Optimization of Product Recommendation Strategies on E-commerce Platforms Based on Association Rule Mining

Baicheng Jiang

Shandong Weifang NO.1 Middle School, Weifang, China

13806368656@163.com

Keywords: Association Rule Mining; E-commerce Product Recommendation; Strategy Optimization; Data-Driven; Precise Matching

Abstract: This paper explores the optimization of product recommendation strategies on e-commerce platforms based on association rule mining. It first outlines the core concepts of association rule mining, the basic logic of e-commerce recommendations, and the value of their integration. It then analyzes existing issues in e-commerce recommendations, such as insufficient accuracy and serious homogenization. Subsequently, it proposes optimization directions from four aspects: improving the quality of data foundations, enhancing rule effectiveness, strengthening the matching degree between rules and user demands, and establishing a real-time iteration mechanism. Implementation paths are provided in terms of data, algorithms, recommendation logic, and dynamic updates. This strategy, by leveraging data-driven mining of implicit product associations, effectively addresses the shortcomings of traditional recommendations and achieves a two-way optimization of user needs and platform growth.

1. Basic Overview of Association Rule Mining and E-commerce Product Recommendation

1.1. Core Concepts of Association Rule Mining

1.1.1. Definition of Association Rules

Association rules describe the regular co-occurrence relationships of items (or events) within a dataset. The core purpose is to uncover hidden associations such as "if A occurs, then B is likely to occur." For instance, in an e-commerce context, transaction data might reveal a pattern like "users who purchase mobile phones are likely to buy phone cases." Association rules do not rely on subjective judgment but instead quantify the strength of item associations through data [1].

These may include intuitive combinations like "toothbrush and toothpaste," as well as cross-category implicit associations such as "strollers and car seats." Essentially, it utilizes data to identify valuable connections from scattered behaviors, providing a basis for predicting users' potential needs.

1.1.2. Core Ideas of Association Rule Mining

The core idea of association rule mining is to automatically discover valuable hidden relationships among items (or events) from massive datasets, rather than depending on human experience [2].

Its logic follows these steps: first, identify frequent itemsets from the data (e.g., e-commerce transaction records) — item combinations that appear together frequently (such as "mobile phone + screen protector"); then analyze the strength of association such as "if A occurs, B also occurs" based on these itemset to extract reliable rules (e.g., "60% of users who buy A also buy B").

The essence is to replace subjective speculation with data-driven insights, uncovering potential relationships beyond intuitive understanding from fragmented behaviors. This transforms item associations from "experience-based judgment" into "quantifiable and verifiable" patterns, thereby providing a foundation for accurate recommendations.

1.1.3. Common Basic Algorithms

The core of commonly used basic algorithms for association rule mining (such as Apriori and FP-Growth) lies in screening valuable association rules through quantitative metrics. The following are

DOI: 10.25236/gemmsd.2025.095

the formulas of key evaluation metrics, which serve as the core criteria for these algorithms:

1) Support: Measures the frequency of an itemset (used to identify frequent itemsets)

Support refers to the probability that a given itemset appears in all transactions, serving as the core indicator of whether the itemset occurs frequently.

Formula: $\text{Support}(\text{A} \cup \text{B}) = \frac{\text{Number of transactions containing both A and B}}{\text{Total number of valid transactions}}$

Notation explanation:

A, B represent different items (e.g., "mobile phone", "phone case");

AB refers to the itemset that contains both A and B;

Numerator: number of orders that include both A and B;

Denominator: total number of valid orders on the platform.

2) Confidence: Measures the reliability of a rule (used to determine whether the association is credible)

Confidence refers to the probability that a user who purchased item A also purchased item B, indicating the reliability of the rule A → B.

Formula: $\text{Confidence}(\text{A} \rightarrow \text{B}) = \frac{\text{Number of transactions containing both A and B}}{\text{Number of transactions containing A}} = \frac{\text{Support(A}\cup\text{B)}}{\text{Support(A)}}$

Notation explanation:

AB represents the rule "if A is purchased, B is likely to be purchased";

Numerator: same as support — number of transactions containing both A and B;

Denominator: number of transactions containing only A, regardless of B.

3) Lift: Measures the actual value of a rule (used to filter valuable associations)

Lift refers to the extent to which recommending B through the rule A → B outperforms randomly recommending B. A lift greater than 1 indicates that the rule is valuable (stronger association than random).

Formula: $\text{Lift}(\text{A} \rightarrow \text{B}) = \frac{\text{Confidence(A}\rightarrow\text{B)}}{\text{Support(B)}} = \frac{\text{Support(A}\cup\text{B)}}{\text{Support(A)}\times\text{Support(B)}}$

Notation explanation:

Numerator: confidence of the rule A → B;

Denominator: support of item B in all transactions (i.e., the probability of randomly purchasing B);

Interpretation:

>1: A and B are positively associated, the rule is valuable;

=1: no association;

<1: negative association (not recommended).

These formulas are the core basis for algorithms like Apriori to screen effective association rules. By calculating support, confidence, and lift, valuable associations with real recommendation value can be extracted from massive datasets.

1.2. Basic Logic of Product Recommendation on E-commerce Platforms

The core logic of product recommendation on e-commerce platforms is “the precise connection between demand and supply”—that is, through data analysis and rule design, enabling the right product to reach the right user in the right context.

This underlying logic can be divided into three dimensions: the first is Preference Capturing Based on User Behavior. By recording user behaviors such as browsing, purchasing, and saving to favorites, the system extracts user preferences for categories, brands, and prices. For example, if a user frequently browses running shoes, the system infers a related demand and recommends similar products; the second is Association Matching Based on Product Attributes. By analyzing product features such as category, function, and usage scenario, the system uncovers complementary (e.g., “phone and phone case”) or substitute (e.g., “different brands of laundry detergent”) relationships, enabling cross-recommendations between products; the third is Dynamic Adaptation Based on Context. Recommendation directions are adjusted based on context factors such as time (e.g., holidays, seasons) and user status (e.g., new users, members). For instance, winter prompts recommendations for warm clothing, while popular low-priced items may be recommended to new users.

These logics work together to meet both explicit and latent user needs, ultimately achieving the goal of "less searching by users, more product exposure, and increased conversions for the platform."

1.3. Value of Integrating Association Rule Mining with E-commerce Recommendations

The core value of integrating association rule mining with e-commerce recommendations lies in using data-driven associative patterns to overcome the limitations of traditional recommendation systems, upgrading from “experience-based judgment” to “precise matching [3].” From a recommendation logic perspective, association rules can directly capture the natural relationships between products, making recommendations more aligned with real consumption scenarios. For example, discovering from data that “70% of users who buy ovens also buy baking molds” is more relevant than merely recommending “the same brand of ovens,” effectively activating latent purchase intentions.

From an implementation perspective, association rules are generated from objective transaction data, reducing subjective bias in manual operations and covering more granular user needs. It can identify both intuitive associations like “mobile phone + phone case” and cross-category implicit associations like “crib + mosquito repellent patch,” enabling the accurate capture of niche or cross-scenario demands.

From an effectiveness perspective, this integration enhances the incremental value of recommendations: not only accelerating the conversion of existing needs but also guiding users to discover new needs through associations. It ultimately achieves a dual optimization where “users spend less time searching for products” and “the platform facilitates more transactions,” allowing the recommendation system to understand both users' explicit preferences and their unspoken potential demands.

2. Current Situation and Problems of Product Recommendation Strategies on E-commerce Platforms

2.1. Common Models of Current E-commerce Recommendation Strategies

Table 1 Summary of Common Recommendation Strategy Models on Current E-commerce Platforms

Recommendation Model	Recommendation Basis	Typical Scenario Description
Personalized Recommendation	User historical behavior (browsing, purchasing, favoriting, etc.)	Homepage “You May Also Like” (e.g., after browsing dresses, recommend similar women’s clothing)
Popularity-based Recommendation	Overall product popularity (sales volume, click rate, growth rate, etc.)	“Top Sales Rankings,” “New Hot Items,” “Trending Products”
Scenario-based Recommendation	Specific scenarios (festivals, usage purposes, etc.), manually curated	“Mid-Autumn Gift Box Zone,” “Back-to-School Stationery Set,” “Home Cleaning Kit”
Similar Product-based Recommendation	Similarity in product category, price range, and function	When browsing a certain brand of shampoo, recommend shampoos with similar function from other brands

As shown in Table 1, current e-commerce recommendation strategies mainly revolve around "visible user behaviors" and "basic product attributes", and typically follow three common models: The first is personalized recommendation, generated based on a user’s historical behaviors (such as browsing, purchasing, and favoriting records). It pushes similar products by analyzing user preference tags [4]. For example, after a user browses dresses, the homepage's "You May Also Like" section continues to recommend women's clothing of similar style. The second is popularity-based recommendation, which focuses on the overall popularity of products, such as "Top Sales Rankings," "Hot Searches," and "New Popular Items." It targets general user needs and suits new users or browsing scenarios without clear purchasing intentions. The third is scenario-based recommendation, in which products are combined by manual operation according to specific scenarios, such as "Mid-Autumn Gift Box Zone," "Back-to-School Stationery Set," or "Home Cleaning Kit." These use scenario tags to guide consumption.

There is also similar product-based recommendation, which recommends products of the same category and price range as the one the user is currently viewing. For example, when browsing a certain brand of shampoo, the system recommends shampoos with the same function from other brands. These models mostly rely on a single data dimension, with direct matching as the core logic.

2.2. Core Issues of the Existing Recommendation Strategies

The core problems of the existing recommendation strategies mainly lie in "insufficient association mining" and "limited adaptability" [5]. Firstly, there is a lack of accuracy. It overly relies on users' explicit behaviors (such as browsing and searching), while ignoring the implicit associations between products. For example, when recommending a mobile phone, it fails to associate it with a phone case, resulting in users having to search again. Secondly, there is a lag in timeliness. The rule update cycle is long, making it difficult to capture short-term demand changes. For instance, after summer arrives, it still continuously recommends winter-related product combinations. Thirdly, there is a serious homogeneity problem. It repeatedly recommends the same category of products (such as after purchasing a T-shirt, continuously recommending similar T-shirts), while ignoring cross-category demands (such as the combination of T-shirts and jeans). Fourthly, the coverage of demands is incomplete. It only focuses on the expressed demands of users and is unable to explore potential demands. For example, users who buy baby milk powder may also need baby wipes, but the existing recommendations do not cover such associations, resulting in limited recommendation value, as shown in Table 2:

Table 2 Summary of Core Problems in Existing E-commerce Recommendation Strategies

Problem Type	Problem Manifestation	Specific Example
Lack of Accuracy	Over-reliance on explicit user behavior; neglect of implicit product associations	Recommending a mobile phone without suggesting a phone case, causing the user to search again
Lag in Timeliness	Long rule update cycles, making it difficult to capture short-term demand changes	Continuing to recommend winter product combinations even after the arrival of summer
Severe Homogenization	Repeatedly recommending products of the same category; ignoring cross-category needs	After buying a T-shirt, continuously recommending similar T-shirts, without suggesting T-shirt + jeans combinations
Incomplete Demand Coverage	Focus only on expressed user needs; difficult to identify potential needs	A user who buys infant formula may also need baby wipes, but such recommendations are not made

3. Optimization Directions for Product Recommendation Based on Association Rule Mining

3.1. Improving the "Data Foundation Quality" of Association Rules

Improving data foundation quality requires optimization from two dimensions: "breadth" and "depth." In terms of breadth, break through the limitation of single transaction data, integrate full-spectrum user behavior data (such as browsing duration, add-to-cart frequency, favoriting preferences, etc.), so that association rules not only reflect "already purchased," but also capture "potential purchase tendencies [6]." In terms of depth, refine data tags and add time dimensions (such as weekdays/weekends, seasons), scenario dimensions (such as commuting/home), and user attributes (such as purchasing power, family structure), for example, distinguishing between "snack associations of student groups" and "snack associations of family users." At the same time, by cleaning abnormal data (such as misoperation orders, test data), avoid noise interfering with the accuracy of association rules and provide a reliable foundation for subsequent mining.

3.2. Optimizing the "Rule Effectiveness" of Association Rule Mining

The core of optimizing rule effectiveness is "filtering high-value rules and removing redundant information [7]." Precisely filter using quantitative indicators: use "support" to filter out low-frequency invalid rules (such as product combinations that appear only a few times a year); use "confidence" to ensure rule reliability (such as the probability of "buying B after buying A" reaching the threshold); use "lift" to determine the actual value of the rule (only retain rules with lift > 1 to avoid "false associations," such as the false association between "umbrella and ice cream" due to seasonal coincidence). Meanwhile, in response to the characteristics of massive e-commerce data, simplify the algorithm process (such as reducing steps for calculating frequent itemsets), improve mining efficiency while ensuring rule quality, and avoid rule delays caused by complex algorithms.

3.3. Strengthening the "Matching Degree between Association Rules and User Needs"

Strengthening the matching degree requires association rules to "fit user profiles and scenario needs." On one hand, combine user tags for precise matching, for example, pushing the association rule "infant formula + nursing pads" to "new mothers," and pushing the association rule "suit + shirt" to "new office workers," avoiding the blindness of general rules [8]. On the other hand, distinguish scenario-based needs. For the "order supplement" scenario (users need low-price products to meet discount requirements), push the association of "cost-effective small items + main products"; for the "bundle" scenario (users pursue one-stop shopping), push "functionally complementary product combinations" (such as "coffee machine + coffee beans + filter paper"), to ensure the rules align precisely with users' current shopping purposes.

3.4. Establishing the "Real-time Iteration Mechanism of Association Rules"

Table 3 Summary of Product Recommendation Optimization Directions Based on Association Rule Mining

Optimization Direction	Core Measures
Improving the "Data Foundation Quality" of Association Rules	Integrate full-spectrum behavior data (browsing, add-to-cart, etc.); refine time/scenario/user attribute tags; clean abnormal data
Optimizing the "Rule Effectiveness" of Association Rule Mining	Filter high-value rules using support, confidence, and lift; simplify algorithm processes to improve efficiency
Strengthening the "Matching Degree Between Association Rules and User Needs"	Precisely match rules with user profiles; differentiate scenario-based needs (e.g., order supplement, bundle) and push corresponding associations
Establishing the "Real-time Iteration Mechanism" of Association Rules	Hierarchical updates (daily for FMCG, weekly for durable goods); emergency response for sudden scenarios; monitor and eliminate inefficient rules

From shown in Table 3, the real-time iteration mechanism needs to consider both "regular updates" and "emergency responses." For regular updates, set the cycle according to product type: fast-moving consumer goods (such as snacks, daily necessities) are updated daily to capture short-term consumption trends; durable goods (such as home appliances, furniture) are updated weekly to balance efficiency and cost [9]. For emergency responses, quickly generate temporary rules for sudden scenarios (such as promotional activities, festivals, hot events), for example, during the "618 Shopping Festival," give priority to pushing "order supplement association rules"; during the "World Cup," temporarily strengthen the association of "beer + peanuts + football-watching pillow." At the same time, monitor the effect of rules (such as click-through rate, conversion rate), automatically eliminate inefficient rules, and ensure recommendations always align with users' dynamic needs.

4. Implementation Strategies for Product Recommendation Based on Association Rule Mining

4.1. Data-Level Implementation Path

At the data level, a “full-link, high-quality” data foundation must be built. First, open up multi-source data channels, integrating transaction data (order details), behavior data (browsing paths, add-to-cart records), user data (age, consumption level), and scenario data (time periods, regions) to form a unified data pool, avoiding data silos. Second, perform data cleaning using tools to automatically identify and remove outliers (such as test orders, misoperation records), and correct missing values (e.g., completing user location tags), ensuring data accuracy. Finally, conduct data tagging: add category and function tags to products; add preference and scenario tags to users [10]. This ensures that subsequent association rule mining can accurately match the “user-product-scenario” dimensions, providing clear data support for rule generation.

4.2. Algorithm-Level Implementation Path

At the algorithm level, a balance must be struck between “rule quality” and “computational efficiency.” First, prioritize lightweight algorithms adapted to e-commerce scenarios, such as an improved Apriori algorithm (reducing redundant itemset computations) or FP-Growth algorithm (suitable for massive datasets), to lower the technical implementation threshold. Second, dynamically set filtering thresholds: raise the support threshold for high-frequency essential goods (e.g., daily necessities) to focus on core associations; lower it for long-tail goods (e.g., niche accessories) to mine fine-grained associations. Meanwhile, use visualization tools to output rule results (e.g., “Product A → Product B, confidence 65%”) so that operation teams can directly understand and apply them, reducing the integration cost between algorithm and business, and ensuring that rules are quickly transformed into recommendation strategies.

4.3. Recommendation Logic-Level Implementation Path

Recommendation logic should achieve “precise alignment of rules with scenarios and users.” Step one: build a tiered association rule library, categorized into “strong associations” (e.g., phone + phone case), “weak associations” (e.g., shirt + tie), and “cross-category associations” (e.g., baby formula + stroller), for flexible rule calling as needed. Step two: match rules to specific scenarios — on the product detail page, emphasize “complementary associations” (e.g., buying a computer, recommend a mouse); in the shopping cart, emphasize “order-supplement associations” (e.g., low-priced items to meet discount thresholds); on the homepage, emphasize “potential associations” (e.g., buying an oven, recommend baking ingredients). Step three: filter rules based on user profiles — for “price-sensitive users,” prioritize high cost-performance combinations; for “quality-oriented users,” recommend brand-linked associations, ensuring that recommendations align with user preferences.

4.4. Dynamic Update-Level Implementation Path

Dynamic updates should establish a dual mechanism of “automatic iteration + manual optimization.” Set a tiered update cycle: for fast-moving consumer goods, automatically update rules daily based on daily transaction data to adjust association strength; for durable goods, update weekly, optimizing rules based on weekly trends. Meanwhile, build a real-time monitoring module to track recommendation performance data (e.g., click-through rate, conversion rate). If the performance of a rule remains below the threshold for 3 consecutive days, the system will automatically deactivate the rule and trigger recalculation. Additionally, reserve a manual intervention channel: for unexpected scenarios (e.g., promotional events, holidays), operators can temporarily add high-priority rules (e.g., “New Year gift boxes + Spring Festival couplets”) to ensure rules respond quickly to short-term demand changes and avoid lag, as shown in Table 4:

Table 4 Implementation Level and Core Execution Path of Associative Rule Recommendation (FMCG & Durable Goods)

Implementation Level	Core Implementation Path
Data Level	Integrate multi-source data (transaction, behavior, user, scenario) into a data pool; clean abnormal data; conduct tagging
Algorithm Level	Use lightweight algorithms (e.g., improved Apriori); dynamically set filtering thresholds; visualize rule outputs
Recommendation Logic Level	Build a tiered association rule library; match rules by scenario; filter recommendations based on user profiles
Dynamic Update Level	Tiered updates (daily for FMCG, weekly for durable goods); real-time performance monitoring; reserve manual optimization channel

5. Summary

In conclusion, the optimization of e-commerce product recommendations based on association rule mining centers on a data-driven approach to address issues in traditional recommendation systems such as insufficient accuracy and severe homogenization. Its value lies in mining implicit associations between products from massive datasets, shifting recommendations from "experience-based judgment" to "quantifiable pattern matching." The key to optimization lies in four aspects: Strengthening the data foundation by integrating full-spectrum behavior and scenario data; Improving rule quality by using support, confidence, and other indicators to filter high-value associations; Enhancing the match between rules and users/scenarios to ensure recommendations align with personalized needs; Establishing a dynamic iteration mechanism to ensure rules can respond to short-term trends and long-term changes.

This strategy not only improves users’ shopping efficiency (reducing decision-making costs) but also creates incremental value for the platform (promoting cross-category purchases and activating long-tail products), ultimately achieving both “precise user demand satisfaction” and “platform business growth.” It is an effective path for e-commerce recommendation systems to evolve from “extensive” to “refined” operations.

References

[1] Miller R J, Yang Y. Association rules over interval data[C]//ACM.ACM, 1997:452-461. DOI:10.1145/253262.253361.

[2] Kim C , Lee H , Seol H ,et al.Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach[J].Expert Systems with Applications, 2011, 38(10):12559-12564.DOI:10.1016/j.eswa.2011.04.042.

[3] Mutter S , Hall M A , Frank E .Using Classification to Evaluate the Output of Confidence-Based Association Rule Mining[J].Springer-Verlag, 2004.DOI:10.1007/978-3-540-30549-1_47.

[4] JOMAA I, POIRSON E, CUNHA C D. A methodology for designing a recommender system based on customer preferences[C]// Proceedings of the 18th International Conference on Engineering Design (ICED 11): Impacting Society through Engineering Design. Vol. 9: Design Methods and Tools, Pt. 1. Lyngby/Copenhagen, Denmark, 15–19 Aug. 2011. Glasgow: Design Society, 2011.

[5] Yongxin Tong, Jieying She, Rui Meng. Bottleneck-aware arrangement over event-based social networks: the max-min approach[J].World Wide Web, 2016.DOI:10.1007/s11280-015-0377-6.

[6] Nass E , Scheibmayer M .Defining a research framework for the business impact of data management[J]. IEEE[2025-07-27].DOI:10.1016/j.ijbiomac.2011.06.015.

[7] Rommers M K , Zwaveling J , Guchelaar H J ,et al. Evaluation of rule effectiveness and positive predictive value of clinical rules in a Dutch clinical decision support system in daily hospital pharmacy practice[J].Artificial Intelligence in Medicine, 2013, 59(1):15-21.DOI:10.1016/j.artmed.2013.04.001.

[8] Xu S , Chen M , Yuan B ,et al.Resident Satisfaction and Influencing Factors of the Renewal of Old Communities[J].Journal of Urban Planning and Development, 2024, 150(1).DOI:10.1061/JUPDDM.UPENG-4632.

[9] Bruns A , Burgess J E .Local and global responses to disaster: #eqnz and the Christchurch earthquake[J].AST Management Pty Ltd, 2012.

[10] Yip R W , Levitt K N .The Design and Implementation of a Data Level Database Inference Detection System[J].Springer US, 1999.DOI:10.1007/978-0-387-35564-1_16.