Vehicle User Identification Model based on Telecom Big Data

Zifan Chen [1], Qingqing Zhang[2]
[1] School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, Beijing, China
[2] Research Institute, China United Network Communications Corporation, Beijing, China
chenzifan@bupt.edu.cn zhangqq102@chinaunicom.cn

***Abstract*—China's automobile market is developing rapidly, high-quality development in the automobile market puts forward high requirements for predicting whether residents have a vehicle or not. This paper designs a user behavior indicator system based on operator big data and introduces the Transductive Support Vector Machine (TSVM) algorithm to predict user status. Under the verification of real sample data, this indicator system and model can be used to predict whether residents are vehicle users with outstanding performance and provides a new reference for the promotion of value-added services of auto service providers and operators.**

Keywords—operator, big data, TSVM, automobile, precision marketing

I. Introduction

Due to the economic development and demographic dividend, China's vehicle ownership is in a leading position in the world, which is expected to reach 1.7 trillion in 2025. However, the automobile service market is relatively backward relative to the prosperity of the sales market. It has only started to develop in the past 10 years, there is still huge room for imagination in the Chinese automotive aftermarket. Facing the vast market space of China's auto aftermarket, both auto brands and auto service providers hope to have a more comprehensive grasp of owner information to formulate more targeted new owner development and old customer maintenance work. Meanwhile, telecom operators, as communication service providers, provide a wide range of communication services for all walks of life, can also in-depth cooperation with the automotive industry to explore the possibility of more value-added services and provide vehicle owners with more convenient and high-quality services.

In [1], it is generally believed that big data can discover the competitive advantages of enterprises, and help enterprises segment customers, evaluate market segments, select target market segments, attack target market and conquer them as well as expand markets through neighboring markets. In [2], authors introduced that in the context of marketing, big data can be used to obtain effective information and propose precision marketing strategies, which will enable companies to gain advantages in market competition. They also believe that the information obtained after in-depth mining of big data is important factors that make precision marketing strategies successful.

By using telecom big data, auto service companies and telecom operators provide more accurate services to vehicle owners, reduce customer acquisition costs and service costs. The first step is accurately to identify vehicle owners. At present, the identification of vehicle users is mainly based on the customer information of the vehicle company itself， cause users' awareness of privacy protection is becoming stronger and stronger, which has brought a heavy burden to operators for data collection and network management.[3]

With the continuous development of mobile Internet and digital technology, more and more new online services that meet market needs are developed, auto related products such as online violation queries are emerging in an endless stream providing various convenient services for vehicle owners. Therefore, it has become a new method to identify vehicle users by analyzing the user behavior of various online products. More information can help auto service providers master more about the use preferences of car owners, facilitate app developers to get user portraits with more analytical value, provide strong support for the update iteration of product functions, and help operators provide users with more personalized customized services. In fact, telecom operators have rich data and research capabilities related to vehicle services, can identify the use of different products in network data through DPI (deep packet inspection) in-depth analysis technology, give full play to the advantages of big data of telecom operators and build prediction models [4-10].

Therefore, based on the pain points of large demand for automobile services and difficult user data collection, this paper proposes a feature extraction method for vehicle users based on telecom operator big data, and introduces semi-supervised machine learning algorithm for user identification for the first time. This method can make full use of various online service data in the automotive aftermarket and effectively solve the problem of too small sample size, get credible prediction results. The specific contributions of this paper are as follows.

- TSVM model is used to predict whether users own vehicles, so as to help automobile sales and after-sales service providers understand user characteristics.
- A user behavior index system based on operator big data is designed to provide users' current vehicle usage and help vehicle service providers adopt more targeted marketing strategies and personalized services. In addition, the system provides operators with the impact of market segments on user needs. Automobile operators can dynamically adjust their marketing efforts according to this index to reduce operating costs.

The following systematically introduces the method from the aspects of mathematical principle, user cases, comparative experimental test and practical application test.

II. Proposal of TSVM model

A. Introduction to TSVM

The semi-supervised learning algorithm is between the supervised algorithm and the unsupervised algorithm, it is the combination of these two algorithms, including the operation of labeled data and the application of unmarked data. The

978-1-6654-6667-7/21/$31.00 ©2021 IEEE
DOI 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00081

transductive learning idea is to use unlabeled samples as the input in the model parameter optimization process so classification information carried in the unlabeled samples can be introduced into the model [15] and train the model with strong generalization ability. This method is more suitable when the sample set is small while the overall scale is huge. With a small amount of labeled data and a large amount of unlabeled data, you can still perform the task of data classification excellently. Considering the high cost and difficulty of collecting vehicle owner sample sets, semi-supervised learning methods can be used for training after a small number of samples are obtained [16]-[26].

Transductive Support Vector Machine(TSVM) is an improvement of the model based on the SVM(Support Vector Machine) model. As a classic direct push learning model, compared with SVM model, unlabeled test set is included in the training process and is labeled at the same time in TSVM model[27].

For *n* labeled sample sets containing positive and negative samples: $S_{train}: (\overrightarrow{x_1}, y_1), (\overrightarrow{x_2}, y_2), \cdots, (\overrightarrow{x_n}, y_n)$, $\overrightarrow{x_n}$ is the feature vector composed of various features, y_n represents the label value. The remaining *k* users are the unlabeled test set:$S_{test}: \vec{x}_1^*, \vec{x}_2^*, \cdots, \vec{x}_k^*$. $n+k$ is the total number of target users. Put *Strain* and *Stest* into the TSVM model at the same time, and the optimized objective function and subject conditions are list as (1a)-(1e):

$$\text{Minimize over} \begin{pmatrix} y_1^*, \cdots, y_n^*, \overrightarrow{\omega}, b, \varepsilon_1, \cdots, \\ \varepsilon_n, \varepsilon_1^*, \cdots, \varepsilon_k^* \end{pmatrix} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \varepsilon_i + C^*\sum_{j=1}^{k} \varepsilon_j^* \tag{1a}$$

$$\text{s.t.} \forall i \in [1, n]: y_i[\omega \cdot x_i + b] \geq 1 - \varepsilon_i \tag{1b}$$

$$\forall j \in [1, k]: y_j^*[\omega \cdot x_j^* + b] \geq 1 - \varepsilon_j^* \tag{1c}$$

$$\forall_{i=1}^{n}: \varepsilon_i \geq 0 \tag{1d}$$

$$\forall_{j=1}^{k}: \epsilon_j^* \geq 0 \tag{1e}$$

The TSVM model mainly requires the four steps for training, as following:

Step 1: Specify the initial parameters C and C*.

Step 2: Use inductive SVM as the initial classifier, perform an initial learning based on labeled samples to obtain an initial classification result. Output *n* unlabeled samples temporarily labeled as positive samples, and the remaining unlabeled samples are labeled as negative samples. At the same time, specify an initial impact factor and respectively for the positive sample and the negative sample;

Step 3: Train all samples (positive samples and negative samples) by exchanging labels for different test samples according to certain rules, and keep looping this step until the objective function can no longer be reduced.

Step 4: Evenly increase the sum of temporary influence factors, and go to step 3 to obtain a lower objective function value. When the sum is equal to C*, the algorithm ends, and the final model parameters are output.

B. Indicators description

The popularity and maturity of mobile Internet technology has made most vehicle owners adopt online services, which are more convenient and efficient. Therefore, all online behaviors related to vehicle services reflect the needs of vehicle use. Telecom operators have a large amount of user online behavior data, which can comprehensively describe the user's occupation, behavior preferences, usage habits, permanent residence and other information. Especially the data in OSS (Operation Support Systems) domain, which has the characteristics of large data volume and short time period can be used to fully excavate, filter out the behavioral data related to vehicle services and build an index system for vehicle users [11]-[14].

From a scenario perspective, vehicle services can meet user needs in different scenarios. Due to the natural differences in the frequency of different scenarios, such as the frequency of maintenance and parking are different, it needs to be further subdivided according to the demand scenario. From a data perspective, users online behavior reflects various active behavior that users take to meet their own needs. This type of data is updated in real time and has a large amount of data.

Telecom big data contains all of the user online data and needs to be classified into categories [28]-[36]. According to the upstream and downstream market of the automobile market and the user's demand for various products in the automobile market, user's online behavior related to automobile services can be further divided into five segmented needs, including vehicle maintenance, refueling, parking, illegal treatment and comprehensive services. These major needs can provide vehicle owners with more comprehensive online and offline services. At the same time, because of the difference in service types, there are differences in frequency and duration of use between different needs. By observing these data to determine the user's dependence on each segmentation demand as the user's vehicle owner attribute characteristics.

By investigating and sorting out the supply products of the target industry, this article evaluates the supply products that meet the segmented demand, and reflects the degree of dependence of users on the segmented demand from the three perspectives of continuity, diversity and contribution.

Fig.1 is the logical structure of the indicator system. In order to achieve the final goal, we further decompose the needs of vehicle users into 5 subdivisions (principal level). Each subdivision demand is represented by three data. Here we draw on the idea of AHP (Analytic Hierarchy Process) method for integration while the weights between different levels are not determined based on expert experience. In this paper, machine learning models are introduced to automatically obtain their nonlinear relationship.

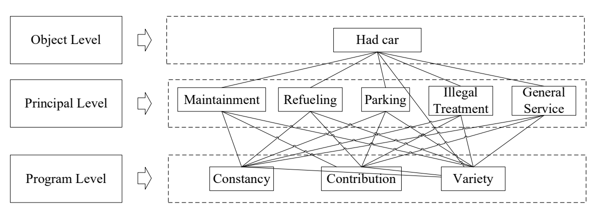

Fig. 1. Indicator system for vehicle users

C. Indicators in program level

OSS data of telecom operators records the time of the user's access to the business in detail, the number of visits and the traffic generated during the visit. In order to fully reflect the user's degree of dependence on each segmentation requirement, three indicators from time, traffic and business diversity are portrayed from three perspectives.

It can be found that users will spend more time browsing and using favorite or necessary online services to meet their own needs. Time and traffic can record users' browsing behavior. On the other hand, in order to meet their own needs, users will actively look for various online services. The type of online services can reflect the demand degree of the demand.

Constancy mainly reflects the degree of dependence of users on demand in the time dimension, which is reflected by the number of visits to products within the segmented demand within a fixed period. The specific calculation process is shown as (2a) (2b):

$$\mathrm{Constancy} = \frac{\sum_{d=1}^{D} I_d}{D} \tag{2a}$$

$$I_d = \begin{cases} 1, \text{used this app in dth day} \\ 0, \text{no used this app in dth day} \end{cases} \tag{2b}$$

where D is the number of days in the observation period, d represents the usage behavior on the *dth* day of observation, and I_d is a dummy variable which represents whether this user has used the relevant app for meeting the segmentation demand on that day.

Variety mainly reflects the user's degree of dependence on this segmented demand from the perspective of business sight. It is reflected by the number of apps accessing products within the segmented demand within a fixed period. The specific calculation formula is as (3):

$$\mathrm{Variety} = \frac{N}{NN} \tag{3}$$

where N represents the number of service apps that users use to meet the segmented demand during the observation period, and NN represents the number of all services that meet the segmented demand.

Finally, *contribution* mainly reflects the user's dependence on demand from the perspective of usage. It is expressed by the average traffic of each APP. The following is the calculation process, as shown in (4) (5):

$$\mathrm{Flow}_d = \frac{\sum_{n=1}^{N'} f_{dn}}{N'} \tag{4}$$

$$\mathrm{Contribution} = \frac{\sum_{d=1}^{D'} \mathrm{Flow}_d}{D'} \tag{5}$$

where *Flow* represents the average access flow of a single APP of the user, d represents the service APP used on the *dth* day, n represents the *nth* service APP, and N' represents the number of service APPs used on the *dth* day. *Contribution* is the user's degree of contribution. D' represents the total number of days the user has access to this type of segmented demand business ($D \leqq D'$).

III. User Case of the TSVM model

A. Sample Selection

In order to test the effectiveness of the TSVM model and indicator framework, we randomly selected 50,000 users of a telecom operator. Then we tracked their behavior from November 1 to November 30, 2021.DPI(deep packet inspection) technology is mainly used to deeply analyze the data generated in the operator's network.Thus, we can obtain the traffic generated by users in the process of using various online APPs, duration of access, content of access, click and other actions.Regarding relevant APPs, we chose 112 APPs based on their downloads in app market which can be related to the market in the indicator framework.

By means of telephone and SMS, we investigated and recorded the vehicle status of these 50000 users in detail, recycled data of nearly 700 users was constructed into structured labels(no car or had car).Finally, we make correspondence between the obtained behavior data and the survey results to construct a satisfactory sample set.

TABLE I. Number of Apps in Different Principal Demand

Principal level	Maintainence	Refuel -ing	Park -ing	Illegal treat- ment	General Service
Total	15	35	10	13	20
Used	10	10	7	12	12
Used(%)	66.7%	28.6%	70%	92.3%	60%

Table 1 shows the number of apps used in different principal demand. From the perspective of APP usage, illegal treatment has the highest proportion of apps, and the absolute number is also the highest. Therefore, the demand for illegal treatment has a low degree of concentration. Users are distributed across multiple apps. While the demand for refueling segmentation is concentrated. There are many products but users are mainly concentrated in the top 10 products. It can be seen that competition for this segmentation demand is very fierce, and it is difficult to develop value-added services or new business development in the future.

This part shows the overall user requirements. Through the analysis of the user's use in the application, the user's centralized demand points can be reflected. Operators can focus on user needs and carry out targeted product reform to provide users with better services. Next, through detailed analysis, we get the differences and respective characteristics of car users and car free users in subdivision needs.

B. Normalization of indicators

Before the analysis, we first need to standardize the traffic indicators. Max-min normalization method is a commonly used method of data normalization. This method converts the indicator to (0,1) to removes the dimension. Then, calculations and comparison can be carried out between different indicators. Considering that the indicator contribution used in the TSVM model represents the visit flow value while other indicators are the ratio, so the max-min method is used to standardize the contribution index to meet the calculation. For the vector x, use the formula(6) to get the corresponding standardized value:

$$\widehat{X_i} = \frac{\max_I X_I - X_i}{\max_I X_I - \min_I X_I} \tag{6}$$

Among them, X_i is the *ith* value of the index vector X_I, and $\widehat{X_i}$ is the eigenvector which is the value after normalization corresponding to the *ith* value of the vector. Table 2 shows the numerical results before and after normalization.

TABLE II. Contribution Values in Different Users

Contribution (Average)		Maintain -ence	Refuel -ing	Park -ing	Illegal treat- ment	General Service
After	**No vehicle**	1.000	0.986	0.984	0.999	0.992
	Had vehicle	0.992	0.984	0.979	0.998	0.995
Before (MB)	**No vehicle**	0.001	0.349	0.079	0.767	7.513
	Had vehicle	0.432	0.404	0.104	2.679	4.575

From Table 2, it can be found that after standardization, the contribution of each subdivision demand in the principal level is between 0 and 1. First, the gap between users with and without vehicles is small. Fig. 2 is a comparison of different demand after standardization. It can be found that in terms of contribution, vehicle-free users are higher than vehicle users in maintenance, refueling, parking and illegal treatment, while vehicle users' performance in the general service segment demand is more prominently. This result is different from our habitual cognition, which may be due to the large fluctuation of the average value due to the small sample size.

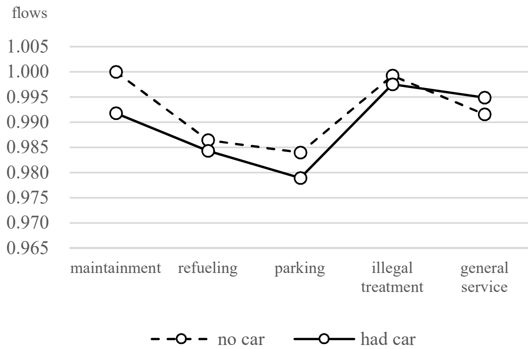

Fig.2. Average of used flows in different demand

Table 3 shows the proportion of users with a value of 0 to overall users in this category. Opposite conclusion can be found in Table 3. The proportion of users without a vehicle with a value of 0 for maintenance, refueling, and parking is significantly higher than that of users with a vehicle, while the proportion of vehicle-free users in terms of illegal treatment and general service is lower than that of users with a vehicle.

TABLE III. NUMBER OF ZERO IN DIFFERENT USERS

0%	Maintainence	Refueling	Parking	Illegal treat-ment	General Service
No vehicle	88.6%	77.3%	88.6%	50%	81.8%
Had vehicle	85.2%	76%	86.1%	52.0%	85.9%

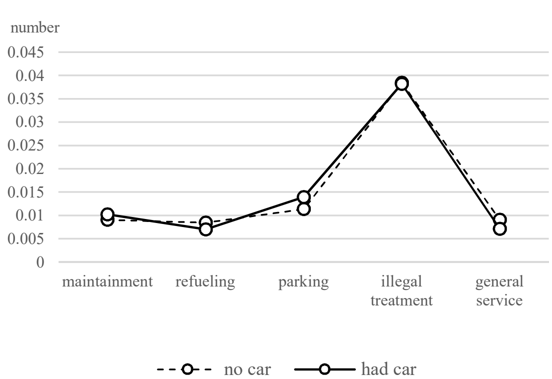

Fig.3. Average of used app in different demand

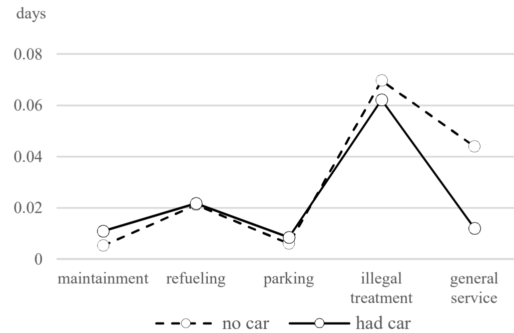

Fig.4. Average of used days in different demand

Fig.3 and Fig.4 are the average situation of the continuity and diversity characteristics of the two groups. The comparison shows that the biggest gap between the vehicle user group and the vehicle-free group is the maintenance service, which is more obvious in terms of continuity and diversity. Users with cars have a greater demand for maintenance services. When distinguishing groups, we can pay more attention to this type of service. Regardless of whether it is a vehicle or a vehicle-free group, illegal treatment and general service services are more frequently visited, and there are more types of access to APP than other segmented needs, which means that these two types of segmented needs can attract more attention. For this type of service, a wide coverage and low-cost marketing strategy can be adopted.

It is found that there are obvious differences in the performance of vehicle users and vehicle-free users in different segmentation needs, which is difficult to observe through the statistical indicators of all users. Thus, it is necessary to fully consider the comparison of different group's own indicators. TSVM algorithm can incorporate the feature value information of unlabeled samples into the model prediction process, so it can supplement the model training and adapt to the current sample status.

Through the separate analysis of the two kinds of users, it can be found that some differences in demand characteristics between car users and car free users are different from what people usually think. These characteristics are difficult to detect in the actual process of user interview, user clustering and demand analysis, so they are difficult to be reflected in product service design. It can be seen that TSVM algorithm has high commercial value for enterprise product operation and contributes to promoting the development of service design.

IV. MODEL EFFECT

A. Model Performance

In order to verify whether TSVM algorithm can effectively improve the accuracy of traditional classification algorithms, this paper uses logistic regression algorithm and SVM algorithm as the basic model. Among the classification methods of linear regression, logistic regression algorithm is one of the most classical machine learning classification algorithms. It is easy to calculate and does not need to consider multicollinearity. It has a wide range of applications. SVM algorithm has strict mathematical theory support, strong interpretability and does not rely on statistical methods, which simplifies the usual classification and regression problems. It is the theoretical basis of TSVM. By comparing with the results of logistic model and TSVM model, we can get the validity of TSVM. Table 4 shows the results of logistic model, SVM model and TSVM model.

TABLE IV. COMPARISON OF RESULTS

Method	Sample	Precision	recall	F1	accuracy
Logistic model	**Train**	0.938	0.998	0.967	0.938
	Test	0.941	0.996	0.968	0.938
SVM model	**Test**	1	0.940	0.969	0.940
TSVM model	**Test**	1	0.948	0.973	0.948

The part of the logistic regression model shows the performance of various performance indicators of the logistic regression model on the training set and the test set. It can be found that logistic regression model performs well both on train set and test set when identifying the user's vehicle use status. The results of SVM model are slightly improved in precision and accuracy compared with logistic regression model.

However, in the training process, it is found that the training time of SVM model is long, and the SVM algorithm directly uses the labeled data samples for training. After training, the test data are classified directly, and there is no iterative process of labeled samples. In contrast, the TSVM model has higher accuracy.

While in the TSVM model section shows the evaluation indicators obtained by the model. By constructing the TSVM model, we incorporate unlabeled samples into the training process. According to the TSVM calculation steps introduced in Section 2 by Python, kernel function is set to use linear and the value of the penalty parameter is set to 0.8.

By comparing various evaluation indicators, it can be found that TSVM has significantly improved in various indicators. In terms of F1 value representing the overall prediction effect of positive and negative samples, it has increased by 0.5% from the relatively high level comparing with logistic regression model，and it also increased by 4% comparing with the SVM model.

B. Actual Effect Test

The constructed TSVM model is used to predict the vehicle status of users in one province, these users are randomly sampled. Similarly, telephone and SMS are used to communicate with users to test the accuracy and precision of the model. Final results show that the accuracy reaches 90.58%, which may be because users are affected by more factors in the real situation.

In the future, we will consider introducing more factors related to vehicle status, and will continue to expand the number of sample sets. After the number increases, we will consider the classification algorithm for analysis and make a new comparative analysis with the semi supervised algorithm.

V. CONCLUSION

This paper systematically proposes a vehicle user identification indicator framework based on telecom operator big data. Based on this, TSVM model is used to predict whether the user has a vehicle or not, it is verified by using real sample data. Results show that the indicator construction and TSVM model proposed in this paper can more accurately identify whether a user has a vehicle which can provide more detailed and precise information for telecom operators and vehicle service providers. Therefore, they can adopt more targeted marketing strategies according to the current vehicle usage status of users. In addition, the intensity of marketing can be dynamically adjusted according to the degree of influence of market segments to further enhance the return on investment of funds.

REFERENCES

[1] Datis K, Mike F, Wilfried M. “Competitiveness in Emerging Markets Contributions to Management Science”, Springer, Cham, 2018.

[2] Ducange P, Pecori R, Mezzina P. “A glimpse on big data analytics in the framework of marketing strategies”, Soft Computing, 2017, 22(5/6):1-18 .

[3] J. Cheng, Y. Wu, E. Yuepeng, J. You, T. Li, H. Li, J. Ge, “MATEC: A lightweight neural network for online encrypted traffic classification,” Computer Networks, 108472, doi: 10.1016/j.comnet.2021.108472.

[4] L. Xu, et al. "A Comprehensive Operation and Revenue Analysis Algorithm for LTE/5G Wireless System based on Telecom Operator Data," in Proc. IEEE ScalCom, Aug 2019, Leicester, United Kingdom, pp.1521-1524.

[5] T. Zhang, et al. "A novel LTE network deployment scheme using telecom big data," in Proc. ICSINC, Beijing, China, October 2015, pp. 261-270.

[6] L. Xu, Y. Chen, "Priority-based resource allocation to guarantee handover and mitigate interference for OFDMA systems," in Proc. IEEE PIMRC, Sept. 2009, Tokyo, Japan, pp.783-787.

[7] W. Wang, et al. "A novel cell-level resource allocation scheme for OFDMA system," in Proc. IEEE CMC, Kunming, China, Jan. 2009, vol.1, pp. 287-292.

[8] L. Xu, et al. "User-vote assisted self-organizing load balancing for OFDMA cellular systems," in Proc. IEEE PIMRC, Sept. 2011, Toronto, Canada, pp. 217-221.

[9] H. Xing, et al. "A PBIL for load balancing in network coding based multicasting," in Proc. ICCSA, July 2016, Beijing, China, pp.34-44.

[10] X. Cheng, et al. "Digging Big Data: Telecom Data Description, Mining and Application, " China Machine Press, Beijing, 2019.

[11] L. Xu, et al. "Self-Organising Cluster-Based Cooperative Load Balancing in OFDMA Cellular Networks," *Wiley Wireless Communications and Mobile Computing*, vol. 15, no. 7, pp.1171-1187, May 2015.

[12] Y. Fan, et al. "An active wideband and wide-angle electromagnetic absorber at microwave frequencies," IEEE Antennas and Wireless Propagation Letters, 15, pp.1913-1916, March 2016.

[13] J. Gao, et al. "A Coverage Self-optimization Algorithm using Big Data Analytics in WCDMA Cellular Networks," in Proc. ICSINC, Oct 2015, Beijing, China.

[14] M. Yuan, et al. "Crowdfunding assisted Cellular System Analysis and Application," in Proc. 5GWN, April 2017, Beijing, China, pp.69-78.

[15] L. Xu, et al. Architecture and Technology of Multi-Source Heterogeneous Data System for Telecom Operator. In Proc. ICSINC, Oct 2020,Rizhao,China.

[16] W. Sun, J. Liu, Y. Yue and Y. Jiang, "Social-Aware Incentive Mechanisms for D2D Resource Sharing in IIoT," IEEE Transactions on Industrial Informatics, vol. 16, no. 8, pp. 5517-5526, Aug. 2020.

[17] W. Sun, P. Wang, N. Xu, G. Wang, Y. Zhang, "Dynamic Digital Twin and Distributed Incentives for Resource Allocation in Aerial-assisted Internet of Vehicles," IEEE Internet of Things Journal, DOI: 10.1109/JIOT.2021.3058213.(Early Access).

[18] L. Zhao, W. Zhao, et al, “Novel Online Sequential Learning-based Adaptive Routing for Edge Software-Defined Vehicular Networks,” IEEE Transactions on Wireless Communications, vol. 20, no. 5, pp. 2991-3004, May 2021.

[19] L. Zhao, G. Han, Z. Li, L. Shu, “Intelligent Digital Twin-based Software-Defined Vehicular Networks”, IEEE Network, vol. 34, no. 5, pp. 178-184, September/October 2020.
[20] Zhang, Q., et al. An Evaluation Model of User Lifetime Value Based on Improved RFM and AHP Method. In Proc. ICSINC,Oct 2020,Rizhao,China..
[21] X. Zhang, et al, "Towards Efficient Battery Swapping Service Operation Under Battery Heterogeneity," *IEEE Transactions on Vehicular Technology*, vol.69, no.6, pp.6107-6118, June 2020.
[22] X. Cheng, et al. "A novel big data based telecom operation architecture," in *Proc. ICSINC*, Beijing, China, October 2015, pp.385-396.
[23] H. Zhang, et al. "Big data research on driving behavior model and auto insurance pricing factors based on UBI," in *Proc. ICSINC*, Sept 2017, Chongqing China, pp.1-8.
[24] Y. Zhang, et al. "A Novel Big Data Assisted Analysis Architecture for Telecom Operator," *in Proc. IEEE IUCC*, Oct 2019, Shenyang, China, pp. 611-615.
[25] G. Cui, et al, "Joint Offloading and Resource Allocation for Satellite Assisted Vehicle-to-Vehicle Communication," IEEE Systems Journal, DOI: 10.1109/JSYST.2020.3017710, online available, 2020.
[26] Y. Chi, et al. "Sorting and Utilizing of Telecom Operators Data Assets Based on Big Data," *in Proc. IEEE IUCC*, Oct 2019, Shenyang, China, pp. 621-625.
[27] Joachims, T.. "Transductive Inference for Text Classification using Support Vector Machines." Sixteenth International Conference on Machine Learning Morgan Kaufmann Publishers Inc. 1999.
[28] B. Yin, Y. Wu, T. Hu, J. Dong, Z. Jiang, "An Efficient Collaboration and Incentive Mechanism for Internet of Vehicles (IoV) With Secured Information Exchange Based on Blockchains," in IEEE Internet of Things Journal, vol. 7, no. 3, pp. 1582-1593, March 2020.
[29] J. Cheng, Y. Wu, E. Yuepeng, J. You, T. Li, H. Li, J. Ge, “MATEC: A lightweight neural network for online encrypted traffic classification,” Computer Networks, 108472, doi: 10.1016/j.comnet.2021.108472.
[30] X. Liu, J. You, Y. Wu, T. Li, L. Li, Z. Zhang, J. Ge, “Attention-Based Bidirectional GRU Networks for Efficient HTTPS Traffic Classification,” Information Sciences, vol. 541, pp. 297-315, 2020, doi: 10.1016/j.ins.2020.05.035.
[31] Y. Wu, H. -N. Dai, H. Wang, “Convergence of Blockchain and Edge Computing for Secure and Scalable IIoT Critical Infrastructures in Industry 4.0,” IEEE Internet of Things Journal, vol. 8, no. 4, pp. 2300-2317, 15 Feb.15, 2021.
[32] M. Eiza, "Towards Sustainable and Economic Smart Mobility: Shaping the Future of Smart Cities," *World Scientific*, London, 2020.
[33] L. Xu, et al. "Research on Telecom Big Data Platform of LTE/5G Mobile Networks," in *Proc. IEEE IUCC*, Oct 2019, Shenyang, China, pp. 756-761.
[34] C. Zhu, et al. "5G Wireless Networks Meet Big Data Challenges, Trends, and Applications," *in Proc. IEEE ScalCom*, Aug 2019, Leicester, United Kingdom, pp. 1513-1516.
[35] J. Yang, et al. "An Insight Scheme for Large-Scale Events Based on Telecom Operators Data," *in Proc. IEEE IUCC*, Oct 2019, Shenyang, China,pp. 626-631.
[36] K. Chao, et al. "Data mining based modeling and application of mobile video service awareness," *in Proc. ICSINC*, Sept 2017, Chongqing China, pp. 389-396.