Optimization Model for Bidding and Procurement Packaging Based on Improved DBSCAN Algorithm

Lu Li*

Beijing Jiaotong University, No.3 Shangyuancun, Beijing, China, 100044
*22120600@bjtu.edu.cn

Abstract. In addressing the packaging issue in tender procurement, this study introduces an enhanced DBSCAN clustering algorithm that incorporates an edge-betweenness-based edge-breaking approach. Utilizing historical tender procurement packaging data, a distance matrix among materials is constructed. The traditional DBSCAN clustering algorithm's density-connectivity characteristic poses challenges to the granularity of clustering results. To address this, an optimization method involving edge-betweenness edge-breaking is devised. By computing the edge-betweenness of connections between data points, edges crucial to the network structure are identified and subsequently removed. This process disrupts the internal connections of "large clusters," causing them to fragment into smaller, more refined clusters. Consequently, the accuracy and resolution of procurement packaging clustering results are enhanced. The proposed method is employed in data experiments involving items requiring tender procurement packaging and is compared with alternative clustering packaging algorithms. The results demonstrate that by optimizing the clustering structure and elevating the silhouette coefficient, this method significantly improves the accuracy and robustness of packaging results.

Keywords: Bidding Procurement, Packaging of Purchased Items, Clustering Problem, DBSCAN Algorithm

1 Introduction

Procurement through tendering, or reverse auction, is an efficient market mechanism where the procuring entity specifies requirements and invites suppliers to bid [1] , with the winner determined by established rules. This method is widely adopted for its effectiveness in price discovery. Centralized procurement, which consolidates goods and services, reflects a trend toward centralized management within enterprises to reduce costs through aggregated purchases [2] .

Procurement bundling, or lotting, is a critical step in centralized tendering, involving the combination of multiple goods or services into single packages to attract more suppliers[3]. However, enterprises face challenges in this process: unclear standards for package division, reliance on inefficient manual methods, and insufficient consolidation of packages. Historical data suggests that manual division often leaves room for further consolidation, which could enhance efficiency and value.

M. Zhang et al. (eds.), *Proceedings of the 2025 5th International Conference on Informatization Economic Development and Management (IEDM 2025)*, Advances in Economics, Business and Management Research 330,
https://doi.org/10.2991/978-94-6463-724-3_40

Therefore, in the context of procurement through tendering, when enterprises face imperfect procurement bundling systems and only have access to historical bundling data, how to quickly, intelligently, and accurately complete the bundling of procurement items—while minimizing the number of procurement packages, reducing costs, and ensuring high relevance among items within the same package to attract supplier participation—remains a critical issue that warrants in-depth research.

2 Literature Review

Procurement bundling strategies are crucial precursors to reverse auctions, significantly influencing procurement results [3], many scholars have proposed packaging strategies and mechanism. Beall et al. [4] put forward the market basket lotting and individual lotting strategies. Linthorst et al. [5] proposed homogeneous and heterogeneous bundling strategies via procurement personnel interviews. Hu et al. [6] set market competitiveness and item similarity variables to assess item - pairing possibilities and proposed the PBBS (Post - Bidding Bundle Strategy) for procurement item bundling. Mansouri and Hassini [7] defined combinatorial procurement auctions winner determination problem, propose a Lagrangian-based approach to solve it.Abbaas and Ventura [8] proposed an iterative procurement combinatorial auction mechanism, the buyer solves a Mixed Integer Non-linear Programming (MINLP) model to determine the winning bids for the current auction iteration.

In the research of algorithms for packaging procurement items, Wang et al. [9] proposed a packaging algorithm based on the quantum algorithm, which was developed on the basis of price complementarity and bid consistency of procurement items. Liao and Zhou [10] designed a packaging algorithm based on the biclustering algorithm to address the problem of library periodical procurement. Zhang and He [11] put forward a packaging algorithm based on transaction cost economics for engineering construction projects.

The DBSCAN algorithm is a density clustering algorithm, but it has the problems of low accuracy and efficiency. Many scholars have improved it. Zhang [12] designed an improved Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm for hazard recognition of obstacles in unmanned scenes. It addresses limitations in the traditional DBSCAN by refining the core point definition using an adaptive density threshold. Dawid and Krzysztof [13] proposed the GrDBSCAN algorithm, which first granularizes the data into fuzzy particles and then performs density based clustering on the obtained particles, shortening the clustering time. Stiphen et al. [14] introduced a new step in the DBSCAN clustering process, iteratively updating feature weights based on the current data partition, improving the limitation of equal treatment of all functions in the DBSCAN algorithm.

Current research reveals that although empirical studies have identified various procurement package attributes, these attributes are difficult to measure and exhibit complex interdependencies, hindering the development of effective quantitative frameworks. Moreover, existing quantitative algorithms often depend on detailed material attribute definitions, weight assignments, or comprehensive supplier information, which diverge from real-world business conditions. Their accuracy and practicality also require further enhancement.

To address these issues, this paper proposes an innovative clustering-based procurement bundling method by integrating domestic and international research findings. This method utilizes an improved DBSCAN algorithm for intelligent material clustering. To tackle the "large cluster" problem commonly encountered in traditional DBSCAN algorithms when processing large-scale data, this paper introduces an optimization mechanism based on edge-betweenness edge-cutting operations, significantly enhancing the algorithm's performance.

3 Model and Solution

3.1 Problem Description and Model Definition

The symbols to be used are shown in Table 1.

Table 1. Symbols and Definitions

Symbol	Definition
$P = \{1,2,...,n\}$	Collection of materials to be tendered
$\mathrm{B} = \{1,2,...k\}$	To meet the collection of all procurement packages for material set P procurement
$PB(b)$	The procurement package b contains a collection of materials that require bidding, $b \in B$
$\mathrm{S}(ij)$	Edge betweenness between materials
$n(ij)$	The frequency at which subclass i and subclass j are jointly divided into the same package in the historical package
r	Neighbor radius
$MinPts$	Minimum number of neighbors
M	Initial correlation value of two materials (taken as 100)
m	Historical data covers the quantities of different material subcategories
$d(ij)$	Distance between two materials
T_i^L, T_i^M, T_i^S	Category information of material i

In order to purchase material $P = \{1,2,...,n\}$, the bidding purchaser needs to create procurement package $\mathrm{B} = \{1,2,...k\}$, make the following assumptions:

1) The set of materials to be purchased $P = \{1,2,...,n\}$ is known;

2) Given the historical packaging status of the set of materials to be purchased, and the materials to be purchased have appeared at least once in the historical procurement packaging;

3) Given the category information of the materials to be purchased (divided into large, medium, and small categories layer by layer);

4) The same purchased item is packaged as a whole and is no longer divided in quantity.

The definition of optimization problem for bidding and procurement packaging scheme is as follows:

$$Min\ |B|; \tag{1}$$

$$Min\ |\mathrm{S}|; \tag{2}$$

$$\mathrm{B} = \{1,2,\ldots k\},\quad k \le n\ ; \tag{3}$$

$$\forall i \ne j, bi \cap bj = \varnothing\ ; \tag{4}$$

$$\cup \mathrm{PB}(b) = P\ . \tag{5}$$

In the context of the formulated equations, Equation (1) establishes the objective of minimizing the total number of procurement packages. Equation (2) defines the goal of maximizing the inter-item relevance within each package, which is equivalent to minimizing the sum of edge betweenness. Equation (4) ensures mutual exclusivity between any two distinct procurement packages in terms of their item composition. Finally, Equation (5) guarantees that the proposed procurement bundling solution B comprehensively covers all procurement items.

3.2 Model Solution

(1) Generation of the Material Distance Matrix

The distance matrix between materials is constructed based on the following two dimensions:

1) Historical Information Correlation Characterization

The subcategory information of materials serves as a critical representation of their attribute characteristics. When calculating the correlation matrix of material subcategories in historical procurement packages, we adhere to the principle that "frequency reflects the strength of correlation." Assuming the historical data covers m distinct material subcategories, the initial correlation value between any two subcategories is uniformly set to M (in this study, $M = 100$). Furthermore, we count the frequency $n(ij)$ at which subcategory i and subcategory j are grouped into the same procurement package in historical data. Based on these statistics, a distance matrix for historical material subcategories is constructed, which quantifies the strength of correlation between different subcategories. Specifically, a higher frequency $n(ij)$ indicates a stronger correlation between the two subcategories, resulting in a smaller corresponding value in the matrix. The formula is as follows:

$$D_H = \begin{matrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mm} \end{matrix} \tag{6}$$

where $d(ij)$ is calculated as:

$$d_{ij} = \begin{cases} \dfrac{M}{n_{ij}+1}, i \neq j \\ 0, i = j \end{cases} \tag{7}$$

When the correlation measure between two material subcategories is less than M, it indicates that materials with these two subcategories have historically been grouped into the same procurement package. Conversely, if the distance measure equals M, it implies that no materials with these two subcategories have ever been grouped into the same procurement package in the historical data. Based on historical learning, it is generally concluded that materials with these two subcategories should not be grouped into the same procurement package.

2) Characterization of Category Information Correlation

Based on material codes and subcategory info, correlation data between any two materials is extracted from the historical subcategory correlation matrix to construct the initial correlation matrix for procured materials, reflecting historical partitioning patterns. Then, considering material category info, the correlation matrix is updated. If two materials are in the same intermediate category, they have a strong correlation; if in the same subcategory, an extremely strong one. Update rules are made to generate the final correlation matrix for procured materials.

Assuming the base value for updates is N, and the material i has a major category T_i^L, intermediate category T_i^M, and subcategory T_i^S, the correlation value between material i and material j is updated according to the following rules:

$$\hat{d}_{ij} = \begin{cases} 3N, T_i^M = T_j^M, d_{ij} > 3N \\ N, T_i^S = T_j^S, d_{ij} > N \\ d_{ij}, \text{Other situations} \end{cases} \tag{8}$$

Additionally, the parameters required for the DBSCAN algorithm are defined, including the neighborhood radius r and the minimum number of neighbors $MinPts$ required to form a core point.

(2) Graph Construction and Edge Betweenness Calculation

Materials' relationships are modeled as a graph. Each material is a node, and an edge represents whether two materials were grouped in the same procurement package. The edge weight corresponds to the distance between materials. Since Dijkstra's algorithm is proven to be universally optimal, it's used to compute all - pairs shortest paths in the graph. For each edge, the number of shortest paths passing through it is counted to calculate the edge betweenness. Edge betweenness, measuring an edge's importance in network flow, is defined as the proportion of all shortest paths in the network traversing the edge[15], with the calculation formula as follows:

$$S_{ij} = \sum_{(l,m)\neq(i,j)} \frac{N_{lm}(e_{ij})}{N_{lm}} \tag{9}$$

Here, S_{ij} represents the betweenness of edge e_{ij}, and (l,m) denotes any pair of nodes in the network, with the condition $l \neq i, m \neq j, l \neq m$, thereby excluding the two endpoints of edge e_{ij}. N_{lm} represents the total number of shortest paths from node l to node m, and $N_{lm}(e_{ij})$ denotes the number of shortest paths from node l to node m that pass through edge e_{ij}. This formula calculates the proportion of shortest paths between all node pairs (excluding the endpoints of edge e_{ij}) that pass through edge e_{ij}. Edges with higher betweenness values play a more critical role in transmitting information or material flows within the network.

(3) Edge Pruning Guided by Betweenness Centrality

To refine the clustering structure, edges are prioritized by betweenness centrality scores. The scores are ranked to find critical edges between clusters or those maintaining oversized clusters. Strategically removing high - betweenness edges disrupts cluster configurations, promoting their decomposition and reformation into more meaningful groupings.

(4) Cluster Formation via DBSCAN

After pruning high-betweenness edges, the DBSCAN algorithm is reapplied to the updated graph.The effectiveness of the clustering is assessed using evaluation metrics such as the silhouette coefficient, which quantifies the quality of clustering by balancing intra-cluster cohesion and inter-cluster separation. The silhouette coefficient assigns a score to each data point, reflecting its alignment with its assigned cluster relative to the nearest alternative cluster[16]. Scores range from [-1, 1], with higher values indicating better-defined clusters. The silhouette coefficient is computed as follows:

$$C(i) = \frac{b(i) - a(i)}{\max(a(i), b(i)} \tag{10}$$

For each data point i, $a(i)$ measures the average distance to all other points within its cluster, while $b(i)$ represents the average distance to points in the nearest neighboring cluster. The resulting coefficient $C(i)$ provides insights into clustering quality: Values approaching 1 indicate strong cluster cohesion and clear separation from neighboring clusters; Scores near 0 suggest ambiguous cluster boundaries, reflecting suboptimal clustering; Values close to -1 imply potential misclassification, where data points may belong to a different cluster.This methodology ensures a robust evaluation of clustering outcomes, balancing theoretical rigor with practical relevance for procurement optimization.

Enterprises can follow the following steps when using this scheme to package bidding procurement packages:

Step 1. Collect historical package procurement information and perform data preprocessing;

Step 2. Based on the distance matrix definition, convert the historical packet partitioning information into a distance matrix;

Step 3. Convert category information into a category distance matrix based on the definition of the category distance matrix;

Step 4. Use the improved DBSCAN algorithm to cluster materials;

Step 5. Professional responsible personnel evaluate the procurement package and merge/further divide the procurement package.

4 Data Experiment

4.1 Multi Algorithm Comparative Experiment

This study used data from Company X, a leading comprehensive energy enterprise engaged in large - scale open - pit coal mining, with operations covering coal mining, deep processing, and diversified energy development. Company X has a wide range of spare parts, large procurement scale, and high tender amounts, making its procurement activities complex and representative.

The research data consists of two datasets: "Contract Details" with 2021 - 2023 historical procurement data (including material bundling and classification codes) and "Pending Centralized Procurement Material Information" listing materials to be procured. The raw data was pre - processed (type conversion, missing value handling) to meet algorithm input requirements. Three algorithms for material bundling were tested using the pre - processed data: the improved DBSCAN algorithm, traditional DBSCAN, and spectral clustering. The key findings are shown in Table 2 and Table 3.

Table 2. Comparison of Packaging Results under Different Algorithms

Algorithm	Silhouette coefficient	Number of Purchase Packages	The total amount of packages exceeding 2 million	The quantity of packages with an amount exceeding 2 million
Improved DBSCAN	0.4688	67	4754.81	7
Double Clustering	0.4027	72	3022.31	5
Chaos Search Clustering	0.3519	56	4183.26	10
Spectral Clustering	0.3010	64	3827.96	14
DBSCAN	0.1345	40	2041.54	4

Table 3. Comparison of Materials in Procurement Packages under Different Algorithms

Algorithm	Typical Package 1	Typical Package 2

Improved DBSCAN	IBC ton drum, industrial alcohol \ 99%, photosensitive resin	Nitrogen, liquid nitrogen
Double Clustering	IBC bucket, stainless steel standard sieve, wide mouth sampling bottle, beaker	Nitrogen, liquid nitrogen
Chaos Search Clustering	IBC ton drum, industrial alcohol \ 99%, photosensitive resin	Nitrogen, liquid nitrogen, acetylene, acetylene, pressure reducing valve
Spectral Clustering	IBC bucket, stainless steel standard sieve, wide mouth sampling bottle, beaker	Nitrogen, liquid nitrogen
DBSCAN	IBC ton drum, industrial alcohol \ 99%, photosensitive resin	Nitrogen, liquid nitrogen, acetylene

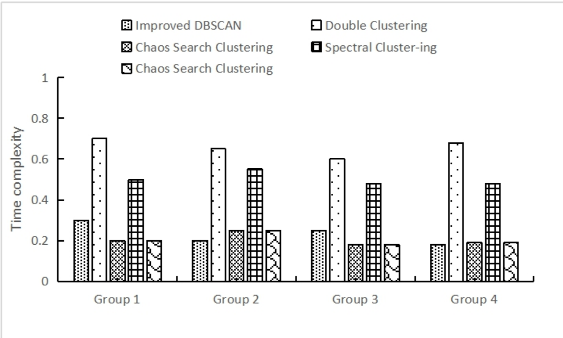

Fig. 1. Comparison chart of time complexity

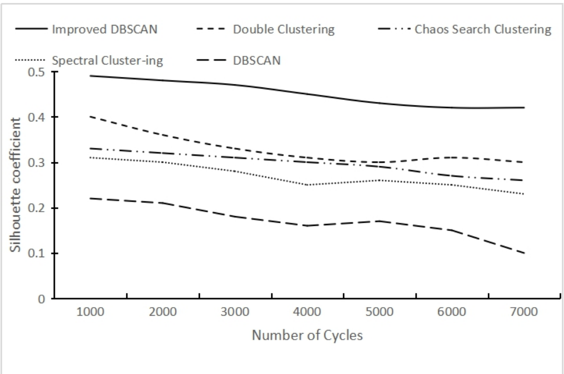

Fig. 2. Comparison chart of contour coefficients

For Typical Package 1, the spectral clustering algorithm couldn't effectively separate laboratory equipment (like stainless - steel standard sieves, wide - mouth sample bottles) from industrial packaging containers (such as IBC ton containers), causing potential procurement and management issues. The improved DBSCAN algorithm, however, achieved professional material categorization, ensuring procurement package quality.For Typical Package 2, the improved algorithm accurately recognized nitrogen and liquid nitrogen procurement needs, avoiding acetylene - related safety and qualification review risks, highlighting its risk - assessment and procurement - optimization capabilities.

Overall, the improved algorithm maintained a silhouette coefficient above 0.4, outperforming Other algorithms (as shown in Figure 1). In computational efficiency, its time complexity stayed below 0.3 (as shown in Figure 2). These results validate the method's effectiveness in improving clustering and bundling, offering a new solution for complex procurement scenarios.

4.2 Multi Dataset Experiment

To verify the effectiveness of the improved DBSCAN algorithm proposed in this article in practical applications and its universality for different datasets, multiple datasets will be selected for further validation. The performance of the improved DBSCAN algorithm proposed in this study on different datasets is shown in Table 4.

Table 4. Performance of Improved DBSCAN Algorithm On Different Datasets (real strategy in parentheses)

Dataset	Dataset I	Dataset II	Dataset IIII
Silhouette coefficient	0.4688	0.5372	0.4043
Number of Purchase Packages	67 (71)	42 (62)	34 (42)
The total amount of packages exceeding 2 million	4754.81 (3542.23)	2144.21 (1892.98)	1572.45 (1143.2)
The quantity of packages with an amount exceeding 2 million	4 (3)	6 (7)	3 (3)
Computing Time	27.2s	24.6s	17.2s

Dataset I is the dataset selected in 4.1, which contains bidding and procurement related data of a coal mine related enterprise over the years, involving a total of 47756 materials over 3 years. Dataset II is the bidding and procurement related data of a coal mining enterprise over the years, involving a total of 38461 types of materials over 5 years. Dataset III is the bidding and procurement related data of a certain new energy vehicle enterprise over the years, involving a total of 15321 materials over 2 years. Based on the experimental data above, it can be seen that the contour coefficient of the algorithm proposed in this article always remains above 0.4, the amount and quantity of procurement package division are also reasonable, and the computation time is also fast. It has good performance in different datasets.

5 Conclusion

Combination auction is an effective centralized procurement mechanism, with its main challenge being the bundling optimization problem in the "bundled combination" two-round bidding framework. However, current research shows significant limitations: most bundling strategies remain conceptual, and quantitative algorithms often rely heavily on material attribute weighting and complete supplier information, failing to align with practical business needs. Traditional DBSCAN clustering tends to form "large clusters," reducing clustering granularity and undermining the accuracy of silhouette coefficients in measuring intra-cluster cohesion and inter-cluster separation.

To address these issues, this study proposes an improved DBSCAN algorithm incorporating edge betweenness edge cutting. By calculating the betweenness centrality of data connections, the algorithm identifies and cuts key edges, refining the clustering structure. Experimental results demonstrate that the proposed algorithm outperforms existing methods in optimizing clustering structures, improving silhouette coefficients, and maintaining computational efficiency.

However, the method proposed in this article has the following limitations: (1) weak processing ability for highly imbalanced and sparse datasets; (2) In terms of computational efficiency, although the algorithm proposed in this article can achieve parallel computing, it relies on the Dijkstra algorithm when calculating betweenness centrality, which may lead to the problem of long processing time for large-scale datasets; (3) The applicable procurement scenarios are relatively single and cannot adapt to the processing of dynamic real-time data. Moreover, only the procurement package division has been completed, and there has been no research on the actual impact of subcontracting results on procurement managers.

Future research can be conducted from the following aspects: (1) Improving distance measurement methods: exploring more suitable distance measurement methods for imbalanced or sparse datasets, such as improved algorithms based on cosine similarity or Mahalanobis distance. (2) Distributed computing and parallelization: Continue to explore distributed computing or parallelization technologies in depth to improve the computational efficiency of algorithms in large-scale datasets. (3) Integrated dynamic clustering technology: Combining the advantages of dynamic clustering algorithms (such as online clustering or incremental clustering) to enhance the algorithm's adaptability to dynamically changing data. (4) Expand application scenarios, such as exploring how to attract more suppliers to participate in bidding by optimizing procurement package division, analyzing its impact on market competitiveness and procurement results, or quantitatively evaluating the actual effects of algorithms in shortening procurement cycles, reducing procurement costs, and improving decision-making efficiency.

References

1. Li T ,Forster P H L ,Bhuiyanmishu S . Transaction Costs, Participation, and the Cost-Effectiveness of Reverse Auctions: Evidence from a Laboratory Experiment [J]. Environmental and Resource Economics, 2024, 88 (2): 1-28.

2. Ding F X ,Liu S F ,Li X W . Pareto Optimality of Centralized Procurement Based on Genetic Algorithm [J]. Tehnički vjesnik, 2022, 29 (6): 2058-2066.
3. Triki C ,Akil J ,Asmakh A H . Optimisation models for the procurement through reverse combinatorial auctions in the logistics and food industries [J]. International Journal of Procurement Management, 2023, 16 (4): 530-548.
4. Carter C R, Kaufmann L, Beall S, et al. Reverse auc-tions—grounded theory from the buyer and supplierperspective[J]. Transportation Research Part E Logis-tics & Transportation Review, 2004, 40(3): 229-254.
5. Linthorst M M, Telgen J, Schotanus F. Buying bun-dles: the effects of bundle attributes on the value of bun-dling[C]. International Euroma Conference, 2008.
6. Hur D, Hartley J L, Mabert V A. Implementing reversee-auctions: a learning process[J]. Business Horizons, 2006, 49(1): 21-29.
7. Mansouri B ,Hassini E . Optimal pricing in iterative flexible combinatorial procurement auctions [J]. European Journal of Operational Research, 2019, 277 (3): 1083-1097.
8. Abbaas O ,Ventura A J . An Iterative Procurement Combinatorial Auction Mechanism for the Multi-Item, Multi-Sourcing Supplier-Selection and Order-Allocation Problem under a Flexible Bidding Language and Price-Sensitive Demand [J]. Mathematics, 2024, 12 (14): 2228-2228.
9. Wang D W, Wang N. Optimization Method for Packaging in Multi-Project Procurement Based on Quantum Evolutionary Algorithm [J]. Industrial Engineering Journal, 2013, 16(1): 45-49.
10. Liao T F, Zhou Y M. Research on the Basic Principles of Book Bidding and the Mathematical Model for Bid Section Division: A Case Study of the Library of Hunan City University [J]. Library Theory and Practice, 2009, (3): 104-106.
11. Zhang Z H, He Q R. Impact of Bid Section Division on Highway Engineering Cost [J]. Journal of Tsinghua University (Science and Technology), 2010, (6): 830-833.
12. Zhang W Y. An improved DBSCAN Algorithm for hazard recognition of obstacles in unmanned scenes [J]. Soft Computing, 2023, 27 (24): 18585-18604.
13. Dawid S ,Krzysztof S . GrDBSCAN: A Granular Density–Based Clustering Algorithm [J]. International Journal of Applied Mathematics and Computer Science, 2023, 33 (2): 297-312.
14. Stiphen C ,Na H ,Renato A D C . Feature weighting in DBSCAN using reverse nearest neighbours [J]. Pattern Recognition, 2023, (137):109314.
15. Mahshid P ,Michael Z ,Paolo M . Edge betweenness centrality as a failure predictor in network models of structurally disordered materials [J]. Scientific Reports, 2022, 12 (1): 11814-11814.
16. Lai H ,Huang T ,Lu B , et al. Silhouette coefficient-based weighting k-means algorithm [J]. Neural Computing and Applications, 2024, (prepublish): 1-15.