applied sciences

MDPI

Article

Dual-Driven Electricity Package Recommendation via Cognitive-Noise Filtration and Consensus Evolution

Xiaofan Ding, Jinhuai Chen and Yuanqian Ma *

School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2024326690034@mails.zstu.edu.cn (X.D.); 2024329301036@mails.zstu.edu.cn (J.C.)
* Correspondence: mayq666@zstu.edu.cn; Tel.: +86-134-380-26258

Abstract

The growing variety of customized tariffs in deregulated electricity retail markets increases customers' decision burden under information asymmetry. Existing recommendation methods generally characterize either objective load patterns or subjective preferences, but rarely account for load-risk heterogeneity and uncertain evaluations within a unified framework. This study proposes a dual-driven electricity package recommendation framework that combines cognitive-noise filtration with dynamic consensus evolution. A risk-driven two-stage affinity propagation (RD-TAP) algorithm first separates customers according to both load-shape similarity and fluctuation risk. Interval-valued intuitionistic fuzzy evaluations are then calibrated using individual-to-subgroup consensus deviations, after which a multi-threshold weighted-fusion mechanism iteratively improves subgroup consensus while retaining moderate individual differences. In a case study based on real-world load data, RD-TAP maintains a clustering quality index above 0.94 for the tested sample sizes, and the complete recommendation framework yields mean RMSE values ranging from 0.1020 to 0.1346 across the four load patterns. These results indicate that jointly modeling objective load risk and subjective evaluation uncertainty can improve the stability and interpretability of electricity package recommendations.

Keywords: electricity retail package; IVIFS; cognitive noise filtration; consensus degree optimization; RD-TAP clustering

1. Introduction

Electricity retail packages have evolved from single-tariff products into combinations of time-of-use pricing [1], green electricity [2], and demand-response services [3] as electricity market reform has progressed [4,5]. Although this diversity provides customers with more options, it also increases the information required to compare package terms [6], while socioeconomic differences may further affect customers' switching behavior in competitive retail markets [7]. Customers who lack domain knowledge may therefore struggle to determine whether a package is compatible with their load characteristics and risk exposure, resulting in choice overload. Reliable package recommendation and personalized retail pricing are consequently important for improving customer choice and retailers' service effectiveness [8,9].

Customer segmentation based on smart-meter data provides an objective foundation for personalized electricity services. Existing studies have combined density-based and hierarchical clustering to segment residential customers [10], developed distributed auto-clustering methods for large-scale advanced metering infrastructure data [11], and

Check for updates

Academic Editor: Alessandro Lo Schiavo

Received: 23 August 2026
Revised: 14 September 2026
Accepted: 21 September 2026
Published: 23 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

compared shape-defining features for load-profile characterization [12]. Earlier studies have also systematically evaluated load-pattern clustering methods [13,14], while recent research has incorporated resident behavior segmentation [15], adaptive customer clustering [16], and self-adapting semi-supervised clustering [17]. Multi-resolution representations have further been used to extract load patterns at different temporal scales [18]. Nevertheless, these approaches mainly emphasize load-shape or behavioral similarity and rarely incorporate customer-specific fluctuation risk directly into pairwise similarity measurement.

Electricity retail package recommendation has mainly relied on collaborative filtering and retail-tariff optimization. Social-information filtering [19], Bayesian hybrid collaborative filtering [20], household usage-pattern filtering [21], and similarity-based collaborative filtering [22] have been applied to personalized plan generation and sparse-data mitigation. More recently, multigranular hesitant fuzzy information has been combined with an improved non-negative latent factor model for electricity retail plan recommendation [23]. Related studies have addressed customized retail pricing [24], retailer participation in demand-response markets [25], green electricity product-menu design under incomplete consumer information [26], and the effects of social learning on tariff choices and electricity package design [27,28]. Pareto optimization, multi-criteria assessment, and differentiated consumer-response modeling have also been used to examine retail-rate and tariff design [29–31]. However, these studies generally do not establish an integrated connection between customers' objective fluctuation risks and their uncertain subjective package evaluations.

Under incomplete preference information, multi-attribute group decision-making methods provide tools for representing bounded rationality and uncertain evaluations [32,33]. Interval-valued intuitionistic fuzzy sets and their aggregation operators can simultaneously describe membership, non-membership, and hesitation in group evaluations [34,35]. Nevertheless, most existing IVIFS-based methods emphasize uncertainty representation, attribute aggregation, or alternative ranking, with limited attention to identifying and correcting individually unconfident evaluations. Meanwhile, recent consensus-reaching studies have developed two-stage subgroup adjustment processes [36] and minimum-adjustment feedback mechanisms [37]. Although these methods improve overall group agreement, they do not jointly coordinate cognitive-noise correction, subgroup consensus improvement, and the retention of acceptable customer preference heterogeneity in electricity retail package recommendation.

To address these limitations, this paper develops an electricity retail package recommendation framework that integrates risk-driven customer profiling, IVIFS-based uncertain-evaluation correction, and dynamic consensus optimization. First, the proposed RD-TAP method incorporates pairwise load fluctuation risk differences into customer similarity measurement, enabling customers with similar average load shapes but different volatility characteristics to be distinguished. Second, individual-to-subgroup consensus deviation is used to detect and adaptively correct unconfident evaluations represented by IVIFSs. Third, a multi-threshold piecewise adjustment mechanism improves subgroup consensus while retaining acceptable preference heterogeneity. Unlike existing clustering-only or decision-only approaches, the proposed framework establishes an interpretable connection between objective electricity-consumption characteristics and subjective package evaluations and produces a complete personalized package ranking.

2. Risk-Driven Customer Profiling and Progressive Clustering Strategy

2.1. Multi-Dimensional Customer Profiling Incorporating Load Fluctuation Risk

To relate customer behavior to electricity retail packages, we construct a multidimensional profile that includes four dimensions: basic load characteristics, period-specific consumption preferences, load fluctuation risk, and consumption efficiency. Let the customer

set $U = \{u_1, u_2, \ldots, u_i\}$ and its discrete load profiles $H = \{h_{i,t} \mid i = 1, 2, \ldots, N; t = 1, 2, \ldots, T\}$ be defined as shown below; the corresponding attributes are extracted for each customer. Here, $h_{i,t}$ denotes the load measurement of customer u_i at the t-th sampling interval, with the index t following the chronological order of the measurements.

The load-fluctuation-risk dimension complements conventional indicators of average load level and stability by quantifying the temporal variability of individual consumption. It therefore supplies information needed to distinguish customers with similar mean profiles but different risk exposure. Table 1 lists the profile labels, physical interpretations, and calculation formulas.

Table 1. Multi-Dimensional Customer Profiling Index System Incorporating Load Fluctuation Risk.

Dimension	Profile Label	Physical Interpretation	Calculation Formula
basic load characteristics	Monthly load factor	Monthly load fluctuation stability	$R_i^{\mathrm{ML}} = \frac{\overline{H}_i}{H_{i,\max}}$
	Monthly maximum utilization hours	Monthly load time utilization efficiency	$M_i^{MU} = \frac{E_i}{H_{i,\max}}$
	Average peak load factor	Load stability during peak hours	$R_i^{PL} = \frac{1}{D}\sum_{d=1}^{D}\frac{H_{i,d}^{peak}}{\overline{H}_{i,d}}$
	Average shoulder load factor	Load stability during shoulder hours	$R_i^{FL} = \frac{1}{D}\sum_{d=1}^{D}\frac{H_{i,d}^{shoulder}}{\overline{H}_{i,d}}$
	Average valley load factor	Load stability during valley hours	$R_i^{VL} = \frac{1}{D}\sum_{d=1}^{D}\frac{H_{i,d}^{\mathrm{valley}}}{\overline{H}_{i,d}}$
	Working day peak-to-valley ratio	Load consistency on Working day	$R_i^{WD} = \frac{1}{D_{wd}}\sum_{d=1}^{D_{wd}}\frac{H_{i,d}^{\max} - H_{i,d}^{\min}}{H_{i,d}^{\max}}$
	Weekend/Holiday peak-to-valley ratio	Load consistency on weekends/holidays	$R_i^{WE} = \frac{1}{D_{we}}\sum_{d=1}^{D_{we}}\frac{H_{i,d}^{\max} - H_{i,d}^{\min}}{H_{i,d}^{\max}}$
load fluctuation risk	Monthly coefficient of variation	Long-term consumption fluctuation range	$V_i^{ML} = \frac{\sigma(b_i)}{\mu(b_i)}$
consumption efficiency	Load efficiency factor	Effective load utilization efficiency	$U_i^{LE} = \frac{T_{\mathrm{eff}}}{T_{\mathrm{total}}}$
period consumption preferences	Peak consumption share	Preference for peak-hour electricity usage	$\theta_i^{\mathrm{peak}} = \frac{E_i^{\mathrm{peak}}}{E_i^{\mathrm{total}}}$
	Valley consumption share	Preference for valley-hour electricity usage	$\theta_i^{\mathrm{valley}} = \frac{E_i^{\mathrm{valley}}}{E_i^{\mathrm{total}}}$
	Working day load deviation	Stability of daily load patterns on weekdays	$D_i^{WD} = \frac{1}{D_{wd}}\sum_{d=1}^{D_{wd}}\frac{\lvert H_{i,d} - \overline{H}_{i,d}\rvert}{\overline{H}_{i,d}}$

Note: *Hi* represents the load value of customer *i*, while $\overline{H}_i$ and Hi, max denote the average load and the peak load, respectively; Ei is the total electricity of customer *i* and D is the total number of days; Dwd and Dwe represent the number of working days and weekend/holidays, respectively; *Hi,d* denotes the average daily load of customer *i* on day *d*, while $H_{i,d}^{\mathrm{peak}}$, $H_{i,d}^{shoulder}$ and $H_{i,d}^{\mathrm{valley}}$ represent the peak, shoulder and valley loads respectively on day *d*, respectively; $H_{i,d}^{\max}$ and $H_{i,d}^{\min}$ denote the maximum and minimum daily loads on day *d*, respectively; b_i is the portrait vector of customer *i*; $\sigma(\cdot)$ and $\mu(\cdot)$ represent the standard deviation operator and the mean operator, respectively; T_{eff} and T_{total} denote the effective duration of the load and the total duration of the metering period, respectively; and E_i^{peak}, E_i^{valkey} and E_i^{total} represent the peak electricity consumption, valley electricity consumption and total electricity consumption, respectively. Detailed parameter derivations and conventional definitions are provided in Appendix A.

2.2. Progressive Customer Clustering via Risk-Driven Affinity Propagation (RD-TAP)

Clustering based only on load shape can merge customers whose average profiles are similar but whose fluctuation risks differ, which may impair subsequent package matching. We therefore introduce the pairwise difference in normalized load-fluctuation risk as a penalty term in the similarity measure and propose RD-TAP. Its two-stage structure first compresses the sample space and then performs risk-aware global clustering, allowing baseline consumption patterns and dynamic risk characteristics to be considered jointly.

2.2.1. Formulation of the Normalized Load Fluctuation Risk Index

To explicitly capture the temporal volatility of individual power consumption, we introduce a normalized risk coefficient ζ_i, derived from the statistical dispersion of the chronologically ordered load measurements defined in Section 2.1. This indicator provides a quantitative basis for the subsequent risk-aware clustering and electricity retail package adaptation. It is formulated as:

$$\zeta_i = \frac{\sigma_i}{\sigma_{\max}}, \sigma_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(h_{i,t} - \overline{h}_i)^2}, \overline{h}_i = \frac{1}{T}\sum_{t=1}^{T} h_{i,t} \quad (1)$$

where $\overline{h}i$ denotes the mean of the preprocessed chronological load measurements of customer u_i, and σ_i represents the corresponding temporal standard deviation, directly reflecting the magnitude of load fluctuations. Furthermore, $\sigma_{\max}$ is the maximum temporal standard deviation observed among all customers. Therefore, the risk coefficient ζ_i is bounded within ([0, 1]), providing a standardized measure for comparing load fluctuation risks across customers.

2.2.2. Two-Stage Progressive Architecture and Risk-Modified Similarity Mechanism

To overcome the computational bottleneck of the standard affinity propagation (AP) algorithm on massive datasets, we construct a two-stage progressive clustering architecture. In the initial local dimensionality reduction stage, the full sample set is randomly partitioned into scale-constrained subsets. Standard AP clustering, driven by the negative Pearson correlation coefficient, is executed independently on each subset to extract a representative cluster centroid ensemble C*.

To incorporate load-fluctuation-risk heterogeneity into the global clustering decision, we construct a risk-modified similarity matrix for the second-stage macro-clustering. For any two candidate centroids c_p and c_q within C*, their pairwise risk difference is defined by the absolute difference between their normalized risk coefficients. The modified similarity $s^{(2)}(c_p, c_q)$ is formulated as:

$$s^{(2)}(c_p, c_q) = s^{(1)}(c_p, c_q) - \kappa\left|\xi_p - \xi_q\right| \quad (2)$$

where $s^{(1)}(c_p, c_q)$ is the initial Pearson-based similarity, ζ_p and ζ_q denote their normalized risk indices, and κ is a cross-validated risk adjustment factor. Because $\kappa \in [0, 1]$, the term $\left|\xi_p - \xi_q\right|$ provides a normalized measure of pairwise risk difference. When two candidate centroids have the same risk level, the risk penalty is zero and their original load-shape similarity remains unchanged. As the difference between their risk levels increases, the modified similarity decreases monotonically. This mechanism therefore reduces the likelihood that customers with similar load shapes but substantially different fluctuation risks are incorrectly assigned to the same cluster.

Global AP clustering is then performed using the modified similarity matrix. The clustering iterations combine an adaptive self-similarity update with a quality index that

considers intra-cluster cohesion and inter-cluster separation. Iteration stops when the quality index converges, producing the macro-level partition. Figure 1 summarizes the workflow; the standard update equations and derivation of the quality index are provided in Appendix B.

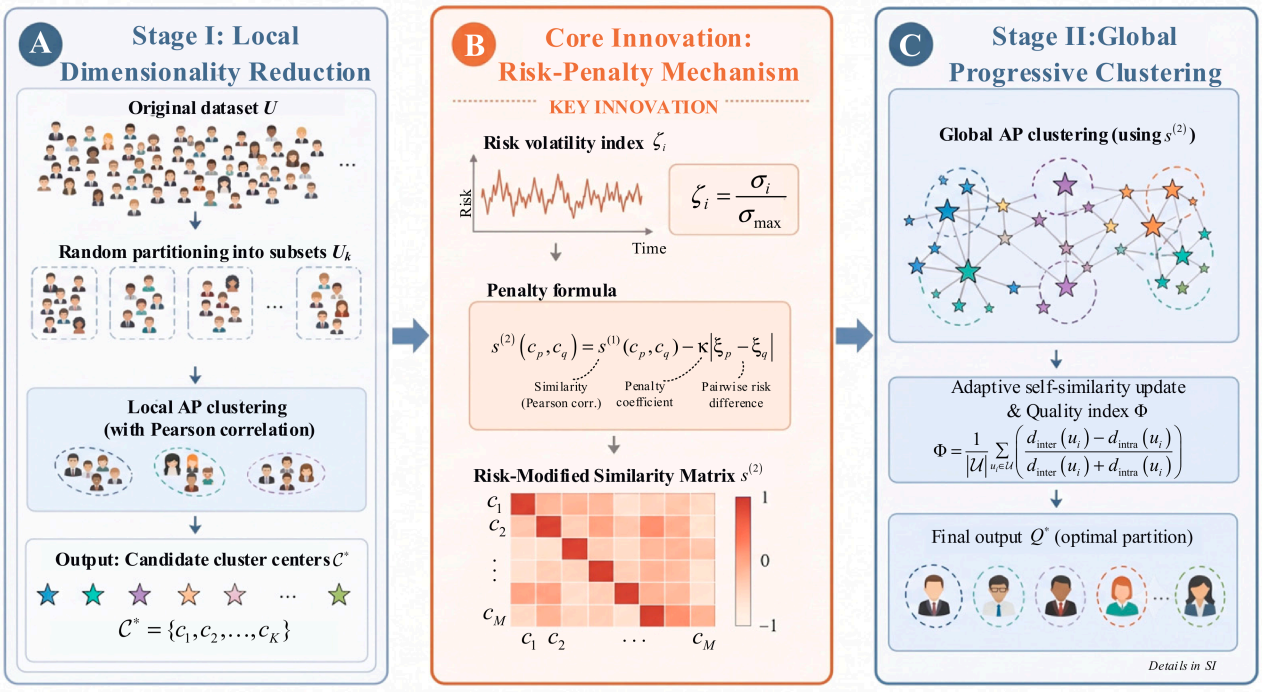

Figure 1. Schematic of the Risk-Driven Two-Stage Progressive Clustering (RD-TAP) Architecture.

3. Cognitive-Noise Filtration via IVIFS-Based Uncertainty Quantification and Adaptive Correction

Information asymmetry can make customers' evaluations of package attributes fuzzy or hesitant. We use IVIFSs to represent this uncertainty and determine attribute weights under incomplete preference information by combining attribute differentiation and influence. Individual-to-subgroup consensus deviation is subsequently used to detect and adaptively correct evaluations classified as unconfident. The corrected evaluations provide the input to the dynamic consensus model.

3.1. IVIFS-Based Cognitive Uncertainty Representation and Multi-Dimensional Distance Formulation

To mathematically characterize the subjective hesitancy and fuzziness inherent in customers' evaluations of complex package attributes (e.g., deviation settlement risks), we formulate an evaluation matrix utilizing Interval-Valued Intuitionistic Fuzzy Sets (IVIFSs). In this framework, the evaluation of customer ui regarding package pj on attribute ak is represented by an IVIFN $\tilde{v}_{ijk}$, which explicitly quantifies the cognitive boundaries of user uncertainty through its membership, non-membership, and hesitancy intervals.

To accurately capture the fine-grained disparities between individual subjective evaluations, we propose a novel multi-dimensional distance measure. Moving beyond conventional distance metrics that often neglect implicit psychological biases, our proposed metric comprehensively evaluates the deviation between any two IVIFNs (α and β) across three specific dimensions:

$$d(\alpha, \beta) = \frac{1}{3}(f_1(\alpha, \beta) + f_2(\alpha, \beta) + f_3(\alpha, \beta)) \quad (3)$$

The component f1 captures the direct deviation of fundamental fuzzy parameters; f2 integrates the balance point of support and opposition to reflect the cognitive offset (implicit biases); and f3 introduces crossover factors to measure the cross-impact of information interaction. By employing this multi-dimensional distance alongside a standardized score function $S(\alpha)$ for quantitative ranking, we establish a rigorous mathematical foundation for the subsequent extraction of attribute influence and consensus deviation. (Detailed mathematical expansions for components f1, f2, f3 and the score function $S(\alpha)$ are provided in Appendix C).

3.2. Dual-Driven Attribute Weight Optimization Under Incomplete Information Constraints

Customers may be unable to specify complete weights for complex package attributes. The proposed weight-determination mechanism therefore combines two complementary signals: the extent to which an attribute differentiates among packages and the extent to which that attribute influences the aggregate evaluation.

First, to fully exploit the limited subjective inputs, an attribute differentiation index (Δ_{ik}) is extracted using the aforementioned multi-dimensional distance. This index quantifies the discriminative power of each attribute across candidate packages. By maximizing the overall differentiation under the customer's incomplete weight constraints (Ω_i), the initial subjective weight $\omega_{ik}^{(0)}$ is robustly anchored.

Next, to objectively calibrate these weights, a "Leave-one-out" strategy is introduced to assess the global influence of individual attributes. By sequentially masking each attribute and measuring the resulting deviation in the comprehensive package evaluation, an objective influence degree (Γ_{ik}) is derived. The final optimized attribute weight (ω_{ik}) is formulated through the synergistic fusion of these two dimensions:

$$\omega_{ik} = \frac{\omega_{ik}^{(0)} \cdot \Gamma_{ik}}{\sum_{k=1}^{L} \omega_{ik}^{(0)} \cdot \Gamma_{ik}}, k = 1, \ldots, L \tag{4}$$

The mathematical programming models for constraint optimization and the standard interval intuitionistic fuzzy aggregation operators are detailed in Appendix C.

3.3. Consensus Deviation-Aware Adaptive Correction Mechanism for Unconfident Behaviors

To effectively mitigate the "unconfident behaviors" arising from customers' cognitive ambiguity, we establish an adaptive correction mechanism driven by individual-subgroup consensus deviation. Utilizing the subgroup partitions generated by the aforementioned RD-TAP algorithm, we first extract the mainstream consensus evaluation ($\tilde{g}_{qj}$) of each subgroup via interval intuitionistic fuzzy weighted aggregation. Subsequently, the cognitive deviation (δ_{ij}) between an individual customer and their corresponding subgroup consensus is quantitatively assessed.

Rather than employing rigid threshold filtering or simple replacement, we propose a piecewise adaptive fusion strategy to softly correct the identified unconfident evaluations. For an unconfident customer whose evaluation deviation exceeds the predefined significance threshold, their initial evaluation vector $\tilde{v}_{ijk}$ is dynamically reconstructed as follows:

$$\tilde{v}'_{ijk} = \left(1 - \eta_{ij}\right) \cdot \tilde{v}_{ijk} \oplus \eta_{ij} \cdot \tilde{g}_{c(u_i)j}^{(k)} \tag{5}$$

Here, the nonlinear fusion coefficient η_{ij} increases with the adjusted deviation. Evaluations that deviate substantially from the subgroup benchmark consequently receive a larger correction, whereas evaluations with small deviations retain more of their original information. The corrected evaluation matrix is then passed to the global consensus-optimization

stage. Detailed formulations regarding the piecewise deviation amplification rules, consensus distance aggregation, and unconfident set judgment conditions are provided in Appendix D.

4. Dynamic Consensus Optimization and Satisfaction-Driven Package Recommendation

While the aforementioned IVIFS correction effectively mitigates unconfident cognitive fuzziness, the subgroup evaluation matrix may still suffer from low consensus due to inherent divergences in personalized preferences. To overcome this bottleneck, we establish a dynamic consensus optimization framework. Rather than forcing rigid group agreement, this framework first quantifies individual-subgroup consistency to precisely identify low-consensus outlier evaluations. Subsequently, a multi-threshold dynamic weighted fusion mechanism is implemented to softly guide these divergent vectors toward the mainstream subgroup consensus, carefully preserving acceptable individual heterogeneity. Ultimately, by coupling this optimized high-consensus matrix with individual load profile similarities, a robust, full-ranking recommendation of electricity retail packages is achieved to maximize personalized customer satisfaction.

4.1. Multi-Level Consensus Quantification and Consistency Evaluation

Based on the corrected evaluation matrix $\tilde{V}'$ and the RD-TAP clustering partitions, we first establish a benchmark evaluation at the subgroup level. Instead of detailing routine aggregation processes, we directly extract the subgroup consensus vector ($\tilde{g}'_{qjk}$) for each package attribute by weighting the interval-valued intuitionistic fuzzy matrices of all customers within subgroup C_q.

To precisely quantify the alignment between an individual's preference and the subgroup's collective rationality, a package-level individual consensus degree (I_{ij}) is formulated. Utilizing the previously defined multi-dimensional distance $d(\alpha, \beta)$ and the dual-driven optimized weights ω_{ik}, it is defined as:

$$I_{ij} = 1 - \sum_{k=1}^{L} w_{ik} d(\tilde{v}'_{ijk}, \tilde{g}'_{qjk}) \tag{6}$$

where a higher I_{ij} indicates a stronger agreement with the mainstream subgroup opinion. Subsequently, by structurally averaging I_{ij} across all candidate packages and all subgroup members, we derive the global individual consensus index I_i and the overarching subgroup consensus index I_q.

Crucially, $I_q \in [0, 1]$ serves as a macro-indicator of group agreement. A predefined consensus threshold ψ (empirically bounded within $[0.85, 0.95]$) is introduced to act as a strict decision boundary: if $I_q \geq \psi$, the subgroup is deemed to have reached sufficient agreement for direct recommendation; conversely, if $I_q < \psi$, a targeted feedback mechanism is sequentially triggered to adjust the divergent outlier evaluations. (Standard arithmetic formulas for subgroup vector aggregation and index averaging are detailed in Appendix D).

4.2. Similarity-Driven Two-Level Identification Mechanism for Low-Consensus Targets

To facilitate efficient consensus convergence, we propose a similarity-driven two-level identification mechanism to precisely lock onto the specific sources of group divergence. Rather than applying a blanket adjustment to the entire subgroup, this mechanism adopts a targeted "macro-to-micro" search strategy to isolate the most critical outlier evaluations.

Specifically, at the macro-customer level, we first evaluate the global evaluation similarity between any individual customer and the rest of their subgroup across all candidate packages. By ranking these overall similarity levels, the primary divergent customer ex-

hibiting the minimum consensus score is isolated. Subsequently, at the micro-package level, we further scan this specific unconfident customer's evaluation vector. By comparing their evaluations against the collective rationality, the exact electricity retail package demonstrating the most severe cognitive deviation is pinpointed. This accurately identified customer–package pair then serves as the primary target for the subsequent dynamic consensus adjustment. (Detailed arithmetic formulations for similarity aggregation and minimum-value search are provided in Appendix E).

4.3. Multi-Threshold Dynamic Weighted Fusion for Subgroup Consensus Evolution

Once the target customer–package pair (u_c, p_c) with the maximum cognitive deviation is isolated, we propose a multi-threshold dynamic weighted fusion mechanism to gently guide the divergent evaluation toward the subgroup consensus. Instead of completely overwriting the outlier's opinion, this approach carefully preserves reasonable personalized preferences.

Initially, we directly extract the upper and lower boundary evaluations (i.e., peers exhibiting the highest and lowest similarity regarding p_c) within the subgroup to establish a reference fusion space. Subsequently, based on predefined consensus segmentation thresholds ($0 < \lambda_1 < \lambda_2 < 1$), the original evaluation is adaptively reconstructed via a piecewise strategy:

$$\hat{\tilde{v}}_{u_c,p_c} = \begin{cases} \varpi \cdot \tilde{v}_{u_c,p_c} \oplus (1-\varpi) \cdot \tilde{v}_{u_{\max},p_c}, & S(u_c,p_c) > \lambda_2 \\ \tilde{v}_{u_c,p_c}, & \lambda_1 \le S(u_c,p_c) \le \lambda_2 \\ \varpi \cdot \tilde{v}_{u_c,p_c} \oplus (1-\varpi) \cdot \tilde{v}_{u_{\min},p_c}, & S(u_c,p_c) < \lambda_1 \end{cases} \tag{7}$$

where $\varpi \in [0,1]$ is defined as the compromise retention coefficient. This pivotal parameter theoretically dictates the trade-off between subgroup conformity and individual autonomy: a smaller ϖ strongly pulls the extreme outlier back to collective rationality, whereas a larger ϖ respects and retains the unique electricity consumption traits of the user.

Following this localized correction, the global consensus index I_q is recursively evaluated. This similarity-ranking and targeted-fusion cycle dynamically iterates until I_q surpasses the termination threshold ψ, ultimately yielding the optimized high-consensus matrix $\widetilde{V}^*$ for the subsequent package recommendation. (Standard iteration pseudocode and boundary search arithmetic are detailed in Appendix E).

4.4. Satisfaction-Driven Full-Ranking Recommendation via Consensus-Similarity Mapping

Once the optimal high-consensus evaluation matrix $\widetilde{V}^*$ is established, it serves as a robust knowledge base for personalized package matching. For any newly entering target customer u_t, we introduce a consensus-similarity mapping mechanism to predict their specific preferences, thereby effectively bypassing the conventional cold-start decision dilemma.

Instead of relying on isolated and unconfident individual judgments, this mechanism maximally leverages the optimized collective rationality. Specifically, we first map the new customer u_t to the most structurally matched sub-group C_q by evaluating their baseline load profile against the pre-extracted cluster centroids. Subsequently, a feature similarity weight, derived from the normalized Pearson correlation, is applied to accurately quantify the behavioral affinity between the target customer and each sample member within C_q.

Ultimately, by performing a similarity-weighted aggregation on the converted single-value satisfaction scores from $\widetilde{V}^*$, a comprehensive predicted satisfaction index Z_{tj} for each candidate package is derived. Ranking these indices in descending order directly yields the optimal full-ranking recommendation list. This terminal mapping step seamlessly translates abstract cognitive evaluations and complex load risks into an actionable, satisfaction-

maximizing decision support scheme for the electricity retail market. (Explicit arithmetic formulations regarding the similarity weighting, score conversion, and aggregation logic are detailed in Appendix E).

5. Empirical Validation in Real-World Electricity Markets

The framework is evaluated using smart-meter load data from a provincial electricity retail market in China. The case study examines three questions: whether RD-TAP separates load patterns with different fluctuation risks, whether the correction and fusion mechanisms reduce constructed cognitive noise and improve consensus, and whether the resulting recommendations agree with the observed package-choice outcomes of real customers.

5.1. Configuration of the Empirical Market Scenario and Dataset Initialization

The dataset comprises 1500 independent retail electricity customers from a regional smart grid in Zhejiang Province, China. Hourly smart-meter readings were collected from 1 January to 31 December 2024, providing up to 8784 observations per customer. The hourly load data cover the full annual observation period. To characterize customers' load behavior at a consistent temporal scale, load characteristics, including the monthly load factor and monthly maximum utilization hours, were calculated on a monthly basis. The monthly indicators were averaged over the 12-month observation period to obtain customer-level features for clustering and subsequent analysis. Negative or meter-error readings were treated as invalid; isolated missing readings were linearly interpolated, and customer days with more than three consecutive missing hours were excluded. Customers were retained when at least 95% of their readings were valid. Readings at the same hour of day were averaged over the monitoring period to obtain an annual average 24 h profile. Each profile was divided by its annual maximum, and the derived features were min-max normalized using the development customers only. Peak periods were defined as [08:00, 11:00) and [18:00, 21:00), the valley period was defined as [00:00, 06:00), and the remaining hours as shoulder periods. RD-TAP identified four empirical load groups: steady load, daytime single-peak load, evening-peak load, and high-frequency fluctuation.

Customers were divided at the customer level into 1275 development customers and 225 held-out test customers using a fixed random seed. Five-fold cross-validation, parameter selection, normalization-parameter estimation, consensus-rule configuration, and ablation design were conducted using the development customers only. The held-out customers were reserved for final recommendation evaluation, and their observed outcomes were not used for fitting or tuning. The n = 1500 experiment in Section 5.5.1 is an independent unsupervised scalability benchmark and does not contribute to parameter selection, recommendation-model fitting, or held-out outcome evaluation. The raw smart-meter records are restricted by data-owner and privacy requirements; therefore, the revised manuscript provides the preprocessing rules and implementation details, while the de-identified raw data are available only subject to data-owner approval.

Figure 2 illustrates differences in both temporal load shape and within-cluster volatility. The steady -load group has a narrow envelope around its prototype, whereas the high-frequency-fluctuation group has a wider and less regular envelope. The daytime- and evening-peak groups show intermediate dispersion. These patterns motivate the risk term in RD-TAP: customers with comparable average profiles can still differ materially in volatility and, consequently, in their exposure to deviation-related charges. Distinguishing this exposure is relevant when matching customers to fixed, time-of-use, risk-sharing, or high-tolerance packages.

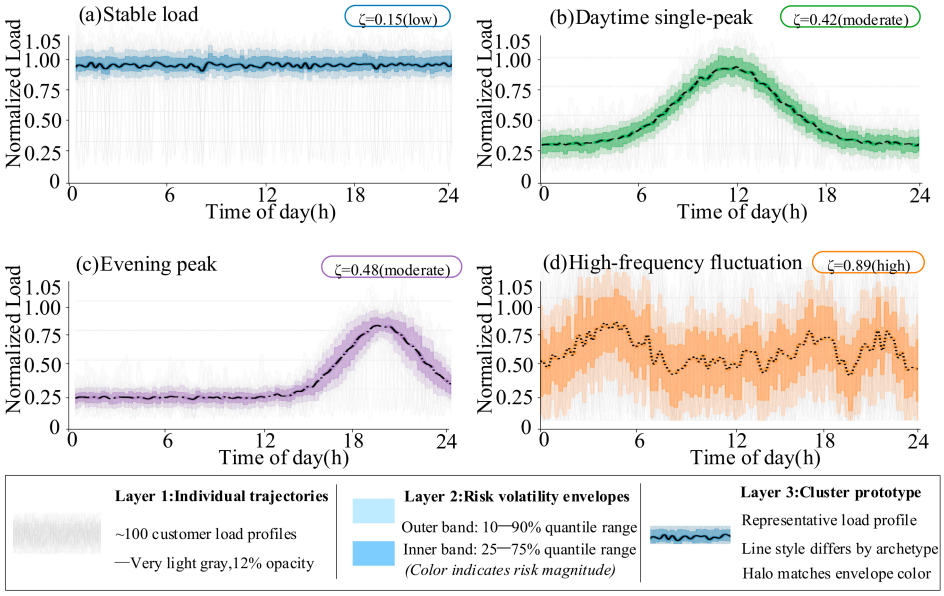

Figure 2. Risk-aware load archetypes revealed by the proposed RD-TAP clustering framework. Four representative daily load patterns are identified, including (**a**) steady load, (**b**) daytime single-peak load, (**c**) evening-peak load, and (**d**) high-frequency fluctuation load. In each panel, faint background trajectories denote individual normalized customer load profiles, the central dark curve represents the cluster prototype, and the shaded envelopes indicate intra-cluster volatility ranges estimated from empirical quantiles. Unlike conventional shape-oriented clustering, which mainly groups customers according to profile similarity, RD-TAP explicitly incorporates the pairwise difference in normalized load-fluctuation risk into the similarity evaluation. The progressively expanded uncertainty envelope from panels (**a**–**d**) illustrates that customers with comparable load levels or profile shapes may still exhibit substantially different fluctuation risks. This risk-sensitive representation provides the empirical motivation for introducing the risk-difference penalty into RD-TAP, thereby reducing the erroneous aggregation of customers with similar load shapes but heterogeneous risk levels.

Four representative electricity retail packages (P1–P4) are defined for the matching experiment. Each package is represented in a normalized four-dimensional attribute space comprising tariff stability, deviation tolerance, eco-friendliness, and value-added-service complexity.

A perturbation rate of 15% is used as the baseline condition for the case-study comparisons. To examine robustness beyond this setting, Section 5.7 varies both the proportion of perturbed customers and the perturbation magnitude, including a zero-injected-noise control. The evaluation matrices before perturbation are retained as fixed references for measuring reconstruction error and unintended changes to unperturbed evaluations. These references represent the evaluations before artificial noise injection and are not assumed to be error-free observations of customers' true preferences. The algorithmic parameters remain fixed at the configuration specified in Appendix E.4.

Figure 3 shows the trade-offs among the four package profiles. P2 has a high eco-friendliness score but low deviation tolerance and is therefore more compatible with steady, environmentally oriented customers under the assumed preference rules. P4 has the highest coverage along the deviation-tolerance dimension and is designed for the high-frequency-fluctuation subgroup (C4). This attribute matrix is used as the basis for the subsequent consensus and recommendation tests.

To evaluate the stability of the proposed method, the complete experiment was repeated using ten random seeds (0–9). For each seed, the Stage-I partition, perturbed-customer selection, and subsequent optimization process were regenerated using the same seed across all compared methods. The perturbation ratio was fixed at 15% for the

baseline comparisons, whereas the robustness analysis in Section 5.7 evaluates multiple perturbation rates and magnitudes. Within each experimental condition, all methods were evaluated on identical data realizations. The reported results are therefore presented as the mean ± standard deviation over ten repeated runs, together with the corresponding 95% confidence intervals.

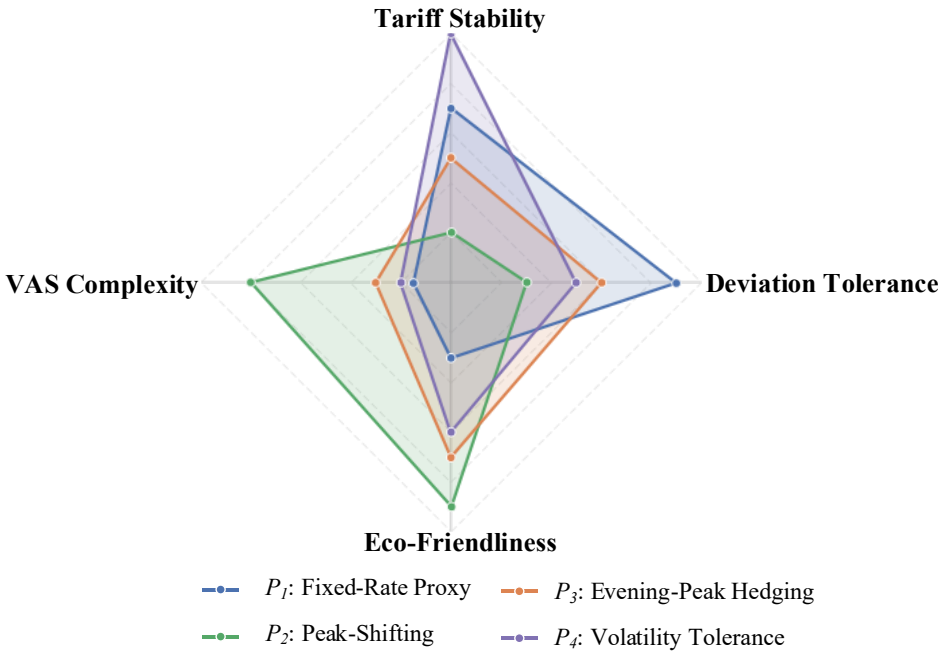

Figure 3. Multi-dimensional attribute configuration and heterogeneous trade-off geometry of the four archetypal electricity retail packages.

In addition to the complete RD-TAP model, two ablated versions were evaluated: RD-TAP without cognitive correction and RD-TAP without the consensus mechanism. K-means and FCM were used as external clustering baselines. The same evaluation data and random seeds were used for all methods.

5.2. Micro-Level Validation of Unconfident Behavior Correction and Cognitive Noise Reduction

A representative customer (U_{001}) from the evening-peak subgroup (C3) is used to illustrate the correction mechanism. The customer has a monthly load factor of 0.3413 and maximum utilization of 238.6 h. Under the package assumptions adopted in this study, this relatively volatile profile gives greater relevance to the deviation-settlement-penalty attribute. The initial membership degree assigned to that attribute is [0.15, 0.30], which is lower than the subgroup reference evaluation.

The consensus-deviation metric identifies this customer–attribute evaluation for correction. After adaptive fusion, the membership interval changes to [0.32, 0.46], and the customer's hesitation interval decreases by 25% on average.

Across the subgroup, the targeted correction reduces the average hesitation degree by 22% and the mean individual-to-consensus distance by 32.4% (Figure 4). These reductions indicate that the correction procedure moves evaluations toward the subgroup benchmark. Whether this adjustment also recovers the evaluations before perturbation, rather than merely increasing agreement, is examined in Section 5.7 using reference-based errors and a zero-injected-noise control.

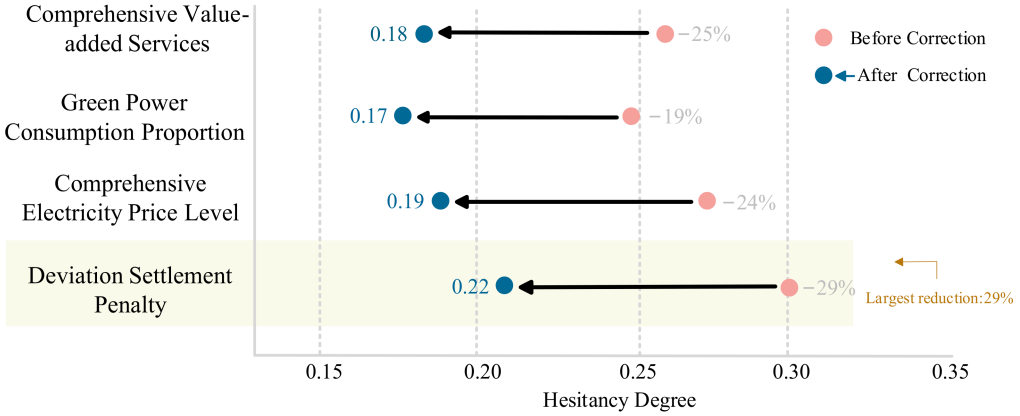

Figure 4. Macroscopic compression of cognitive noise: A dumbbell plot illustrating the significant reduction in the average hesitancy degree across multi-dimensional package attributes before and after the adaptive fusion correction.

5.3. Convergence Validation of the Dynamic Consensus Evolution Mechanism

Figure 5 traces the group consensus index (GCI) for the four subgroups during weighted fusion. The initial GCIs of C2 and C4 are 0.765 and 0.712, respectively, both below the prescribed consensus threshold ($\psi = 0.9$). One iteration consists of updating the evaluation vector of a single customer–package pair.

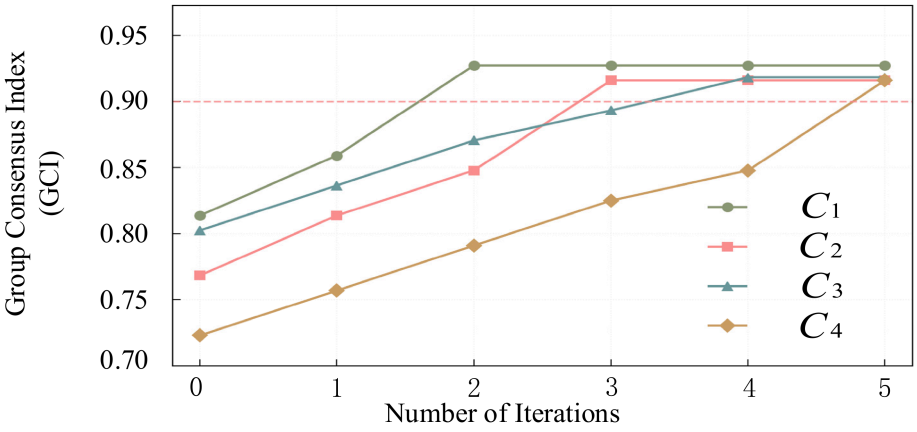

Figure 5. Evolution of Consensus Degree.

Under the baseline parameter configuration, all four subgroups reach the threshold within five iterations. C1 reaches a GCI of 0.915 after two iterations, while C4 increases monotonically and reaches 0.902 after five iterations.

The different trajectories indicate that the number of required adjustments depends on the initial level of disagreement. Under the tested settings, the mechanism increases consensus in both low- and high-volatility subgroups without assigning identical evaluations to all members. A separate controlled simulation in Section 5.8.2 quantifies both the increase in consensus and the retained dispersion, with explicit reporting of non-attainment cases.

5.4. Personalized Package Matching and Archetypal Preference Validation

A newly entering customer (U_{new}) from the evening-peak subgroup (C3) is used to illustrate the final matching step. Table 2 shows that P3 has the highest predicted satisfaction. The customer's consumption is concentrated between 18:00 and 22:00 and is characterized by a low monthly load factor and substantial variability. Under the package definitions used here, P3 combines evening-peak pricing with a 5:5 deviation-risk-sharing rule; P1 and P2 therefore receive lower predicted scores.

Table 2. Predicted Satisfaction of U_{new}.

Electricity Retail Package	P1	P2	P3	P4
Predicted Satisfaction	0.4358	0.5124	0.8975	0.7642

For an additional archetype-level check, observed package-choice outcomes were aggregated within each load archetype to obtain the empirical package-preference ranking: steady loads are associated with P1, high environmental preference with P2, evening-peaking and fluctuating loads with P3, and high-frequency fluctuations with P4.

As shown in Table 3, the recommended rankings agree with the observed customer choices, providing evidence of predictive validity for the package-matching stage. Broader validation across additional markets and time periods remains necessary.

Table 3. Comparison Between Recommended Ranking Results and Observed Customer Choices.

Electricity Retail Packages	Recommended Ranking (Proposed Method)	Observed Customer Choice
P1	4	4
P2	3	3
P3	1	1
P4	2	2

5.5. Benchmarking Algorithmic Superiority and Recommendation Robustness

5.5.1. Superiority of the Risk-Aware Clustering Architecture in High-Volatility Scenarios

To address the risk-sensitive nature of the comprehensive clustering quality index, denoted by Z_Q, the revised benchmark compares RD-TAP with standard affinity propagation (AP), two-stage AP without the risk-penalty term ($\kappa = 0$), K-means, fuzzy C-means (FCM), and Ward-linkage hierarchical agglomerative clustering. Silhouette, Davies–Bouldin (DB), and Calinski–Harabasz (CH) scores are reported as meth-od-independent internal clustering metrics. Higher Silhouette and CH scores and lower DB scores indicate better geometric separation. All methods use the same normalized customer feature vectors, target four clusters, and matched random seeds.

The benchmark was evaluated at input sizes of n = 500, 750, 1000, and 1500, with ten matched random seeds at each size. Figure 6 reports the mean Z_Q values and their standard deviations across input sizes. At n = 1500, RD-TAP obtained $Z_Q = 0.958 \pm 0.009$, a Silhouette score of 0.71 ± 0.02, a DB score of 0.48 ± 0.03, and a CH score of 1690 ± 85. The corresponding Z_Q values were 0.944 ± 0.011 for two-stage AP with $\kappa = 0$, 0.901 ± 0.018 for standard AP, 0.713 ± 0.095 for K-means, 0.742 ± 0.095 for FCM, and 0.872 ± 0.020 for Ward hierarchical clustering. The repeated-run comparisons are summarized in Table 4.

Table 4. Neutral clustering metrics and the comprehensive quality index at n = 1500 (ten matched runs).

Method	Silhouette	DB	CH	ZQ	Wilcoxon p vs. RD-TAP
RD-TAP	0.71 ± 0.02	0.48 ± 0.03	1690 ± 85	0.958 ± 0.009	—
Two-stage AP ($\kappa = 0$)	0.68 ± 0.02	0.54 ± 0.04	1540 ± 90	0.944 ± 0.011	0.004
Standard AP	0.64 ± 0.03	0.62 ± 0.05	1305 ± 110	0.901 ± 0.018	0.001
K-means	0.63 ± 0.03	0.65 ± 0.05	1280 ± 105	0.713 ± 0.095	0.001
FCM	0.65 ± 0.03	0.61 ± 0.05	1340 ± 100	0.742 ± 0.095	0.001
Ward hierarchical	0.60 ± 0.02	0.69 ± 0.04	1180 ± 80	0.872 ± 0.020	0.001

The neutral metrics support the same ordering as Z_Q while reducing dependence on the authors' risk-sensitive construction: RD-TAP has the highest Silhouette and CH scores

and the lowest DB score among the compared methods. Table 5 reports runtime and peak memory. At n = 1500, RD-TAP requires 33.2 ± 1.8 s and 0.93 ± 0.05 GB, compared with 176.3 ± 12.0 s and 6.42 ± 0.31 GB for standard AP. All methods converge in 10/10 runs except standard AP at n = 1500, which converges in 9/10 runs. The two-stage AP without the risk penalty remains computationally efficient but has slightly lower clustering quality than RD-TAP. These results support computational scalability within the tested input range; broader generalization to larger and more heterogeneous customer populations requires additional independent data.

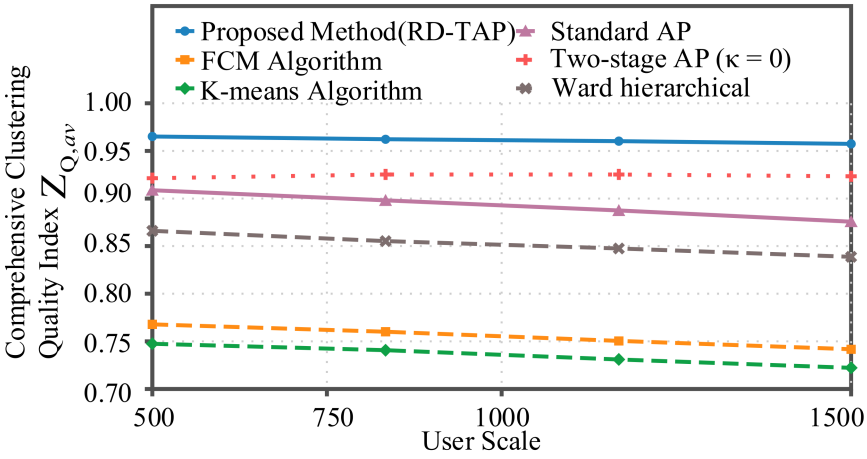

Figure 6. Mean Z_Q across input sizes for RD-TAP and the comparison methods; error bars denote standard deviations over ten matched runs.

Table 5. Runtime (s) and peak memory (GB) across input sizes (ten matched runs).

Method	n = 500	n = 750	n = 1000	n = 1500
RD-TAP	8.4 ± 0.5 s 0.36 ± 0.02 GB	13.1 ± 0.7 s 0.47 ± 0.03 GB	19.4 ± 1.0 s 0.62 ± 0.04 GB	33.2 ± 1.8 s 0.93 ± 0.05 GB
Two-stage AP (κ = 0)	9.1 ± 0.5 s 0.38 ± 0.02 GB	14.9 ± 0.8 s 0.49 ± 0.03 GB	22.3 ± 1.1 s 0.65 ± 0.04 GB	38.7 ± 2.3 s 1.01 ± 0.06 GB
Standard AP	17.8 ± 1.2 s 0.82 ± 0.05 GB	41.7 ± 3.0 s 1.64 ± 0.10 GB	78.5 ± 5.4 s 2.98 ± 0.18 GB	176.3 ± 12.0 s 6.42 ± 0.31 GB
Ward hierarchical	12.6 ± 0.8 s 0.46 ± 0.03 GB	28.9 ± 1.9 s 0.93 ± 0.06 GB	54.8 ± 3.6 s 1.61 ± 0.10 GB	128.7 ± 8.4 s 3.59 ± 0.18 GB

5.5.2. Ablation Study: Contributions of Cognitive Correction and Consensus Evolution

An ablation study evaluates the contributions of unconfident-behavior correction and dynamic consensus optimization across the four customer groups. RMSE is calculated for the complete framework and two variants: Scheme 1 removes the cognitive-bias correction, and Scheme 2 removes consensus optimization. Additional ablations of the risk term and the three distance components are reported in Section 5.8.1.

To assess stochastic stability, the complete framework and both variants were rerun using ten random seeds, with identical seed-specific partitions and perturbation patterns for all methods. Because the satisfaction scores are evaluated on a normalized scale, RMSE is dimensionless and is reported on the same normalized scale. Lower RMSE values indicate better agreement between the predicted and reference satisfaction values.

Figure 7 shows that removing either component increases RMSE, although the increases differ among customer groups. The bars represent the mean RMSE over ten runs, and the error bars represent the standard deviations. Scheme 1 retains the perturbed fuzzy evaluations, while Scheme 2 leaves low-consensus evaluations unadjusted; the latter

produces larger errors for the high-fluctuation group. The repeated-run RMSE results for the four load patterns are summarized in Table 6.

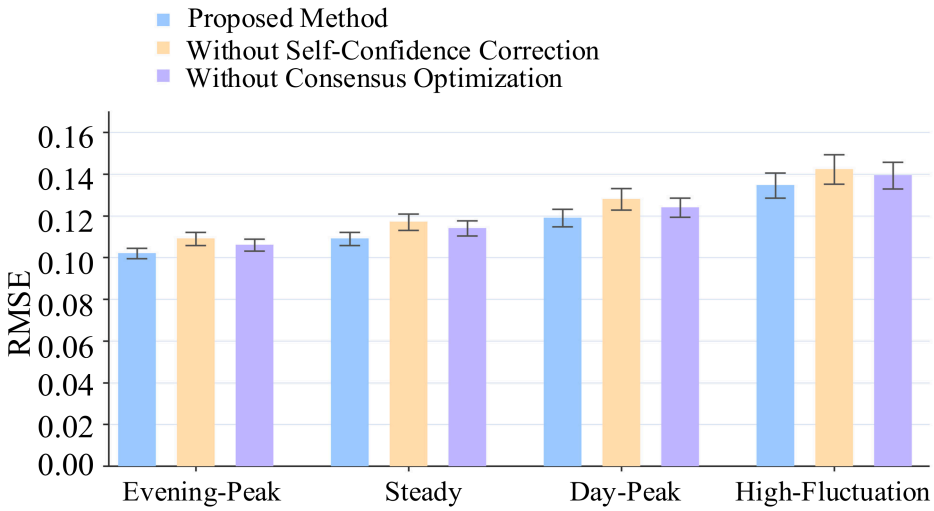

Figure 7. Comparison of RMSE for different recommendation methods.

Table 6. Repeated-Run RMSE Results for the Ablation Study across Load Patterns.

Load Pattern	Complete	Scheme 1	Scheme 2
Evening-Peak	0.1020 ± 0.0025	0.1090 ± 0.0032	0.1060 ± 0.0028
Steady	0.1090 ± 0.0032	0.1170 ± 0.0040	0.1140 ± 0.0036
Day-Peak	0.1190 ± 0.0042	0.1280 ± 0.0052	0.1240 ± 0.0046
High-Fluctuation	0.1346 ± 0.0060	0.1423 ± 0.0070	0.1394 ± 0.0064

Together, the ablation, sensitivity, and runtime analyses identify the modules that contribute under the tested setting without implying that every component is universally necessary.

Across the four load patterns, the mean RMSE ranges from 0.1020 to 0.1346 for the complete framework, from 0.1090 to 0.1423 for Scheme 1, and from 0.1060 to 0.1394 for Scheme 2, as reported in Table 6. The complete framework achieves a lower mean RMSE than both ablated variants in each of the four load patterns. These results indicate that both cognitive-noise correction and consensus optimization contribute to prediction stability under the adopted perturbation and package-preference settings.

5.6. Parameter Selection and Sensitivity Analysis

Cross-validation and sensitivity analyses are conducted using the customer dataset described in Section 5.1. The dataset is divided into 1275 development customers and 225 held-out test customers. Five-fold cross-validation is performed within the development set using ten matched random seeds. Customer partitions and evaluation inputs are kept identical across parameter configurations, and validation and test evaluations do not contribute to training subgroup consensus. Among $\kappa \in \{0, 0.25, 0.5, 1, 2, 4\}$, $\kappa = 2$ yields the lowest validation RMSE of 0.1128 ± 0.0044. The corresponding held-out test RMSE is 0.1242 ± 0.0011. Validation standard deviation is calculated across five-fold means after averaging seeds, whereas test standard deviation describes variability across ten seeds on the fixed test set. RMSE is calculated on the common dimensionless score scale generated by the IVIFS score function, using the same score conversion for the predicted and reference values.

Using the baseline configuration specified in Appendix E.4, each parameter is varied individually while the remaining parameters are held fixed. With κ fixed, $\tau \in \{0.20, 0.25, 0.30, 0.35, 0.40\}$, $\psi \in \{0.80, 0.85, 0.90, 0.925, 0.95\}$, and the compromise retention coefficient $\varpi \in \{0.20, 0.30, 0.40, 0.50, 0.60\}$ are varied individually, while the remaining parameters retain their baseline values. Performance is assessed using RMSE, consensus iterations, subgroup attainment rates, and Kendall's τ_b relative to the baseline package rankings.

Across the tested τ values, mean RMSE ranges from 0.110 to 0.125, and Kendall's τ_b ranges from 0.90 to 1.00. All tested parameter configurations reach their prescribed consensus thresholds. The mean number of iterations required for convergence ranges from 4 to 18 across ψ and from 6 to 24 across ϖ. These results characterize the effects of parameter variation on consensus convergence, recommendation error, and ranking stability within the tested ranges.

5.7. Robustness to Evaluation Perturbations and Over-Correction

We evaluated the proposed method under perturbation rates of 0%, 5%, 15%, 30%, and 50% and perturbation magnitudes $\alpha \in \{0.10, 0.30, 0.50\}$. The zero-noise control was evaluated once per seed, yielding 13 conditions in total. The complete model and a variant without cognitive correction used identical customer partitions, perturbations, and ten matched random seeds, with all parameters held fixed across conditions. Recommendation RMSE was measured against fixed reference satisfaction scores, while reconstruction error and unintended distortion were evaluated before consensus against the original, unperturbed evaluations.

As shown in Table 7, under zero injected noise, the complete model exhibited negligible distortion (0.00023 ± 0.00010), and recommendation RMSE was 0.0986 ± 0.0031 for the complete model and 0.1019 ± 0.0030 without correction, indicating that the correction module did not materially distort clean evaluations. At a perturbation rate of 15% and $\alpha = 0.30$, reconstruction error decreased from 0.0750 to 0.0341, while recommendation RMSE was 0.1168 ± 0.0040 with correction and 0.1247 ± 0.0050 without correction.

Table 7. Recommendation accuracy and correction-induced distortion under different noise conditions.

Noise Condition	Recommendation RMSE with Correction	Recommendation RMSE Without Correction	Reconstruction Error Before → After Correction	Over-Correction Rate	Distortion of Unperturbed Evaluations
No injected noise	0.0986 ± 0.0031	0.1019 ± 0.0030	-	N/A	0.00023 ± 0.0001
Low perturbation (5%, $\alpha = 0.10$)	0.1038 ± 0.0035	0.1112 ± 0.0040	$0.0200 \rightarrow 0.0116$	$1.4\% \pm 0.5\%$	0.00108 ± 0.0003
Moderate perturbation (15%, $\alpha = 0.30$)	0.1168 ± 0.0040	0.1247 ± 0.0050	$0.0750 \rightarrow 0.0341$	$3.8\% \pm 1.0\%$	0.00128 ± 0.0005
High perturbation (50%, $\alpha = 0.50$)	0.1592 ± 0.0080	0.1871 ± 0.0100	$0.1250 \rightarrow 0.1142$	$14.3\% \pm 2.0\%$	0.00625 ± 0.0015

N/A: not applicable. Since no noise was injected under this condition, no over-correction occurs and the over-correction rate is undefined. "-" indicates that the reconstruction error remains unchanged before and after correction.

The benefit diminished under severe contamination. At a perturbation rate of 50% and $\alpha = 0.50$, reconstruction error decreased from 0.1250 to 0.1142, with an over-correction rate of $14.3 \pm 2.0\%$. The corresponding RMSE values were 0.1592 ± 0.0080 with correction and 0.1871 ± 0.0100 without correction. Overall, cognitive correction improved recommendation RMSE and reduced reconstruction error under the tested perturbation conditions, although its benefit became smaller and the over-correction rate increased as perturbation

intensity increased. These results support the effectiveness of the correction mechanism within the tested range while indicating reduced reliability under severe perturbation.

5.8. Component Ablation and Consensus-Fusion Validation

5.8.1. Component Ablation and Held-Out Ranking

Seven matched configurations were evaluated: the complete model, $\kappa = 0$, endpoint Hamming distance, removal of f1, f2, or f3, and no fusion. The fixed split contains 1275 development and 225 held-out customers; test evaluations and reference scores are excluded from fitting and prediction. Hit Rate@1 (HR@1) measures whether the highest-reference package is ranked first; NDCG@3 uses nonnegative satisfaction as linear relevance with logarithmic discounts; and Kendall's τ-b evaluates all four packages. Table 8 reports the mean $\pm$ standard deviation across ten matched seeds against the observed reference outcomes.

Table 8. Component ablation and ranking performance on 225 held-out customers.

Configuration	RMSE	Hit@1 (%)	NDCG@3	τ-b
Complete	0.1242 ± 0.0011	75.3 ± 0.9	0.9502 ± 0.0018	0.4840 ± 0.0293
No risk penalty	0.1272 ± 0.0018	66.1 ± 1.8	0.9430 ± 0.0026	0.3900 ± 0.0403
Hamming distance	0.1249 ± 0.0010	75.2 ± 1.0	0.9507 ± 0.0013	0.4937 ± 0.0125
Without f1	0.1249 ± 0.0010	75.2 ± 1.0	0.9507 ± 0.0013	0.4937 ± 0.0125
Without f2	0.1237 ± 0.0012	75.3 ± 0.9	0.9504 ± 0.0016	0.4860 ± 0.0273
Without f3	0.1237 ± 0.0012	75.3 ± 0.9	0.9505 ± 0.0016	0.4863 ± 0.0273
No consensus fusion	0.1242 ± 0.0011	75.3 ± 0.9	0.9502 ± 0.0018	0.4840 ± 0.0293

Removing the risk penalty raises RMSE from 0.1242 to 0.1272 and lowers Hit@1 from 75.3% to 66.1%. Distance ablations are mixed, and the complete and no-fusion variants coincide because this reconstructed pipeline makes no updates. Active fusion is therefore evaluated separately below with controlled post-correction inputs.

5.8.2. Controlled Validation of Consensus Gain and Heterogeneity Retention

The controlled experiment uses four subgroups of 100 customers, four packages, and four attributes. Package utility centers are 0.35, 0.45, 0.55, and 0.65; attribute offsets are -0.03, -0.01, 0.01, and 0.03; uniform deviations have a half-width of 0.28. IVIFN bounds follow the original construction, and no additional cognitive correction is triggered.

Following Equation (7), $\lambda 1 = 0.35$, $\lambda 2 = 0.75$, $\psi = 0.90$, and the retention coefficient 0.40 are unchanged. After a pilot, 30 new seeds (10,000–10,029) were run without exclusions. The no-fusion and identical-evaluation controls isolate fusion effects; stopping uses a 200-update limit or an unchanged state. Dispersion D is the mean absolute endpoint difference, and R is post-/pre-fusion D. Results are averaged over four subgroups within each seed, with seeds as replicates.

As summarized in Table 9, fusion raises mean GCI from 0.8860 $\pm$ 0.0008 to 0.9001 $\pm$ 0.0003, whereas no fusion remains at 0.8860. The mean gain is 0.01405 (95% bootstrap interval [0.01375, 0.01434]); all 120 groups improve, none enter the middle no-update branch, and 119 reach 0.90. One group stops at a fixed point below the target. Mean updates are 56.98 $\pm$ 3.41 per group.

Dispersion decreases from 0.1595 $\pm$ 0.0007 to 0.1451 $\pm$ 0.0008, retaining 90.96 $\pm$ 0.54%; top-choice retention is 96.01 $\pm$ 0.93% and the ranking tau-b is 0.9554 $\pm$ 0.0072. Identical evaluations give GCI = 1 and D = 0. Figure 8 summarizes the consensus trajectory and the retained dispersion.

Boundary tests retain all outcomes: at a half-width of 0.35, 0/40 groups reach 0.90; at 0.30, 39/40 do; for 25- and 50-member groups at 0.28, attainment is 38/40 and 39/40,

respectively. The gain is therefore conditional on the disagreement structure. All primary and boundary runs, branch events, matrices, and verification records are retained in the reproducibility package.

Table 9. Consensus and retained dispersion in the controlled fusion experiment (30 seeds).

Configuration	GCI Before	GCI After	D After	R (%)	Attained Groups
No fusion	0.8860 ± 0.0008	0.8860 ± 0.0008	0.1595 ± 0.0007	100.00 ± 0.00	0/120
Fusion	0.8860 ± 0.0008	0.9001 ± 0.0003	0.1451 ± 0.0008	90.96 ± 0.54	119/120
Identical evaluations	0.8860 ± 0.0008	1.0000 ± 0.0000	0.0000 ± 0.0000	0.00 ± 0.00	120/120

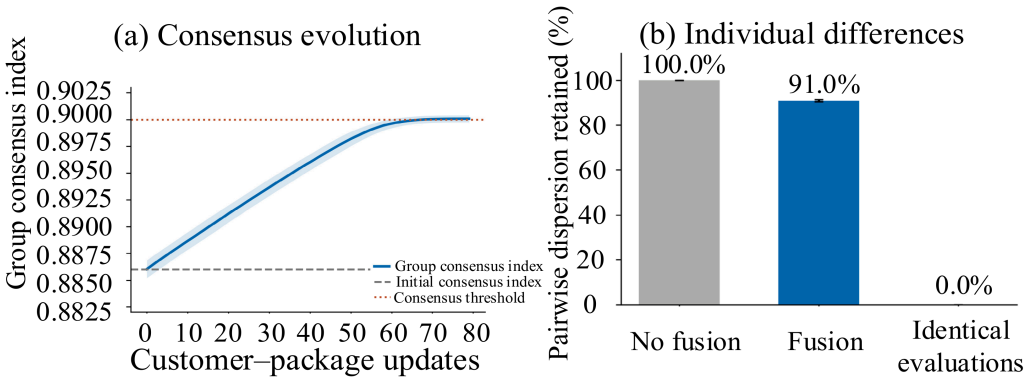

Figure 8. Controlled consensus fusion: mean GCI trajectories and dispersion retention across 30 seed-level averages; error bars denote one standard deviation.

6. Conclusions

This study proposes a dual-driven electricity retail package recommendation framework that integrates cognitive-noise filtration with dynamic consensus evolution.

In the case study, the RD-TAP algorithm maintains a clustering quality index above 0.94 for sample sizes of 500–1500. The IVIFS-based correction reduces the injected evaluation noise, and the multi-threshold fusion procedure brings the four case-study subgroups above the prescribed consensus threshold within five iterations under the baseline parameter configuration. Across the four load patterns, the complete framework achieves mean RMSE values ranging from 0.1020 to 0.1346, consistently lower than those of the variants without cognitive-noise correction or consensus optimization under the tested conditions.

The framework provides an interpretable way to connect load-profile risk with uncertain package evaluations. The current validation combines empirically selected thresholds and controlled perturbation stress tests with a held-out evaluation based on observed customer data, which provides evidence of predictive validity within the studied market. Broader cross-market and temporal validation remains necessary. Future work should estimate adjustment thresholds from data, test sensitivity across markets and time periods, and extend validation to additional customer populations and market settings.

Author Contributions: Conceptualization, X.D. and J.C.; Methodology, X.D. and J.C.; Software, X.D.; Formal analysis, X.D.; Data curation, J.C.; Writing—original draft, X.D.; Writing—review & editing, Y.M.; Visualization, J.C.; Supervision, Y.M.; Project administration, Y.M.; Funding acquisition, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China, grant number 72301248.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

U	The sample customer set
P	The set of candidate electricity retail packages
A	The set of package evaluation attributes
b_i	The portrait vector of customer u_i
ζ_i	The load fluctuation risk coefficient of customer u_i
C^*	The ensemble of candidate cluster centroids
$s^{(1)}(c_p, c_q)$	The initial similarity between candidate centroids c_p and c_q
$s^{(2)}(c_p, c_q)$	The risk-modified similarity between candidate centroids
κ	The risk adjustment factor
Z_Q	The clustering quality evaluation index
$\tilde{v}_{ijk}$	The interval-valued intuitionistic fuzzy evaluation of customer u_i
$d(\alpha, \beta)$	The multi-dimensional distance measure between two IVIFNs
$S(\alpha)$	The standardized score function for an IVIFN
Δ_{ik}	The attribute differentiation index of attribute a_k for customer u_i
Γ_{ik}	The objective influence degree of attribute a_k
ω_{ik}	The optimized attribute weight of customer u_i
$\tilde{g}_{qj}$	The mainstream consensus evaluation vector of subgroup C_q
δ_{ij}	The cognitive consensus deviation of customer u_i
η_{ij}	The nonlinear fusion coefficient for soft correction
$\tilde{v}'_{ijk}$	The adaptively corrected IVIFN evaluation of customer u_i
$\tilde{g}'_{qjk}$	The subgroup consensus vector of corrected evaluations
I_{ij}	The package-level individual consensus degree of customer u_i
I_q	The overarching subgroup consensus index (GCI)
ψ	The consensus threshold
(u_c, p_c)	The target customer–package pair with maximum deviation
λ_1,λ_2	The consensus segmentation thresholds
ϖ	The compromise retention coefficient
$\tilde{V}^*$	The optimal high-consensus evaluation matrix
u_t	The newly entering target customer
Z_{tj}	The predicted satisfaction index of customer u_t for package p_j

Appendix A

Detailed parameter derivations and conventional definitions for the index system.

Let $H_i(t)$ represent the raw continuous load reading of customer i at a time interval within the total metering horizon T.

Dimension 1. *Basic Load Characteristics*

(1) Monthly load factor (R_i^{ML}): The monthly average load is derived as

$$\overline{H}_i = \frac{1}{T} \sum_{t=1}^{T} H_i(t) \tag{A1}$$

which quantifies global demand stability to help retailers identify customers with stable baselines and lower peak-shaving costs.

(2) Monthly maximum utilization hours (M_i^{MU}): The total monthly consumption is derived as:

$$E_i = \sum_{t=1}^{T} H_i(t) \cdot \Delta t \tag{A2}$$

where Δt is introduced to represent the smart-meter sampling interval, thereby reflecting plant-operational scale for contract capacity optimization.

(3) Average peak load factor (R_i^{PL}): The daily peak-period average load is derived as:

$$H_{i,d}^{\text{peak}} = \frac{1}{\left|T_{\text{peak}}\right|} \sum_{t \in T_{\text{peak}}} x_i(d,t) \tag{A3}$$

where T_{peak} is introduced as the peak-tariff hour set, evaluating load retention or peak-shaving potential during high-tariff intervals.

(4) Average shoulder load factor (R_i^{FL}): The daily shoulder-period average load is derived as:

$$H_{i,d}^{\text{shoulder}} = \frac{1}{|T_{\text{shoulder}}|} \sum_{t \in T_{\text{shoulder}}} x_i(d,t) \tag{A4}$$

where T_{shoulder} is introduced as the transitional tariff hour set, capturing consumption consistency during mid-tier price periods.

(5)Average valley load factor (R_i^{VL}): The daily valley-period average load is derived as:

$$H_{i,d}^{\text{valley}} = \frac{1}{\left|T_{\text{valley}}\right|} \sum_{t \in T_{\text{valley}}} x_i(d,t) \tag{A5}$$

where T_{valley} is introduced as the low-price valley hour set, assessing low-tariff baseline retention for energy storage integration.

(6) Working day peak-to-valley ratio (R_i^{WD}): For any weekday $d \in D_{wd}$, the daily boundaries are extracted via localized peak searches as:

$$H_{i,d}^{\max} = \max_{1 \le h \le 24} x_i(d,h) \tag{A6}$$

$$H_{i,d}^{\min} = \min_{1 \le h \le 24} x_i(d,h) \tag{A7}$$

where h is the hourly index, capturing the extreme weekday load gradient and peak-shaving elasticity.

(7) Weekend/Holiday peak-to-valley ratio (R_i^{WE}): Formulated identically to R_i^{WD}, the day loop is strictly restricted to within the complementary non-working day set D_{we}, establishing the behavioral load contrast to separate lifestyle-driven users from rigid industrial schedules.

Dimension 2. *Load Fluctuation Risk*

(8) Monthly coefficient of variation (V_i^{ML}): The daily average load profile vector is defined as: $b_i = \left[\overline{H}_{i,1}, \ldots, \overline{H}_{i,D}\right]$ across D days, expanding the daily mean operator as: $\mu(b_i) = \frac{1}{D} \sum_{d=1}^{D} \overline{H}_{i,d}$ to capture long-term macroscopic volatility boundaries for clustering risk penalties.

Dimension 3. *Consumption Efficiency*

(9) Load efficiency factor (U_i^{LE}): The effective duration parameter is derived as:

$$T_{\text{eff}} = \sum_{t=1}^{T} f(H_i(t)) \cdot \Delta t \tag{A8}$$

where $f(\cdot)$ is introduced as a piecewise indicator function outputting 1 if $H_i(t) \geq H_{\text{threshold}}$ and 0 otherwise, evaluating effective power utilization relative to grid losses.

Dimension 4. *Period Consumption Preferences*

(10) Peak consumption share (θ_i^{peak}): The specific peak energy volume is derived as:

$$E_i^{\text{valley}} = \sum_{d=1}^{D} \sum_{t \in T_{\text{valley}}} x_i(d,t) \cdot \Delta t \tag{A9}$$

which is normalized by the total monthly volume E_i^{total} to determine the customer's financial risk exposure under steep time-of-use tariffs.

(11) Valley consumption share (θ_i^{valley}): The responsive valley energy volume is derived as:

$$E_i^{\text{valley}} = \sum_{d=1}^{D} \sum_{t \in T_{\text{valley}}} x_i(d,t) \cdot \Delta t \tag{A10}$$

which quantifies the user's load-shifting flexibility and willingness to respond to low-tariff incentives.

(12) Working day load deviation (D_i^{WD}): The historical weekday reference baseline curve is derived for each hour $h \in [1,24]$ as:

$$\overline{H}_{i,h} = \frac{1}{D_{wd}} \sum_{d \in D_{wd}} x_i(d,h) \tag{A11}$$

which measures micro-level day-to-day behavioral uncertainty and direct exposure to market deviation settlement penalties.

Appendix B

Comprehensive index derivations and standard cyclic update formulas for RD-TAP clustering.

(1) Initial Similarity Measure (Pearson Correlation):

During the local dimensionality reduction stage, the linear correlation between customer profiling vectors is evaluated. The Pearson correlation coefficient is derived as:

$$r(u_i,u_j) = \frac{\sum_{d=1}^{G} \left(b_{i,d} - \overline{b}_i\right)\left(b_{j,d} - \overline{b}_j\right)}{\sqrt{\sum_{d=1}^{G} \left(b_{i,d} - \overline{b}_i\right)^2} \sqrt{\sum_{d=1}^{G} \left(b_{j,d} - \overline{b}_j\right)^2}} \tag{A12}$$

where $b_{i,d}$ is the normalized value of the d-th profiling label for customer u_i, $\overline{b}_i$ is the mean value, and $G = 12$. Define the similarity as a negative correlation coefficient $s^{(1)}(u_i,u_j) = r(u_i,u_j)$ and take the median of the similarity among all customers in the subset is taken as the initial self-similarity of each customer $s^{(1)}(u_i,u_i)$.

(2) Clustering Quality Evaluation Index

To objectively evaluate the clustering quality, a comprehensive evaluation index that balances intra-class cohesion and inter-class separation is defined as:

$$Z_Q = \frac{1}{|\mathcal{U}|} \sum_{u_i \in \mathcal{U}} \left(\frac{d_{\text{inter}}(u_i) - d_{\text{intra}}(u_i)}{d_{\text{inter}}(u_i) + d_{\text{intra}}(u_i)} \right) \tag{A13}$$

where $d_{inter}(u_i)$ is the average distance between customer u_i and all customers not in its class, and $d_{intra}(u_i)$ is the average distance between customer u_i and other customers in its class. The larger the Z_Q value, the better the clustering effect.

(3) Adaptive Self-Similarity Update Mechanism

To reduce the impact of initial self-similarity on the number of clusters, an adaptive update mechanism is formulated as:

$$s^{(2)}\left(c_p, c_p\right)^{(t+1)} = s^{(2)}\left(c_p, c_p\right)^{(t)} + \lambda \cdot \left(Q_{\text{target}} - Q^{(t)}\right) \tag{A14}$$

where λ is the preset step size, $Q^{(t)}$ is the number of clusters in the t-th iteration, and Q_{target} is the median of the target cluster number range. Risk-driven clustering, quality evaluation, and self-similarity updates are executed cyclically until the number of clusters stabilizes or the Φ value converges. Finally, the number of clusters that maximizes Φ is selected as the optimal partition Q^*.

Appendix C

IVIFN-based evaluation and attribute weight optimization.

Appendix C.1

IVIFN definition and multi-dimensional distance formulation.

(1) Let the customer set be $U = \{u_1, u_2, \ldots, u_N\}$, the electricity retail package set be $P = \{p_1, p_2, \ldots, p_M\}$, and the package attribute set be A = $\{a_1, a_2, \ldots, a_L\}$. The evaluation of customer u_i regarding package p_j on attribute a_k is represented by an interval-valued intuitionistic fuzzy number (IVIFN) as $\tilde{v}_{ijk} = \left(\left[\mu_{ijk}^L, \mu_{ijk}^U\right], \left[\nu_{ijk}^L, \nu_{ijk}^U\right]\right)$, where $\left[\mu_{ijk}^L, \mu_{ijk}^U\right] \subseteq [0,1]$ and $\left[\nu_{ijk}^L, \nu_{ijk}^U\right] \subseteq [0,1]$, represent the membership and non-membership intervals, respectively, satisfying $\mu_{ijk}^U + \nu_{ijk}^U \leq 1$.

The hesitancy interval is defined as:

$$\left[\pi_{ijk}^L, \pi_{ijk}^U\right] = \left[1 - \mu_{ijk}^U - \nu_{ijk}^U, 1 - \mu_{ijk}^L - \nu_{ijk}^L\right] \tag{A15}$$

This definition facilitates the rigorous quantification of user uncertainty regarding specific attributes like deviation settlement and deviation charges, effectively capturing the hesitancy interval associated with complex ToU sales packages.

(2) To accurately measure the fine-grained disparities between individual subjective evaluations, let us denote any two generalized IVIFNs as $\alpha = \left(\left[\mu_\alpha^L, \mu_\alpha^U\right], \left[\nu_\alpha^L, \nu_\alpha^U\right]\right)$ and $\beta = \left(\left[\mu_\beta^L, \mu_\beta^U\right], \left[\nu_\beta^L, \nu_\beta^U\right]\right)$. The overall multi-dimensional distance measure $d(\alpha, \beta)$ is defined in Equation (3) of the main text as the arithmetic mean of three specific deviation components $f1, f2$, and $f3$.

(3) $f1, f2$, and $f3$ mathematically capture the evaluation deviation across three distinct dimensions:

i. Direct Deviation ($f1$): Measures the fundamental geometric differences between direct membership, non-membership, and hesitancy using a normalized Manhattan distance framework:

$$f_1(\alpha, \beta) = \frac{1}{6}\left(\begin{array}{l} \left|\mu_\alpha^L - \mu_\beta^L\right| + \left|\mu_\alpha^U - \mu_\beta^U\right| + \\ \left|\nu_\alpha^L - \nu_\beta^L\right| + \left|\nu_\alpha^U - \nu_\beta^U\right| + \\ \left|\pi_\alpha^L - \pi_\beta^L\right| + \left|\pi_\alpha^U - \pi_\beta^U\right| \end{array}\right) \tag{A16}$$

ii. Cognitive Offset ($f2$): Integrates indirect membership (the balance point of support and opposition) to reflect the implicit cognitive biases of customers in the evaluation of electricity sales packages:

$$f_2(\alpha,\beta)=\frac{1}{2}\left(\begin{array}{|l}\left|\frac{\mu_\alpha^L+1-\nu_\alpha^L}{2}-\frac{\mu_\beta^L+1-\nu_\beta^L}{2}\right|+\\ \left|\frac{\mu_\alpha^U+1-\nu_\alpha^U}{2}-\frac{\mu_\beta^U+1-\nu_\beta^U}{2}\right|\end{array}\right) \tag{A17}$$

iii. Cross-Impact Factor ($f3$): Introduces crossover multiplication factors to capture the information interaction bias:

$$f_3(\alpha,\beta)=\frac{1}{4}\left(\begin{array}{|l}\left|\min\left\{\mu_\alpha^L,\nu_\beta^L\right\}-\min\left\{\mu_\beta^L,\nu_\alpha^L\right\}\right|+\\ \left|\max\left\{\mu_\alpha^L,\nu_\beta^L\right\}-\max\left\{\mu_\beta^L,\nu_\alpha^L\right\}\right|+\\ \left|\min\left\{\mu_\alpha^U,\nu_\beta^U\right\}-\min\left\{\mu_\beta^U,\nu_\alpha^U\right\}\right|+\\ \left|\max\left\{\mu_\alpha^U,\nu_\beta^U\right\}-\max\left\{\mu_\beta^U,\nu_\alpha^U\right\}\right|\end{array}\right) \tag{A18}$$

(4) Finally, to convert the interval-valued intuitionistic fuzzy evaluations into quantitative single values for package ranking, the score function $S(\alpha)$ is defined as:

$$S(\alpha)=\frac{\mu_\alpha^L+\mu_\alpha^U-\nu_\alpha^L-\nu_\alpha^U+\left(1-\pi_\alpha^L\right)+\left(1-\pi_\alpha^U\right)}{4} \tag{A19}$$

A higher score $S(\alpha)$ indicates a higher level of customer acceptance for the corresponding electricity sales package.

Appendix C.2

Procedures for dual-driven attribute weight optimization.

Detailed steps to optimize attribute weights under incomplete information constraints and determine the influence degrees are as follows.

(1) Calculate the sum of deviations for different electricity retail package evaluations under each attribute using the multi-dimensional distance measure. This attribute differentiation index Δ_{ik} is formulated as:

$$\Delta_{ik}=\sum_{j=1}^{M}\sum_{j'=1}^{M}d\left(\tilde{v}_{ijk},\tilde{v}_{ij'k}\right),k=1,2,\ldots,L \tag{A20}$$

(2) Construct an optimization model combined with the set of incomplete weight information Ω_i provided by the customer to solve for the initial subjective weights $\omega_i^{(0)}=\left(\omega_{i1}^{(0)},\omega_{i2}^{(0)},\ldots,\omega_{iL}^{(0)}\right)^T$:

$$\begin{aligned}&\max\sum_{k=1}^{L}\Delta_{ik}\omega_{ik}^{(0)}\\ &\text{s.t.}\sum_{k=1}^{L}\omega_{ik}^{(0)}=1,\\ &\omega_{ik}^{(0)}\geq 0,k=1,2,\ldots,L,\\ &\omega_i^{(0)}\in\Omega_i\end{aligned} \tag{A21}$$

This maximizes the sum of the discriminations of attributes while satisfying the customer's subjective preferences.

(3) Aggregate all attributes using the interval intuitionistic fuzzy weighted average (IIFWA) operator to obtain the overall evaluation $\tilde{z}_{ij}$ of customer u_i for package p_j:

$$\tilde{z}_{ij}=\oplus_{k=1}^{L}\omega_{ik}^{(0)}\tilde{v}_{ijk} \tag{A22}$$

where $\oplus$ denotes the interval intuitionistic fuzzy weighted average operator, which is defined as follows:

$$\oplus_{k=1}^{L}\omega_{ik}\tilde{v}_{ijk} = \left(\begin{array}{c} \left[1-\prod_{k=1}^{L}\left(1-\mu_{ijk}^{L}\right)^{\omega_{ik}}, 1-\prod_{k=1}^{L}\left(1-\mu_{ijk}^{U}\right)^{\omega_{ik}}\right], \\ \left[\prod_{k=1}^{L}\left(\nu_{ijk}^{L}\right)^{\omega_{ik}}, \prod_{k=1}^{L}\left(\nu_{ijk}^{U}\right)^{\omega_{ik}}\right] \end{array} \right) \tag{A23}$$

(4) Assess the objective influence degree of each attribute using the "Leave-one-out" method. Attribute a_k is sequentially masked, and the comprehensive evaluation $\tilde{z}_{ij}^{(-k)}$ is recalculated after normalizing the remaining weights. The objective influence degree Γ_{ik} is defined as:

$$\Gamma_{ik} = \frac{1}{M}\sum_{j=1}^{M} d\left(\tilde{z}_{ij}, \tilde{z}_{ij}^{(-k)}\right) \tag{A24}$$

Ultimately, the optimized attribute weights ω_{ik} are obtained by synergistically fusing the initial weights with Γ_{ik}, as explicitly formulated in Equation (4) of the main text.

Appendix D

Consensus evolution and adaptive adjustment mechanism.

Appendix D.1

Derivations of consensus distance aggregation and piecewise deviation amplification.

Detailed steps and mathematical formulations regarding the consensus distance aggregation, piecewise deviation amplification rules, and unconfident set judgment conditions are as follows.

(1) Extract the mainstream consensus evaluation matrix for each subgroup C_q. For package p_j, the subgroup consensus evaluation vector $\tilde{g}_{qj}$ is defined via interval intuitionistic fuzzy weighted aggregation:

$$\tilde{g}_{qj} = \oplus_{u_i \in C_q}\phi_i\tilde{\mathbf{v}}_{ij} \tag{A25}$$

where $\tilde{\mathbf{v}}_{ij} = \left(\tilde{v}_{ij1}, \tilde{v}_{ij2}, \ldots, \tilde{v}_{ijL}\right)$ is the evaluation vector of customer u_i for package p_j. ϕ_i is the customer weight within the subgroup (equal weight $\phi_i = 1/|C_q|$ in this article).

(2) To identify whether customer u_i has any unconfident behavior towards package p_j, the weighted distance between the individual evaluation and the subgroup consensus is calculated:

$$\delta_{ij} = \sum_{k=1}^{L}\omega_{ik}\cdot d\left(\tilde{v}_{ijk}, \tilde{g}_{c(u_i)j}^{(k)}\right) \tag{A26}$$

where ω_{ik} is the weight of customer u_i to attribute a_k, and $\tilde{g}_{c(u_i)j}^{(k)}$ is the component of the subgroup consensus evaluation on the k-th attribute.

(3) Define the adjusted deviation coefficient to magnify the difference value to highlight the potential lack of confidence:

$$\hat{\delta}_{ij} = \begin{cases} \rho_1 \times (\delta_{ij})^{0.4}, & 0 < \delta_{ij} \le \varepsilon \\ \rho_2 \times (\delta_{ij})^{0.8}, & \varepsilon < \delta_{ij} \le \tau \\ \delta_{ij}, & \delta_{ij} > \tau \end{cases} \tag{A27}$$

where ρ_1 and ρ_2 are the adjustment coefficients for the low and medium deviation intervals, respectively; ε is the segmentation threshold for the low deviation interval, and τ is the deviation significance judgment threshold. If $\delta_{ij} > \tau$, then it is determined that customer u_i exhibits unconfident behavior in the evaluation of package p_j.

(4) Establish the unconfident set judgment conditions. If $\delta_{ij} > \tau$, then it is determined that customer u_i exhibits unconfident behavior in the evaluation of package p_j. The specific unconfident set $\mathcal{N}_j$ is defined as:

$$\mathcal{N}_j = \{u_i \mid \delta_{ij} > \tau, i = 1, \ldots, N\} \tag{A28}$$

(5) For the identified unconfident evaluations within the set $\mathcal{N}_j$, the non-linear fusion coefficient η_{ij} used for soft correction is explicitly derived as:

$$\eta_{ij} = \frac{\dot{\delta}_{ij}}{\tau + \dot{\delta}_{ij}} \tag{A29}$$

This coefficient dynamically guides the robust reconstruction of the initial evaluation vector, heavily pulling extreme outliers toward the subgroup's collective rationality as explicitly formulated in Equation (5) of the main text.

Appendix D.2

Subgroup consensus aggregation and index averaging.

To quantify the consensus level within a customer subgroup, the following sequential steps are performed.

(1) For attribute a_k of package p_j, the subgroup consensus evaluation $\tilde{g}'_{qjk}$ of subgroup C_q is defined as:

$$\tilde{g}'_{qjk} = \oplus_{u_i \in C_q} \phi_i \tilde{\mathbf{v}}'_{ijk} \tag{A30}$$

where $\tilde{\mathbf{v}}'_{ijk}$ denotes the corrected interval-valued intuitionistic fuzzy evaluation of customer u_i for attribute a_k of package p_j. ϕ_i is the weight assigned to customer u_i.

(2) The individual consensus degree I_{ij} of customer u_i for package p_j is defined in Equation (6) of the main text using the multi-dimensional distance $d(\alpha, \beta)$ and the optimized attribute weights ω_{ik}.

(3) By averaging I_{ij} over all candidate packages, the global individual consensus index I_i of customer u_i for the entire package set P is obtained as:

$$I_i = \frac{1}{M} \sum_{j=1}^{M} I_{ij} \tag{A31}$$

where M is the total number of electricity retail packages.

(4) Finally, the overall consensus index I_q of subgroup C_q is calculated as:

$$I_q = \sum_{u_i \in C_q} \phi_i I_i \tag{A32}$$

Appendix E

Recommendation process and parameter configuration.

Appendix E.1

Similarity aggregation and minimum-value search.

(i) The overall evaluation similarity $s(u_i, u_h)$ between customer u_i and other customers u_h is evaluated as follows:

$$s(u_i, u_h) = \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{L} w_{ik} s(\tilde{v}'_{ijk}, \tilde{v}'_{hjk}) \tag{A33}$$

(ii) Then, the overall similarity level $\bar{s}(u_i)$ of customer u_i within the subgroup is evaluated as follows:

$$\bar{s}(u_i) = \frac{1}{|C_q| - 1} \sum_{u_h \in C_q, h \neq i} s(u_i, u_h) \tag{A34}$$

(iii) The customer u_c with the lowest similarity score is selected first for adjustment:

$$u_c = \underset{u_i \in C_q}{\operatorname{argmin}}\, \bar{s}(u_i) \tag{A35}$$

(iv) For the customer u_c with the lowest similarity score, the similarity $S(u_c, p_j)$ is evaluated across each electricity retail packages p_j as follows:

$$S(u_c, p_j) = \frac{1}{|C_q| - 1} \sum_{u_h \in C_q, h \neq c} \sum_{k=1}^{L} w_{ck} s(\tilde{v}'_{cjk}, \tilde{v}'_{hjk}) \tag{A36}$$

(v) By comparison, the electricity retail packages p_c with the lowest similarity and the largest cognitive deviation is identified:

$$p_c = \underset{p_j \in P}{\operatorname{argmin}}\, S(u_c, p_j) \tag{A37}$$

Appendix E.2

Boundary search arithmetic.

Once the target pair (u_h, p_c) is identified, the two boundary reference customers are located as follows. The customer with the highest similarity for p_c is:

$$u_{\max} = \underset{u_h \in C_q, h \neq c}{\operatorname{argmax}}\ S(u_h, p_c) \tag{A38}$$

and the customer with the lowest similarity (excluding u_c) is:

$$u_{\min} = \underset{u_h \in C_q, h \neq c}{\operatorname{argmin}}\ S(u_h, p_c) \tag{A39}$$

Their evaluation vectors $\tilde{v}_{u_{\max}, p_c}$ and $\tilde{v}_{u_{\min}, p_c}$ form the reference fusion space for the piecewise adjustment in Equation (7) of the main text, where $\varpi \in [0, 1]$ is the compromise retention coefficient. The adjusted evaluation is substituted back, I_q is recalculated, and the cycle repeats until $I_q \geq \psi$, yielding the final high-consensus evaluation matrix.

Appendix E.3

Similarity weighting, score conversion, and aggregation logic.

- Step 1: Based on the set of cluster centers $C = \{c_1, c_2, \ldots, c_K\}$ output by the RD-TAP algorithm, the target subgroup C_q with the highest similarity to the new customer is determined:

$$C_q = \underset{c_k \in C}{\operatorname{argmax}}\, \mathrm{r}(u_t, c_k) \tag{A40}$$

- Step 2: For sample customer u_i in subgroup C_q, the interval-valued intuitionistic fuzzy evaluation $\tilde{v}^*_{ijk}$ for package p_j is converted to a single numerical score using the score function $S(\alpha)$:

$$y_{ij} = \sum_{k=1}^{L} w_{ik} S(\tilde{v}^*_{ijk}) \tag{A41}$$

- Step 3: The feature similarity weight γ_{ti} between the target customer u_t and sample customer u_i is calculated using the normalized Pearson correlation coefficient:

$$\gamma_{ti} = \frac{1 + r(u_t, u_i)}{2} \tag{A42}$$

where $r(u_t, u_i) \in [-1, 1]$ represents the Pearson correlation coefficient between the portrait vectors of customers u_t and u_i.

- Step 4: The predicted satisfaction Z_{tj} of target customer u_t for package p_j is obtained by weighted aggregation of the sample customers' scores:

$$Z_{tj} = \frac{\sum_{u_i \in C_q} \gamma_{ti} y_{ij}}{\sum_{u_i \in C_q} \gamma_{ti}} \tag{A43}$$

Appendix E.4

Parameter configuration for the empirical market scenario.

To ensure the stability of the unconfident behavior correction and the subsequent dynamic consensus evolution process, the key algorithmic parameters governing both mechanisms are empirically initialized prior to the case study. Specifically, the model parameters are uniformly specified as follows: the adjustment coefficients for the low- and medium-deviation intervals are set to $\rho_1 = 0.313$ and $\rho_2 = 0.786$, respectively; the thresholds for unconfident behavior correction are set to $\varepsilon = 0.1$ and $\tau = 0.3$; the overall consensus threshold is set to $\psi = 0.9$; the segmentation thresholds for dynamic adjustment are set to $\lambda_1 = 0.35$ and $\lambda_2 = 0.75$; and the compromise retention coefficient is set to $\varpi = 0.4$. These values define the baseline parameter configuration. The risk-adjustment coefficient κ was selected by five-fold cross-validation on the development customers. The correction coefficients, correction thresholds, and segmentation constants were fixed before the final evaluation according to the definitions of the corresponding deviation intervals and dynamic-adjustment rule; they were not tuned using the held-out customers. Section 5.6 reports the one-factor sensitivity of the segmentation threshold τ, consensus threshold ψ, and compromise retention coefficient, together with the resulting RMSE, consensus iterations, subgroup attainment rates, and final package-ranking stability.

References

1. Celebi, E.; Fuller, J.D. Time-of-use pricing in electricity markets under different market structures. *IEEE Trans. Power Syst.* **2012**, *27*, 1170–1181. [CrossRef]
2. Mulder, M.; Zomer, S.P.E. Contribution of green labels in electricity retail markets to fostering renewable energy. *Energy Policy* **2016**, *99*, 100–109. [CrossRef]
3. Jin, M.; Feng, W.; Marnay, C.; Spanos, C. Microgrid to enable optimal distributed energy retail and end-user demand response. *Appl. Energy* **2018**, *210*, 1321–1335. [CrossRef]
4. Guo, H.; Davidson, M.R.; Chen, Q.; Zhang, D.; Jiang, N.; Xia, Q.; Kang, C.; Zhang, X. Power market reform in China: Motivations, progress, and recommendations. *Energy Policy* **2020**, *145*, 111717. [CrossRef]
5. Cheng, X.; Liu, P.; Zhu, L. The impact of electricity market reform on renewable energy production. *Energy Policy* **2024**, *194*, 114306. [CrossRef]
6. Belton, C.A.; Lunn, P.D. Smart choices? An experimental study of smart meters and time-of-use tariffs in Ireland. *Energy Policy* **2020**, *140*, 111243. [CrossRef]
7. Esplin, R.; Best, R.; Scranton, J.; Chai, A. Who pays the loyalty tax? The relationship between socioeconomic status and switching in Australia's retail electricity markets. *Energy Policy* **2022**, *164*, 112904. [CrossRef]
8. Li, S.; Luo, F.; Yang, J.; Ranzi, G.; Wen, J. A personalized electricity tariff recommender system based on advanced metering infrastructure and collaborative filtering. *Int. J. Electr. Power Energy Syst.* **2019**, *113*, 403–410. [CrossRef]
9. Qiu, D.; Wang, Y.; Wang, J.; Jiang, C.; Strbac, G. Personalized retail pricing design for smart metering consumers in electricity market. *Appl. Energy* **2023**, *348*, 121545. [CrossRef]

10. Michalakopoulos, V.; Sarmas, E.; Papias, I.; Skaloumpakas, P.; Marinakis, V.; Doukas, H. A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs. *Appl. Energy* **2024**, *361*, 122943. [CrossRef]
11. Vahedi, S.; Zhao, L. Distributed auto-clustering for residential load profiling using AMI data from the U.S. High Plains. *IEEE Trans. Smart Grid* **2023**, *14*, 4530–4541. [CrossRef]
12. Yilmaz, S.; Chambers, J.; Patel, M.K. Comparison of clustering approaches for domestic electricity load profile characterization—Implications for demand side management. *Energy* **2019**, *180*, 665–677. [CrossRef]
13. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *42*, 68–80. [CrossRef]
14. McLoughlin, F.; Duffy, A.; Conlon, M. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Appl. Energy* **2015**, *141*, 190–199. [CrossRef]
15. Lee, E.; Baek, K.; Kim, J. Customer targeting for load flexibility via resident behavior segmentation. *IEEE Trans. Smart Grid* **2024**, *15*, 1574–1583. [CrossRef]
16. Meng, F.; Ma, Q.; Liu, Z.; Zeng, X.-J. Multiple dynamic pricing for demand response with adaptive clustering-based customer segmentation in smart grids. *Appl. Energy* **2023**, *333*, 120626. [CrossRef]
17. Zhang, X.; Ramírez-Mendiola, J.L.; Li, M.; Guo, L. Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study. *Appl. Energy* **2022**, *308*, 118335. [CrossRef]
18. Li, R.; Li, F.; Smith, N.D. Multi-resolution load profile clustering for smart metering data. *IEEE Trans. Power Syst.* **2016**, *31*, 4473–4482. [CrossRef]
19. Luo, F.; Ranzi, G.; Wang, X.; Dong, Z.Y. Social information filtering-based electricity retail plan recommender system for smart grid end users. *IEEE Trans. Smart Grid* **2019**, *10*, 95–104. [CrossRef]
20. Zhang, Y.; Meng, K.; Kong, W.; Dong, Z.Y.; Qian, F. Bayesian hybrid collaborative filtering-based residential electricity plan recommender system. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4731–4741. [CrossRef]
21. Zhao, P.; Dong, Z.Y.; Meng, K.; Kong, W.; Yang, J. Household power usage pattern filtering-based residential electricity plan recommender system. *Appl. Energy* **2021**, *298*, 117191. [CrossRef]
22. Cho, Y.-H.; Chae, J.-G.; Lee, D. Similarity-based optimization framework for curtailment service providers through collaborative filtering and generalized dynamic factor model. *IEEE Trans. Smart Grid* **2023**, *14*, 1056–1069. [CrossRef]
23. Ma, Y.; Zheng, R.; Lu, Y.; Zhang, Z.; Wang, Y.; Lin, Z.; Yang, L.; Liang, H.; Liu, P.X. Electricity retail plan recommendation method based on multigranular hesitant fuzzy sets and an improved non-negative latent factor model. *IEEE Trans. Energy Mark. Policy Regul.* **2024**, *2*, 146–161. [CrossRef]
24. Yang, J.; Zhao, J.; Wen, F.; Dong, Z.Y. A framework of customizing electricity retail prices. *IEEE Trans. Power Syst.* **2018**, *33*, 2415–2428. [CrossRef]
25. Zugno, M.; Morales, J.M.; Pinson, P.; Madsen, H. A bilevel model for electricity retailers' participation in a demand response market environment. *Energy Econ.* **2013**, *36*, 182–197. [CrossRef]
26. Jiang, Y.; Fang, D.; Lei, L. Green electricity product menu design for retailers without knowing consumer environmental awareness. *Energy Econ.* **2024**, *139*, 107928. [CrossRef]
27. Wang, P.; Fang, D.; Cao, G. How social learning affects customer behavior under the implementation of TOU in the electricity retailing market. *Energy Econ.* **2022**, *106*, 105836. [CrossRef]
28. Fang, D.; Wang, P. Optimal real-time pricing and electricity package by retail electric providers based on social learning. *Energy Econ.* **2023**, *117*, 106442. [CrossRef]
29. Saumweber, A.; Wederhake, L.; Cardoso, G.; Fridgen, G.; Heleno, M. Designing Pareto optimal electricity retail rates when utility customers are prosumers. *Energy Policy* **2021**, *156*, 112339. [CrossRef]
30. Vaughan, J.; Doumen, S.C.; Kok, K. Empowering tomorrow, controlling today: A multi-criteria assessment of distribution grid tariff designs. *Appl. Energy* **2023**, *341*, 121053. [CrossRef]
31. Ancel, J. Tariffs time-dynamics in competitive electricity retail markets with differentiated consumer reactions. *Energy Econ.* **2025**, *148*, 108615. [CrossRef]
32. Jin, F.; Gao, X.; Zhou, L. Bounded rationality consensus reaching process with regret theory and weighted moment estimation for multi-attribute group decision making. *Inf. Fusion* **2025**, *115*, 102778. [CrossRef]
33. Mathebula, J.; Mbuli, N. Application of TOPSIS for multi-criteria decision analysis (MCDA) in power systems: A systematic literature review. *Energies* **2025**, *18*, 3478. [CrossRef]
34. Rani, P.; Mishra, A.R.; Deveci, M.; Alrasheedi, A.F.; Alshamrani, A.M.; Pedrycz, W. Interval-valued intuitionistic fuzzy Yager power operators and possibility degree-based group decision-making model. *Cogn. Comput.* **2025**, *17*, 37. [CrossRef]
35. Wang, C.; Zhou, H.; Dinçer, H.; Yüksel, S.; Ubay, G.G.; Uluer, G.S. Analysis of electricity pricing in emerging economies with hybrid multi-criteria decision-making technique based on interval-valued intuitionistic hesitant fuzzy sets. *IEEE Access* **2020**, *8*, 190882–190896. [CrossRef]

36. Wan, S.-P.; Dong, J.-Y.; Zhang, Z.-H. Two-stage consensus reaching process in social network large group decision-making with application to battery supplier selection. *Inf. Sci.* **2024**, *668*, 120526. [CrossRef]
37. Yang, W.; Zhang, L.; Shi, J.; Lin, R. New consensus reaching process with minimum adjustment and feedback mechanism for large-scale group decision making problems under social trust networks. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108230. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.