PAPER • OPEN ACCESS

Charging Station Recommendation for Electric Vehicle Based on Federated Learning

To cite this article: Xiaohui Wang *et al* 2021 *J. Phys.: Conf. Ser.* **1792** 012055

View the article online for updates and enhancements.

Charging Station Recommendation for Electric Vehicle Based on Federated Learning

Xiaohui Wang[1*], Xiaokun Zheng[1]and Xiao Liang[1]

[1]Artificial Intelligence on Electric Power System State Grid Corporation Joint Laboratory (GEIRI), Global Energy Interconnection Research Institute Co. Ltd, Beijing, 102209, China

[*]wangxiaohui@geiri.sgcc.com.cn

Abstract. At present, the usage of EV charging facilities is unbalanced. The accuracy of the charging station recommendation does not meet the demand. Due to the limitation of user privacy protection, charge point operators and vehicle enterprises cannot provide data to each other for joint analysis. Therefore, we proposed recommendation method of EV charge point based on federated learning. The federated factorization machine is implemented to make use of data features in both sides and cross features between them. We build the model by encrypted entity alignment, secure federated training and predicting. The experimental results show that the federated model improves the AUC of the model by 6% over those built with features only from the charge point operators. The model is superior to centralized LR-based and RF-based models. While the data does not need to leave the original platform, the model realizes the secure and precise federated charging point recommendation based on more comprehensive features.

1. Introduction

In recent years, driven by policy support of the government and new energy technologies, the industry of electric vehicles and charging facilities has developed rapidly. According to the statistics of the China Charging Alliance, by the end of December 2019, the number of charging piles in China reached 1.219 million, of which 516,000 were public charging piles. However, the number of new-energy vehicles exceeded 4.1 million. The overall vehicle-to-pile ratio is around 3.4:1, but the utilization rate of charging piles is less than 15% generally[1]. According to the development plan of the new-energy automobile industry, the sales of new-energy vehicles will reach 7 million at a conservative estimate in 2025[2]. On one hand, the number of charging stations is insufficient. On the other hand, the usage of charging facilities is uneven. In reality, electric vehicle users often come up against problems such as malfunctioning charging piles, long waiting times for charging and parking space occupancy by gasoline vehicles. Therefore, the intelligent and effective charging service recommendation is of great significance to the improvement of the users' charging experience. It also benefits the overall operation of charging facilities, and the healthy development of the new energy industry.

There has been some researches on the recommendation for electric vehicle charging facilities. Bu et al.[3] proposed collaborative filtering algorithm to recommend charging piles for users based on the similarity of historical charging behaviors which only considered charging action, but didn't take into account the characteristics of the vehicle and operation features of the piles. It is not enough to support real-time recommendation for charging facilities. YAN et al.[4] proposed the optimization method of

charging path recommendation that considering electric vehicles, power distribution network and road network as a whole, but the optimization goal was to relieve local traffic congestion and pressure on the grid, rather than improving user experience and charging pile utilization. Ding et al.[5-8] carried out path planning of the shortest path or combined optimization methods, considering the trip rule and energy consumption prediction of electric vehicle. However, there are not large-scale centralized scheduling and charging prediction strategies within the current domestic charging facilities. Those methods achieved the optimization goal in a local simulation environment, but are in lack of practical value. Some foreign scholars have studied the combined application of edge computing and charging facility recommendation[9,10], but this kind of theoretical research is not applicable to the foundation of the construction and operation of domestic charging facilities.

More comprehensive and related features will be more conducive to improving the accuracy of the charging recommendation model, especially involving vehicle status and operation of charge points. With increasingly stringent data security supervision system and user privacy protection requirements, vehicle and charge points data owners cannot provide data to each other by data transactions or sharing transmission. Therefore, it is indispensable to establish a cross-platform joint data analysis mechanism that enable federated model training with data not leaveing its original platform, only the intermediate results and parameters are exchanged in encrypted way.

This paper is application-oriented, which makes the first attempt to import federated learning in the field of power data analysis and design the federated factorization machine (FM)[11] algorithm to solve the problem of intelligent recommendation for charging facilities. The major contributions are as follows.

- We design the cross-platform federal learning architecture to realize EV charging point recommendation. Data from vehicle data platform and charge point data platform is used together to train the recommendation model while keeping the original data in the local platform to effectively protect data privacy.

- We propose an encrypted entity alignment method for different IDs from different platform based on hash and RSA encrypting algorithm.

- We implement the vertical federated factorization machine algorithm by homomorphic encryption, achieving secure federated training and predicting. It has proved that the federated model improves the AUC (Area Under the Curve) of the FM model by 6% over those built with data from the charge point operators merely. Compared with centralized training, the model is almost lossless and superior to centralized LR-based(Logistic-based) and RF-based(RandomForest-based) models.

2. Federated leraing

2.1. Federated learning concept
Artificial intelligence is booming, but data quality and quantity has been restricting its further development and application in many fields. Data privacy and security has gained more attention with the promulgation of a series of privacy protection laws and policies in various countries. Federated learning was proposed by Google in 2016[12,13] to realize the establishment of a machine learning model based on distributed mobile data sets and prevent data leakage simultaneously. WeBank has further expanded federated learning to an encrypted collaborative machine learning framework and mechanism[14], consequently various organizations can jointly build models without disclosing the underlying data and its encrypted form. According to the distribution characteristics of data, federated learning can be divided into horizontal and vertical one and federated transfer learning. At present, a large number of scholars have studied the application of federated learning in the field of private computing and distributed machine learning[15,16].

Among them, horizontal federated learning is for scenarios where the feature spaces of multi-party data sets overlap but the samples are different. Vertical federated learning is prefered when the sample spaces of the data set overlap but the feature spaces are distinct. The train process of vertical federated learning requires a trusted third party in general, which can be an authority, a security platform server,

or an SGX (software guard extensions) trusted computing environment. The process contains sample alignment, federated training and federated inference. The federation implementation of different algorithms could be made by distinctive encryption algorithms and aggregation methods. The detailed description of the implemented federated vertical factorization machine is in section 2.3-2.5.

2.2. Federated learning application

Federated learning is being applied in the distributed learning of voice and text models on smartphones, credit evaluation and anti-money laundering detection in the financial field. There are also a large number of potential application scenarios in the fields of IoT (Internet of Things), autonomous driving, visual security, smart retail and auxiliary diagnosis of smart medical care[14]. At present, there is no application scenario of federated learning in the power field, and federated learning will be of great significance to power data security analysis between different fields and data integration and data share with external enterprises.

3. Implementation of recommendation algorithm based on federated factorization machine

3.1. Cross-platform federal learning architecture

As shown in Figure1, the cross-platform federated learning architecture for charging facility recommendation mainly involves three platforms, including charge point data platform (CP), vehicle data platform (VP) and third-party platform.
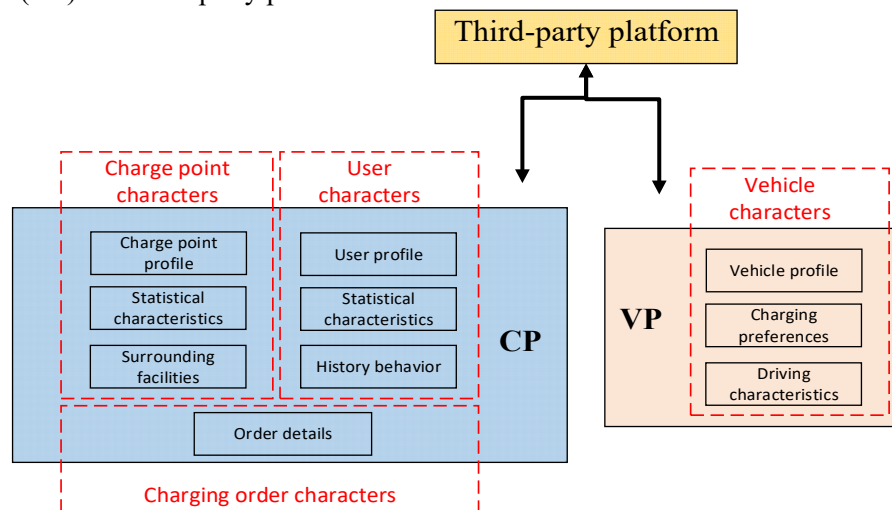

Figure 1. Cross-platform federated learning architecture for vehicle charging station recommendation.

The data of CP comes from charging records of charging service APP(Application) users and using records of charge points. We extract characteristics of charge point, user and charging order. The VP provides electric vehicle characteristics. The data on both sides of the platform jointly supports personalized recommendations for electric vehicle charging services, neither the original data not leaving the local platform nor transmitting to the third party. The trusted third-party platform manages the keys in the subsequent federal learning process to summarize the encrypted features. Encryption methods and aggregation strategies ensure that attributes and original data will not be divulged.

3.2. Dataset generation

Dataset generation includes two steps: feature extraction and label generation.

- Feature Extraction. We make use of the historical data of the charge point data platform and the vehicle data platform to extract the relevant characteristics for the charging recommendation.

- Label Generation. We generate labels (whether charging or not) on the side of the charge point data platform from the historical charging orders for model training.

Part of data features we extrated from VP and CP are shown in Table 1.

Table 1. Data features.

	Platform	Description	example
UID_i	CP	The user identification	5037988
$Cost_i$	CP	Charging cost in the last week of the user	29.62
Cap_i	CP	Charging capacity in the last week of the user	37.03
$CID_{i,k}$	CP	ID of the charge point used by the user in the last k time	5040490000000457
CID_x	CP	Average number of charging hourly of the charge point	1.2
F_x	CP	Average charging cost hourly of the charge point	0.53
C_x	CP	Average charging capacity hourly of the charge point	15.7
(Lat_x, Lng_x)	CP	Latitude and longitude of the charge point	(31.334762,109.907606)
P_x	CP	Parking fee of the charge point	0
S_x	CP	Service fee of the charge point	1.0
CT_x	CP	Charging type of the charge point	quick
R_x	CP	Number of restaurants in the surrouding	15
VID_z	VP	Electric vehicle identification	8542
VC_z	VP	Vehicle battery capacity	82
SP_z	VP	Average speed of the vehicle	47.5
CI_z	VP	Charge interface type of the vehicle	2015
VT_z	VP	Main purpose of the vehicle	Private car

The features extracted from CP include the user profile and charge point profile, statistical characteristics of charge points and users, the number of surrounding facilities like restaurants, sports facilities, life services. The features extracted from VP are consist of vehicle profile, charging perferences and driving characteristics.

For charging order characteristics, explicit information is extracted from CP side, including charging start and stop time, charging cost, charging quantity, etc., while charging start and stop time is deduced from VP side through vehicle position reported every 5 seconds and state of charging and running. In order to facilitate entity alignment during federal training, the charging start time, end time and the charging duration calculated from both platforms are retained to the minute dimension.

We mark the sample labels as 1 for the explicit charging order as positive samples of the data set. Negative samples are generated by a heuristic strategy. The charge point which is idle at the beginning of charging and within a certain range (σ_1KM) around the charge point in the charging order is choosen. Then we use the same user identification, user characteristics and charging order features corresponding to positive samples as supplement for negative samples whose labels are marked as 0, just replacing the features of the charge point.

3.3. Entity alignment

Alignment methods of encrypted sample ID are usually implemented based on the same type of ID[17], but the user IDs in CP are different from the vehicle IDs in VP. This paper proposes an encrypted entity alignment method for different IDs innovatively, which realizes the sample alignment between the electric vehicle and the users of charging service. The steps are shown in Table 2.

Table 2. Entity alignment steps.

No	Description
1	Calculate the sample ID (Hid) in CP and VP side respectively.
2	Generate RSA key pair on VP side, send public key (E,n) to CP, keep private key (d,n).

3	Generate E with hash algorithm based on random number encryption for Hid on CP side. Send E to VP.
4	Decrypt E to get D. Generate F by hashing the Hid on VP side. Pass D and F to CP.
5	Hash D which is divided by the random number to get G on CP side. Take the intersection of G and F to get I, the set of Hid ids shared on both sides. CP sends the I back to the VP and keeps the Hid and UID mappings.
6	Deduce the common set of ids and retained the Hid and VID mappings on VP side.
7	Deduce the map relationship between UID and VID in the common ID set on both sides.

The number of samples are mcp and mvp on the CP side and VP side respectively. We can ascertain a matched charging order uniquely by common attributes including the charging location (latitude and longitude), charging start time and charging duration (in minutes). As shown in formula (1), the sample ID (Hid) is generated through the hash algorithm.

$$Hid = H(T_{start}, Lng, Lat, T_{duration}) \tag{1}$$

The calculation method of E, D, F and G is shown in formula (2) -(5).

$$E_i = rand_i^e * H(Hid_i), i \in \{i, 2, 3, \cdots\cdots, \ m_{cp}) \tag{2}$$

$$D_i = (E_i)d = rand_i * H(Hid_i)^d \%n \tag{3}$$

$$F_j = H(H(Hid_j)^d), j \in \{1, 2, 3, \cdots\cdots, \ m_{vp}\} \tag{4}$$

$$G_i = H(D_i / rand_i) = H((H(Hid_i))^d) \tag{5}$$

3.4. Algorithm implementation of federated factorization machine

The recommendation model for electric vehicle charging facilities is built on factorization machine, which is good at learning cross features and context information, and improving generalization ability and efficiency compared with recommendation algorithms such as collaborative filtering and GBDT[18]. We implement the federated FM by decomposing the cross feature calculation and loss calculation. And the homomorphic encryption algorithm and safe aggregation strategy are used to ensure secure multi-party federation training.

The calculation decomposition method of cross features is divided into two cases. When two features are on the same data platform, they are calculated on one side. And the cross-platform features are summed by a third-party platform based on homomorphic encryption.

The characteristic decomposition of the cross features in two platforms is described in formula (6):

$$[[\sum_i \sum_j <V_i, V_j> x_{q,j} x_{p,i}]] = [[\sum_j V_j^{vp} x_{q,j}^{vp}]] + [[\sum_i V_i^{cp} x_{p,i}^{cp}]] \tag{6}$$

"[[ ]]" is the symbol for homomorphic encryption calculation.

The loss function is decomposed into two parts, one can be calculated separately and the other one needs to be calculated jointly. Based on homomorphic encryption, the third-party platform passes parameters.

The calculation formula for the overall loss of the factorization machine is shown in (7)

$$Loss = \sum_i (y_i - \hat{y}_i)^2 + \lambda_w \|W\|^2 + \lambda_v \|V\|^2 \tag{7}$$

y_i is the prediction, $\hat{y}_i$ is the true value. The latter two terms are the regularization terms for the weight of single feature and cross features separately.

As shown in formula (8), the calculation of the prediction is decomposed into that on both sides.

$$y_i = u_i^{cp} + u_i^{vp} \tag{8}$$

It can be described by formula (9) after the loss is split.

$$[[Loss]] = [[Loss^{cp}]] + [[Loss^{vp}]] + [[Loss^{cp-vp}]] \tag{9}$$

With reference to the reformation of the longitudinal linear regression model in [12], this paper proposes the calculation methods for the loss of the federated FM overall and each part shown in formulas (10) -(13).

$$[[Loss]] = [[\sum_i (u_i^{cp} + u_i^{vp} - \hat{y}_i)^2 + \lambda_{w^{cp}} \| W^{cp} \|^2 + \lambda_{w^{vp}} \| W^{vp} \|^2$$
$$+ \lambda_{v^{cp}} \| V^{cp} \|^2 + \lambda_{v^{vp}} \| V_{vp} \|^2 + \lambda_{v^{cp-vp}} \| V_{cp-vp} \|^2]] \tag{10}$$

$$[[Loss_{cp}]] = [[\sum_i (u_i^{cp} - \hat{y}_i)^2 + \lambda_{w^{cp}} \| W^{cp} \|^2 + \lambda_{v^{cp}} \| V^{cp} \|^2]] \tag{11}$$

$$[[Loss_{vp}]] = [[\sum_i (u_i^{vp})^2 + \lambda_{w^{vp}} \| W^{vp} \|^2 + \lambda_{v^{vp}} \| V_{vp} \|^2]] \tag{12}$$

$$[[Loss^{cp-vp}]] = [[\sum_i 2(u_i^{cp} - \hat{y}_i)[[u_i^{vp}]] + \lambda_{v^{cp-vp}} \| V_{cp-vp} \|^2]] \tag{13}$$

The specific steps are shown in Table 3.

Table 3. Federated factorization machine steps.

No	Description
1	Initialize the weight vector of single feature **W** and cross weight vector **V** on CP and VP sides respectively.
2	Calculate the estimated value, encrypted loss and characteristic gradient of VP side and sent it to CP. ($[[u_i^{vp}]]$, $[[Loss_{vp}]]$, $[[\sum_j V_j^{vp} x_{q,j}]]$)
3	Calculate the encrypted loss and characteristic gradient of CP side and sent it toVP.($[[u_i^{vp}]] + [[u_i^{cp} - \hat{y}_i]]$, $[[\sum_i V_i^{cp} x_{p,i}^{cp}]]$)
4	Send gradient that encrypted and added secure aggregate mask on VP and CP side seperatedly to third-party platform. Send loss to third-party platform on CP side.
5	Decrypte and summarize the calculated gradient on the third party platform. Send the updated gradient which added the corresponding security aggregation mask to VP and CP respectively.
6	Remove the mask and update the model on VP and VP side. Go back to step 2.

3.5. Federated predicting

When the model is called on CP side, we first pull a list of charge points within a certain range (σ_2KM) nearby based on the users' UID and current location. Then, we calculate the the single features, cross features both on CP side and features that crossed with VP according UID and the CIDs in the charge point list. After calculation, we send u^{cp} to the third party and send Hid to VP side. Based on the Hid ,we get the VID list on VP side. Then we calculate u^{vp} based on the local features, and sent it to the third party. The third party summarizes the calculation results of the entire feature to obtain the prediction, which is encrypted and sent to CP side.

After decrypting, we get the recommended prediction, the range of which is from 0 to 1. We sorted the charge points in the list in reverse order according to the recommended predicted value, and recommend the top N to the user in order.

4. Experimental results and evaluation

4.1. Experimental Dataset and Parameters

Our dataset is collected from the charge point data platform and the vehicle data platform of a certain southern city from September to October in 2019. There are over 650,000 charge point operating samples of 30,000 users and 1,800 charge points as well as more than 620,000 vehicle status samples of 50,000 electric vehicles in the same period. Finally, we screened the data of 800 active users in September as the training set and took the relevant data of those users in October as the test set.

Based on the method in section 2.2, we carry out feature extraction and label generation as well as the training set and test set construction. We extract 42 single features in the charge point platform (including label), and 11 features in the vehicle platform. σ_1 is set to 5KM. Based on the method in 2.3, the training set has achieved encryption alignment of about 49,000 Hid samples, involving 651 matches between UID and VID. The federal reasoning σ_2 is set to 5KM, and number of charge points to recommend N is set to 10.

4.2. Model Evaluation

Model performance is evaluated from three aspects. First is comparision between federated FM model and CP-side FM model. As shown Figure 2, the AUC of the federated FM model reaches 0.98 in training set, and CP-side FM is 0.92. On test data, AUC of the federated FM model is 0.94, while CP-side FM model is 0.08 less than it.
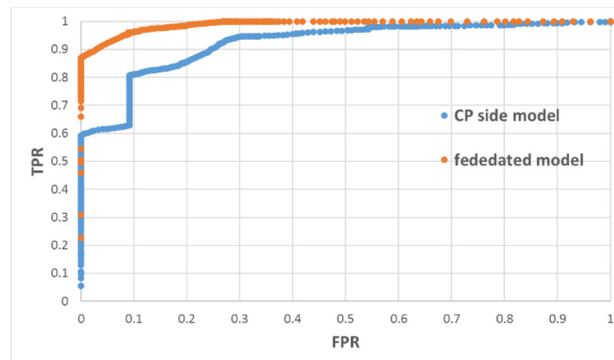

Figure 2. AUC comparison of federated FM model and CP-side FM model.

Secondly, we compare the federated training model with the model of centralized training. The federated model is nearly lossless. Finally, the FM model under the centralized training is compared to Linear Regresion (LR)model and Random Forest(RF) model which are also centralized trained. Results are shown in table 4.

Table 4. Model performance comparison.

	Federated FM	Centralized FM	Centralized LR	Centralized RF
AUC	0.94	0.95	0.86	0.88

5. Conclusion

In this paper, we design an application framework of federated learning in the field of electric power, and realize the recommendation method of EV charging point based on federated learning. We use heuristic strategy to generate training sets and propose an encrypted entity alignment method for different IDs, implement a federated factorization machine on the basis of homomorphic encryption and completes federated training and predicting. This paper builds an experimental environment to verify the usability of the framework of federation learning with real data. In consideration of data features in both the charge point data platform and the vehicle data platform and introduction of recommendation model with cross features, the federated model improves the AUC over those with

builting features only from the charge point operators. The model is almost lossless in comparison with the centralized training and superior to centralized LR-based and RF-based models. The work of this paper is the first application of federated learning in the field of electric power. Based on protecting the privacy of electric power data and electric vehicle user data, it explores the safe way of collaborative computing and external sharing with electric power data. The current data set does not have the users' behavior monitoring data after recommending, therefore, the feedback of the hit rate indicator of the recommendation will be collected to optimize the model in the future. In addition, how to reduce the communication and computing overhead for federated training and apply power data to other federated learning scenarios is also a momentous issue in the future research work.

Acknowledgments

References
[1]   F Zhang, K Li, X L Qin, et al. 2020. 2019-2020 Annual Report on China's Charging Infrastructure Development. China Electric Vehicle Charging Infrastructure Promotion Alliance, Beijing, BJ, CHN.
[2]   2019. New Energy Automobile Industry Development Plan (2021-2035). Ministry of Industry and Information Technology, Beijing, BJ, CHN.
[3]   F P Bu, S M Tian, J J Gao, et al (2017). Recommendation method for electric vehicle charging based on collaborative filtering. Science & Technology Review, 35(021), 61-67.
[4]   Y Y YAN, Y G LUO, T ZHU, et al (2015). Optimal Charging Route Recommendation Method Based on Transportation and Distribution Information. Proceedings of the CSEE, 35(2), 310-318.
[5]   D Ding, Research and Implementation of New Energy Vehicle Charging Path Planning and Service System, https://cdmd.cnki.com.cn/Article/CDMD-10700-1018835760.htm
[6]   H C Zhang, Z C Hu, Y H Song, et al (2014). The Forecasting Method of Electric Vehicle Charging Load Considering Temporal and Spatial Distribution. Automation of Electric Power Systems, 38(1), 13-20.
[7]   S SU, T T YANG, Y J LI, et al (2014). Charging Route Planning for Electric Vehicles Considering Real-time Dynamic Energy Consumption. Automation of Electric Systems, 38(1), 13-20.
[8]   X Y Liu, Optimal Dispatch of Electric Vehicles and Planning of Charging Facilities in Power Distribution Systems with Photovoltaic Access, https://cdmd.cnki.com.cn/Article/CDMD-10335-1019132446.htm
[9]   Y Cao, O Kaiwartya, Y Zhuang, et al (2019). A Decentralized Deadline-Driven Electric Vehicle Charging Recommendation. IEEE Systems Journal, 13(3), 3410-3421.
[10]   C K Wen, J C Chen, J H Teng, et al (2012). Decentralized plug-in electric vehicle charging selection algorithm in power systems. IEEE Transactions on Smart Grid, 3(4), 1779-1789.
[11]   S Rendle. 2010. Factorization machines. IEEE International Conference on Data Mining. IEEE, 995-1000.
[12]   J Konečný, H B McMahan, D Ramage, et al (2016). Federated optimization: Distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527.
[13]   J Konečný, H B McMahan, F X Yu, et al (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.
[14]   Q Yang, Y Liu, T Chen, et al (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.
[15]   J X Liu, X F MENG (2020). Survey on Privacy-Preserving Machine Learning. Journal of Computer Research and Development, 57(02), 346-362.

[16] Y Y Jia, Z Zhang, J FENG, et al (2020). The Application of Federated Learning Model in Confidential Data Processing. Journal of CAEIT, 15(1), 43-49.

[17] M Scannapieco, I Figotin, E Bertino, et al. 2007. Privacy preserving schema and data matching. Proceedings of the 2007 ACM SIGMOD international conference on Management of data. ACM, 653-664.

[18] K K Zhao, L F Zhang, J Zhang, et al (2019). A Summary of Research on Factorization Machine Model. Journal of Software, (3), 19.