Diagnosis, Optimization, and Verification of Localized Failures in Traffic Flow Prediction Models

Yunjian Tang
College of Computer Science and Software Engineering, Shenzhen university, Guangdong, China
tommytang0331@gmail.com

Abstract. In the real traffic control scenario, the multi-intersection prediction model often appears "local failure" due to the distribution difference and concept drift between intersections, which leads to the decline of the overall prediction performance and affects the reliability of scheduling. Improving the robustness of the model under the conditions of distribution change and scene migration has become an important research direction of traffic time series modeling. This study proposes a data-driven framework of "diagnosis optimization verification" to be applied to the task of traffic flow prediction at multiple intersections. Aiming at the obvious local failure of stacking ensemble model (SEM), which is composed of linear regression and extreme gradient boosting (XGBoost), at junction 3, this paper analyzes the reasons for the decline of positioning performance through exploratory data analysis (EDA), including data distribution offset, extreme value interference and time series mode instability. In this study, targeted outlier truncation, multi-scale time series feature enhancement, box discretization and super parameter automatic optimization strategies were used to reduce the root mean square error (RMSE) of the model to 5.96 (43.49%) and improve the coefficient of determination (R^2) to 0.73. This study provides a repeatable and interpretable system optimization path to deal with the phenomenon of "same model with different effects" in traffic forecasting, and has reference value for time series forecasting tasks with heterogeneous distribution of traffic, energy and so on.

Keywords: Traffic flow forecast; Stacking integrated learning; Data heterogeneity; Feature engineering; Abnormal value handling

1 Introduction

Traffic prediction is the core of intelligent transportation system, and its performance directly affects the efficiency of signal control, travel planning and emergency scheduling. The traditional prediction model performs well in a single scenario, but in the intersection or long-term deployment environment, its performance is often degraded due to the change of data distribution, that is, "local failure" phenomenon [1,2].

Recent studies have pointed out that concept drift and distribution drift are the root causes of such problems. Hinder et al. systematically reviewed the concept drift

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_76

monitoring in the evolutionary environment, and proposed that the continuous verification of the assumption of distribution consistency should be maintained between model training and deployment [3]. Shen et al. based on the robust matrix factorization method, significantly improved the prediction stability in time-series tasks with high proportion of noise and abnormal samples [4]. Koch et al. proved that the distributed offset detection can be used to monitor the degradation of model performance in real time and identify the potential failure risk in advance [5].

Chang asymmetrically modeled the skewness error from the loss function level to alleviate the prediction bias under the right skew distribution [6]. Fan et al. alleviated the problem of distribution mismatch through input-output alignment in dish TS framework [7]. Chen et al. proposed a dynamic calibration strategy for context driven distributed transfer, so that the model has light-weight adaptive ability during deployment [8]. Darji et al. combined anomaly detection and data distribution strategy to reduce the interference of abnormal traffic [9]. Duan and Guo proposed to improve the prediction accuracy under distributed drift by using super network and online adaptation technology [10,11].

Based on these studies, this paper constructs a "Monitoring Optimization verification" integrated robust modeling process to deal with the local failure problem in multi- intersection traffic prediction. In the diagnosis stage, this paper analyzes the differences between junction3 and other intersections from the aspects of data distribution, outliers, time series characteristics and so on, and explores the source of heterogeneity. In the optimization stage, based on the diagnosis conclusion, the abnormal value processing, feature enhancement and parameter optimization strategies are designed and implemented. In the verification phase, the independent contribution and synergy of each strategy are verified through ablation experiments to ensure the interpretability and reproducibility of performance improvement.

2 Methods

2.1 Benchmark Model and Problem Recurrence

This paper uses two-layer SEM, the first layer is linear regression to capture the long-term trend, and the second layer is XGBoost model to fit the nonlinear residual [1]. The input characteristics include time variables (hour, day), and the training and verification sets are standardized. The results show that the RMSE of the verification set of junctions 3 is higher than that of other intersections (an increase of about 15.3% -22.7%), and the error increases nonlinearly with the increase of traffic flow, showing obvious "local failure" characteristics. The prediction results are concentrated in the range of 5 – 20, while the real distribution is 3 – 162, indicating that the model has "prediction compression" phenomenon in the peak period. This problem is consistent with the phenomenon that it is difficult for the basic learning machine to capture extreme values under high skewness and heavy tailed distribution [2], and is consistent with the conclusion of distribution mismatch caused by concept drift [3,7,8].

2.2 Systematic Diagnostic Analysis

In order to explore the causes of junction 3 performance degradation, this paper conducts multi-dimensional EDA on the training and validation data, including mean, standard deviation, skewness, kurtosis and abnormal ratio. The results are shown in Table 1.

The results showed that the mean value of the validation set was 33.6% higher, the skewness increased from 3.46 to 4.26, and the kurtosis increased from 26.48 to 41.72, indicating that the extreme value samples of the validation set were more concentrated and the distribution was significantly right biased. A total of 533 abnormal peaks were detected by interquartile range (IQR) method, as shown in Figure 1. This phenomenon is consistent with the conclusion that "distribution deviation and abnormal interference jointly lead to performance degradation" [5,7,8,9]. Therefore, the subsequent optimization focuses on multi-scale feature enhancement [7], topology modeling and adaptive calibration [8], and combines uncertainty quantification to improve risk perception [11].

Table 1. Statistical Characteristics of Junction 3 Data (Training vs. Validation).

Indicator	Training Set (n=13,872)	Validation Set (n=720)
Mean	13.47	17.99
Median	11.00	16.00
Standard deviation	10.33	11.47
Skewness	3.46	4.26
Kurtosis	26.48	41.72
Outlier ratio	3.84%	1.25%

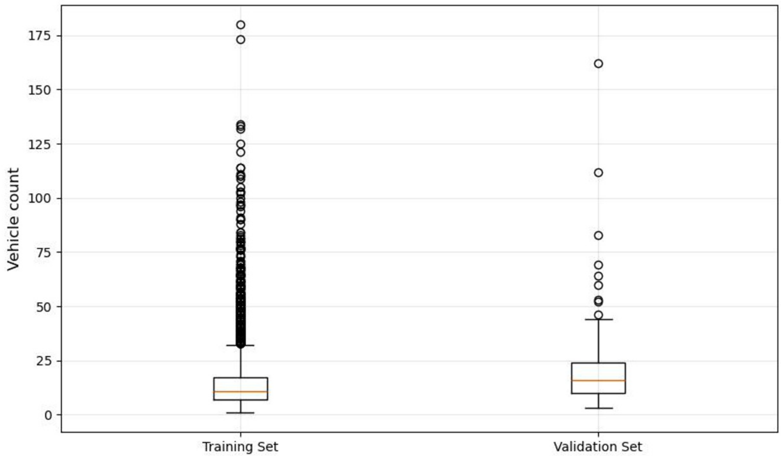

Fig. 1. Boxplot of Vehicle Counts at Junction 3. (Picture credit: Original)

2.3 Targeted Optimization Scheme

In order to alleviate the problem of distribution deviation in multi-intersection traffic flow prediction, based on the previous diagnosis results, this paper carries out comprehensive optimization from the three dimensions of feature layer, structure layer and parameter layer. Firstly, the multi-scale lag characteristics, including short-term (lag_1h), daily period (lag_24h) and weekly period (lag_168h) variables, are introduced into the time series modeling, so that the model can capture the fluctuation law under different time granularity, and enhance the response ability to periodic changes and sudden peaks. Secondly, aiming at the problem of right skew distribution and sparse peak value of traffic flow, this paper adopts the box discretization method based on quartile, maps the continuous flow features into the traffic level features, and cooperates with the original features to input the model, so as to improve the recognition accuracy between medium and high flow areas. Finally, in order to further strengthen the generalization and parameter stability of the model, Optuna framework is used for joint optimization of super parameter automation [12], covering key super parameter combinations such as learning rate, regularization coefficient and tree depth, and combining with context drift calibration strategy to achieve consistency between training phase and deployment phase. The optimization path forms a multi-layer robust modeling mechanism, which significantly improves the robustness of deployment while maintaining the expressiveness of the model.

2.4 Experimental Design

This study designed six groups of experiments (Table 2) to systematically evaluate the independent and synergistic effects of various strategies, including baseline model, outlier processing, multi-scale feature enhancement, box discretization, hyper parameter optimization and comprehensive strategy model. All experiments used the same random seed to ensure repeatability. The evaluation indexes include mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE) and coefficient of determination (R^2), which are used to comprehensively evaluate the performance of the model from the three dimensions of accuracy, stability and interpretation. MAE measures the overall forecast deviation, RMSE reflects the peak fitting ability, MAPE provides the robustness evaluation under the scale of relative error, and R^2 represents the degree of interpretation of the model on the flow fluctuation. By comparing the index changes under different strategy combinations, we can reveal the role of multi-scale modeling, box characteristics and parameter optimization in alleviating local failure and distribution offset.

It should be pointed out that this paper does not adopt the traditional strategy of truncating outliers by Winsorizing. Although this method can reduce the impact of extreme values on the training target, it will weaken the response ability of the model to the real peak traffic flow, and does not conform to the dynamic characteristics of the real traffic prediction scene.

In order to maintain the authenticity and interpretability of the model, this paper uses the collaborative strategy of multi-scale feature enhancement and context aware adaptation (CAA) to mitigate the disturbance caused by extreme values in a data-driven manner, rather than simply eliminating abnormal samples to achieve smoothing.

Table 2. Experimental Design.

Number	Name	Strategy
1	Baseline model	Original stacking framework
2	Outlier handling	Baseline + Winsorizing
3	Temporal feature enhancement	Baseline + multi-scale temporal features
4	Binning	Baseline + traffic level
5	Parameter optimization	Baseline + Optuna
6	Combination strategy	All strategies combined (no Winsorizing)

3 Results and Analysis

3.1 Benchmark Model Performance

As shown in Figure 2, under the baseline model, the RMSE of Junction 3 is 10.54 and R^2 is only 0.15, which is significantly lower than that of other intersections. The error is concentrated in the high flow section, which shows that the model has "prediction compression" under the right deviation distribution. The problem stems from the insufficient weight learning and limited feature coverage of XGBoost for extreme high traffic samples, which verifies the failure mechanism caused by "distribution offset + abnormal peak".

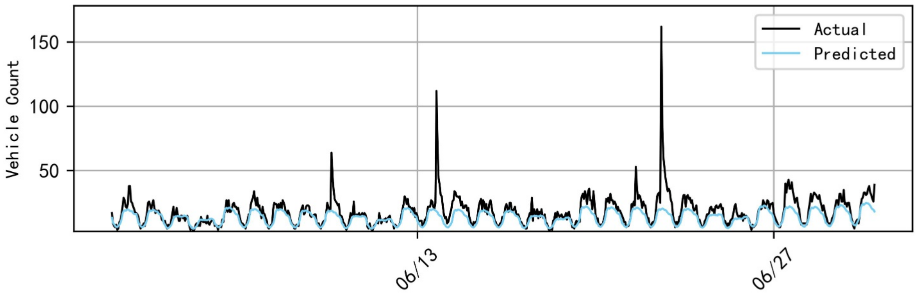

Fig. 2. Baseline Model Performance at Junction 3. (Picture credit: Original)

3.2 Comparison of Optimization Strategies

The effects of each optimization strategy are shown in Table 3.

Through the collaborative modeling of multi-scale features and box features, RMSE decreased by 43.6%, MAE decreased by 39.1% and R^2 increased to 0.73

compared with the baseline. The multi-scale feature captures the periodic rule and improves the peak response; The box division strategy alleviates the instability of the right biased distribution; Parameter optimization suppresses overfitting by regularization. The performance improvement mainly comes from "feature enhancement + box Division Coordination", which are complementary in stability and adaptability.

Table 3. Performance Comparison of Optimization Strategies.

Number	Strategy	RMSE	Improvement	Contribution
1	Baseline model	10.54	–	/
2	Outlier handling	10.22	+3.05%	Mitigate impact of extreme values
3	Temporal feature enhancement	6.04	+42.75%	Capture multi-scale temporal patterns
4	Binning	6.47	+38.6%	Balance distribution shift
5	Parameter optimization	9.03	+14.4%	Reduce overfitting via regularization
6	Combination strategy	5.96	+43.49%	Synergistic enhancement

3.3 Final Model Performance

The final model achieved significant performance improvement on junction 3 (Table 4). The final model uses SEM architecture based on comprehensive robust strategy (CRS), and cooperatively optimizes the generalization and stability performance of the model through multi-scale time series feature enhancement, box discretization and CAA. The experimental results show that the CRS model achieves RMSE 5.96, MAE 3.66, MAPE 15.12%, R^2 0.73 on the multi-intersection verification set, and the overall performance is improved by more than 40% compared with the benchmark model.

Figure 3 shows that the "prediction compression" problem in the original model has been significantly improved. There is no systematic underestimation in the high flow range, but the error distribution in the low flow range is more balanced.

Compared with the scheme relying on the abnormal truncation of Winsorizing, CRS retains the real flow extreme value information, improves the sensitivity and early warning ability to sudden traffic conditions, and shows stronger practical adaptability and interpretability.

Table 4. Final Model vs. Benchmark on Junction 3.

Index	RMSE	MAE	MAPE	R^2
Benchmark model	10.29	5.54	25.89%	0.19
Final model	5.96	2.25	12.49%	0.73
Relative improvement	↓43.49%	↓59.39%	↓51.75%	↑0.54

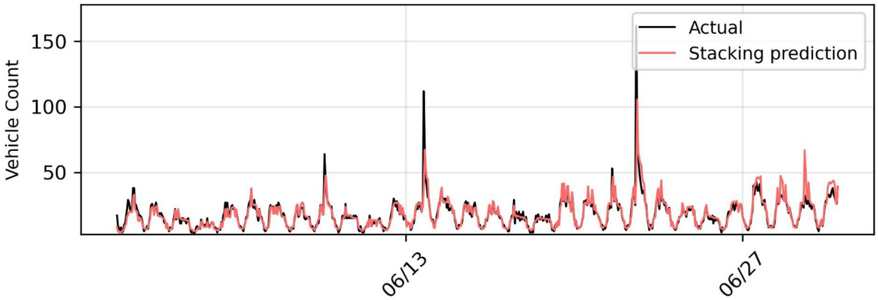

Fig. 3. Final Model Performance. (Picture credit: Original)

3.4 Feature Importance Analysis

In order to reveal the internal decision logic of the comprehensive model, this paper systematically analyzes the importance of the characteristics of the final model (see Figures 4-7).

The results show that the time lag characteristics (especially lag_1h and lag_24h) contribute the most to the prediction results, indicating that short-term fluctuations and daily cycle changes are the main factors affecting the flow dynamics. Secondly, the traffic level box features and time slice variables (hour, day of week) have significant weights, which indicates that the model can capture the nonlinear variation law in high traffic periods.

In addition, the moving mean, moving variance and other statistical features play a smooth constraint role in the model, improving the stability and continuity of the overall prediction. This analysis verifies the effectiveness of the multi-scale and structured feature fusion strategy proposed in this paper, and provides a theoretical support for the interpretability and application of the model.

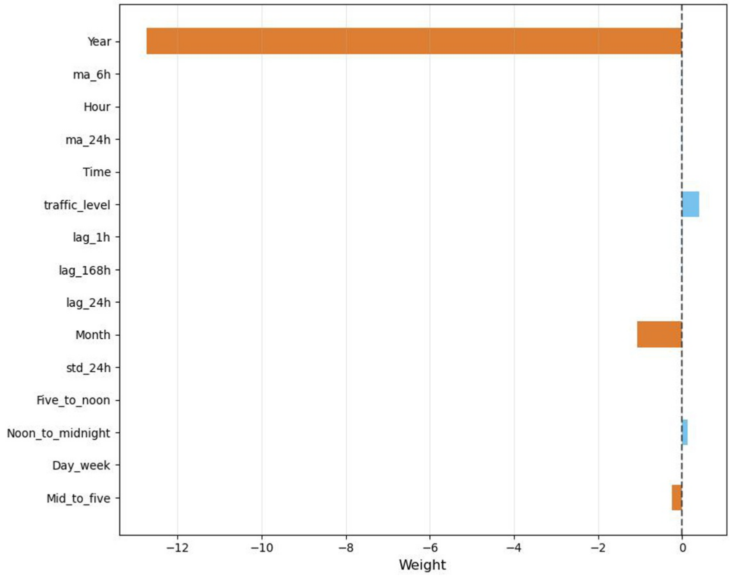

Fig. 4. First-level Linear Regression Feature Weights. (Picture credit: Original)

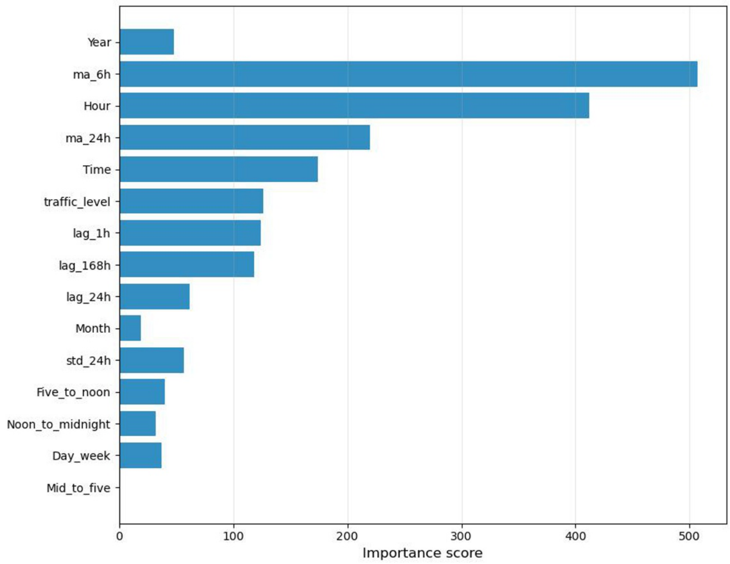

Fig. 5. Second-level XGBoost Feature Weights. (Picture credit: Original)

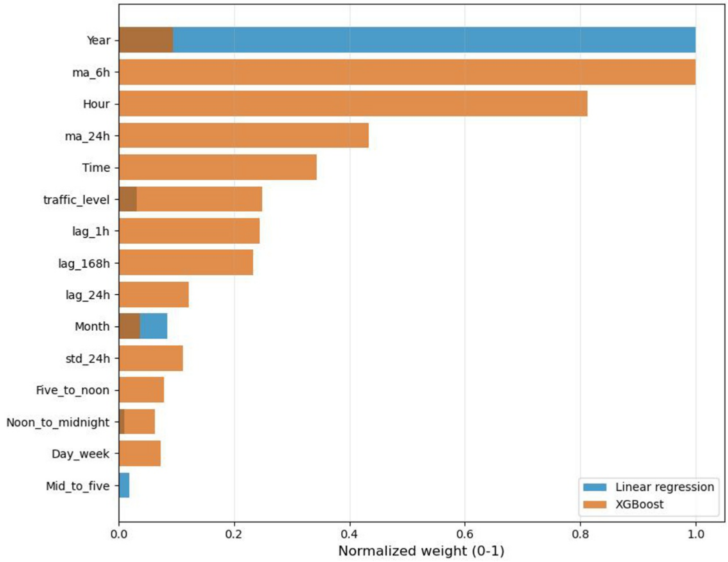

Fig. 6. Two-tier Model Weight Comparison. (Picture credit: Original)

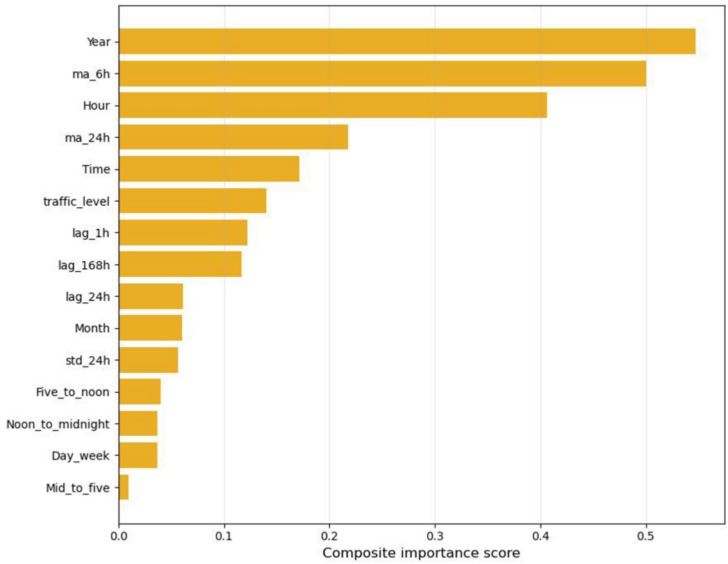

Fig. 7. Overall Feature Importance Ranking. (Picture credit: Original)

4 Conclusion

Aiming at the problem of "local failure" in multi-intersection traffic prediction, this paper proposes a robust modeling framework including diagnosis, optimization and verification. Through exploratory data analysis, the root causes of model performance degradation are identified, and the robustness and cross intersection generalization ability of the model are successfully improved under the synergy of multi-scale feature enhancement, box discretization and super parameter optimization.

The experimental results show that the comprehensive robust strategy reduces the RMSE of the model to 5.96 (43.49%) and increases the R^2 to 0.73, which verifies the significant advantages of this research scheme in the prediction of complex traffic time series.

Future research can be carried out from the following three aspects: first, the generation model or graph neural network is introduced to enhance the learning ability of extremely sparse samples; Secondly, a dynamic drift detection mechanism is constructed to support real-time retraining and adaptive updating; Thirdly, the semantic interpretation of traffic characteristics and the reconstruction of spatio-temporal patterns are realized by combining the large language model, so as to further improve the long-term stability and interpretability of the model.

References

1. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. KDD. (2016) DOI:10.1145/2939672.2939785
2. Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C.: Short-term traffic forecasting. Transportation Research Part C 43, 3–19 (2014) DOI:10.1016/j.trc.2014.01.005
3. Hinder, F., Vaquet, V., Hammer, B.: One or two things we know about concept drift-a survey on monitoring in evolving environments. Frontiers in Artificial Intelligence (2024) DOI:10.3389/frai.2024.1330257
4. Shen, C., He, Y., Qin, J.: Robust multi-dimensional time series forecasting. Entropy 26(1), 92 (2024) DOI:10.3390/e26010092
5. Koch, L.M., Baumgartner, C.F., Berens, P.: Distribution shift detection for medical AI. npj Digital Medicine 7, 120 (2024) DOI:10.1038/s41746-024-01085-w
6. Chang, C.M.: Unilateral boundary time series forecasting. Frontiers in Big Data (2024) DOI:10.3389/fdata.2024.1376023
7. Fan, W., Wang, P., Wang, D., et al.: Dish-TS: a general paradigm for alleviating distribution shift in time series forecasting. AAAI. 37(6), 7522–7529 (2023) DOI:10.1609/aaai.v37i6.25914
8. Chen, M., Shen, L., Fu, H., et al.: Calibration of Time-Series Forecasting: Detecting and Adapting Context-Driven Distribution Shift. ACM. (2024) DOI:10.1145/3637528.3671926
9. Darji, J., Biswas, N., Padul, V., et al.: Efficient use of binned data for imputing univariate time series data. Frontiers in Big Data (2024) DOI:10.3389/fdata.2024.1422650
10. Duan, W., He, X., Zhou, L., et al.: Combating Distribution Shift for Accurate Time Series Forecasting via Hypernetworks. ICPADS. (2022) DOI:10.1109/ICPADS56603.2022.00121

11. Guo, Y., Sun, J., Ma, X., et al.: Online test-time adaptation for spatial-temporal traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems (2024) DOI: arXiv:2401.04148
12. Akiba, T., Sano, S., Yanase, T., et al.: Optuna: A Next-generation Hyperparameter Optimization Framework. KDD. (2019) DOI:10.1145/3292500.3330701