International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 5, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n5.04
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on the Layout Algorithm of Automobile Charging Piles Based on Density Clustering

Chao Zhang [1], Yixin Zuo [2], Ru Tian [1, *], Fan Zhang [1]

[1] China Auto Information Technology (Tianjin) Co., Ltd., Tianjin, China
[2] CATARC Huacheng Certification (Tianjin) Co., Ltd. Tianjin, China
*Corresponding Author: tianru@catarc.ac.cn

ABSTRACT

To address major industry challenges in China's urban electric vehicle charging infrastructure, including unbalanced spatial layout of electric vehicle charging piles, mismatch between supply and demand, strong subjectivity in site selection, and limited distribution resources in older urban districts, this paper proposes a two-layer layout planning algorithm for charging piles based on improved DBSCAN density clustering + dual-objective constrained optimization. First, multi-source data, including urban POI locations, residents' travel GPS stay-point data, road network data, and distribution network capacity data are integrated to quantify the spatial distribution of charging demand, and a mathematical model for charging-demand density is constructed; Second, the density-based clustering process is improved, and the centroids of the identified clusters are used as candidate locations for EV charging facilities; Finally, with the optimization objective of minimizing total construction cost and minimizing the average charging detour distance for users, a mixed-integer programming mathematical constraint model is established to optimize the number of chargers allocated to each candidate site. The proposed method verifies the scientific validity and practical applicability of the improved density-clustering algorithm in EV charger layout planning.

KEYWORDS

Electric vehicles; EV charging station layout; DBSCAN; Density clustering; Dual-objective optimization

1. INTRODUCTION

The State Council has issued the "Guiding Opinions on Further Building a High-Quality Charging Infrastructure System (2023)", which explicitly called for improving the urban public charging infrastructure network system, coordinating and optimizing the spatial siting and layout of EV charging piles, and setting the construction control target that the service radius of public charging facilities in the central urban area should not exceed 1km. By the end of 2025, the number of new energy vehicles in China had exceeded 35 million. The rapid expansion of the industry has created an urgent need for accelerated deployment of supporting charging infrastructure. However, structural imbalances between supply and demand in the industry have become increasingly prominent. In urban commercial districts and large contiguous residential communities, charging-station parking spaces are often in short supply during peak hours, making queues for charging a regular occurrence. Meanwhile, suburban areas and older built-up districts face insufficient charging points and gaps in facility coverage. At present, most charging stations still follow the manual experiential site selection model. Their planning processes often lack quantitative constraints involving multiple factors such as the spatial distribution of charging demand, the carrying capacity of local distribution networks, and the cost of construction land. As a result, some charging stations operate under excessive load, while others remain underutilized for long periods after construction, leading to resource misallocation. Although the current national planning specifications require the layout of charging piles should be based on fundamental factors such as population distribution, urban road networks, and public parking resources, a standardized quantitative mathematical model for site selection has not yet been established, and the conventional extensive site selection approaches can no longer meet the irequirements of refined and scientific planning in the era of big data. Density clustering algorithms can autonomously identify demand agglomeration areas according to the spatial density characteristics of charging demand, without requiring the number of stations to be predetermined. This feature is consistent with the planning principle of "location based on demand and pile placement based on demand" for charging facilities, and have now become an important research approach in the field of intelligent location of charging piles.

At present, with regard to the siting of charging facilities, uropean and American researchers were among the first to use vehicle GPS trajectory data for charging-facility location planning. Some studies adopted DBSCAN+K-Means combined clustering to dig vehicle dwell hotspots, and using hotspot centers as candidate locations for fast charging stations. However, these studies did not incorporate constraints related to power grids or construction costs. The EV-Planner algorithm formulates charging-station siting as a multi-objective optimization problem and uses improved clustering for optimized site selection. Nevertheless, its parameters rely on stochastic search, resulting in relatively low convergence efficiency. Foreign studies mainly focus on fast-charging station deployment along highway networks, lacking domestic models for adapting to complex land use scenarios in old urban areas and residential communities. Domestic scholars have implemented the location of charging stations based on density peak clustering, but often consider only a single dimension, such as traffic flow, while neglecting land-use and distribution-network constraints. Some studies have used K-Means clustering for site selection, but K-Means the number of stations to be specified in advance, making it unsuitable for planning scenarios with unknown demand distributions. Existing DBSCAN-based layout studies generally have two major limitations: manual parameter assignment and the lack of volumetric optimization models. These limitations constitute the main research gap addressed in this paper.

Therefore, this paper proposes a charging pile layout algorithm based on density clustering., which can improve the theoretical system of charging framework for charging-facility siting based on spatial density characteristics, and can also establish a two-tier mathematical planning model that integrates preliminary site selection through demand-density clustering with capacity sizing through multi-objective optimization. It enables integrated quantitative modeling of both site selection and capacity sizing, thereby enriching the mathematical methods used in electric vehicle infrastructure planning. The charging pile layout scheme can help municipal planning departments and charging operators deploy charging piles scientifically based on urban travel data, reduce construction investment, improve the convenience of charging for vehicle owners, optimize distribution-grid load allocation, and contribute to the intensive construction of urban charging networks.

2. RESEARCH ROUTE

This paper investigates EV charging piles layout planning by following a research route consisting of data processing, density-based algorithm optimization, siting model optimization, and summary and analysis. First, multi-source basic data, including population distribution, road networks, parking locations, and vehicle stay records, are integrated. Data cleaning and normalization are then performed, and a charging-demand density quantification model under a spatial grid framework is constructed to provide a refined quantitative representation of regional charging demand.; Second, to address the limitation of manually selected parameters in the traditional DBSCAN algorithm, the silhouette coefficient and the elbow method are introduced to achieve adaptive optimization of the two parameters, ε and MinPts. Based on the improved density-clustering algorithm, charging-demand aggregation hotspots are identified, and candidate sites for EV charging facilities are selected. Then, with the optimization objectives of the lowest construction cost, minimizing users' charging detour distance, and balanced distribution network load, a multi-objective integer programming model is constructed to complete the capacity optimization of charging piles at candidate sites. Finally, case-based simulation and comparative analysis across multiple algorithms are conducted to verify the effectiveness of the proposed layout model and improved algorithm, and the research conclusions and subsequent optimization directions are summarized. The research roadmap is shown in Figure 1.

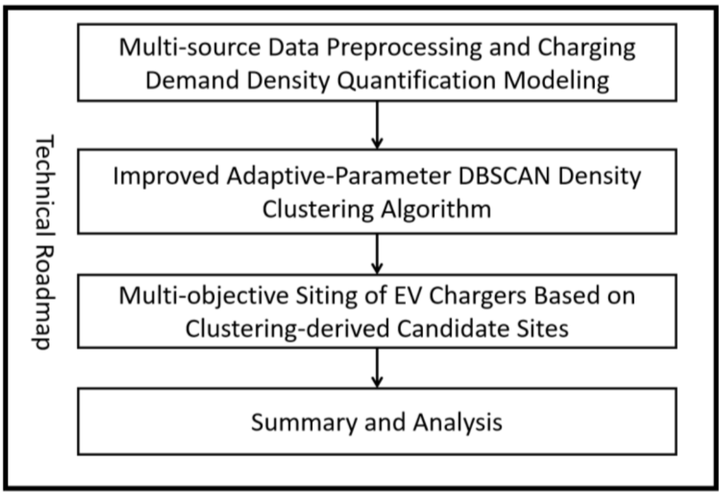

Figure 1. Technical Roadmap

2.1. Factors Influencing Charging Pile Layout and Data Preprocessing

2.1.1. Factors Influencing the layout of Charging Piles

(1) Demand-side factors

Regional charging demand is constrained by three key factors: population distribution, vehicle parking patterns, and the penetration rate of new energy vehicles. First, the population size of residential communities determines the baic scale of slow-charging demand in the area. Second, the average daily parking duration and parking frequency of vehicles in shopping malls, office buildings, and public parking lots directly affect the degree of concentration of fast charging demand; Third, the regional ownership rate of new energy vehicles, denoted as μ, refers to the ratio of new energy vehicles to the total number of motor vehicles within the jurisdiction, and determines the upper limit of effective charging demand in the region.

(2) Supply-side constraints

In the process of candidate sites for charging piles, three types of objective constraints should be considered: land availability, distribution network capacity, and construction cost. First, at the land-use level, ompliant public parking lots should be prioritized, while areas unsuitable for construction, such as low-lying flood-prone areas, areas near flammable, explosive hazard sources, and zones subject to school or cultural relic protection controls, should be excluded. Second, at the distribution-network level, the rated capacity of the distribution transformer at candidate site i, denoted as Smax, i, constrains the maximum accessible power of the station and directly determines the maximum number of charging piles that can be deployed at a single site. Third, at the economic-cost level, site rental fees, distribution engineering renovation expenses, charging pile equipment purchase and annual operation and maintenance expenses jointly constitute the full life-cycle cost of station construction, thereby constraining the construction scale of each station under the investment ceiling.

(3) Service constraints

To achieve balanced deployment and efficient utilization of public charging facilities, two spatial constraints should be followed. First, the service radius R of a single station should not exceed 1 km, so as to ensure convenient access for users to charging stations. Second, the distance D between stations within the same cluster should be no less than 0.5km, and the proportion of overlapping service areas should be strictly controlled. These two thresholds are promote the formation of a charging-station cluster pattern characterized by moderate competition and limited overlapping coverage.

2.1.2. Data Preprocessing

Data preprocessing transforms raw, disordered, and potentially non-compliant charging-station data into standardized datasets that meet spatial constraints and can be directly used for site planning and resource allocation.

(1) Outlier processing

To ensure the accuracy of subsequent spatial planning analysis, systematic outlier identification and processing were conducted on the original charging-station data. First, invalid records, such as missing longitude and latitude, coordinates drifting into areas unsuitable for station construction, and records with zero charging guns, were removed. Second, stations with insufficient coordinate accuracy or duplicate spatial locations were merged. Then, the distance between stations was calculated, and records with inter-station distances below the specified threshold were marked as anomalous data. Meanwhile, range checks are performed on the operational parameters, and values exceeding reasonable thresholds were marked as suspicious and submitted for manual review. Finally, an outlier-processing log was generated to record the type, the basis for judgment, and the processing result to ensure that the data ensuring that the data-cleaning process is traceable and reviewable.

(2) Missing value handling

To retain as many valid sites as possible while ensuring data usability, a hierarchical missing value handling strategy was applied to the original charging-station data First, longitude and latitude were used as key fields for spatial computation, records with missing coordinate values were directly removed and logged without imputation. Second, the service radius field was uniformly filled according to planning standards to ensure that all sites meet the basic assumptions of coverage constraints. For operational parameters involved in the comprehensive evaluation, such as rated power and the number of charging guns, missing values were imputed using the median values of stations in the same administrative region and of the same type, with an “imputed” label added. For highly uncertain fields, such as the operation commencement date, missing values were retained to avoid errors introduced by inappropriate imputation. All imputation operations were accompanied by corresponding tag fields in the output dataset, and a missing value handling log was generated to ensure that the data-cleaning process is transparent and reviewable.

2.2. Quantified Mathematical Model of Charging Demand

2.2.1. Quantitative Mathematical Model of Charging Demand Density

Construct a two-dimensional planar coordinate system [1] based on the longitude and latitude coordinates of the planning area, divide the urban area into M×N uniform grid cells. assume the grid cells are G_i^j, the grid center coordinates are (x_i^j, y_i^j), and quantify the average daily charging demand of the grid as the original input feature data D_i^j for density clustering. The calculation formula of D_i^j is shown in Eq. (1).

$$D_i^j = N_i^j \cdot \mu \cdot \frac{T_{avg}}{T_{one}} \tag{1}$$

In Eq. (1), D_i^j denotes the average daily effective charging demand of grid G_i^j. N_i^j denotes the average daily number of parked motor vehicles within the grid. μ denotes the proportion of new energy vehicles in the region. T_{avg} denotes the average daily parking duration of vehicles, and T_{one} denotes the time required for a single vehicle to complete one full charging process.

The spatial density of charging demand in each grid is calculated as shown in Eq. (2).

$$\rho_i^j = \frac{D_i^j}{S_g} \tag{2}$$

In Eq. (2), ρ_i^j denotes the charging demand intensity per unit area. A larger value indicates a higher priority for charger deployment in the corresponding area. S_g denotes the geographical area of a single grid cell.

The original dataset for clustering is constructed, with each sample containing three-dimensional features of geographic coordinates and demand density, which is $X = \{(x_i^j, y_i^i, \rho_i^j) \mid \mathrm{i} = 1,2...,\mathrm{m}, \mathrm{j} = 1,2...,\mathrm{n}\}$ based on the input data source of density clustering.

2.2.2. Preliminary Selection Rules for Candidate Stations

DBSCAN clustering generates demand cluster sets, $C = \{C_1, C_2, \cdots, C_n\}$ a single cluster, where the geometric center of the cluster is the $C_k = \{p_1, p_2 \cdots, p_m\}$ coordinate of the candidate site, as shown in Eq. (3). (X_k, Y_k) (X_k, Y_k)

$$X_k = \frac{\sum_{t=1}^{m} x_t}{m}, Y_k = \frac{\sum_{t=1}^{m} y_t}{m} \tag{3}$$

In Eq. (3), m denotes the number of effective demand grids within the cluster. Noise points identified during clustering are not treated as independent station-construction sites. Instead, they are assigned to nearby candidate stations in adjacent clusters to provide unified charging services.

2.3. Improve Adaptive DBSCAN Density Clustering

2.3.1. Density-based Clustering

DBSCAN is a representative density-based clustering algorithm. Its core idea is to identify high-density regions separated by low-density regions [2, 3]. In other words, in the data space, points in a sufficiently dense region are assigned to the same cluster, while sparse regions serve as boundaries between clusters or are identified as noise.

DBSCAN involves several key concepts [1-4]:

(1) $\varepsilon-$Neighborhood: The neighborhood of a point refers to a circular region centered at that point with radius ε.

(2) Core point: A point is defined as a core point if its $\varepsilon-$ neighborhood contains at least MinPts points. MinPts is a predefined density threshold.

(3) Boundary point: A point is not a core point (the number of points in its neighborhood <MinPts), but it is located in the neighborhood of a core point. $\varepsilon-$

(4) Noise point: A point that is neither a core point nor a boundary point.

(5) Density direct: If point q lies within the $\varepsilon-$ neighborhood of core point p, then q is directly density-reachable from p.

(6) Density connection: If there exists a core point o such that both points p and q are density-reachable from o, then p and q are density-connected. In DBSCAN, each cluster is the maximal set of density-connected points.

2.3.2. Improving Adaptive Parameters for the DBSCAN Model

(1) Data normalization

To eliminate the influence of different dimensions among the three feature parameters x, y, and ρ, min–max normalization is applied. The calculation formula is shown in Eq. (4).

$$x_s = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

In Eq. (4), x_s denotes the normalized value, while $x_{\max}$ and $x_{\min}$ denote the maximum and minimum values before normalization, respectively.

(2) Determination of the optimal ε

The k-distance method is used to traverse candidate MinPts values and determine the corresponding optimal ε.

(3) Calculation of the silhouette coefficient

For each combination of MinPts and ε, DBSCAN clustering is performed, and the corresponding silhouette coefficient, denoted as SC, is calculated.

(4) Selection of optimal parameters

The parameter combination (MinPts, ε_{opt}) corresponding to the maximum SC value is selected as the optimal parameter set of the algorithm.

(5) Density clustering

Perform density clustering at optimal parameters and output the coordinates of each demand cluster center (candidate station locations).

2.4. Multi-objective Siting Model Based on Clustering Candidate Points

Based on the candidate station coordinates (K as the total number of clusters) as the optimization variable, a $\{(X_i, Y_i) \mid i = 1, 2, \cdots, k\}$ dual-objective mixed integer programming model is constructed, where the decision variable $x_k \in \{0,1\}$ indicates that 0 means that no station is built at candidate site k, while 1 means that a station is built at candidate site k. The variable n_k denotes the number of EV charging piles configured at site k.

(1) Objective function 1: Minimization of total construction cost

$$F_1 = \min \sum_{i=1}^{K} x_i (C_f + n_i \cdot C_u + C_{elec}^i) \tag{5}$$

In Eq. 5, F_1 denotes the total construction cost of EV charging facilities in the region; C_f denotes the fixed land and civil engineering cost of a single station; C_u denotes the average annual cost of procurement, operation, and maintenance of a single EV charger; and C_{elec}^k denotes the distribution-network renovation cost at site k.

(2) Objective function 2: Minimization of users' average charging detour distance

$$F_2 = \min \frac{\sum_{i=1}^{k} \sum_{\rho \in C_i} \rho_p \cdot x_i \cdot dist(p, (X_i, Y_i))}{\sum_{p \in X_{valid}} \rho_p} \tag{6}$$

In Eq. (6), ρ_p denotes the demand density of grid p. The demand-density-weighted average distance ensures that areas with higher charging demand are given higher priority in reducing detour distance.

(3) Normalized weighted solution of the dual-objective model

A linear weighting method is used to transform the multi-objective optimization problem into a single-objective optimization problem. The objective normalization function is shown in Eq. (7).

$$f_l = \frac{F_l - F_l^{\min}}{F_l^{\max} - F_l^{\min}}, l = 1,2 \tag{7}$$

The comprehensive objective formula is as shown in Eq. (8).

$$F = \min(w_1 f_1 + w_2 f_2), w_1 + w_2 = 1 \tag{8}$$

In this paper, the integer Particle Swarm optimization (PSO) algorithm is used to solve for x_i the optimal sum n_i.

3. RESEARCH CONCLUSIONS

This paper constructs a two-layer charging pile layout mathematical model for EV charging piles layout planning, integrating charging demand density quantification, adaptive DBSCAN-based preliminary site selection, and multi-objective mixed-integer optimization for charger capacity sizing. First, the proposed model enables refined station siting based on spatial charging demand density, overcoming the inherent limitations of traditional manual site selection, fixed parameter DBSCAN, and K-Means clustering; Second, the improved DBSCAN achieves dual-parameter adaptive optimization through contour coefficients + elbow rule, with precise division of demand clusters and candidate stations in line with actual charging hotspots; Third, compared with traditional algorithms, the proposed method achieves comprehensive optimization in three core indicators: construction cost, user detour distance, and distribution-grid load balance. It also satisfies the multiple constraints, including the national standard charging service radius, land-use requirements, and distribution-network capacity, demonstrating strong feasibility for practical implementation.

In the simulation stage, only static daily average demand data was used, while dynamic charging demand during morning and evening rush hours was not introduced. In addition, the differentiated layout of fast-charging and slow-charging facilities was not further refined. Future research can be extended in three directions. First, time-series GPS data from ride-hailing vehicles and taxis can be integrated to construct a spatio-temporal dynamic demand-density model and enable differentiated layout planning across different time periods. Second, photovoltaic energy storage and V2G vehicle-grid interaction constraints can be incorporated into the optimization model to support the layout optimization of integrated photovoltaic-storage-charging stations. Third, GIS technology can be integrated to develop visualized intelligent planning software for EV charger deployment.

REFERENCES

[1] Xu, Z., Zhao, M., Zhang, X., et al. (2026). Electric vehicle charging station layout method based on discrete whale algorithm. Power Electronics Technology, 60(1), 133–138.

[2] Wang, X., & Peng, Y. (2026). ICESat-2 photon data denoising algorithm for two-level DBSCAN. Beijing Surveying and Mapping, 40(5), 594–599.

[3] Zhang, C., Li, X., & Wang, T. (2026). Passenger cabin comfort control optimization based on DBSCAN clustering algorithm. SAIC Motor, (5), 39–43.

[4] Liu, X., Zhu, Z., Zang, Q., et al. (2026). Analysis of ride-hailing demand based on DBSCAN algorithm. Highway and Road Transport, 42(2), 15–20.