Optimization Analysis of Business Circle Based on Machine Learning

Yuyang Fang

Beijing University of Posts and Telecommunications, Beijing, China

PattonCharles0503@outlook.com

Abstract. This paper begins with a discussion of data preprocessing in order to improve the data quality, standardize the feature ratio, and comprehend the data pattern and trend. The preprocessing technology utilized in this paper is feature aggregation, which aggregates the transaction data based on the transaction time and date to determine the transaction type (weekday or weekend) and calculates the transaction frequency and average amount for each store as the final feature of cluster analysis. Then, for the processed data, this paper compares the advantages and disadvantages of three different clustering algorithms, including K-means clustering, spectral clustering, and DEC deep clustering, in terms of their clustering effects. The conclusion of the paper is that deep clustering is superior to traditional clustering algorithms in terms of clustering precision, but it requires more computing resources and offers practical suggestions for existing shop clustering.

Keywords: Machine Learning, K-means clustering, spectral clustering, DEC deep clustering.

1. Introduction

Since the turn of the 21st century, the rapid growth of the global economy and the advancement of science and technology have brought about significant changes in the commercial sector. The increasing prevalence of emerging technologies such as network technology, big data, and artificial intelligence enables all industries to comprehend the significance of innovation and technology combination in enhancing competitiveness. In the retail industry, shopping malls are the primary shopping destinations for consumers, and their store layout and business planning are crucial for enhancing the customer experience and shopping satisfaction. Historically, layout and business planning of shopping malls relied heavily on manual experience and conventional market research techniques. Designers and planners develop layout strategies based on years of experience and the findings of consumer surveys. These methods include the division of mall interior space, the design of product display methods, and the planning of promotional activities. However, these traditional methods are limited and may rely too heavily on the experience and intuition of individual designers or planners, resulting in results that are heavily influenced by subjective personal factors. In addition, traditional market research data are frequently inaccurate and out-of-date, and it is difficult to capture the subtle distinctions between market changes and consumer demand. Therefore, the industry has shifted its focus to how to use advanced technology to improve the operational efficiency and profitability of shopping malls. In this context, an increasing number of researchers and businesses began investigating the application of artificial intelligence technology in the layout and business planning of shopping malls in order to overcome the limitations of traditional methods and achieve more efficient, precise, and intelligent business operations.

Machine learning is a vital subfield within artificial intelligence (AI). It can efficiently solve numerous problems by enabling computer systems to automatically learn and improve from data. In recent years, machine learning has made remarkable advancements in fields such as natural language processing, image recognition, and recommendation systems, resulting in revolutionary changes across multiple industries. Successful machine learning applications in natural language processing include speech recognition, machine translation, and emotion analysis. Google Translate, for instance, uses machine learning technology to provide users with real-time, high-quality translation services, thereby minimizing the difficulties caused by language barriers [1]. Intelligent voice assistants such as Apple's Siri and Amazon's Alexa use machine learning technology to interpret users' voice

commands and carry out related tasks, thereby facilitating people's daily lives [2]; In many aspects, the deep learning method based on machine learning has surpassed human recognition accuracy in the field of image recognition. For instance, deep learning techniques have made remarkable strides in medical image analysis, enabling doctors to diagnose diseases such as cancer and diabetic retinopathy with greater precision [3]. Moreover, in the field of self-driving cars, machine learning plays a crucial role in environmental perception and behavior prediction, laying the groundwork for the future realization of intelligent transportation systems [4].

The application of machine learning technology has advanced in the retail sector. With the development of big data and artificial intelligence in recent years, a growing number of shopping malls have begun to focus on how to use machine learning and other technologies to improve marketing effects and optimize business planning. To this end, the academic community has conducted extensive research on consumer behavior, shopping habits, and preferences [5] in order to provide shopping malls with targeted marketing strategies and optimization recommendations. However, there is currently no research on the cluster analysis of stores in large shopping malls and the optimization of marketing strategies based on the results of such an analysis.

Based on this, the primary objective of this paper is to cluster the stores in a shopping mall using three distinct clustering algorithms. We hope that by implementing these algorithms, we will be able to provide store managers with a more transparent store classification structure, allowing them to better meet customer needs and optimize store operations. We will compare and evaluate the effects of K-means clustering, spectral clustering, and DEC deep clustering to determine the optimal clustering technique. After completing the clustering process, we will provide store managers with practical suggestions based on the characteristics of each category to assist them in developing more appropriate marketing strategies for various types of stores.

The following are the main contributions of this paper:

(1) Comprehensive comparative analysis: By comparing the application effects of K-means, spectral clustering, and DEC deep clustering in shopping malls, we can provide more comprehensive evaluation results of clustering methods to managers of shopping malls, as well as determine the most appropriate and effective clustering algorithm for various application scenarios.

(2) Innovative method design: This paper applies DEC deep clustering algorithm to the cluster analysis of shopping malls and stores, and discusses the application of a new method based on deep clustering in actual scenes, which provides new ideas and references for similar research in the future.

(3) Verifiable experimental cases: This paper makes an empirical study based on the actual shopping mall data, which increases the practical application cases of clustering algorithm in the commercial field and improves the practical value of the research.

2. Related work

Numerous researchers have utilized the machine learning scheme for business simulation and retail management planning. Razmochaeva [6] and others have utilized a number of techniques to reduce the dimension of feature space in retail sales management, focusing on the t-SNE algorithm, a nonlinear dimension reduction algorithm whose purpose is to map high-dimensional data to low-dimensional space for visualization. T-SNE captures similarity in input data by establishing a Gaussian similarity matrix and a T similarity matrix in low-dimensional space. T-SNE minimizes Kullback-Leibler divergence between high-dimensional and low-dimensional transformations, allowing data points to maintain maximum similarity in both high- and low-dimensional spaces. Calculating the similarity matrix between high-dimensional data points is the initial step.

$$p_{ij} = \frac{\exp\left(-\| x_i - x_j \|^2 / 2\sigma_i^2\right)}{\sum_{k \neq l} \exp\left(-\| x_k - x_l \|^2 / 2\sigma_i^2\right)} \tag{1}$$

Among them, the similarity matrix $p$is a symmetric matrix, which indicates the similarity between each point and other points. Next, calculate the similarity matrix in the low-dimensional space$q$:

$$q_{ij} = \frac{\left(1+\| y_i - y_j \|^2\right)^{-1}}{\sum_{k \neq l}(1+\| y_k - y_l \|^2)^{-1}} \tag{2}$$

Similarity matrix$q$ is used to describe the similarity of data points in low-dimensional space and $y$is a variable that needs to be optimized. Finally, calculate the relative entropy of high-dimensional and low-dimensional space, that is, Kullback-Leibler divergence:

$$C = KL(P| \quad |Q) = \sum_{i \neq j} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) \tag{3}$$

The position of each data point is updated along the mapping from high-dimensional space to low-dimensional space until the goal of minimizing KL divergence is achieved using optimization algorithms such as gradient descent. It is stated that reducing the dimension of the feature space will not result in the loss of information, but will lessen the load on the computing system.

Tug C e Bilen [7] and others have developed a smart city application utilizing machine learning technology for the purpose of urban retail location planning. This application is capable of collecting the key characteristic values of a specific business, learning the prediction model for future data, estimating the characteristics, and recommending the optimal regional cluster for business location. This application's fundamental algorithms include hierarchical clustering. Hierarchical clustering is used in this application to classify the estimated features and identify urban areas with similar features. This aids in providing entrepreneurs with the best advice for establishing businesses in locations similar to their own. Specifically, each characteristic is standardized and input into hierarchical clustering to reduce the error rate. Then, hierarchical clustering of features and regions is employed to categorize all regions. London is divided into three main clusters based on housing prices and population size, with the Web client API providing access to specific regional clusters. The benefit of this algorithm is that it can cluster the data set without making too many assumptions, resulting in more accurate and stable clustering results. Martin Prause[8] and others proposed a benchmark trade model suite for testing machine learning algorithms and decision simulation. This model benchmark suite is capable of simulating a particular economic market environment, which includes numerous enterprises with distinct strategies and a complex decision-making and product trading procedure. In this benchmark suite, the author successfully applied deep reinforcement learning to the dynamic enterprise decision-making market environment and evaluated the performance of various machine learning algorithms, including deep Q-learning, e- greedy strategy, and target network in this environment. The research findings indicate that these algorithms are capable of accurately simulating the behavior of the actual economic market. However, the author also notes the method's limitations and suggests that a mixed approach combining economic principles and deep learning technology may yield superior results. In addition, these studies are geared toward macroeconomic market analysis and provide optimization recommendations and effective strategies, but there are still limitations for store planning and the economic operation of a particular shopping mall.

3. Solution

In order to solve the above problems, we will build a mathematical model to cluster the shops in a specific shopping mall. In this section, we divide the proposed method into two steps: data preprocessing and data modeling.

3.1 Data preprocessing

Data preprocessing is a crucial step in projects involving machine learning and data analysis. Before applying the machine learning algorithm, this refers to the process of cleansing, transforming, and standardizing the original data. To improve data quality and reduce inconsistencies in data. This procedure can help us enhance the performance of the machine learning model and gain a deeper

understanding of the data's potential patterns. In practice, the data we've collected may contain numerous issues, such as missing values, abnormal values, noise, and inconsistent measurement units. These issues may affect the subsequent analysis and modeling negatively. Through data preprocessing, we must improve the data quality and performance of the subsequent machine learning model. In addition, various characteristics may have distinct scales and distributions. One feature may have a range of 0 to 1, while another feature may have a range of 1 to 1,000. This difference in scale will result in an unbalanced weight distribution of different features in the machine learning algorithm, which will negatively impact the performance of the model. Through data preprocessing, we can transform features to the same scale, allowing machine learning algorithms to more accurately identify patterns in data.

Moreover, data preprocessing can help us better comprehend the data and identify potential issues. During the data preprocessing phase, we can visually examine the data to identify anomalous values and potential trends. This will help us make better decisions during the modeling process that follows.

In our original data set, we included the following attributes: store ID, transaction date, transaction time, transaction amount, transaction bank, and the last four digits of the transaction bank card. Nonetheless, some features are superfluous when the shops are divided into distinct levels. If these characteristics are considered in the clustering model, it will result in redundant data or a subpar clustering effect. Therefore, we use the preprocessing method of feature aggregation, based on the transaction time and date, to determine the type of transaction (whether it is a transaction on a working day or a transaction on a weekend), count the number of transactions on a working day and on a weekend in different stores, as well as the average transaction amount, as the final features, and establish a mathematical model for clustering.

3.2 Data Modeling

3.2.1 Clustering algorithm based on K-means

The K-means clustering algorithm is a popular and widely employed unsupervised learning technique that discovers hidden structures in data sets by grouping data points with similar characteristics. Its central concept is to find the optimal clustering division by minimizing the sum of square errors between data points in each cluster and its cluster center. The output of K-means clustering is K clusters, each of which has a cluster center that represents the mean value of the data points within it. This algorithm is applicable to numerous fields, including market segmentation, text classification, and image segmentation, among others.

The K-means clustering algorithm selects K data points at random from the data points to serve as the initial cluster center. These cluster centers are typically represented as vectors with the same dimensions as data points.

In each iteration, the following two steps are performed:

(1) Assignment: Assign each data point to the nearest cluster center. The distance here is usually measured by Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4}$$

Where $x$and $y$are two data points, $n$is the dimensions of data points.

(2) Update: Recalculate the center of each cluster and update it to the average of all data points in the cluster:

$$c_j = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_i \tag{5}$$

Where c_j is the center of the j cluster, S_j is the set of data points belonging to the j cluster, and $|S_j|$represents the number of data points in the j cluster.

Repeat the allocation and update operations in step 2 until the convergence condition is satisfied. Typically, the convergence condition is satisfied when the change in cluster center is less than a predetermined threshold or when the maximum number of iterations is reached. When the algorithm converges, we obtain k clusters and the centers of each cluster. These clusters group data points with similar characteristics into groups. Notably, because the K-means clustering algorithm uses a random initialization, it may converge on a local optimal solution. Typically, it is necessary to execute the algorithm multiple times and select the result with the smallest sum of squares of intra-cluster errors in order to achieve a superior clustering effect.

Although the K-means clustering algorithm performs well in a variety of applications, it is not without its limitations. First, it is necessary to determine the value of k in advance, which can be difficult in real-world situations. Second, the algorithm is sensitive to the choice of initial clustering centers, and different initial values may produce distinct outcomes. In addition, the K-means clustering algorithm may be impacted by noise data and outliers, and its performance may suffer when dealing with non-spherical clusters or clusters of varying sizes and densities.

3.2.2 Clustering algorithm based on spectral clustering

Spectral clustering algorithm is a graph-theory-based clustering method. By spectral decomposition of the similarity matrix between data points, the data in high-dimensional space is mapped to low-dimensional space, allowing conventional clustering methods (such as K-means) to be used for clustering. By analyzing the structure of the graph, the spectral clustering algorithm identifies potential clusters by treating data points as nodes in a graph and similarity between them as the weight of edges. The performance of the spectral clustering algorithm is excellent, particularly when dealing with non-convex or complex clusters.

Initial construction of the similarity matrix of data points. Each element of this symmetric matrix represents the similarity between two data points. The most common measure of similarity is the Gaussian kernel function:

$$w_{ij} = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \tag{6}$$

Where w_{ij} represents the similarity between data points x_i and x_j, and σ is the width parameter of Gaussian kernel.

Then, the degree matrix and Laplace matrix are constructed. The degree matrix is a diagonal matrix, and the elements on the diagonal are equal to the sum of the elements in the corresponding rows (or columns) in the similarity matrix. Laplace matrix calculation formula is:

$$L = D - W \tag{7}$$

Where$D$ is the degree matrix and W is the similarity matrix.

The eigenvectors and eigenvalues of the Laplacian matrix are then determined. These eigenvectors and eigenvalues provide crucial data regarding the data structure. To create a new matrix, select the eigenvectors corresponding to the first k minimum eigenvalues. Each data point is mapped to a low-dimensional space based on the values of these k eigenvectors. Traditional clustering algorithms (such as K-means) are used to cluster data points in this low-dimensional space. This will generate k clusters along with their respective cluster centers. In order to obtain the final clustering result, these clusters are mapped back to the initial data space. By capturing the geometric structure of the data in a low-dimensional space, spectral clustering reveals the cluster structure in the original data.

Numerous fields, such as image segmentation, social network analysis, and so on, have made extensive use of spectral clustering algorithms. Compared to other clustering methods, the spectral clustering algorithm is better able to capture the global structure of data, allowing it to handle complex

data with greater efficiency. However, the spectral clustering algorithm has some drawbacks, including high computational complexity and noise sensitivity.

3.2.3 Deep clustering algorithm based on DEC

Deep clustering algorithm is a technique that combines deep learning and conventional clustering algorithm. It can discover potential clusters in high-dimensional space after learning the complex, nonlinear structure of data. In comparison to conventional clustering algorithms, deep learning technology can learn the abstract, high-level characteristics of data and reveal its potential structure. This improves the algorithm's performance when dealing with large-scale, high-dimensional data and allows it to model complex data structures more precisely. The deep clustering algorithm based on DEC(Deep Embedded Clustering)[9] is an example of a typical deep clustering technique.

DEC algorithm aims to cluster N data points $x_i \in X * i = 1^n$into K clusters, each of which has a centroid $u_j, j = 1, \dots, k$. DEC does not cluster data space X directly, but maps data space X to feature space Z through nonlinear mapping $f * \theta: X \to Z$, where θ is a learnable parameter. Nonlinear mapping $f_\theta$is usually approximated by neural network.

Then, DEC algorithm has two goals: 1) learning K cluster centers$u_j \in Z * j = 1^k$ in feature space Z; 2) Learn the network parameters θthat map data to the feature space Z .. In order to achieve these two goals, DEC uses unsupervised alternating two-stage method to improve the clustering effect. In the first stage, the soft allocation of embedded nodes and cluster centers is calculated; In the second stage, the mapping is updated $f * \theta$and the cluster center is refined from the current high confidence distribution by using the auxiliary target distribution.

When calculating soft allocation, we use T distribution as a measure of the similarity between embedded nodes and cluster center, and the specific formula is as follows:

$$q_{ij} = \frac{\left(1+\| z_i - u_j \|^2/\alpha\right)^{-\frac{\alpha+1}{2}}}{\sum_{j'}\left(1+\| z_i - u_{j'} \|^2/\alpha\right)^{-\frac{\alpha+1}{2}}} \tag{8}$$

Where $z_i = f_\theta(x_i) \in Z$ is the vector after $x_i \in X$ is embedded; α is the degree of freedom of t distribution (let $\alpha = 1$); $q_{ij}$is considered as the probability of assigning sample I to cluster J.

Next, in the stage of minimizing KL divergence, we train the model by comparing the soft distribution obtained above with the target distribution. In order to achieve this goal, we define a loss function based on KL divergence to measure the gap between soft distributionq_i and auxiliary distributionp_j. The specific formula is as follows:

$$L = KL(P \parallel Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{9}$$

Among them, q_{ij} is the soft distribution obtained above, andp_{ij} is a target distribution. Target distribution should have the following properties: it can improve the cohesion of clusters in clustering; Can pay more attention to data points with high confidence distribution; The contribution of each cluster center to the loss is standardized to prevent large clusters from distorting the feature space. We choose to square the soft allocation probability q_i, so as to realize the target distribution, that is:

$$p_{ij} = \frac{q_{ij}^2/f_j}{\sum_{j'} q_{ij'}^2/f_{j'}} \tag{10}$$

Where$f_j = \sum_i q_{ij}$ is the soft frequency.

Through the alternate optimization of these two stages, the DEC algorithm can identify a suitable cluster division in the feature space Z and discover a neural network that maps the original data to the feature space. Thus, the clustering results in the feature space are able to capture the nonlinear

structure of the original data, resulting in a more effective clustering effect. By updating the mapping and cluster center, we can obtain the final clustering result.

The primary benefit of the DEC algorithm is its ability to effectively capture complex data structures and cluster them in low-dimensional space. In addition, the DEC algorithm is suitable for large-scale, high-dimensional data sets due to its scalability. However, the DEC algorithm's convergence speed is slow and requires many iterations. As a method for deep clustering, the DEC algorithm has great potential and benefits. When dealing with complex data, it can provide more detailed modeling capabilities and valuable insights for data analysis and interpretation. By combining deep learning technology with a clustering algorithm, the DEC algorithm can not only maintain good performance when dealing with large-scale, high-dimensional data, but also mine data for complex structures and potential relationships. In addition, the DEC algorithm has strong generalization capabilities and is applicable to numerous data types. By adjusting the structure and parameters of the self-encoder, we can adapt to various application scenarios involving images, texts, and data from other fields. This broadens the potential applications of the DEC algorithm to fields such as computer vision, natural language processing, and bioinformatics. In practical applications, the DEC algorithm may face challenges such as selecting an appropriate self-encoder structure and adjusting super-parameters. Moreover, the defined clustering loss may destroy the feature space and result in meaningless and unrepresentative features, thereby affecting the clustering effect. Consequently, grid search and Bayesian optimization can be utilized to determine the optimal model configuration. Moreover, the DEC algorithm can be further improved, such as with the IDEC algorithm [10] to safeguard the data structure.

4. Results and Discussion

We run our model on a MacBook with a m1 chip using Python 3.8. We cluster the preprocessed data using the K-means algorithm, with the goal of classifying the stores into K distinct categories. We extract the features'weekday_count' (number of working days),'weekend_count' (number of weekends), and'avg_amount' (average amount) after reading the data set. Then, these characteristics are standardized to make the data comparable across characteristics. In order to determine the division number k, we must quote an evaluation index to evaluate the clustering effect for various values of k, and we choose the contour coefficient. Contour coefficient is an index for assessing the quality of clustering results, with a range of -1 to 1. The clustering effect is greater the larger the value. We calculate the contour coefficient for each clustering result by employing the K-means algorithm to traverse different cluster numbers (from 2 to 7). The cluster number with the greatest contour coefficient is then identified as the optimal cluster number, and this optimal cluster number is utilized for K-means clustering once more. Figure 1 depicts the contour coefficients obtained by varying cluster numbers.

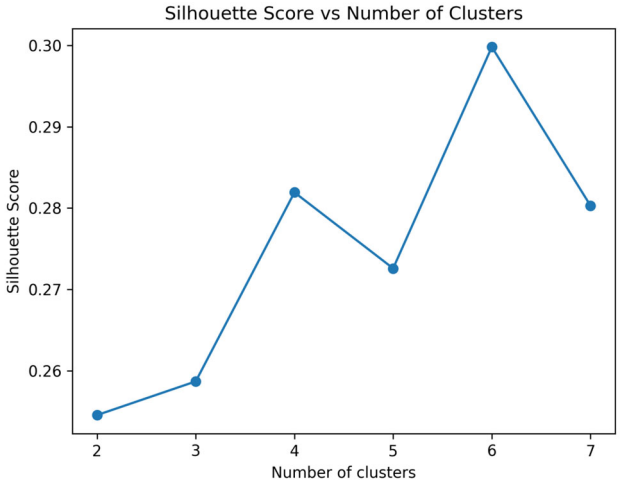

Figure 1. contour coefficient of k-means after clustering with different cluster numbers

When K = 6, the classification effect is optimal and the contour coefficient is 0.2998. After determining the value of K, we cluster the data using K-means, as depicted in Figure 2.

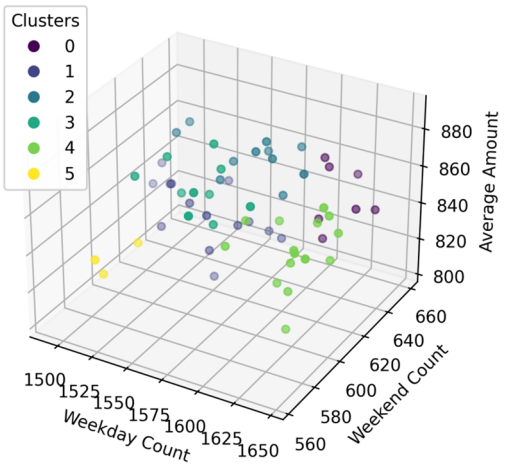

Figure 2. clustering effect of K-means when k = 6

Similarly, in order to determine the optimal cluster number of spectral clustering, we quoted the contour coefficient as our evaluation index, and the contour coefficient after clustering with different cluster values is shown in Figure 3.

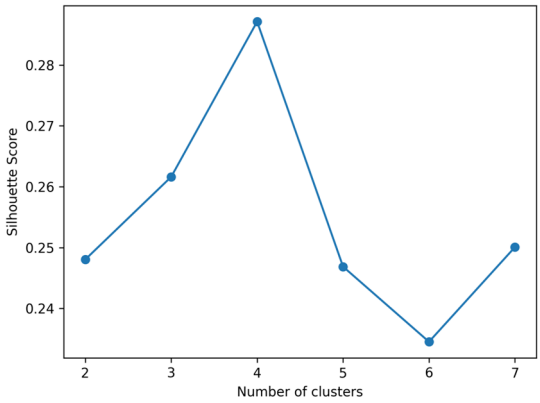

Figure 3. Profile coefficients of spectral clustering with different cluster numbers.

The classification effect is optimal when there are four clusters. Currently, the contour coefficient is 0.2871, and Figure 4 depicts the clustering outcome.

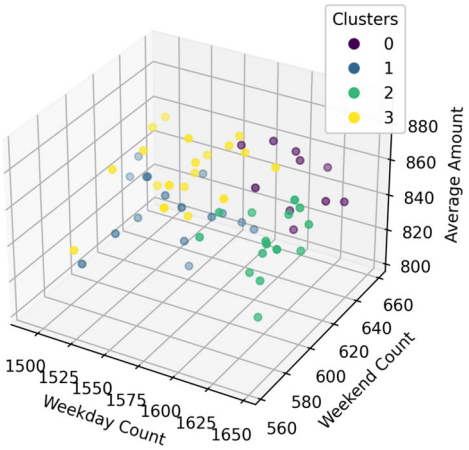

Figure 4. clustering effect of spectral clustering when the number of clusters is 4.

After data feature extraction and standardization for DEC deep clustering, we define a simple self-encoder and extract the new feature representation of the data via the coding portion. The DEC algorithm is then used to cluster the newly added features, and the evaluation index is computed. Here, the self-encoder is primarily used to extract the low-dimensional representation of the original data, allowing the DEC algorithm to cluster more effectively in this low-dimensional space. Coding and decoding comprise the self-encoder. The coding component is responsible for mapping the original data to a low-dimensional space, while the decoding component restores the low-dimensional representation to the original data. By training the self-encoder to minimize reconstruction error, we can obtain a low-dimensional representation that is capable of capturing the data's internal structure. Specifically, we define a self-encoder with two hidden neurons and the mean square error as the loss function. The gradient descent optimizer is then used to train the self-encoder to minimize the difference between the original and reconstructed data. Upon completion of training, the original data is transformed into a low-dimensional representation using the coding portion.

After obtaining the new feature representation, we experiment with different cluster numbers using the DEC clustering algorithm. Calculating the contour coefficient yields the optimal cluster number, which is then added to the data set. Figure 5 depicts the clustering outcome.

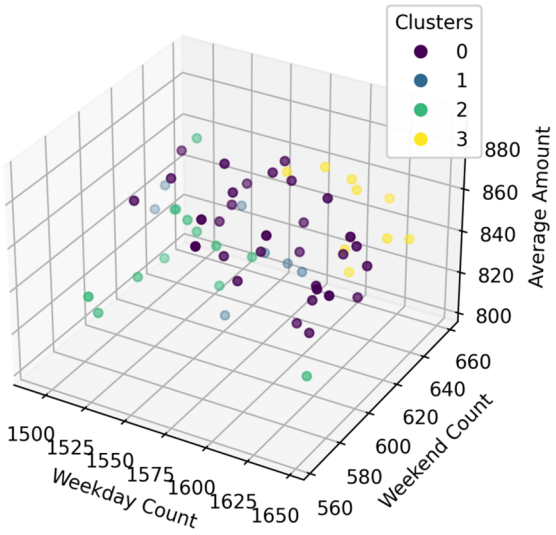

Figure 5. DEC depth clustering effect

The contour coefficients of three different clustering methods are shown in Table 1.

Table 1. Contour coefficients of three different clustering methods

Clustering method	Contour coefficient
K-means clustering	0.2998
Spectral clustering	0.2871
DEC deep clustering	0.5552

According to the data set, the preprocessed data has the characteristics of high similarity, low data dimension, and small data volume; however, all three clustering methods can effectively cluster the data set, with the performance of the traditional K-means clustering algorithm and spectral clustering algorithm being comparable, and the clustering result having been significantly enhanced by the introduction of deep clustering. Consequently, compared to the traditional clustering algorithm, the deep clustering algorithm is not only capable of handling more modal and complex data, but also performs better with smaller data sets. However, compared to traditional clustering, deep clustering frequently requires more system resources and longer processing time, so clustering algorithms suited to particular scenarios must be chosen based on varying requirements.

5. Conclusion

Using machine learning algorithms to improve operating efficiency and profitability in shopping malls has become the preferred method as a result of the continuous advancement of commercial technology code. This paper demonstrates the application of the machine learning algorithm to the clustering of shopping malls and compares the benefits and limitations of the traditional clustering algorithm and the deep embedding clustering algorithm. This paper discusses the significance of store clustering in shopping malls and the shortcomings of traditional methods, such as inaccurate, incomplete, and subjective data. Comparing the experimental outcomes of K-means clustering, spectral clustering, and depth clustering algorithms applied to shopping malls, we find that the depth clustering algorithm has significant advantages in terms of clustering accuracy, but it has deficiencies in terms of computing resource consumption and algorithm efficiency.

The results indicate that machine learning technology can increase the operational efficiency and profitability of shopping malls, but different clustering algorithms have advantages and disadvantages. In shopping mall clustering, traditional methods can effectively cluster shopping malls and stores. These algorithms are simple, straightforward, and quick to execute. Spectral clustering algorithm can adapt to complex data and irregular data points, but it is slower and more sensitive to noise than K-means clustering algorithm. However, the application of machine learning algorithm in shopping mall clustering is not yet mature, and the application effect in various shopping mall data and situations requires further research and investigation, particularly in the processing and application of large-scale shopping mall data.

Deep learning and analysis of original data, better extraction of feature representation, and nonlinear clustering are advantages of the deep clustering algorithm, allowing for more accurate clustering of shopping malls and stores. Experiments demonstrate that deep clustering has significant clustering accuracy advantages over conventional clustering methods. However, the computational requirements and algorithmic performance of the deep embedding clustering algorithm are relatively high, posing challenges for the management and application of large-scale shopping mall data. Therefore, when selecting clustering algorithms, shopping malls must strike a balance between efficiency and accuracy, and use these algorithms in the most effective manner to improve operational efficiency.

This research provides new ideas and methods for clustering shopping malls, which are crucial for enhancing the operational efficiency and profitability of shopping malls. Commercially, machine learning technology has become a hot topic. Shopping malls must continuously investigate the optimization and application of machine learning methods based on the premise of optimizing data collection, processing, and application methods in order to better respond to market shifts and consumer demand.

References

[1] Castelvecchi, D. (2016). Deep learning boosts Google Translate tool. Nature, 10.

[2] Bouguettaya, A., Zarzour, H., Kechida, A., & Taberkit, A. M. (2022). Machine Learning and Deep Learning as New Tools for Business Analytics. In Handbook of Research on Foundations and Applications of Intelligent Business Analytics (pp. 166-188). IGI Global.

[3] Abdou, M.A. Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput & Applic 34, 5791–5812 (2022).

[4] Zhang X, Li Z, Gong Y, et al. OpenMPD: An Open Multimodal Perception Dataset for Autonomous Driving[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 2437-2447.

[5] Klieštik, T., Zvarikova, K., & Lăzăroiu, G. (2022). Data-Driven Machine Learning and Neural Network Algorithms in the Retailing Environment: Consumer Engagement, Experience, and Purchase Behaviors. Economics, Management, and Financial Markets, 57–69.

[6] N. V. Razmochaeva, D. M. Klionskiy and V. V. Chernokulsky, "The Investigation of Machine Learning Methods in the Problem of Automation of the Sales Management Business-process," 2018 IEEE

International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS), St. Petersburg, Russia, 2018, pp. 376-381, doi: 10.1109/ITMQIS.2018.8525008.

[7] T. Bilen, M. Erel-Özçevik, Y. Yaslan and S. F. Oktug, "A Smart City Application: Business Location Estimator Using Machine Learning Techniques," 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 2018, pp. 1314-1321, doi: 10.1109/HPCC/SmartCity/DSS.2018.00219.

[8] M. Prause and J. Weigand, "Market Model Benchmark Suite for Machine Learning Techniques," in IEEE Computational Intelligence Magazine, vol. 13, no. 4, pp. 14-24, Nov. 2018, doi: 10.1109/MCI.2018.2866726.

[9] Xie, J., Girshick, R. & Farhadi, A.. (2016). Unsupervised Deep Embedding for Clustering Analysis. Proceedings of The 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research 48:478-487 Available from https://proceedings.mlr.press/v48/xieb16.html.

[10] Guo, X., Gao, L., Liu, X., & Yin, J. (2017, August). Improved deep embedded clustering with local structure preservation. In Ijcai (pp. 1753-1759). https://proceedings.mlr.press/v48/xieb16.html.