A Semantic and Multidimensional Analysis-based Charging Station Classification and Operation Method and its Application

Weijia Zheng[1, a], Lingji Kong[2], Zihan Xu[1], Sijia Chen[2], Zhijuan Qu[1]

[1] STATE GRID SMART INTERNET OF VEHICLES CO., LTD, Beijing, China

[2] State Grid Shandong Electric Power Company Heze Power Supply Company, Heze, China

[a]lyjtxb@163.com

Abstract

To address challenges in charging station operations such as low efficiency, resource imbalance, and inadequate fault prediction during the large-scale development of new energy vehicles, this study proposes a hierarchical operation and maintenance (O&M) method integrating semantic analysis, grid clustering, and uncertainty modeling. By consolidating multi-source O&M data, the method establishes a four-dimensional evaluation index system. It utilizes semantic models to extract information from unstructured customer service texts, employs grid clustering to identify regional service pressure, and incorporates polyhedral uncertainty theory to enhance evaluation robustness, ultimately forming a tiered O&M strategy. Practical applications demonstrate that this approach enables precise allocation of O&M resources, improves fault response efficiency, effectively reduces excessive O&M costs, and provides technical support for intelligent charging station management.

Keywords

Charging station, hierarchical operation and maintenance, semantic analysis, grid clustering, uncertainty modeling.

1. Introduction

With the advancement of the "dual carbon" goals, the new energy vehicle industry has entered a phase of explosive growth, driving the charging infrastructure to enter a new stage of large-scale deployment. By 2025, the number of public charging piles nationwide is expected to exceed 6 million, with daily usage frequency per pile generally reaching over 8 times. In high-traffic areas such as expressway service areas and core commercial districts, the daily usage frequency per pile even surpasses 15 times. The high-load and high-frequency operational characteristics have led to a 40% increase in equipment failure rates compared to initial levels, making the traditional "periodic inspections + reactive repairs" maintenance model increasingly inadequate for industry development needs [1].

Current charging station operations face four core challenges: First, fragmented maintenance coverage-some cities have operational radii exceeding 50 kilometers per area, with maintenance personnel spending over 60% of their time commuting between inspection trips, severely limiting effective working hours. Second, data silos: operational data from charging stations, user charging patterns, and environmental monitoring are scattered across different systems, lacking unified integration and in-depth analysis. Third, incomplete evaluation frameworks: traditional maintenance grading focuses solely on individual equipment status while neglecting spatial connectivity and regional service pressure transmission effects. Fourth, questionable data quality: multi-source data contains measurement errors and omissions, with conventional methods failing to account for data uncertainty impacts on evaluation results [2-3]. These issues lead to imbalanced resource allocation, hinder service quality improvement,

and urgently require developing an efficient, precise, and proactive tiered maintenance approach [4-7].

This paper proposes a semantic and multidimensional analysis-based classification method for charging station operation and maintenance. The approach utilizes multi-source data collection as its foundation, forming a structured evaluation dataset through semantic analysis and quantitative metrics. Grid clustering is employed to achieve regional pressure classification, while robust comprehensive operation and maintenance indices are calculated using uncertainty modeling. Ultimately, differentiated operation and maintenance strategies are formulated based on the classification results.

2. Construction of Hierarchical Operation and Maintenance Model for Charging Stations

Step 1: Collect multi-source operational data from the e-Charging Platform, 95598 hotline, and WeChat Official Account through standardized interfaces to build a comprehensive dataset. The dataset includes: transaction serial numbers, charging equipment IDs, vehicle VIN codes, transaction completion reasons, charging start/end times, and charging volumes from order records. It also captures real-time voltage, current, and power data with millisecond-level waveforms during charging, along with 95598 customer service logs, environmental temperature/humidity, seasonal conditions, equipment operational history, manufacturer details, and installation locations.

Step 2: Construct an O&M evaluation index system based on multi-source data, comprising three key components: Location Indicators (describing critical charging station areas, distance to local city center, and target city center), Customer Demand Indicators (repair requests frequency, consultation/complaint data, and service pressure in surrounding areas), and Operational Indicators (annual average monthly charging volume, maximum charging station concurrency rate). The equipment status analysis includes three metrics: core component health status, fault prediction probability, and communication stability.

Step 3: Utilize semantic models to analyze 95598 customer service records. By controlling generation parameters and prompts, the system automatically extracts keywords to construct structured data samples. This includes: 1) The customer repair frequency metric, where semantic models automatically identify repair records in texts, categorize monthly repair counts by charging station names, and generate monthly sample data; 2) The customer consultation and complaint metric, which identifies consultation and complaint records, converts them into work order counts according to predefined rules, and produces sample data for this indicator. (Table.1)

Table 1. Operation and maintenance evaluation indicators and quantitative rules

Dimensional metrics	Specific indicator name	Metric Definition	Quantitative rule (value ranges from 0 to 100, where higher scores indicate better operation performance)
Location indicator	Description of key areas in charging stations	Assess the importance of the charging station area (such as core business district, expressway service area, etc.)	100 points for key locations (e.g., downtown business districts, highway service areas) and 60 points for non-key locations
	Distance to local city center	Measure the geographical convenience of charging stations and local city centers	Distance ≤5km gets 100 points, 5-10km gets 80 points, 10-15km gets 60 points,>15km gets 40 points (distance calculated according to the spherical distance of the earth)
	Distance to target city center	Assess cross-regional service capabilities	Distance ≤50km gets 100 points, 50-100km gets 80 points, 100-200km gets 60 points,>200km gets 40 points (the distance is calculated by straight line distance)
Customer demand metrics	Customer repair request frequency	The frequency of customer reports for equipment failures during the statistical period	Urban charging stations: 40 points for m=0, 50 points for m=1, 60 points for 2≤m<4, 70 points for 4≤m<8, and 100 points for m≥8. High-speed charging stations: 40 points for n=0, 50 points for n=1, 60 points for n=2, and 100 points for n≥3 (where m/n represents the number of reported repairs in the previous year).
	Customer consultation and complaint indicators	The combined frequency of customer inquiries and complaints during the statistical period (complaints are converted to three times the number of regular work orders)	The work order calculation formula (consultation + suggestion + opinion + complaint × 3) is as follows: m=0 gets 100 points, 1≤m<3 gets 80 points, 3≤m<5 gets 60 points, and m≥5 gets 40 points.
	Surrounding service pressure indicator	Average workload pressure across all charging stations in the same administrative area/highway section	The average total number of work orders per shift is 100 points for m≥9, 80 points for 5≤m<7, 70 points for 4≤m<5, 60 points for 3≤m<4, 50 points for 2≤m<3, 40 points for 1≤m<2, and 30 points for 0≤m<1.
Metrics	Monthly average charging volume	The monthly average annual charging capacity of a single pile reflects the equipment's usage intensity.	100 points for monthly average charging ≥1500kWh, 80 points for 1000-1500kWh, 60 points for 500-1000kWh, and 40 points for <500kWh (20 points for <10kWh).
	Maximum charging rate for charging piles	The percentage of charging stations used simultaneously during the statistical period, reflecting equipment load	A score of 100 is awarded for p=100%,85 for 75%≤p<100%,70 for 50%≤p<75%,60 for 25%≤p<50%,40 for 0<p<25%, and 20 for p=0.
Device status analysis metrics	Health index of core components of charging pile	Assess the aging/failure risk of core components	The component temperature is normal (-10℃-40℃) and no maintenance in the past year will receive 100 points. Temperature exceeding 40℃ or 1-2 maintenance will receive 70 points. Temperature exceeding 60℃ or ≥3 maintenance will receive 40 points.
	Charging pile fault warning probability index	Predict the probability of functional failure of charging piles in the next 7 days	The higher the warning probability, the higher the score: 100 points for <10%,80 points for 10%-30%,60 points for 30%-50%, and 40 points for>50%.
	Charging network communication stability index	Check the communication quality between the charging pile and the backend system	0 times of monthly disconnection: 100 points, 1-2 times: 80 points, 3-5 times: 60 points,>5 times: 40 points; data transmission delay <1s: +20 points,>3s: -20 points

Step 4: Based on the service pressure index of a single station, the grid clustering algorithm is introduced to construct a grid-charging station mapping table to identify and quantify the regional service pressure. The specific steps are as follows:

(1) Grid division and mapping table construction

The target area is divided into a grid system with coordinates (latitude and longitude) using a 2km×2km grid cell size, ensuring no more than 5 charging stations per grid. GPS coordinates of all charging stations are collected and matched to corresponding grid cells to generate a grid-charging station mapping table. This table includes: grid ID, station name, number of charging piles, location metrics, last year's repair reports, work order counts, monthly average charging volume, and average charging pile utilization rate.

(2) Grid-based Cluster Analysis for Regional Service Pressure Identification: Using charging station characteristics such as average repair reports, average work order counts, monthly average charging volume, and average charging pile occupancy rate as clustering features, the algorithm identifies high-density grids when the number of stations with service pressure exceeding regional averages meets a predefined threshold. The system categorizes results into three pressure levels: high-pressure, medium-pressure, and low-pressure grids. The specific quantification method is as follows:

Let the feature value of the i-th charging station in the grid be x_{ij} , where j represents the average number of maintenance requests, average work order count, average monthly charging capacity, and average simultaneous utilization rate of charging piles respectively, and n is the total number of charging stations in the grid.The average feature value in the grid is $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$; the average feature value of all grids in the region is $\mu_j = \frac{1}{m}\sum_{k=1}^{m} \bar{x}_{kj}$, where m is the total number of grids in the region, and $\bar{x}_{kj}$ is the j-th feature average of the k-th grid; the standard deviation of all grid features in the region is $\sigma_j = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(\bar{x}_{kj} - \mu_j)^2}$; the standardized feature value is: $z_{kj} = \frac{\bar{x}_{kj} - \mu_j}{\sigma_j}$ .Based on the standardized feature values, calculate the comprehensive service pressure index of each grid using the weighted sum method. The weights are set according to the impact of each feature on service pressure. The comprehensive service pressure index of the k-th grid is $Q_k = \sum_{j=1}^{4} \omega_j * z_{kj}$

where a higher value indicates greater pressure. The weights of the average number ω_1、ω_2、ω_3、ω_4 of maintenance requests, average work order count, average monthly charging capacity, and average simultaneous utilization rate of charging piles are 0.3, 0.3, 0.2, and 0.2 respectively, summing to 1.

Set the density threshold d, and perform clustering division combined with the comprehensive service pressure index:

High-pressure grid: Determined when the grid's comprehensive service pressure index $Q_k \geq \mu_Q + 0.5\sigma_Q$ (where μ_Q is the average pressure index of all grids in the region, and σ_Q is the standard deviation) and the number of grids meeting this condition is ≥ density threshold d;

Low-pressure grid: Determined when the grid's comprehensive service pressure index $S_k \leq \mu_Q - 0.5\sigma_Q$ and the number of grids meeting this condition is ≥ density threshold d;

Medium-pressure grid: Grids with comprehensive service pressure index in the interval $\mu_Q - 0.5\sigma_Q \leq Q_k \leq \mu_Q + 0.5\sigma_Q$ are automatically classified as medium-pressure grids.

Step 5: Utilize the Analytic Hierarchy Process (AHP) to calculate initial weights for location indicators, customer demand indicators, operational indicators, and equipment status $W = \{W_1, W_2, \ldots, W_{11}\}$ $W = \{W_1, W_2, \ldots, W_{11}\}$ indicators. Employ the entropy method to determine secondary indicator weights, then integrate these to form a weight vector for 11 operation and maintenance (O&M) metrics. Subsequently, introduce polyhedral uncertainty theory to construct an error-robust model, thereby reducing data error interference in hierarchical evaluation. For O&M metrics, analyze historical data error ranges. Using the sum of actual values and error ranges as vertices, construct a polyhedral uncertainty set. Within this set, calculate the worst-case comprehensive O&M index S, which represents the value that minimizes index impact from errors.

$$S = \min_{X \in U} \sum_{i=1} (X_i \times W_i)$$

Here, S denotes the Robust Comprehensive Operation and Maintenance Index, U represents the polyhedral uncertainty set, Xi is the possible value of the i-th indicator, and Wi is the weight of the i-th indicator.

Based on the robust integrated operation index, the charging station classification logic is defined as follows: Level 1 charging station: S≥80; Level 2 charging station: 60≤S<80; Level 3 charging station: S<60.

Step 6: Develop operation and maintenance strategies based on the robustness index classification and grid pressure levels, and form an O&M strategy matrix. (Table 2)

Table 2. Charging station operation and maintenance control strategy

Robust operation and maintenance level	Grid service pressure level	Operation and maintenance cycle	Fault response time	Special policy
one-level	High Pressure Grid	Inspect every 3 days	≤ 1.5 hours	Daily Monitoring
	Medium Pressure Grid	Inspect every 4 days	≤ 2 hours	Weekly Trend Alert
	Low-pressure grid	Inspect every 5 days	≤ 2.5 hours	Automatically recognize video alerts
two stage	High Pressure Grid	Inspect every 5 days	≤ 2 hours	Dynamic promotion: If the status remains abnormal for three consecutive days, the system will temporarily upgrade to Level 1 response.
	Medium Pressure Grid	Check every 7 days	≤ 3 hours	Monthly health assessment
	Low-pressure grid	Inspect every 10 days	≤ 4 hours	Replacement parts after a failure
three-level	High Pressure Grid	Check every 7 days	≤ 3 hours	Force promotion
	Medium Pressure Grid	Inspections every 14 days	≤ 6 hours	Quarterly deep maintenance
	Low-pressure grid	Inspections every 20 days	≤ 8 hours	Customer repair request drive response

3. Practical Application Cases

3.1. Scenario Overview

This case study examines 60 charging stations in and around a provincial capital city in eastern China, covering six typical application scenarios: core commercial districts (9 stations), highway service areas (2 stations), transportation hubs (10 stations), large residential communities (12 stations), industrial parks (20 stations), and suburban towns (7 stations). The stations represent charging infrastructure operating under varying loads (high, medium, low). Equipment analysis includes chargers from five major manufacturers (A, B, C, D, E), with operational lifespans distributed as follows: 1-3 years (25 stations), 4-6 years (20 stations), and 7-8 years (15 stations). The aging patterns demonstrate a gradient distribution, aligning with real-world operational characteristics of large-scale charging networks. In 2024, the region had

450,000 new energy vehicles, with an average daily usage frequency of 8.7 per charger. High-traffic areas like commercial districts and highway service zones recorded peak usage of 16.3 daily chargers per station, accompanied by 8.2% equipment failure rates. In contrast, suburban towns showed only 2.1 daily chargers per station with 3.5% failure rates. These stark operational differences between scenarios provide a representative sample for validating tiered maintenance methodologies.

3.2. Data Collection and Preprocessing

3.2.1. Data Collection Scope

The project will collect multi-source data from January 1 to December 31,2024, forming a comprehensive dataset comprising: 1.2 million charging order records from the e-Charging platform (including transaction serial numbers, charging device IDs, charging timestamps, and charge volumes); 50 million operational parameters of charging devices (voltage, current, power, and millisecond-level waveform data).

Customer service text data: 12,000 pieces of user feedback text recorded by the 95598 customer service system, including problem description, occurrence time, charging station involved, processing result, etc.;

Environmental perception data: 8,760 pieces of environmental temperature and humidity data collected by sensors built into charging piles, combined with 32 extreme weather (high temperature, rainstorm, cold wave) records supplemented by meteorological department interface;

Equipment basic data: static information such as GPS coordinates, equipment model, manufacturer, and operation years of 120 charging stations, as well as 1,860 equipment maintenance records of the year (including maintenance parts, fault causes, and maintenance time).

3.2.2. Data Preprocessing Process

The study eliminated 32,000 abnormal data points using the 3σ rule, removing records with negative charge values or voltages/currents exceeding equipment specifications. Missing operational parameters (18,000 records) were filled using linear interpolation to ensure data integrity. Additionally, 95,598 customer service texts underwent tokenization, de-duplication, and synonym substitution processing, eliminating meaningless characters and duplicate entries. Ultimately, 11,200 valid texts were retained for analysis.

3.2.3. Grid Division and Mapping

The target area was divided into a 2km×2km grid using geographic coordinates, resulting in 28 grids. Among these, 25 contained valid charging stations while 3 were vacant (suburban areas without charging stations). A spatial matching algorithm was applied to assign 60 charging stations to their corresponding grids, generating a grid-to-charging station mapping table.

3.2.4. Grid Pressure Classification

Four core metrics were selected: average repair reports, average work order count, monthly average charging volume, and average charging station concurrency rate. After Z-score normalization, the comprehensive service pressure index for each grid was calculated using weighted summation (0.3, 0.3, 0.2, 0.2). With a density threshold of d=3, the grid classification results are as follows: -High-pressure grids: 7 grids (3 core business districts, 2 highway service areas, 2 transportation hubs). These grids exhibit significantly higher service pressure than others, with average repair reports at 4.8 per station, work order count at 10.2 per station, monthly average charging volume at 1680kWh per station, and concurrency rate at 82%.

Medium-voltage grids: 13 in total, mainly located in large residential communities (5), industrial parks (4), core business district peripheries (2), and highway service areas (2). The

stations recorded an average of 2.3 repair requests per station, 5.6 work orders per station, 1,050 kWh of monthly charging per charging pile, and an average 65% charging concurrency rate.

The low-pressure grid comprises 8 stations, primarily located in suburban towns (6) and industrial park peripheries (2). These stations exhibit an average of 1.1 repair requests per station, 2.3 work orders per station, 420 kWh/month charging capacity, and 38% average concurrency rate. The pressure classification results closely match real-world operational scenarios, with high-pressure grids concentrated in densely populated, high-traffic core areas, validating the effectiveness of the grid clustering methodology.

3.3. Grading Results and Verification

A combined weighting method integrating Analytic Hierarchy Process (AHP) and entropy method was employed to calculate the composite weights of 11 operation and maintenance indicators. The primary-level indicator weights (AHP method) were: Location Index (0.2), Customer Demand Index (0.3), Operational Index (0.25), and Equipment Status Analysis Index (0.25) (Consistency Ratio CR=0.06<0.1, meeting consistency requirements). The secondary-level indicator weights (entropy method) showed the highest weight for Customer Repair Reports (0.095) and Customer Consultation & Complaints (0.090), followed by Charging Station Key Area Description (0.052) and Core Component Health Index (0.065), with Distance to City Center being the lowest weight (0.040).

By applying polyhedral uncertainty theory, we calculated the historical data error ranges (±3% to ±8%) for each indicator to construct an uncertainty set. The robust comprehensive operation index S was then computed under worst-case scenarios. Following the classification criteria: Level 1 (S≥80), Level 2 (60≤S<80), and Level 3 (S<60), the results identified 16 Level 1 charging stations (26.7% of total stations). These include 4 core commercial districts, 3 highway service areas, 3 transportation hubs, 2 large residential communities, 2 industrial parks, and 1 suburban township. All stations are classified as high-load, high-importance, and high-failure-risk facilities.

Secondary charging stations: 29 (48.3%), distributed in various scenarios, including 7 large communities, 6 industrial parks, 3 core business districts, 2 expressway service areas, 2 transportation hubs, and 9 suburban towns, with medium level of operation and service demand;

Three-level charging stations: 15 (25.0%), mainly 6 in suburban towns, 2 in industrial parks, 2 in large communities, 1 in core business districts, 2 in expressway service areas, and 2 in transportation hubs, most of which are low-load and low-importance stations.

A charging station in a core business district (equipment from Manufacturer A, 5 years of operation) previously recorded 14.8 daily uses per charging pile, 12 annual maintenance requests, 5.3-hour fault response times, and a 76% equipment integrity rate. After implementing a tiered maintenance strategy featuring "three-day inspections, 1.5-hour fault response, and daily monitoring," the station proactively resolved 18 charging module failures and 12 interface issues throughout the year. This reduced fault response time to 1.2 hours and decreased customer complaints from 120 to 35. In suburban charging stations (Level 3 ,low-pressure grid), pre-implementation data showed 2.3 daily uses per pile, 1 annual maintenance request, and 16.5-hour response times. The new tiered maintenance approach-implementing biweekly inspections, 8-hour fault response, and customer-driven repair protocols-has significantly reduced unnecessary inspection frequency.

4. Conclusion

This invention establishes a scientific classification and operation system for charging stations through multi-source data integration, semantic text mining, grid clustering partitioning, and uncertainty modeling. It achieves a transformation from "passive response" to "proactive prediction" and from "single-point management" to "regional coordination" in operations and maintenance. Practical applications demonstrate that this approach effectively enhances operational efficiency, optimizes resource allocation, and provides technical support for high-quality operation of new energy charging networks, demonstrating broad application and promotion value.

Acknowledgement

This work was supported by the State Grid Science and Technology Project “Key Technologies Research and Application of Intelligent Operation & Maintenance and Auxiliary Decision-Making for Electric-Vehicle Charging Equipment” (Project No. 5400-202371588A-3-2-ZN).

References

[1] Cui, D., Wang, Z., Liu, P., Wang, S., Dorrell, D. G., Li, X., & Zhan, W. (2023). Operation optimization approaches of electric vehicle battery swapping and charging station: A literature review. Energy, 263, 126095. https://doi.org/10.1016/j.energy.2022.126095

[2] Fang, H., Liao, J., Huang, S., & Zhang, M. (2024). Research on status assessment and operation and maintenance of electric vehicle DC charging stations based on XGboost. Smart Cities, 7(6), 3055–3070. https://doi.org/10.3390/smartcities7060119

[3] Ghanbari Motlagh, S., & Li, L. (2026). A review on electric vehicle charging station planning: Infrastructure placement, sizing, upgrades, and uncertainties. Journal of Energy Storage, 141, 119325. https://doi.org/10.1016/j.est.2025.119325

[4] Singh, P. P., Wen, F., Palu, I., Sachan, S., & Deb, S. (2022). Electric vehicles charging infrastructure demand and deployment: Challenges and solutions. Energies, 16(1), 7. https://doi.org/10.3390/en16010007

[5] Ghanbari Motlagh, S., & Li, L. (2026). A review on electric vehicle charging station planning: Infrastructure placement, sizing, upgrades, and uncertainties. Journal of Energy Storage, 141, 119325. https://doi.org/10.1016/j.est.2025.119325

[6] Hai, X., Yu, Q., Bai, O., Wu, S., Zheng, W., & Yuan, S. (2021). Operation and maintenance system of electric vehicles’ charging and discharging facilities based on repository. In 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 894–897). https://doi.org/10.1109/ICCASIT53235.2021.9633471

[7] Yu, H., Li, L., & Wen, X. (2023). A two-layer optimization method of station area-charging facilities for orderly charging distributed energy storages. In 2023 International Conference on Power System Technology (PowerCon) (pp. 1–6).