Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2026.060116
ISSN 2790-0983 Vol. 6, Issue 1: 140-150

Research on Personalized Route Recommendation Methods Combining Driving Behavior Profiles and Multi-Dimensional Constraints

Yuntian Shen

College of Engineering, Industrial Engineering & Operations Research, University of California, Berkeley, 94720, California, USA

Keywords: Driving behavior profiling; personalized route recommendation; time and distance constraints; net increment prediction; candidate route generation

Abstract: To address the challenges of assessing the value of driving behavior information under fixed candidate budgets and the potential for personalized candidates to replace effective general routes, this study proposes a multi-dimensional constrained path recommendation method driven by net coverage increment. The study constructs a road selection profile based on vehicle trajectories. Under unified constraints of traffic flow, time, and distance, it calculates positive, zero, and negative coverage increments for different candidate allocation schemes and predicts the pre-trip supplementary benefit. Results show that the proposed method achieves a RR@5 of 33.44%, which is 1.98 percentage points higher than the fixed-quota method and 0.63 percentage points higher than the strong preference-perceived baseline, but the difference interval does not show a stable advantage. Under a 30 ms response budget, the proposed method's coverage is lower than the preference-perceived baseline, and the Top-1 overlap does not improve proportionally. The study indicates that the effectiveness of the behavior profile depends on the assessment of the supplementary benefit, and its performance is also influenced by the incremental label distribution, road constraints, and computational budget.

1 Introduction

Urban navigation recommends paths according to distance, time and traffic situation, but the selection of the route behavior of vehicles does not fully correspond with minimization of an only one cost. Route choice may be dependent on knowledge of the route, steering load and travel situation. There are also significant differences in the factors affecting route selection at different time and among different group, as Zhang et al. pointed [1]. Li et al. discovered that there are both interactions between driver attributes and navigation information that impact the route selection decision [2]. Based on these studies, historical behaviour can serve to provide the additional information needed for route recommendation, however, it is not necessarily suitable for all trips.

The data from the vehicle trajectories is used for analysis of the route selection difference. Deng et al. grouped route selection based on the historical trajectories, and concluded a correlation

Copyright: © 2026 by the authors. This is an Open Access article distributed under the Creative Commons Attribution License (CC BY 4. 0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons. org/licenses/by/4. 0/).

between behavioural characteristics and the actual routes [3]. Wan and Dodge surveyed the preferences like time efficiency, simplified steering operations and habitual routes and offered a preference-aware candidate generation procedure [4]. It is important to note, however, that this profile is stable, meaning that using personalized candidates may not elicit a greater effective coverage - as each commercial vehicle travels through the same collection of roads due to the type of service operations they are carrying out.

The candidate determined the range of routes for the final ranking as per the following. Liu et al. used to integrate the two models, namely the candidate generation and the route selection models, and they suggested that these two would affect each other [5]. When the candidate number is small, the route covering effect of different generation algorithms is very different in high resolution road network map (Wang et al., 2008) [6]. Malhotra et al noted the necessity of choosing between authenticity of candidate routes, redundancy and applicability of the algorithm [7]. Some previous work includes behavior aware path generation, but there remains a particular issue: based on a specific trip, is it beneficial to replace a generic path generation with a personalized one, and will the benefit outweigh the extra computation required to generate the personalized path?

The first is about predicting the coverage increments that the candidate is able to provide for a fixed budget, the second is about studying how the quota given for the candidates change according to the driving behavior profile in the presence of multiple types of road constraints. The model builds three types of labels (positive, zero and negative) for every quota and assigns search quota according to the pre-departure profiles and road conditions based on historical trips. The study assesses the method's effectiveness based on prediction error, simple allocation rules, strong baseline comparison and discusses its applicability under a response time constraint.

2 Research Methods

2.1 Research Design and Data Processing

This study used 10-second resampled trajectories of 500 commercial vehicles in a city in September 2020, 5-minute traffic statistics, and a formal road network containing 578 road segments. The study reconstructed trips by vehicle and time, and examined road connectivity, timestamps, origin and destination, and minimum mileage. Included vehicles had at least 3 valid historical trips during the training period and at least 6 qualified trips during the testing period; ultimately, 160 vehicles were selected as the analysis subjects. For each vehicle, the last 6 qualified trips from the testing period were retained based on departure time. The remaining test trips were not included in the evaluation denominator and were not used for global model training or parameter tuning, but records earlier than the target trips could be used to update the online profile. Anonymous vehicle identifiers are not equivalent to natural person drivers.

2.2 Behavioral profiling and multi-dimensional constraint path generation

The profile includes relative detour rate, turning frequency density, road repetition rate, and conditional driving efficiency. The relative detour rate is the ratio of the observed distance to the shortest distance between the same origin and destination minus one; a change in direction exceeding 45° is considered a valid turn, and the density is calculated by kilometer. The road repetition rate is the proportion of the side length of the road segment repeatedly traversed within 14 days, and the driving efficiency is based on the historical time of the same road segment and similar time period before departure. The profile uses a maximum of 30 valid trips within the most recent 14 days. When there are fewer than 3 historical trips, the model uses the group features and missing markers during the training period, and does not learn individual weights separately. Each feature is

normalized according to the median and interquartile range during the training period; when the interquartile range is zero, it is replaced by 1. Trip selection is based on road connectivity and abnormal jump control of road segment matching error [8].

The study denotes the directed road network as $G = (V, E)$and sets the road segment directions and connectivity as hard constraints. For the origin and destination od, the candidate set that simultaneously satisfies both time and distance budgets is defined as .

$$\mathcal{F}_{od} = \{r \in \mathcal{P}_{od}: T(r) \le (1+\epsilon_t)T_{min},\ D(r) \le (1+\epsilon_d)D_{min}\}. \quad (1)$$

In the formula, $T_{min}$and $D_{min}$are calculated independently from the same road network, with basic tolerances set to $\epsilon_t = 0.18$and $\epsilon_d = 0.20$. Path time is estimated based on a five-minute traffic window available before departure, without using measured speeds after departure. The study separately handles complete feasibility determination and resource-constrained search. If no compliant route is found after 2500 queries or a 30 ms response deadline, it is recorded as "no candidate found within the resource limit," rather than directly judging the road constraint as infeasible. The shortest time route exceeding the budget can be used as an independent fallback solution, but it is not included in the constrained candidate set or RR@K statistics.

General search uses non-negative time, distance, and turning costs; behavior-aware search adjusts the weights of similar attributes based on the user profile. Let be a non-negative road segment feature, $\phi_e\psi_{ef}$be the turning feature of adjacent directed edges, and define the path cost as...

$$C_u(r) = \sum_{e\in r} \mathrm{softplus}\,(Ap_u + b)^\top \phi_e + \sum_{(e,f)\in r} \mathrm{softplus}\,(Bp_u + d)^\top \psi_{ef}. \quad (2)$$

The model learns weights by pairwise ranking loss of observed routes and feasible alternative routes with the same OD: $\mathcal{L} = \sum_i \log\left[1 + \exp\left(C_u(r_i^*) - C_u(r_i^-)\right)\right] + 10^{-3} \parallel \theta \parallel_2^2$. Each training trip can extract a maximum of 5 feasible alternatives at historical time points. Training uses the Adam optimizer with a learning rate of 0.01, a batch size of 128, and a maximum of 50 rounds; the internal validation loss stops early if it does not decrease for 5 consecutive rounds. For the evaluation trip t, the model only uses the rolling history estimate of standardized parameters before departure, fills in missing values, and learns cost weights. Inner layer training fits the first 80% of the time prefix and stops early for the last 20%. When history is insufficient, a fixed non-learning cost and population missing labels are used, and the parameters are not updated using the current trip. Algorithm alternatives do not represent routes that the driver has explicitly rejected. Wu et al.'s research on trajectory and road network learning provides a basis for road state representation, but cannot replace vehicle-level profiling [9].

The candidate engine employs Yen's off-path enumeration, calculating the turning cost by expanding the state through the "current node - previous directed edge" relationship. Output routes must not contain duplicate nodes; identical route sequences are deduplicated before further enumeration. Each request accesses the expanded state a maximum of 2500 times. The two search methods alternately generate candidates according to allocated quotas; duplicate routes across sources are counted only once. Once a source is exhausted, remaining quotas are transferred to the other source until K100 candidates are obtained or the common expansion limit is reached. If a timeout occurs or there are insufficient candidates, the algorithm retains previously generated compliant routes and their actual number, without using duplicate routes to fill the gaps. Neither experiment includes coverage of over-budget fallback routes.

2.3 Net Increment Forecast and Candidate Budget Allocation

The total candidate limit is denoted as K, where a $\in \{0,1,\ldots,K-1\}$is the number of personalized

search candidates, and the remaining candidates are used for general search. A fixed-ratio baseline is set $a = \mathrm{round}(0.8K)$; when , 4 personalized candidates and 1 general candidate are generated; $K = 5K = 10$when , the two types of candidates are 8 and 2 respectively. For different conditions K, candidates and labels are constructed separately, and the predictor is retrained. Under $K = 5$the condition , the fixed ratio corresponds $a = 4$, therefore $a = 4$, when selecting the adaptive method, both methods should return the same candidate set.

For historical itineraries i, the observed routes are denoted as and r_i^* the candidate sets corresponding to $\mathcal{C}_i(a)$the quota allocation are denoted as a. The proportion of the length of the common directed edge to the length of the observed route is used to determine whether the route is reproduced.

$$Y_i(a) = \mathbb{I}\left[\max_{r \in \mathcal{C}_i(a)} \frac{L_\cap(r, r_i^*)}{L(r_i^*)} \geq 0.80\right]. \tag{3}$$

If the candidate set is empty, then let $Y_i(a) = 0$, and retain that trip in all test samples. When calculating the common edge length, repeated segments are counted based on the smaller number of passes between the two routes. The study also calculates symmetric overlap $O_{sym} = 2L_\cap/[L(r) + L(r_i^*)]$to reflect potential additional detours for candidate routes. Top-1 refers to the compliant candidate ranked first after unified sorting; when a candidate is empty, its one-way and symmetric overlap are both recorded as 0. Both overall means and RR@K are calculated with all 960 test trips as the denominator. Compared to a purely generic configuration, net coverage increment is defined as

$$\Delta_i(a) = Y_i(a) - Y_i(0) \in \{-1, 0, 1\}. \tag{4}$$

A positive net increment indicates new coverage, a negative increment indicates that a previously valid general candidate has been replaced, and a zero increment indicates that the coverage status of the two configurations is the same. During training, the profile and cost are estimated for each evaluation trip using only the earlier historical window, and then the incremental labels for different configurations are calculated offline. The LightGBM regressor takes the profile, historical data quality, OD network characteristics, departure time, and aother factors as input to predict...

$$\widehat{\Delta}_u(a) = \widehat{\mathbb{E}}[\Delta_i(a) \mid p_u, c_i, a]. \tag{5}$$

LightGBM employs squared error regression (objective=regression, metric=l2), with 150 trees, a learning rate of 0.05, a maximum depth of 3, a maximum leaf node count of 7, and a minimum leaf sample count of 30. Training labels are generated using the inner-layer rolling start point during the training period; information used for standardization, missing data imputation, profiling cost model fitting, and early stopping is all prior to the target run. The outer-layer validation period from September 22nd to 25th is used only for selecting regressor hyperparameters and unifying quotas, and does not participate in training label generation. Online quota decision is defined as...

$$a_u^* = \arg\max_{0 \leq a < K} \widehat{\Delta}_u(a). \tag{6}$$

For $a = 0$direct settings $\widehat{\Delta}_u(0) = 0$, no regressor is trained. Other configurations are selected only when their prediction gains are no greater than zero $a = 0$; in case of a tie, the configuration with fewer personalized slots is prioritized. Only candidates for the selected configuration are generated during the online phase. All methods are sorted in ascending order of the same normalized profile cost $C_u(r)$; when costs are the same, the route with the shorter prediction time is selected to differentiate the role of candidate generation from the sorting. The cost of offline traversal of slots is calculated separately and is not included in the online single-run time.

2.4 Controlled Experiment and Statistical Methods

The study designated September 1-21, 22-25, and 26-30, 2020, as the training, validation, and testing periods. Valid trips must contain at least three directed edges and a distance of at least 0.8 km; trips were segmented when continuous observation was interrupted for more than 120 seconds. Inner-layer rolling training retained at least three historical trips to ensure that the current trip did not participate in its own feature and cost learning. All methods used the same OD set, road network, and time and distance budgets, with comparisons including shortest time, K-shortest time, Link Penalty, general diversity search, static profiling, fixed ratio, unified optimal quota during validation, profiling quality allocation, and preference-aware candidate generation. The preference-aware baseline referenced a publicly available implementation. During training, route categories were identified based on road attributes; during testing, the category was predicted using the pre-departure profiling, and routes were then selected from candidates generated by the unified candidate engine. Its deduplication, completion, and backoff rules were consistent with the proposed method. Category prediction and online candidate pool generation were both included in the time consumption, and the observed routes of the test trips were not used. Wang et al.'s research on route selection integration showed that model performance is affected by sample conditions, so all methods use the same time division[10].

This study evaluates RR@K, Top-1 unidirectional and symmetric overlap, actual number of candidates, query-level double-budget compliant candidate return rate, and single request latency. For trips without compliant candidates, RR@K and both overlap factors are recorded as 0, and the trip is still included in all test samples; no returned candidates are not directly equivalent to constraint infeasibility. Since the same number of candidates may correspond to different computational costs, a 30 ms response cutoff experiment is included in the study. The timing environment is Intel Core i5-12400, single-threaded, Python 3.11, NetworkX 3.4, and LightGBM 4.5. Cross-request test route caching is disabled for each method. After 10 warm-ups, the measurement is repeated 5 times, and the median latency is summed. Timing includes profile reading, category or quota prediction, candidate pool generation, path enumeration, deduplication, and sorting, but does not include offline training. After the response timeout is reached, the search terminates, only the previously generated compliant candidates are returned, and no further rollback is performed. Statistics are based on 10,000 pairing bootstraps performed on anonymous vehicles. During the training, outer-layer validation, and out-of-time testing phases, the MAE and MSE of the incremental prediction and the constant-zero predictor were calculated only $a = 1,2,3,4$for each trip-action pair, excluding those defined as always zero $a = 0$. The experiments also report athe label distribution, final action frequency, history length sensitivity, and $K = 10$results for each.

3. Research Results

3.1 Samples and Incremental Labels

The study included 160 anonymized vehicle entities and 6150 trips, with 4550, 640, and 960 trips allocated to training, validation, and testing sets, respectively. Each entity retained 6 test trips. 750 observation routes met the basic joint budget, accounting for 78.13%. Table 1 lists the label configurations during training and the number of trips selected during testing. $a = 0$ The incremental value was always zero, while the zero-label ratio for the other four configurations ranged from 82.81% to 91.41%. Compared to pure general search, the final strategy added coverage 64 times and incurred losses 27 times, while the coverage status of 869 trips remained unchanged, accounting for 90.52%. The label distribution during training is detailed in Table 1. An additional 36 trips did not obtain compliant candidates; those exceeding the budget were not included in the

reproduction count.

Table 1. Selection of Training Increment Labels and Testing Period Quotas

Quota A	Training positive increment	Training with zero increment	Training negative increment	Test the number of selections
0	0	4550	0	200
1	218	4159	173	185
2	307	4009	234	220
3	372	3890	288	213
4	430	3768	352	142

3.2 Recommended performance and response budget

Table 2 shows that the pure general method, fixed ratio method, unified optimal method during the validation period method, profile quality allocation method, preference perception method, and the proposed method reproduced 284, 302, 306, 309, 315, and 321 runs, respectively. The proposed method improves performance by 1.98 percentage points compared to the fixed ratio method, with a subject pairing 95% range of 0.00–3.85 percentage points; it also improves performance by 0.63 percentage points compared to preference perception method, with a range of −1.25–2.60 percentage points. Both ranges include zeros, indicating that a stable advantage cannot yet be established.

Table 2 Comparison of candidate size and recommendations with a 30 ms cutoff

method	RR@K/%	Top-1/%	symmetry/%	Average number of candidates	Average time/ms	30msRR@K/%
Shortest time (K=1)	19.79	50.1	48.5	0.96	2.1	19.79
K Shortest time	25.52	50.1	48.5	4.75	14.4	25.31
Link Penalty	29.27	53.5	51.3	4.70	19.6	29.17
Purely general-purpose and diversified	29.58	53.8	51.5	4.72	20.3	29.38
Static image search	29.17	54.9	52.6	4.67	23.8	28.85
Fixed quota	31.46	**56.4**	**54.0**	4.67	27.5	30.94
Unified Optimization during Verification Period	31.88	56.3	53.9	4.69	27.9	30.42
Image quality allocation	32.19	56.2	53.8	4.67	30.2	31.35
Preference-aware candidates	32.81	56.0	53.7	4.67	34.1	**31.88**
Net incremental allocation in this study	**33.44**	56.1	53.8	4.54	35.8	30.31

Note: All results are taken except for the shortest time single path $K = 5$; Top-1 and symmetric overlap are expressed as percentages. Under a unified resource limit, 924 OD queries returned at least one compliant candidate, and 36 returned none. Since the latter did not undergo a complete feasibility check, it cannot be concluded that there are no feasible routes in the road network. All 960 trips are included in the denominator of RR@5, Top-1, and symmetric overlap; empty sets, early deadlines, or only budget overruns are recorded as 0. The compliant candidate return rate is 96.25%, which is a different metric from RR@5 and the average number of candidates for each algorithm.

Compared to a fixed quota, the average execution time of the proposed method increased from 27.5 ms to 35.8 ms, an increase of 30.18%. Under a uniform 30 ms response deadline, the proposed method and the preference-aware baseline reproduced 291 and 306 trips respectively, with a coverage difference of 1.56 percentage points, and no coverage advantage was observed in the

proposed method. Meanwhile, the Top-1 overlap of the proposed method was slightly lower than that of the fixed quota, indicating that the improvement in candidate coverage did not simultaneously improve recommendation quality and computational efficiency. Routes returned before the deadline still need to meet the time and distance budget; routes not generated after the timeout are not counted as valid candidates.

3.3 Incremental Prediction and Mechanism Testing

The study divided 160 test subjects into four groups based on predicted returns, with 40 subjects and 240 runs in each group. Figure 1 shows that the observed net gains for the four groups were −0.83, 0.42, 5.42, and 10.42 percentage points, respectively, while the predicted mean gains were 0.30, 1.80, 4.50, and 8.20 percentage points, respectively. The highest actual gain in the group was higher than the predicted value, indicating that the group ranking trend does not necessarily mean that the prediction has been accurately calibrated. The MAE of the constant-zero predictor and the incremental estimator during the training period were 0.1304 and 0.1356, respectively; during the outer validation period, they were 0.1367 and 0.1435, respectively; and during the out-of-time testing period, they were 0.1349 and 0.1515, respectively. The MAE of the incremental estimator was higher than that of the constant-zero predictor in all three periods. The MSEs during the testing period were 0.1349 and 0.1346, respectively, with limited differences. Errors were only counted up $a = 1$ to configuration 4. Since MAE is biased towards the conditional median, while quota decisions focus on conditional expectations and returns, the two types of indicators need to be evaluated separately.

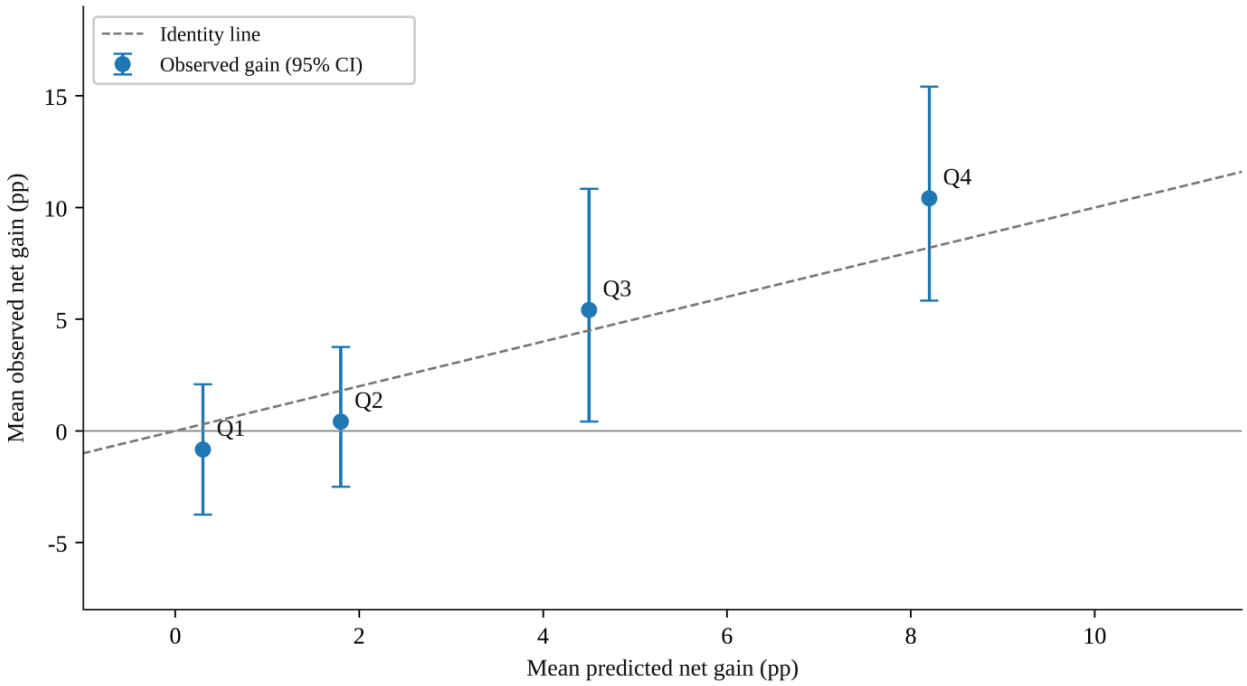


Figure 1. Grouped comparison of predicted and observed net coverage gains (95% interval)

Table 3 compares candidate coverage under different historical run counts. The test runs for 3–7, 8–15, and at least 16 historical runs were 216, 330, and 414, respectively. The reproducibility rates of the proposed method compared to the fixed-quota method differed by 0, 3.64, and 1.69 percentage points, respectively, and did not monotonically increase with historical run length. The

number of historical runs cannot solely explain the allocation payoff; prediction error and decision payoff still need to be evaluated separately.

Table 3. Comparison of candidate coverage for historical trip count groupings

Historical travel number	Test trip	Fixed quota recurrence number	Recurrence count in this study	This study RR@5/%
3 to 7	216	72	72	33.33
Articles 8-15	330	99	111	33.64
No fewer than 16	414	131	138	33.33

3.4 Multidimensional Constraints and K-Scale

Figure 2 compares the route reproducibility rates when the distance tolerance is 20%. With a time tolerance of 8%, the RR@5 for the fixed proportion, preference-aware, and proposed methods are 28.13%, 27.92%, and 28.02%, respectively; at 18%, they are 31.46%, 32.81%, and 33.44%; and at 25%, they are 38.02%, 39.58%, and 40.52%. Under strict budget constraints, the coverage differences among the methods are small; however, the differences increase after the budget is relaxed.

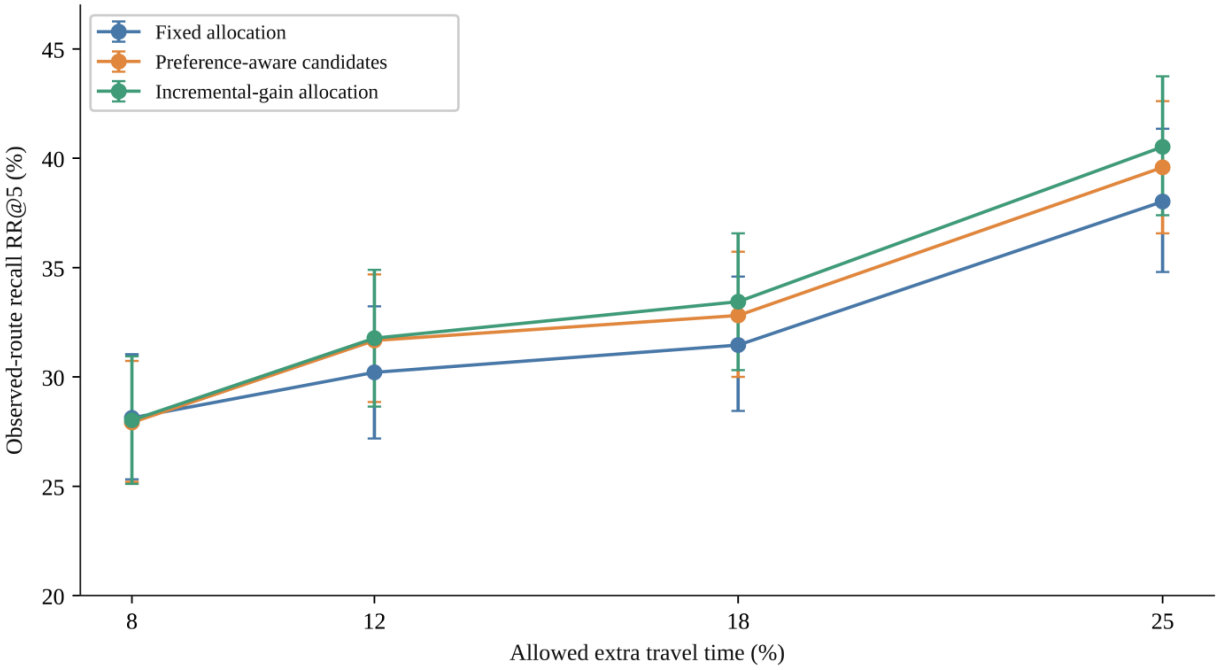


Figure 2. RR@5 (95% interval) of three candidate generation methods under time budget sensitivity.

$K = 10$At that time , the fixed-ratio quota was adjusted to 8 personalized candidates and 2 general candidates, and the proposed model $a = 0, \ldots, 9$was relearned based on the gain of each configuration. The fixed-ratio and the proposed method reproduced 412 and 416 test runs, respectively, with RR@10 of 42.92% and 43.33%, respectively. The difference was 0.42 percentage points based on the original number of successes, which was less than the $K = 51.98$ percentage points of the original method. Under the condition of K=10, the coverage gap between the two methods narrowed, indicating that increasing the candidate size may reduce the marginal benefit of quota allocation, but path enumeration and deduplication still incur computational costs.

4. Discussion

This separation of stability of historical behavioural profiles and net incremental value of candidate searches enables personalised quota allocation to take into account coverage and replacement losses of candidate searches. Table 2 indicates that while the proposed approach achieves a small average improvement at the same candidate size, the pairing interval contains also zero whereas other baselines (with strong Preference Perception) do not. This small advantage is not carried over to uniform response-cutoff-condition of 30ms which has to be seen together with online computation costs and is not sufficient for the candidate coverage benefit to be evaluated when compared with RR@5 without a time limit.

In training, many of the non-zero configurations had a large proportion of zero-increment labels and the constant-zero predictor was able to have low absolute errors for many of the trips. The incremental estimator was more accurate than a constant-zero predictor (CZP) with greater MAE despite not having an absolute error gain. But MAE argument is in terms of conditional median and prediction error is not the same as the argument of selection of the candidate based on the conditional expectation. More historical trips did not necessarily result in increasing betterment. We may have a mixture of key historical features, OD structure, number of incremental positive and negative samples, etc. influencing the prediction performance and out-of-time calibration under various route complexities is still to be tested.

Results can only be interpreted within the range of data and evaluation tools. A road record of commercial vehicles re-sampled for public using. Those repeated routes are not necessarily considered carte blanche preferences of natural persons themselves but might be impacted by operational tasks. Not uniform time/distance acceptance tolerance, which is not an explicit constraint suggested by the driver. Hao et al. discovered the need to consider the dual optimization of safety and traffic efficiency in vehicle routing from vehicle trajectories, which suggests the fixed dual budget will not be sufficient to deal with various factors for route selection [11]. The unidirectional nature of the road overlap may be skewed towards candidates that have additional detours, and a combination of symmetric overlap, a combined dual budget compliance rate and final cost analysis are needed. With that said, despite compliant routes in OD, the scale of search and the response time may not allow the algorithm to come up with the candidates in time.

This paper uses historical routes to assess the coverage of the candidates and does not take into consideration feedback from real navigation usage (or deviation) on the route. From route selection experiments, Han et al. discovered that the route reproduction in the future cannot be regarded as a substitute of the real-time response in the route following, since the tendency to follow the suggested routes also varies with the participants' goals [12]. Navigation feedback is important, but when it exists it's important to evaluate Top-1 quality, adoption rate and driving cost comprehensively. The scenario of applicability may also vary due to differing road network structures and vehicle groups, however, and requires further test work.

5. Conclusion

This work tries to solve the problem of limited allocation of personalized candidates under the multidimensional road constraints, presents a candidate allocation method to predict the net coverage increment when adding coverage while considering the losses when replacing general candidates. It is seen that although personalized slots will enhance good candidate slots in some trips, superiority over strong preference-aware baseline is not yet demonstrated; with a uniform 30 ms response deadline, no increase in coverage over the strong preference-aware baseline is observed. So, the usage policy for the profile should be based on a mixture from the candidate value and response time budget (RTB) side and candidate coverage and final recommendation quality

should be verified on separate grounds.

References

[1] *ZHANG N, LIANG Q, WU Y, RONG J. What influences drivers' travel route choice behavior? A literature review of the last 5 years' studies[J]. Research in Transportation Business & Management, 2026, 67: 101701. DOI: 10.1016/j.rtbm.2026.101701.*

[2] *LI W, AN Q, JI Y. Enhanced navigational insights and their impact on driver route choice: A hybrid utility-regret analysis with heterogeneity[J]. Research in Transportation Economics, 2025, 110: 101539. DOI: 10.1016/j.retrec.2025.101539.*

[3] *Zhang, J. (2026). Research on Financial Data Governance Framework and Data Quality Evaluation in the Big Data Environment.*

[4] *Su, J. (2026, July). Research on Machine Learning-based Abnormal Traffic Detection and Anti-Interference Methods for Android End-to-End Encrypted Transmission in High-Reliability Application Scenarios. In 2026 5th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-6). IEEE.*

[5] *Wang, N. (2026, May). Research and Development of an Enterprise Operation Load Forecasting Model Optimization and Resource Dynamic Allocation System based on A Temporal Attention Mechanism. In 2026 4th International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.*

[6] *Wang, Z. (2026). Study on Improving Equity Valuation Accuracy Driven by LightGBM Algorithm Regularization Optimization and Sample Balancing Strategy. Procedia Computer Science, 285, 621-630.*

[7] *Ma, W. (2026). Research on Reliability Enhancement of Cloud Data Platform Driven by Adaptive Sharding and Fault Tolerance Algorithms. Procedia Computer Science, 285, 1020-1029.*

[8] *Chi, C. (2026, July). Credit Risk Dynamic Pricing Representation Learning Algorithm based on Long-Term SelfSupervised Learning. In 2026 5th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-7). IEEE.*

[9] *Zhang, Y. (2026). A Framework for Semantic Alignment and Privacy Compliance of Large-Scale Language Models for Cross-Border Data Governance. Procedia Computer Science, 285, 1039-1047.*

[10] *Sun, L. (2026, August). Optimization and Practice of High-Performance Advertising Data Compression Based on LZ4 and ZSTD Algorithm Fusion. In 2026 IEEE 5th World Conference on Applied Intelligence and Computing (AIC) (pp. 326-332). IEEE.*

[11] *Sun, L. (2026). Optimization of Resource Scheduling for Multi Tenant Advertising Systems.*

[12] *Xin, H. (2026). Freight Operation Automation and State-Machine Processing Mechanism Based on a Distributed Workflow Engine.*

[13] *Qu, Z. (2026). Biocompatibility Evaluation of Nanomedicine Delivery Systems and Their Safety Research in Biomedical Engineering.*

[14] *Yuan, Y. (2026). High Performance Load Balancing Method for Distributed Inference of Large-Scale Language Models.*

[15] *Shi, C. (2026). Research on AI Driven Big Data Platform Intelligent Operation and Anomaly Detection Technology.*

[16] *Sun, X. (2026). Design of Infrared Thermal Imaging Fault Detection and Early Warning System for Data Center Equipment.*

[17] *Xu, Z. (2026). Dynamic Resource Allocation and Energy Optimization for Virtualized Data Center.*

[18] Xu, Z. (2026). Server Energy Consumption Monitoring and Adaptive Control in Large-Scale Data Center.

[19] Zhang, L. (2026). Research on the Construction of International Tax Compliance Systems and Tax Risk Control Strategies for Multinational Enterprises.

[20] Shao, D. (2026). Research on Real-time Data Synchronization Technology for Database Migration Based on CDC.

[21] Song, W. (2026). Research on Memory Leakage Detection and Optimization Methods for Android Applications.

[22] Liu, M. (2026). Evaluation of the Effectiveness of Financial Risk Management Tools Under the New Regulatory Environment.

[23] Gao, C. (2026). Organizational Change and Management Innovation in the Construction of Ai Data Centers under the Background of" Artificial Intelligence+.

[24] Li, S. (2026). Research on Corporate Expense Allocation and Budget Execution Variance Analysis Based on Multi-Dimensional Operational Indicators.

[25] Hu, J. (2026). Research on Fault Tolerant Mechanism of Payment Order State Machine Based on Raft Consensus Algorithm.

[26] Lin, Q. (2026). Research on Multi-source Demand Perception and Inventory-Capacity Co-optimization of Sales and Operation Planning in Discrete Manufacturing Enterprises.

[27] Li, Z. (2026). Design of Collaborative Architecture for Android Mobile Smart Systems Based on Edge Computing.

[28] Fei, W., Xue, W., & Wang, B. (2025). Research on the innovative path of rural education teaching methods from the perspective of PBE and PBL integration. Int. Acad. J. Humanit. Soc. Sci.

[29] Wendan Xue, Wen Fei. An Analytical Study on the Differences and Determinants of Basic Education Resource Allocation in Rural China and the United States. International Journal of Educational Research and Development, 1(3), 13-17.

[30] Boli Wang, Wen Fei. Innovations and practices in the classroom for rural aesthetic education in the perspective of school integration. International Journal of Educational Research and Development, 3(1), 36-41.