Demand Prediction by Incorporating Internet-of-Things Data: A Case of Automobile Repair and Maintenance Service

Jieyi Zhang
Xiamen University
jyee_zhang@outlook.com

Cenying Yang
City University of Hong Kong
cenying.yang@cityu.edu.hk

Yihao Feng
University of Texas at Austin
yihao.ac@gmail.com

Abstract

While anecdotal evidence highlights the value of Internet-of-Things (IoT) data for business operations, rigorous empirical validation is still limited. The key challenge lies in integrating IoT analytics into business evaluation. To address the issues, we focus on the automotive industry and study the value of telematics data, an important IoT application in this domain, in terms of predicting maintenance, repair, and operations (MRO) service demands. Our approach involves building a prediction system with users' driving behavior, MRO service records, and environmental data (weather and traffic). We show a substantial improvement in prediction performance upon incorporating user behavior information derived from IoT data. Specifically, we find that hard acceleration, hard braking, and speeding rank the third, fifth, and sixth, respectively, in terms of their contribution to the MRO prediction. Our results shed light on the design of product-service systems (PSS), an emerging trend to integrate product offerings with service offerings.

Keywords: Internet of Things, demand prediction, product-service systems

1. Introduction

The rise of Internet-of-Things (IoT) has generated voluminous sensor data that can inform user behavior. Wearables reveal users' daily activities, like running, walking and sitting. Smart cars reveal users' driving behavior, like speed. Insights into user behavior can be further used to improve product and service design. Despite the hype over the big volume of sensor data, there is dearth of research on the value of such large-scale sensor data. The research vacuum is probably due to the unavailability of the sensor data and a lack of an integration of IoT analytics into business evaluation. Our work attempts to bridge the gap by examining the value of massive sensor data when applying predictive analytics.

Specifically, we place our investigation under the automotive context where automobile manufacturers have been collecting large-scale and fine-grained telematics data through various embedded sensors in vehicles (Ho et al., 2022, Choudhary et al., 2020). The telematics data provides driver performance data such as speed and acceleration as well as vehicles' location (see Figure 1 for an illustration of telematics data). Speed and acceleration can be further used to capture driving behavior, like hard braking, hard acceleration, and speeding. The assessment of these driving behaviors contributes to the measurement of driving risks and enables insurance companies to innovate risk assessment (Ho et al., 2022) and personalized insurance pricing, like usage-based insurance (UBI) (Choudhary et al., 2020, Soleymanian et al., 2019). The telematics data also demonstrates value in fleet management, as operators can efficiently manage their fleets by monitoring drivers' driving behaviors such as route selection and timing (Wolski, 2016).

However, this granular level of telematics data has not been utilized for predicting maintenance, repair, and operations (MRO) service requests, possibly due to a lack of collaboration between manufacturers and dealers.

As an essential part of after-sales services, MRO services are critical for the reliability and durability of vehicles and customer satisfaction (Williams, 2007). Most automotive manufacturers use franchised dealers to provide MRO services. Thus, dealers are at the front

URI: https://hdl.handle.net/10125/106986
978-0-9981331-7-1

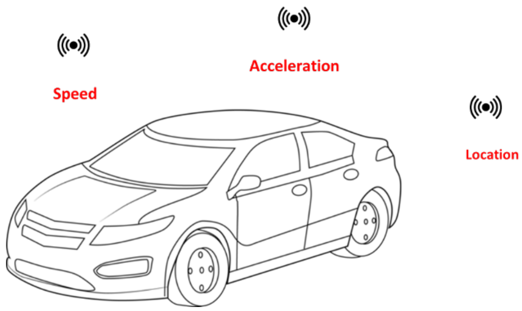

Figure 1. Automotive telematics data

line of customer contact and the quality of services provided by dealers may affect product sales, repeat sales, and brand reputation (Golara et al., 2021). As a response, manufacturers are increasingly extending traditional offerings to the provision of integrated product service bundles in recent years (Resta et al., 2017). It represents a gradual transformation from a "pure product" orientation towards an integrated "product-service system" (PSS) perspective (Gaiardelli et al., 2010). Although scholars have theoretically advocated advantages of integrating products and services, like enhanced customer satisfaction and improved competitive edge (Adam et al., 2017, S. Kim et al., 2019), the practical implementations of PSS are still in its nascent stage. In automotive industry, the management of supply chains for product-oriented PSS requires collaboration between manufacturers and dealers. However, the heterogeneity of dealer-originated information, concerns about protecting sensitive information (including issues of data ownership) hinder the achievement of closer collaboration between dealers and manufacturers (Infosys, 2018). Despite the challenges in fostering direct collaboration, we have devised a strategy to address this by manually integrating telematics data from manufacturers with MRO records from dealers, allowing us to predict the MRO demand. This collaboration is crucial for optimizing supply chain management and driving the implementation of PSS in the automotive industry.

Previous studies on MRO demand prediction have primarily relied on historical MRO records (Dangut et al., 2021, Kobayashi et al., 2017, Patil et al., 2017) and environmental factors such as weather and traffic (Chen et al., 2021). However, no work yet has incorporated the telematics data into demand prediction. Telematics data provides valuable insights into user driving behavior, such as hard braking, hard acceleration, and speeding, which have been shown to be good indicators of unsafe driving that may affect the frequency of vehicles' repair and maintenance (Choudhary et al., 2020, Soleymanian et al., 2019). Intuitively, observing that a driver frequently engages in hard braking provides substantial information about the wear out of the vehicles' brakes, and thus the subsequent MRO service requests. Thus, driving behavior could greatly affect a vehicle's lifecycle and thus contribute to the MRO demand at dealers.

Our work examines expanding the data in MRO prediction to IoT data and, specifically, to large scale fine-grained telematics data which informs users' driving behavior. We attempt to answer the following research question. *Does this large scale fine-grained telematics data add value to automobiles' MRO demand prediction?* We collaborate with a large automotive manufacturer and obtain the telematics data that informs user driving behavior, including hard braking, hard acceleration, and speeding. We further combine this telematics data with (1) historical MRO records obtained from our collaborative manufacturer's authorized dealers, and (2) environmental data including weather and traffic conditions. To pre-process the data, we slice the time-series data for each vehicle by a time window of 4-week. By doing so, we transform the problem of time series prediction into a binary classification problem. Note that our final dataset is highly imbalanced, as the number of negative samples (i.e., no MRO request) is approximately 25 times higher than the number of positive samples (i.e., MRO request). We then employ light gradient-boosting machine (LightGBM) for the classification problem and address the imbalance issue by adjusting the weights of positive samples during training.

The findings reveal that the inclusion of large-scale fine-grained telematics behavioral data leads to a discernible improvement in MRO prediction. Specifically, the precision, recall, and f1-score improve by 1.86%, 8.36%, and 2.37% respectively. This improvement confirms that driving behaviors, including hard braking, hard acceleration, and speeding, offer valuable insights into the vehicle's condition and potential damage, and thus the MRO demand prediction at dealers. This conclusion is further supported by the results of feature importance. We find that hard acceleration, hard braking, and speeding rank the third, fifth, and sixth, respectively, in terms of their contribution to the MRO prediction. To validate the robustness of the outcomes, we adjust the time window to 1 week and 8 weeks and results on prediction improvement remain.

This study makes significant contributions in several ways. We first add to the rapidly growing literature on the value of the large-scale fine-grained data (Fu and Fisher, 2023, Zhang and Moe, 2021, Adomavicius and Tuzhilin, 2005, Martens et al., 2016) by demonstrating the predictive capability of IoT-based telematics data in demand forecasting. By doing so, we also contribute to the literature on the value of telematics data in automotive industry. Previous works have revealed the value of IoT-based telematics technology in assessing driving risks in the context of Usage-based Insurance (UBI) (Choudhary et al., 2020, Soleymanian et al., 2019) and fleet management (Wolski, 2016). However, our focus shifts to examining the issue of MRO prediction from the perspective of manufacturer-dealer relationship. We establish the predictive prowess of behavioral data obtained from IoT-based telematics data in predicting MRO service requests. This provides concrete approaches for manufacturers and dealers to engage in delivering "product service" bundles, which have been theoretically advocated by scholars (Williams, 2007). Thus, we also contribute to the literature on PSS by taking a step forward to facilitate the practical implementation of PSS from the conceptual level to the operational level.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 describes the data and empirical models and methodology. We present results in Section 4, and conclude the paper in the final section.

2. Literature review

Our work integrates three streams of literature: (1) prediction based on large-scale fine-grained data, (2) value of telematics data, and (3) product-service systems.

2.1. Prediction based on large-scale fine-grained data

In recent years, considerable attention has been directed towards predictive analysis utilizing extensive fine-grained data that informs user behavior. This attention stems from the availability of granular data ranging from social media interactions (Fu and Fisher, 2023, Zhang and Moe, 2021), online purchase transactions (Adomavicius and Tuzhilin, 2005), to payment records (Martens et al., 2016). These works primarily focus on applications in the marketing domain. By integrating these large-scale fine-grained user behavior data with structured demographic information, researchers have examined consumer perceptions of brands (Zhang and Moe, 2021), predicted short-term market trend changes (Fu and Fisher, 2023), and determined the likelihood of consumer purchases for specific products or services (Martens et al., 2016). Previous studies have empirically confirmed the predictive effectiveness and scalability of such large-scale fine-grained data analysis approaches.

The rise of the IoT opens up greater opportunities to gather detailed user behavior information through various sensors. Compared to non-IoT data, IoT technologies offer greater diversity and real-time capabilities in capturing user behavior information. Non-IoT data sources are often limited to specific platforms or service providers, constraining data coverage, dimensions, and granularity. In contrast, IoT data can be collected from a wider range of physical devices and sensors, providing multi-dimensional information (Brous et al., 2020). Furthermore, while non-IoT data relies on users' active behavior and may introduce delays, IoT data can be collected and transmitted in real-time through sensors, offering more timely insights into user behavior without manual intervention (Boos et al., 2013). Thus, the large-scale, fine-grained data generated by IoT devices holds tremendous potential for predictive analysis. Combining it with non-IoT data enhances value compared to using non-IoT data alone.

For instance, wireless technologies, such as the Global Positioning System (GPS), radio-frequency identification (RFID) chips, and contactless smart cards, give rise to large-scale human-sensing data that captures spatiotemporal dynamics of movement behavior (Wang et al., 2022). Based on such data, the prediction of urban mobility can contribute to the development of efficient transportation management systems (Wang et al., 2022).

Telematics data, an important IoT application in automotive industry, has been growing rapidly in recent years. Based on telecommunication components, vehicular sensors, wireless networking, and data dashboards, telematics technology enable the long-distance transmission of data from moving transportation devices (Writer, 2022). Thus, telematics data can offer insights into users' driving behavior and vehicles' motion trajectories (Longhi and Nanni, 2020, Ho et al., 2022). The availability of such granular data has the potential to disrupt the automotive industry. This study aims to enrich research on the predictive capabilities of large-scale, fine-grained data by focusing on telematics data in automotive industry.

2.2. Value of telematics data

The telematics data typically measures vehicle's physical features such as speed and acceleration, from which we could infer user driving behavior. Additionally, it captures the motion trajectories of vehicles equipped with GPS technology. This large-scale fine-grained information about vehicles and users is crucial for optimizing services in mobility-related companies (Longhi and Nanni, 2020). By monitoring driving behaviors such as route selection and timing, fleet operators can manage their fleets more efficiently (Wolski, 2016). Using comprehensive measurements of driver behaviors and trajectory characteristics, insurance companies can evaluate risks at both the trip and driver levels(Ho et al., 2022)and promote the adoption of UBI (Soleymanian et al., 2019). UBI brings a host of benefits: insurers gain product distinction and cost savings, consumers enjoy premium control and behavior-based perks, and society sees improved road safety (Soleymanian et al., 2019).

The availability of data on driver behavior and driving trajectories enabled by IoT-based telematics technology could benefit not only insurance companies and fleet management, but also MRO services at dealers. Wolski (2016) highlights that telematics technology can steer driver behavior towards greater safety, thereby reducing fleet maintenance costs such as windshield replacement. However, research in this field remains limited, particularly regarding the use of telematics data for predicting MRO service requests at dealers. Effective dealership management, as a crucial link between manufacturers and end customers, enables manufacturers to provide high-quality after-sales service to customers and contributes to the overall success of their business (Resta et al., 2017). Thus, this study aims to explore the value of telematics data on after-sales service. Specifically, we investigate whether incorporating driving behavior that is obtained from telematics data at the manufacturer side could improve MRO prediction at dealers. Our results could drive the delivery of a “product service” bundle rather than a single product.

2.3. Product-service systems

PSS is defined as combining marketable products and services to satisfy customer needs (van Halen MSc and te Riele MSc, 1999). It can be categorized into three types: product-oriented, use-oriented, and result-oriented according to Tukker (2004). Through blurring the traditional boundary between products and services (Mont, 2002), PSS strives to provide customers with holistic and customized solutions, facilitating the cultivation of robust customer relationships (S. Kim et al., 2019). Concurrently, PSS endeavors to optimize resource allocation and enhance production efficiency (Brehm and Klein, 2017), thereby engendering sustainable development within organizations (Williams, 2007). Constructing a PSS requires an in-depth analysis and understanding of customer needs, encompassing actual, latent, and psychological demands, as well as expectations regarding product and service quality, pricing, and convenience (K.-J. Kim et al., 2012). In the automotive industry, customers seek more than just vehicle ownership through traditional sales; they desire comprehensive post-sales services, such as maintenance contracts or extended warranties (Williams, 2007), to ensure long-term functionality. Specifically, MRO services aid customers in addressing potential faults, maintaining vehicle performance, and extending its lifespan. This product-oriented PSS could effectively satisfy customer requirements throughout the entire automotive life-cycle and improve customer satisfaction.

The construction of a PSS often requires collaboration with partners who possess the necessary capabilities and resources, particularly in areas where the company may be lacking (K.-J. Kim et al., 2012). For instance, in the automotive industry, manufacturers may partner with dealers who provide after-sales services. By integrating telematics data from manufacturers with MRO records from dealers, accurate forecasting of MRO demand can be achieved. This enables dealers to effectively plan and manage their inventory, ensuring timely availability of repair and maintenance parts (Chen et al., 2021). Moreover, within the vertical cooperation relationships in a product-oriented PSS, manufacturers can adjust production plans and logistics arrangements based on the predicted results. This could further facilitate efficient supply chain management. Thus, our work attempts to bridge the gap in previous research that has predominantly remained at the conceptual stage regarding PSS supply chain design and management (Resta et al., 2017) and promote the transition and development of PSS into practical implementation.

3. Research data and methods

3.1. Data description

Previous studies largely depend on traditional MRO data, like historical MRO records, to predict the demand of MRO at dealers (Dangut et al., 2021, Kobayashi et al., 2017, Patil et al., 2017). More recently, Chen et al. (2021) integrate environmental data with MRO data to build a prediction model. They validate that weather and traffic are key factors in influencing vehicle lifecycle, and thus could add predictive power to service requests at dealers. The major contribution of this project is to further incorporate driving behavior data which are obtained from telematics data into the prediction model. See Figure 2 for an illustration of the data we use in our model. Below we will describe (1) driving behavior data; (2) MRO data; and (3) environmental data in details.

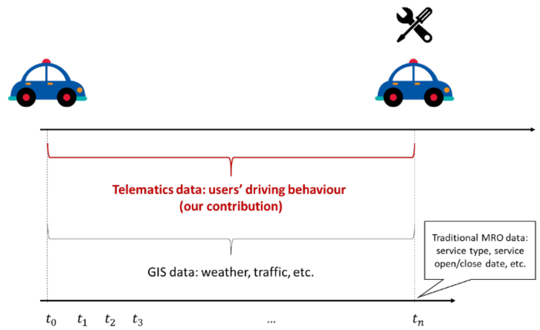

Figure 2. Illustration of the data we use in model prediction

Driving behavior data. We collaborate with a large automotive manufacturer and use its IoT based remote monitoring service to collect telematics data which is collected every 10 seconds for *unidentified* vehicles. The manufacturer started remote monitoring service in 2018. The telematics data reveals (1) driving behavior measured in terms of car performance, like speed and acceleration, and (2) vehicle trips, like when (time) and where (location) vehicles are running. Due to the sheer volume of the telemetry data, we first propose a sampling approach to select representative vehicles for analysis. Then, we extract data on aggregate driving behavior. We detail the major steps as below.

Step 1: Car Sampling. We are interested in examining whether vehicle use data can help improve predicting MRO service request. Since we infer vehicle use data from automotive telematics data, we focus on cars whose drivers subscribe to the manufacturer's telematics service between 2018 and 2019. Due to the sheer amount of the telematics data, we randomly select 10% of the total vehicles. Next, we extract aggregate driving behavior for those sampled vehicles.

Step 2: Aggregate driving behavior data. Hard braking, hard acceleration, and speeding have been shown to indicate unsafe driving behavior (Choudhary et al., 2020, Simons-Morton et al., 2009, Soleymanian et al., 2019). Hard braking occurs when a driver uses more force than necessary to stop a vehicle; hard acceleration occurs when a driver uses more power than necessary to pull off from a dead stop [1].

In this study, we use hard braking, hard acceleration, and speeding as proxies for risky driving behavior. For each sampled car from 2018 to 2019, we aggregate daily driving behavior in terms of the total number of hard braking events, total number of hard acceleration events, and the total number of speeding events. A hard-braking event is identified if the deceleration is *6.5 MPH/S*. A hard acceleration event is identified if the acceleration is *8.5 MPH/S*. A speeding event is identified if the speed of the car is greater than the sum of the average speed and one standard deviation within an area based on 3 decimals of latitude and longitude.

Besides, we extract the daily mileage for each vehicle by calculating the difference between the mileage record at the end of the day and the mileage record at the start of the day.

MRO data. We extract MRO records for sampled vehicles between 2018 and 2019 from our collaborative manufacturer's authorized dealers. The MRO data includes the open / close date of the service, the service type, the cost paid by the customer, and the basic information of the vehicle. On average, each vehicle conducts 1.25 times MRO service during our observation time. We consider a dummy variable for MRO, which is equal to 1 is a service request is recorded on a day for a vehicle. This is our dependent variable we would like to predict.

We integrate the MRO data with the driving behavior data based on the day and location. Vehicles that do not have any records in the MRO data are removed.

Environmental data. Finally, we extract environmental data on weather and traffic. Given a certain location (combination of one decimal of latitude and longitude) and a date, we obtain (1) average daily temperature from *OpenWeather*[2] and (2) traffic density from *National Neighborhood Data Archive*[3]. The traffic density refers to the average volume of traffic passing through a ZIP Code Tabulation Area (ZCTA) in a given period of time (Finlay et al., 2021).

Since the traffic density data includes ZCTA and our data includes latitude and longitude. We first obtain ZCTA for our data based on the latitude and longitude, then merge our data with the traffic density data. As the platform only provides data on an annual basis, we obtained and matched the average, maximum, and minimum traffic density values for each location-year combination. Subsequently, employing a normal distribution, we generated daily traffic density data for each geographic coordinate.

Additionally, we also obtain data on car attributes and driver attributes from our collaborative automotive manufacturer. Car attributes contains vehicle model, vehicle trim, purchase time, and engine size. And driver attributes contains gender, age, and income level. We consider a dummy variable for gender, where a value

[1] Mix Telematics: How Harsh Braking and Acceleration Impacts Your Fleet

[2] OpenWeatherMap

[3] National Neighborhood Data Archive (NaNDA)

of 1 represents male. The final dataset combines user driving data, MRO data, environmental data, as well as car attributes and driver attributes. Table 1 summarizes statistics of major variables at vehicle-day level.

Table 1. Statistics summary of sampled cars (16,287,690 observations)

	Mean	Min	Max
MRO	0.01	0.00	1.00
Hard braking	4.98	0.00	26.00
Hard acceleration	1.06	0.00	7.00
Speeding	3.04	0.00	18.50
Mileage	53.73	0.02	202.00
Engnine size	4.66	1.40	6.60
Driver age	51.87	16.00	99.00
Driver gender	0.76	0.00	1.00
Temperature	15.30	-40.00	43.10
Traffic Density	19383.23	1.00	359348.00

3.2. Models and methods

Data preprocess. We first aggregate the original vehicle-day level data into vehicle-week level data. We do so for several reasons. First, our dependent variable is highly imbalanced, with an average MRO request ratio of 100:1 compared to non-request. Aggregating the data to the weekly level helps mitigate this imbalance (approximately 25:1), leading to more reliable and accurate prediction outcomes. Second, despite the aggregation, the weekly-level data retains a relatively fine-grained granularity, providing valuable temporal and periodic information compared to longer time spans, such as monthly or quarterly aggregation. Finally, predicting MRO demand for the upcoming week holds greater significance for PSS, enabling longer-term decision-making in supply chain aspects, including logistics arrangements, inventory management, and distribution planning. This augmentation improves the operational efficiency of PSS by ensuring the timely availability of necessary maintenance and repair components. Regarding the specific aggregation process, temporal features, such as driving behavior data and MRO indicator, are aggregated through summation, while weather and traffic density are averaged. Constant features, such as vehicle attributes and driver attributes, remain unchanged.

Then, we divide the variables of hard braking, hard acceleration, and speeding by mileage to eliminate the correlation among variables and improve the model accuracy. Intuitively, as a vehicle travels a greater distance, it is expected to exhibit more instances of bad driving behavior. After this transformation, the driving behavior information derived from telematics data is summarized into the following three variables: $Hard\ braking\ per\ mile$, $Hard\ acceleration\ per\ mile$, and $Speeding\ per\ mile$.

Currently, each vehicle i has a weekly-level time series. Every timestamp (week t) in the records includes information about the vehicle's MRO demand, users' driving behavior information, environmental factors, and other constant features such as vehicle attributes and driver attributes for that particular week. Denoting temporal features (i.e. driving behavior and environmental factors) as s_{it} and constant features as h_i, each entry can be expressed by:

$$MRO_{it} = \{s_{it}, h_i\}$$

Since training and predicting on each individual vehicle's time series would be time-consuming due to the large number of vehicles, we transform this time series problem for multiple vehicles into a binary classification problem by reshaping the dataset. Specifically, we slice the temporal part by time window N $(N < t)$, which can be expressed as:

$$s_{it}^{'} = \{s_{i(t-N)}, s_{i(t-N+1)}\ , ..., s_{i(t-1)}\}$$

Then, we combine these reshaped temporal features with constant features. In this case, $MRO_{it} = \{s_{it}^{'}, h_i\}$

After this reshaping process, we enrich the information of each entry from containing the current week to containing the past N weeks. In other words, we utilize information from the past N weeks to predict the MRO demand of the current week. The whole data reshaping process is shown in Figure 3.

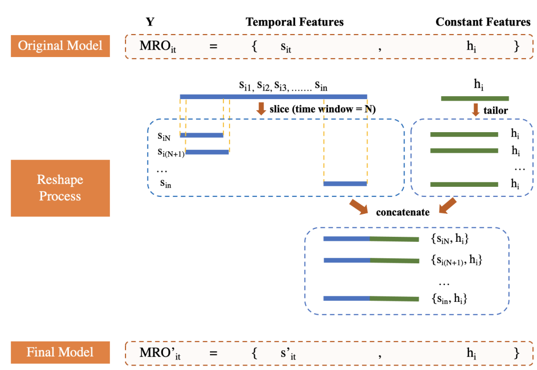

Figure 3. Process of data integration

In our context, we set time window N as 4-week. The final expression for vehicle i is:

$$MRO_{it} = \{s_{i(t-4)}, s_{i(t-3)}, s_{i(t-2)}, s_{i(t-1)}, h_i\}$$

Model Establishment. Our goal is to expand the data in MRO prediction to IoT-based telematics data, and further confirms the value of such large scale fine-grained behavioral data within the automobile context. Thus, in the base model, we use car attributes, driver attributes, and environmental data to predict MRO service as in previous studies (Dangut et al., 2021, Kobayashi et al., 2017, Patil et al., 2017, Chen et al., 2021). We add driving behavior as predictors in our model. The comparison of features used in the base model and our model is shown in Table 2. $Hard\ braking\ per\ mile_{it}$, $Hard\ acceleration\ per\ mile_{it}$, and $Speeding\ per\ mile_{it}$ indicate the times of hard braking, hard acceleration, and speeding every mile of car *i* on week *t*, respectively.

CAR_{ic} represents vehicle attributes *c* for car *i*, including model, trim, purchase time, and engine size. $DRIVER_{id}$ represents driver attributes *d* for car *i*, including gender, age, and income level. Car attributes and driver attributes do not change over time. $ENVIRONENT_{ite}$ represents environmental attributes *e*, including temperature and traffic density, in the area (i.e., in terms of one decimal of latitude and longitude) where car *i* is running on week*t*.

Table 2. Feature comparison between base model and our model.

Model	Features
Our model	$\sum_{t-N}^{t-1}$Hard braking per mile$_{it}$ $\sum_{t-N}^{t-1}$Hard acceleration per mile$_{it}$ $\sum_{t-N}^{t-1}$Speeding per mile$_{it}$ $\sum_{c}$CAR$_{ic}$ $\sum_{d}$DRIVER$_{id}$ $\sum_{e}$ENVIRONMENT$_{ie}$
Base model	$\sum_{c}$CAR$_{ic}$ $\sum_{d}$DRIVER$_{id}$ $\sum_{e}$ENVIRONMENT$_{ie}$

We employ LightGBM, a fast and high-performance gradient boosting framework based on decision tree algorithms, to predict the MRO service request. It offers advantages such as faster training speed, lower memory consumption, and support for parallel learning (Li et al., 2022). With its improved efficiency and higher model accuracy, LightGBM is an ideal choice for predicting repair and maintenance requests at the vehicle-week level. Despite aggregating daily-level data to the weekly level, the data imbalance issue persists, with a substantial disparity between the number of negative (i.e., MRO = 0) and positive (i.e., MRO = 1) samples. Specifically, the number of negative samples is approximately 25 times higher than the number of positive samples.

To address the issue, we adjust the weights of the samples during training. Specifically, we increase the weights of the positive samples (minority class) to enhance their influence on the loss calculation. In our experimental context, positive samples represent occurrences of MRO events. Thus, this weighting strategy enables the model to focus more on accurately classifying the minority class, thereby better capturing MRO demand.

We randomly split the vehicles into 80% of training data, and 20% of testing data. Best hyper parameters are determined using five fold cross validation on the training data. After the training, we apply the model to the testing data to obtain results.

4. Results

4.1. Experimental results

As time window is set to 4 weeks when reshaping the temporal features, we predict the MRO demand for the current week using the information during the last 4 weeks (i.e., last month). After training the model on the training set and determining the optimal model parameters through cross-validation, we evaluated the model on the test set. We use precision, recall, and f1-score as evaluation metrics because our binary dependent variable is highly imbalanced.

To showcase the value of IoT data in MRO prediction, we compared model with and without the incorporation of user driving behavior information inferred from telematics data. The results are shown in the Table 3.

Table 3. Results of MRO prediction.

	Precision	Recall	F1-score
Our model	**0.049**	**0.566**	**0.091**
Base model	0.048	0.522	0.089

Our results reveal that the incorporation of user driving behavior information has yielded notable improvements in the precision, recall, and f1-score of the classification model. Specifically, precision, recall, and f1-score improve by 1.86%, 8.36%, and 2.37% respectively. It is crucial to highlight that our main focus is to accurately capture the entire MRO demand for these vehicles. Therefore, we place significant importance on the recall score of our model, which has shown substantial improvement. Thus, our results corroborate the effectiveness of integrating user driving behavior information in predicting MRO demand within the classification framework.

Additionally, user behavior are top features in predicting MRO demand. We calculate the importance of each feature based on the number of times the feature is used in the model. This means that each time a feature is chosen as the splitting criterion in the tree, its importance score increases. For temporal features, we aggregate the importance scores of multiple features generated by slicing. For example, we sum up the importance scores of $Temperature_1$, $Temperature_2$, $Temperature_3$, and $Temperature_4$ to obtain the overall importance score of the $Temperature$ variable in the model. According to Figure 4, all three driving behavior features are among the top six features out of the total twelve features, with $Hard\ acceleration\ per\ mile$, $Hard\ braking\ per\ mile$, $Speeding\ per\ mile$ ranking the third, fifth, and sixth. This emphasizes the significance of including driving behavior features in our predictive models.

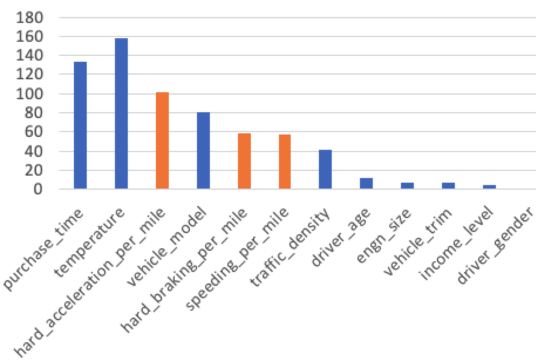

Figure 4. Feature importance.

4.2. Robustness checks

To ensure the robustness and generalizability of our findings, we conducted additional experiments to examine the impact of varying time windows on the performance of our MRO prediction model. Specifically, we explored two different time window settings: a shorter duration of 1 week and a longer duration of 8 weeks. By doing so, we sought to gain insights into the temporal dynamics and stability of the model's performance.

Our results consistently demonstrate the efficacy of integrating user behavior and vehicle usage information in MRO prediction (see Table 4). The precision, recall, and f1-score increase by 1.09%, 9.55%, and 1.83% respectively when using the past week's data to predict the next week's MRO demand. The precision, recall, and f1-score increase by 1.07%, 2.81%, and 1.21% respectively when using the past 8 weeks' data to predict the next 8 weeks' MRO demand.

Thus, we demonstrate that the inclusion of user behavior information derived from telematics data yields significant improvements in predicting MRO demand. These findings reinforce the notion that

Table 4. Results of MRO prediction with various time windows (incremental gains).

Time window	δPrecision	δRecall	δF1-score
1 week	1.09%	9.55%	1.83%
8 weeks	1.07%	2.81%	1.21%

incorporating large-scale fine-grained IoT data can provide valuable insights into the MRO demand in automotive industry.

5. Conclusions and discussions

Recent years have witnessed the tremendous growth of IoT data, which has opened up new opportunities for business operations[4]. In automotive industry, telematics data is an important IoT application and has grabbed attention of different players in this field. Previous works have studied the value of telematics data in innovative insurance design (Choudhary et al., 2020, Soleymanian et al., 2019) and fleet management (Wolski, 2016). However, little work has focused on the dealership management, which is critical to the success of high-quality after-sales service in automotive industry (Resta et al., 2017).

Our work attempts to bridge the gap. We integrate telematics data from the manufacturer into MRO service prediction at dealers. The telematics data informs important information of user driving behavior. Our empirical findings unequivocally highlight the significant improvement achieved by incorporating user driving behavior information facilitated by IoT technology. This improvement highlights the value of large-scale fine-grained IoT data in demand prediction.

Additionally, the implications of our findings extend beyond the domain of demand prediction and have profound implications for the broader field of PSS provision. With PSS increasingly recognized for its potential to revolutionize customer satisfaction and loyalty, dealers are confronted with new challenges in managing their supply chains effectively (Resta et al., 2017). The design and optimization of these supply chains, which lie at the heart of PSS implementation, are currently in their nascent conceptual stages. By leveraging our research insights to predict MRO demand at the dealer level, we empower dealers to proactively design efficient and rational supply chains that align with the demands of PSS. This practical guidance facilitates the transition of PSS from a conceptual idea to a tangible reality, enabling dealers to deliver superior services and stay ahead in a competitive market

[4]McKinsey Report: The Internet of Things: Catching up to an accelerating opportunity

landscape.

Looking ahead, our research holds tremendous potential for expansion and application beyond the realm of MRO demand prediction. The methodology and insights gained from integrating IoT data can be leveraged in diverse domains to unlock new opportunities for decision-making and innovation. Industries such as healthcare, sports, and beyond can benefit from harnessing the power of large-scale fine-grained IoT data to gain actionable insights, optimize operations, and enhance overall performance. By venturing into these uncharted territories, we aim to further contribute to the advancement of IoT analytics and foster fruitful collaborations with industry practitioners and researchers across various domains.

References

Adam, M., Strähle, J., & Freise, M. (2017). The interaction of product-service systems (pss) and corporate environmental management (cem): Can pss drive today's fashion industry toward more environmental sustainability? *Service Science*, *9*(3), 235–249.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, *17*(6), 734–749.

Boos, D., Guenter, H., Grote, G., & Kinder, K. (2013). Controllable accountabilities: The internet of things and its challenges for organisations. *Behaviour & Information Technology*, *32*(5), 449–467.

Brehm, L., & Klein, B. (2017). Applying the research on product-service systems to smart and connected products. *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers 19*, 311–319.

Brous, P., Janssen, M., & Herder, P. (2020). The dual effects of the internet of things (iot): A systematic review of the benefits and risks of iot adoption by organizations. *International Journal of Information Management*, *51*, 101952.

Chen, C., Liu, Y., Sun, X., Di Cairano-Gilfedder, C., & Titmus, S. (2021). An integrated deep learning-based approach for automobile maintenance prediction with gis data. *Reliability Engineering & System Safety*, *216*, 107919.

Choudhary, V., Shunko, M., & Netessine, S. (2020). Does immediate feedback make you not try as hard? a study on automotive telematics. *Manufacturing Service Operations Management*, *23*.

Dangut, M. D., Skaf, Z., & Jennions, I. K. (2021). An integrated machine learning model for aircraft components rare failure prognostics with log-based dataset. *ISA transactions*, *113*, 127–139.

Finlay, J., Melendez, R., Esposito, M., Khan, A., Li, M., Gomez-Lopez, I., Clarke, P., & Chenoweth, M. (2021). National neighborhood data archive (nanda): Traffic volume by zip code tabulation area, united states, 1963-2019.

Fu, Y., & Fisher, M. (2023). The value of social media data in fashion forecasting. *Manufacturing & Service Operations Management*, *25*(3), 1136–1154.

Gaiardelli, P., Resta, B., Songini, L., Pezzotta, G., et al. (2010). Aligning the servitization level of a company with its organizational configuration. In *Proceedings of 2011 apms "value networks: Innovation technologies and management"*. University of Stavanger.

Golara, S., Dooley, K. J., & Mousavi, N. (2021). Are dealers still relevant? how dealer service quality impacts manufacturer success. *Production and Operations Management*, *30*(10), 3560–3578.

Ho, Y.-J., Liu, S., Pu, J., & Zhang, D. (2022). Is it all about you or your driving? designing iot-enabled risk assessments. *Production and Operations Management*, *31*(11), 4205–4222.

Infosys. (2018). Automotive dealer- manufacturer data collaboration. https : / / www . infosys . com / industries / automotive / white - papers / Documents / dealer - manufacturer - data - collaboration.pdf

Kim, K.-J., Lim, C.-H., Lee, D.-H., Lee, J., Hong, Y.-S., & Park, K. (2012). A concept generation support system for product-service system development. *Service Science*, *4*(4), 349–364.

Kim, S., Christiaans, H., & Baek, J. S. (2019). Smart homes as product-service systems: Two focal areas for developing competitive smart home appliances. *Service Science*, *11*(4), 292–310.

Kobayashi, S., Fukuda, K., & Esaki, H. (2017). Mining causes of network events in log data with causal inference. *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 45–53.

Li, Y., Li, Y., & Liu, Y. (2022). Research on used car price prediction based on random forest and lightgbm. *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, 539–543.

Longhi, L., & Nanni, M. (2020). Car telematics big data analytics for insurance and innovative mobility services. *Journal of Ambient Intelligence and Humanized Computing*, *11*(10), 3989–3999.

Martens, D., Provost, F., Clark, J., & de Fortuny, E. J. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS quarterly*, *40*(4), 869–888.

Mont, O. K. (2002). Clarifying the concept of product–service system. *Journal of cleaner production*, *10*(3), 237–245.

Patil, R. B., Patil, M. A., Ravi, V., & Naik, S. (2017). Predictive modeling for corrective maintenance of imaging devices from machine logs. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1676–1679.

Resta, B., Gaiardelli, P., Cavalieri, S., & Dotti, S. (2017). Enhancing the design and management of the product-service system supply chain: An application to the automotive sector. *Service Science*, *9*(4), 302–314.

Simons-Morton, B. G., Ouimet, M. C., Wang, J., Klauer, S. G., Lee, S. E., & Dingus, T. A. (2009). Hard braking events among novice teenage drivers by passenger characteristics. *Proceedings of the... International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, *2009*, 236.

Soleymanian, M., Weinberg, C. B., & Zhu, T. (2019). Sensor data and behavioral tracking: Does usage-based auto insurance benefit drivers? *Marketing Science*, *38*(1), 21–43.

Tukker, A. (2004). Eight types of product–service system: Eight ways to sustainability? experiences from suspronet. *Business strategy and the environment*, *13*(4), 246–260.

van Halen MSc, C. J., & te Riele MSc, H. R. (1999). Product service systems, ecological and economic basics.

Wang, Y., Currim, F., & Ram, S. (2022). Deep learning of spatiotemporal patterns for urban mobility prediction using big data. *Information Systems Research*, *33*(2), 579–598.

Williams, A. (2007). Product service systems in the automobile industry: Contribution to system innovation? *Journal of cleaner Production*, *15*(11-12), 1093–1103.

Wolski, C. (2016). Shaping driver behavior with telematics. https://www.automotive-fleet.com/156870/shaping-driver-behavior-with-telematics

Writer, C. B. T. (2022). Telematics meaning, working, types, benefits applications. https://www.spiceworks.com/tech/iot/articles/what-is-telematics/

Zhang, K., & Moe, W. (2021). Measuring brand favorability using large-scale social media data. *Information Systems Research*, *32*(4), 1128–1139.