the modeling assumptions or the uncertainty of the parameter estimation, and other factors[2]. With the development of information technology, especially the wide application of electronic police data, it is possible to obtain more accurate OD. The feasibility of the OD matrix obtained by using the electronic police passing data with intersections as the starting and ending points of the traffic cell is discussed as follows:

(1) Traffic demand management theory: signalized intersections, as the regulation point of traffic flow, have a significant impact on the distribution of traffic flow and demand satisfaction. By implementing measures such as left-turn light strip going and bus-only lanes, the traffic management of the intersection can be improved, which in turn has an impact on the formation of the OD matrix;

2) Traffic distribution theory: the distribution of traffic flow depends on the travel impedance between traffic cells, such as time and distance. Signal intersection as the boundary of the cell, its signal cycle, red light duration and other parameters will directly affect the travel impedance between the cells, which will have an impact on the prediction of OD matrix;

(3) Traffic flow theory: signalized intersections, as key nodes of traffic flow, have parameters such as capacity and saturation that are crucial for assessing traffic flow characteristics, and analyzing these characteristics helps to predict the OD matrix more accurately;

4) Traffic modeling: using intersections as the starting and ending points of a traffic cell allows for a more accurate simulation of the generation, distribution and allocation of traffic flows. This approach takes into account the actual traffic connections between the cells, which improves the OD matrix prediction accuracy.

In summary, the advantage of using intersections as traffic cells is that they can directly reflect the actual traffic flow distribution reducing the errors caused by model assumptions and parameter estimation.

4. Data Pre-processing

4.1. Road Network Selection

This study aims to explore the OD generation method of urban road network, therefore, the following influencing factors should be considered when selecting the example road network: 1) the study road network should be an urban road network, with traffic flow having morning peak, flat peak and evening peak characteristics, containing congested road sections or areas; 2) the topology of the road network should cover a wide range of land use types, which needs to cover a large geographic space, and be able to depict a complex vehicular travel trajectory; and 3) The electronic police equipment in the road network has comprehensive coverage and works properly, and can provide reliable measured traffic data.

Based on the above considerations, the road network form is selected as follows Figure 1:

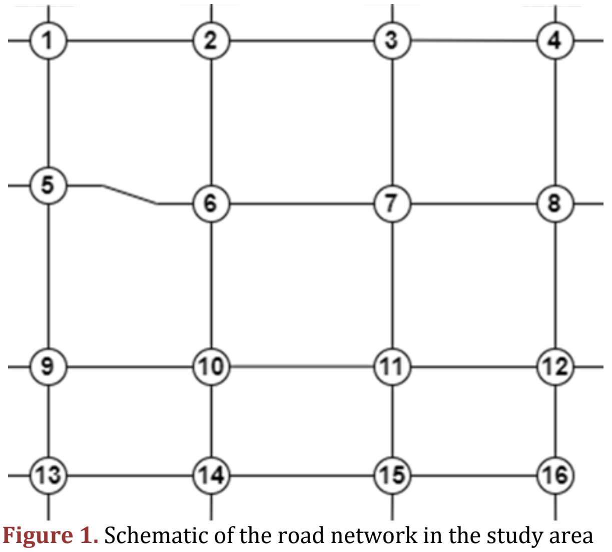

Figure 1. Schematic of the road network in the study area

4.2. Data Properties

The OD matrix within the urban road network is extracted by dividing the traffic allocation area in terms of intersection nodes and simulating the starting and ending points of vehicles. Extract the data collected by the electronic police equipment from the traffic database, the dataset contains information such as the license plate number of the vehicle and the time it passed through the intersection(Table 1). Taking the license plate number as the first identification, segmenting and sorting according to the time interval, determining the start and end points of the vehicles in the time period through group aggregation, and constructing the OD matrix using the cross-table function. The OD matrix within the urban road network is extracted by dividing the traffic distribution area in terms of intersection nodes and simulating the starting and ending points of vehicles. Due to the large amount of data, in order to be able to check the data processing process, a particular license plate number can be printed to demonstrate the processing, the specific steps are as follows:

Step1: Import the dataset into the database after preprocessing and filter the invalid data;

Step2: Create a database connection engine with a function;

Step3: Convert the timestamp form to datetime type;

Step4: Match the location index information in the dataset with the electric police points;

Step5: Match the location index information in the dataset with the point locations;
Step6: Grouping and aggregating, obtaining the starting point and end point, printing the starting point and end point information, and constructing the OD matrix;
Step7: Filter specific license plate number data and capture path information for checking;
Where Step5 in the acquisition of each vehicle travel O point and D point, that is, a vehicle's O point position in the time period of the first time in the intersection of the location of the intersection cell, D point for the last time in the time period in the intersection of the location of the intersection cell.

Table 1. Partial vehicle travel information.

License plate number	Date	Pass_time	Location_code
BJ***1	20241022	07:51:21	10
BJ***1	20241022	07:53:18	14
BJ***1	20241022	07:56:25	13
BJ***1	20241022	07:59:03	9

4.3. Completion of Missing Data

The data collected by the electronic police have obvious time series characteristics, including periodicity (e.g., daily cycle, weekly cycle) and trend, and are subject to data missing problems due to equipment failures, data transmission problems, or external factors. This indeed not only affects the complete analysis of traffic patterns, but also may lead to bias in traffic flow prediction models, which in turn affects traffic management and decision making.

Seasonal-Trend Decomposition using LOESS Model (STL) is able to decompose time series data into trend component, seasonal component and residual component[3]. Through this decomposition the structured information in the data can be identified without the interference of discrete cluster points, and the trend and seasonal characteristics of known data points can be used to fill in the missing values, thus restoring the integrity of the time series. In the time series analysis of electronic police data, STL decomposition can identify and separate trend components and seasonal components, such as daily peak hours, weekday and weekend traffic differences and long-term trends, and can more accurately analyze and predict changes in traffic flow.STL decomposes a time series into three parts:

$$y(t) = T(t) + S(t) + R(t) \tag{4}$$

$T(t)$-Trend components ; $S(t)$- Seasonal component; $R(t)$- Residual component.

The trend component represents the long-term variation of the time series, the seasonal component represents the cyclical variation of the time series, and the residual component represents the stochastic fluctuation component after removing the trend and seasonality.

(1) Estimation of trend component $T(t)$

LOESS regression is a nonparametric regression method for fitting the trend component of time series data, which estimates the trend by fitting a polynomial in the local neighborhood of each data point of the form:

$$T(t)=\sum_{i=1}^{n} w_i(t)\cdot y(t_i) \tag{5}$$

$$w_i(t)=\left(1-\left(\frac{|t-t_i|}{\delta}\right)^3\right)^3 \tag{6}$$

$w_i(t)$ - the weight of time t out on $y(t_i)$; $y(t_i)$ - the original observation at time t_1; δ - the bandwidth parameter that controls the rate of decay of the weights.

The bandwidth parameter δ can be chosen through cross-validation, with smaller δ ones making the fit closer to the data points and reflecting more details, and larger δ ones making the fit smoother and ignoring more local fluctuations.

(2) Estimation of the seasonal component $S(t)$

The trend component is removed from the original time series to obtain the series Y, preserving seasonality and random fluctuations:

$$y_d(t)=y(t)-T(t) \tag{7}$$

The series is grouped into seasonal periods m. If grouped according to hourly data, m is taken as 24 for the daily period, which contains data points: data points of 1, 25, 49,⋯; if according to daily data, m is taken as 7 for the weekly period. For each grouping $j(j=1,2,\ldots,m)$), the mean value of the data points is calculated and used as an estimate of the seasonal component:

$$S(t)=\frac{1}{N_j}\sum_{t\in G_j} y_d(t) \tag{8}$$

G_j - The set of time points in group j; N_j - number of data points within the group.

In some cases, the direct use of group averages may result in large biases in the estimates of the seasonal components, so they are corrected for centering, i.e., the adjustment is achieved by subtracting the average value of each seasonal component from its mean.

(3) Estimation of the residual component $R(t)$

For missing values, the estimated trend and seasonality components can be reconstructed using the estimated trend and seasonality components, and if the residual component is unknown, it can be estimated using its average value over the same period since the residuals are usually considered to follow fluctuations around a mean of zero.

5. OD Matrix Acquisition with Intersection as Origin-destination

5.1. OD Matrix Acquisition Method

After data preprocessing, segmentation and sorting according to time intervals, determining the start and end points of vehicles in the time period through group aggregation, and constructing the OD matrix by using the cross-table function. The OD matrix within the urban road network is extracted by dividing the traffic allocation area with the intersection node as a unit, simulating the starting and ending points of vehicles, and the specific steps are as follows, the specific steps are as follows:

Step1: Read the original data file, add row and column labels, and filter out invalid data;

Step2: import SQL alchemy library and create database connection engine with functions;

Step3: Match the location information in the dataset with the road network number;

Step6: using the license plate number as the keyword and each vehicle as a unit, sort the vehicle elapsed time and separate them according to the time period;

Step7: grouping and aggregating, obtaining the start point and end point of each vehicle, and printing the start point and end point information;

Step8: Construct and print the OD matrix.

The OD matrix of each ten-minute trip during peak hours on weekdays is selected as an example, and the OD matrix of trips starting and ending at intersections can be derived according to the above steps. Since the experiment is not divided into time intervals, the vehicle endpoint is allowed to be an arbitrary node, i.e., there is a situation where the starting point and the end point are at the same point. Part of the data is shown in Figure 2:

O/D	1	2	3	4	5	6	7	8
1	12	48	11	0	112	2	0	0
2	44	5	55	7	2	122	0	0
3	18	65	5	87	0	3	61	0
4	1	4	109	0	0	0	6	71
5	90	3	0	0	10	70	0	0
6	1	116	1	0	71	5	63	0
7	0	0	76	3	2	73	5	74
8	0	0	0	94	0	0	69	2

Figure 2. Example of OD Matrix.

Conclusion

With the data captured by the electronic police system, the research needs can be met after the data cleaning and processing using the STL model. The study explores the division of traffic allocation areas based on intersection nodes, and extracts the OD matrix of the starting and ending points of intersections as a comparative reference value for the subsequent OD generation method. The results show that the method can accurately and efficiently generate OD matrices, which provides strong support for transportation planning and management.

References

[1] Han L, Ma X, Sun L, et al. Continuous-time and multi-level graph representation learning for origin-destination demand prediction[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022: 516-524.

[2] Wang R, Guo Q, Dai S, et al. An OD time prediction model based on adaptive graph embedding[J]. Physica A: Statistical Mechanics and its Applications, 2025, 657: 130217.

[3] Cleveland R B, Cleveland W S, McRae J E, et al. STL: A seasonal-trend decomposition[J]. J. off. Stat, 1990, 6(1): 3-73.