Application of Cluster Analysis in Optimizing Financial Product Recommendation System

Ni Rui

Department of Computer Science, Hong Kong Baptist University, Hong Kong, China
nirui_nory@outlook.com

Abstract: A financial product recommendation system is an important tool for financial institutions in financial services to enhance customer experience and satisfaction. Such systems provide personalized recommendations, assisting users in identifying and selecting appropriate financial products. However, traditional recommendation methods, including collaborative filtering and content-based recommendation, frequently suffer from limited personalization capabilities due to challenges like sparse data and user cold-start problems. This study investigates the application of cluster analysis in financial product recommendation systems and the role of various clustering algorithms (e.g., K-means, DBSCAN, and hierarchical clustering) in customer segmentation and personalized recommendations and evaluates their advantages and challenges in optimizing recommendation strategies. The results show that clustering analysis can effectively group users and products, significantly improving product matching and, thus, recommendation accuracy and customer satisfaction. In addition, this paper discusses the main challenges of applying clustering techniques to financial recommendation systems, such as ensuring the scalability of clustering models and accurately evaluating validation clustering results.

Keywords: Financial product recommendation system, clustering algorithms, personalization

1. Introduction

With the continuous growth of the financial market and the increasing diversification of customer needs, financial product recommender systems have become one of the most effective ways for financial institutions to enhance competitiveness and optimize customer experience. By using personalized recommendation technologies, institutions can effectively match customers with suitable financial products, significantly enhancing customer satisfaction, engagement, and conversion rates. According to the Global Fintech Report 2019 [1], more than 50% of financial institutions have applied data-driven recommender systems to optimize product promotion and have found that personalized suggestions can significantly increase customers' product purchase intention. However, traditional methods, like collaborative filtering and content-based recommendation, encounter significant limitations. These limitations include data sparsity issues, cold-start problems—where insufficient historical data on new users reduces recommendation accuracy—and inadequate personalization due to overreliance on past behavior data [2].

To overcome these challenges, increasing research is focusing on the application of cluster analysis in financial product recommender systems. Cluster analysis is an unsupervised learning method that can classify customers into different groups based on their age, income, consumption habits, and other

aspects to segment customers effectively and deliver highly personalized financial product recommendations without extensive historical data reliance [3]. It can effectively distinguish the potential needs of customer groups and optimize the recommended content according to customer groups' features, thereby significantly enhancing the accuracy of product matching and improving customer satisfaction.

This paper systematically reviews existing research on the integration of cluster analysis methods, specifically focusing on popular clustering algorithms such as K-means, DBSCAN, and hierarchical clustering within financial recommendation systems. It evaluates their respective strengths and weaknesses in customer segmentation, analyzes their potential to optimize recommendation strategies, and identifies ongoing challenges such as model scalability and accurate evaluation of clustering performance. Ultimately, this review aims to provide clear insights into how clustering-based techniques can effectively address personalization shortcomings in traditional recommendation approaches and outline directions for future research in the field.

2.   Traditional Methods and Their Limitations

Financial product recommendation systems play a vital role in modern financial services. They aim to recommend appropriate financial products to users based on their historical behaviors and preferences, thereby improving user experience and satisfaction [4]. However, traditional recommendation methods have some limitations in practical applications, prompting researchers to explore new technical means to optimize the performance of recommendation systems. Cluster analysis, as an unsupervised learning method, can effectively segment customers and provide new ideas for optimizing recommendation systems.

2.1.   Traditional Recommendation Methods

Traditional financial product recommendation methods mainly include collaborative filtering, content-based recommendation, and hybrid recommendation models. Collaborative filtering analyzes the user's historical behavioral data to identify similar user preferences and then recommends products that other similar users have selected. This approach has the advantage of being able to discover potential user interests, but it also suffers from problems such as data sparsity and cold start. Content-based recommendation, on the other hand, utilizes the feature information of a product to recommend products similar to its historical preferences for the user. This approach performs well when dealing with new products but may lead to a lack of diversity in recommendation results [5]. Hybrid recommendation models combine both collaborative and content-based approaches to address the weaknesses of each method individually. Such hybrid systems integrate multiple data sources and algorithms, thus improving recommendation accuracy, personalization, and product diversity.

2.2.   Limitations of traditional recommendation methods

Although traditional recommendation methods satisfy users' needs to a certain extent, they still have challenges in dealing with data sparsity, cold-start problems, and personalized recommendations. For example, collaborative filtering is hindered by data sparsity, resulting in reduced accuracy when historical interaction data is insufficient. Content-based recommendation, on the other hand, may not be able to adequately capture the diverse needs of users, often fail to provide sufficient recommendation diversity, leading to overly homogeneous suggestions that limit user satisfaction [6]. Therefore, the introduction of new technical tools, such as cluster analysis to segment customers, has become an important direction to improve the performance of recommender systems.

3. Application and Advantages of Cluster Analysis for Personalized Recommender Systems

Cluster Analysis is an unsupervised learning method that aims to classify objects into groups based on the similarity between data points. The core principle is to group individuals with similar characteristics into the same cluster by measuring the distance or similarity between data points, while objects between different clusters exhibit large differences. It can classify customers into different groups based on their behavioral and demographic characteristics. This customer segmentation strategy enables financial institutions to provide more targeted financial product recommendations based on the characteristics of different groups, thereby improving recommendation accuracy and customer satisfaction [3]. The primary advantages of cluster analysis in recommendation systems include the ability to handle high-dimensional, complex datasets, accurately capturing intricate relationships among diverse customer attributes and preferences [7]. Additionally, segmenting customers into distinct clusters allows recommender systems to deliver more diversified recommendations, enhancing user experience and satisfaction [8]. In addition, for new users or new products, clustering analysis can alleviate the cold-start problem and enhance the recommendation effect by identifying groups with similar characteristics [9].

3.1. Comparative Analysis of Clustering Algorithms in Financial Recommendation Systems

3.1.1. K-means

K-means works by dividing the dataset into K clusters so that the data points within each cluster have a high degree of similarity and a high degree of variation across clusters. The process of the K-means algorithm consists of three steps: initialization, assigning data points, and iteration. First, initialization determines the number of clusters K by randomly selecting K data points as the initial clustering centers. Second, data points are assigned by calculating the distance between each data point and each cluster center, assigning it to the nearest cluster, and then updating the clustering centers and re-calculating the center of each cluster, i.e., the mean value of all the data points in the clusters. Finally, in the iteration step, repeat the above steps of assigning and updating until the clustering centers no longer change significantly or reach a predetermined iteration. Centers no longer change significantly or a predetermined number of iterations are reached [10]. This process ensures the similarity of data points within clusters and the difference of data points between clusters.

In a financial product recommendation system, the K-means clustering algorithm is mainly used for clustering users or items to achieve personalized recommendations. In user grouping, K-means divides users into different clusters based on their behavioral characteristics (e.g., browsing history, purchase records, etc.). For users in the same group, the recommender system can provide similar products or services to improve the relevance of recommendations and user satisfaction. In item clustering, K-means divides items into different clusters based on their characteristics (e.g., category, price, brand, etc.) [9]. In this way, the system can recommend other items belonging to the same cluster according to the user's interest, increasing the diversity of recommendations.

3.1.2. DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that discovers arbitrarily shaped clusters and efficiently handles noisy points. In recommender systems, DBSCAN improves the accuracy of recommendations by identifying data points in high-density regions and grouping them into the same cluster.

DBSCAN achieves clustering mainly through the following steps: first, identify the core points, and for each data point, calculate the number of its neighboring points within a radius ε. The number

of points in the radius ε is calculated if the number is not less than the minimum number of points. If the number is not less than the minimum number of points MinPts, the point is regarded as a core point; secondly, the cluster expansion is carried out, starting from the core point, and the points in its neighborhood are grouped into the same cluster. For a new core point, continue to expand its neighborhood until the cluster no longer grows. Finally, noise points are handled; points that are not included in any cluster are considered noise points [11]. This approach allows DBSCAN to discover clusters of arbitrary shape and is robust to noise.

In recommender systems, DBSCAN can also be used for user clustering and item clustering, but its mechanism to improve the accuracy is mainly the following two points: dealing with noisy data and discovering arbitrarily shaped clusters. DBSCAN is able to identify and exclude the noisy points and reduce the influence of abnormal data on the recommendation results. Compared with other clustering algorithms, DBSCAN is able to recognize clusters of arbitrary shapes, adapt to complex data distribution, and improve the diversity of recommendations.

3.1.3. Hierarchical clustering

Hierarchical Clustering is a method used to analyze the similarity of objects in a dataset by constructing a hierarchical structure and gradually merging or splitting data points to form clusters. Hierarchical clustering is mainly divided into the following two strategies: The first one is Agglomerative Hierarchical Clustering (AGNES), using a bottom-up approach, where each data point is first considered as an independent cluster and then gradually merged based on similarity until all data points are merged into a single cluster or a predefined number of clusters is reached. The second is Divided Hierarchical Clustering (DIANA): a top-down approach is used, where all data points are first considered as a whole cluster and then gradually split according to the differences until each data point becomes an independent cluster or reaches a predetermined number of clusters.

This hierarchical clustering method does not require the number of clusters to be pre-specified and can generate a tree-like hierarchy (dendrogram) that visually demonstrates the relationships between data points [12].

In recommender systems, DBSCAN can also be used for user clustering and item clustering, but its mechanism to improve the accuracy is mainly the following two points: The first point is to reveal the hierarchical structure of the data. Hierarchical clustering can show the hierarchical relationship of the data and help to recognize the inner structure of the user or the item to provide more targeted recommendations. The second point is to deal with complex data distribution where hierarchical clustering does not need to pre-specify the number of clusters and can adapt to complex data distribution and improve the diversity of recommendations.

3.2. Challenges and Future Directions of Clustering Algorithms in Applications

Although clustering algorithms are widely used in financial product recommendation systems, there are still some challenges that highlight research gaps and areas for future exploration.

First, as the size and complexity of datasets increase, traditional clustering algorithms are often overwhelmed in terms of efficiency and performance. For example, algorithms such as hierarchical clustering become computationally prohibitive when dealing with large-scale data due to high time complexity. Therefore, there is a need to develop more scalable clustering techniques to handle big data efficiently [13]. For example, parallel computing and distributed frameworks can be introduced. Techniques like MapReduce and Apache Spark are widely used to implement parallel versions of traditional clustering algorithms such as K-Means and Fuzzy C-Means. These frameworks allow for distributed computing, which significantly improves scalability and efficiency by processing data across multiple nodes [14].

Second, incorporating more financial domain expertise into the clustering process can improve the relevance and accuracy of the results. However, formalizing and integrating this knowledge remains challenging. Semi-supervised clustering methods using constraints based on expert inputs are being explored to address this problem, but practical applications require further research [15].

Finally, assessing the quality of clustering results is inherently challenging due to the unsupervised nature of the clustering task. Establishing standardized evaluation metrics and validation techniques is essential for comparing and improving clustering algorithms [15].

4. Conclusion

In this paper, we have addressed the application of cluster analysis in financial product recommendation systems with a focus on customer segmentation and the optimization of recommendation strategies. Traditional methods such as collaborative filtering and content-based recommendation often struggle with challenges like data sparsity, cold-start problems, and a lack of personalization. Cluster analysis provides an effective alternative by segmenting customers based on behavioral and demographic characteristics, enabling more personalized recommendations that enhance user experience and business efficiency.

We discussed three of the most popular clustering algorithms—K-means, DBSCAN, and hierarchical clustering—highlighting their concepts, advantages, and drawbacks in improving financial product recommendations. K-means clustering is widely used for its efficiency in segmenting customers based on predefined numerical features, though its reliance on a fixed number of clusters can be a limitation. DBSCAN offers more flexibility by identifying clusters of arbitrary shapes and handling noisy data, making it suitable for complex financial datasets. Hierarchical clustering, on the other hand, provides a structured and visual representation of relationships between data points, allowing for adaptive and interpretable customer segmentation.

Despite their benefits, clustering methods also face certain limitations when applied to financial recommender systems. As financial data continues to grow in volume and complexity, more scalable clustering techniques are needed to process large datasets efficiently. Additionally, integrating domain knowledge into clustering models remains an ongoing research challenge, with semi-supervised clustering presenting a potential solution. Evaluating clustering quality in an unsupervised setting also requires standardized validation techniques to ensure the effectiveness of recommendation strategies.

Future research should focus on improving the scalability of clustering algorithms, incorporating financial domain expertise, and developing more effective evaluation metrics. Combining clustering with deep learning and hybrid recommendation approaches could further enhance personalization, accuracy, and adaptability in financial services. By leveraging advanced clustering techniques, financial institutions can provide smarter, more user-centric recommendations, ultimately driving customer engagement and business growth.

Acknowledgements

References

[1] EY, Global FinTech Report 2019: Reaching the Tipping Point. Ernst & Young Global Limited, 2019.

[2] Y. Xu et al., "The improvement of personalized recommendation system based on sparse data of financial products," World Journal of Research and Review, vol. 7, no. 2, 2018. doi: 10.31871/wjrr.7.2.18.

[3] W. Luo et al., "Application of improved clustering algorithm in investment recommendation in embedded system," Microprocessors and Microsystems, vol. 75, p. 103066, 2020. doi: 10.1016/j.micpro.2020.103066.

[4] X. Yue et al., "Application of AI technology in personalized recommendation system for financial services," Applied Mathematics and Nonlinear Sciences, vol. 9, 2024. doi: 10.2478/amns-2024-1349.

[5] C. Musto, G. Semeraro, P. Lops, M. De Gemmis, and G. Lekkas, "Financial Product Recommendation through Case-based Reasoning and Diversification Techniques," presented at the 8th ACM Conference on Recommender Systems (RecSys), Foster City, CA, USA, 2014.

[6] L. Wang, Y. Liu, and J. Wu, "Research on financial advertisement personalised recommendation method based on customer segmentation," International Journal of Wireless and Mobile Computing, vol. 14, pp. 97–101, 2018. doi: 10.1504/IJWMC.2018.10011092.

[7] V. Kumar, S. Ramesh, and S. Gupta, "High-dimensional clustering techniques for customer behavior analysis," U niversity of Minnesota, Department of Computer Science and Engineering, 2019. [Online]. Available: https://www -users.cse.umn.edu/~kumar/papers/high_dim_clustering_19.pdf.

[8] L. Zhang, Y. Wang, and J. Chen, "Enhancing recommendation diversity using cluster-based techniques in financia l applications," Sustainability, vol. 15, no. 4, p. 2947, 2023.

[9] M. Liu, Q. Zhao, and T. Yang, "Addressing the cold-start problem in recommender systems using clustering-based user profiling," arXiv preprint, 2021. [Online]. Available: https://arxiv.org/pdf/2109.12839.

[10] I. Met, A. Erkoç, S. Seker, M. Erturk, and B. Ulug, "Product Recommendation System With Machine Learning," Jo urnal of Intelligent Systems, vol. 33, no. 1, pp. 45–58, 2024. [Online]. Available: https://onlinelibrary.wiley.com/d oi/10.1155/2024/5585575

[11] P. K. S. Patil and S. S. Shrimant, "Clustering by DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clearly explained," Medium, 2021. [Online]. Available: https://medium.com/@satpatishrimanta/clustering- by-dbscan-density-based-spatial-clustering-of-applications-with-noise-clearly-f93c5c72f706

[12] A. Javari and M. Jalili, "A probabilistic model to resolve diversity-accuracy challenge of recommendation systems," arXiv preprint arXiv:1501.01996, Jan. 2015. [Online]. Available: https://arxiv.org/abs/1501.01996.

[13] I. Assent, "Clustering high dimensional data," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 2, 2012. doi: 10.1002/widm.1062.

[14] Z. Dafir, Y. Lamari, and S. Slaoui, "A survey on parallel clustering algorithms for Big Data," Artif. Intell. Rev., vol. 54, pp. 2411–2443, 2020, Available: https://doi.org/10.1007/s10462-020-09918-2.

[15] M. Mahnoor, I. Shafi, M. Chaudhry, E. C. Montero, E. S. Alvarado, L. A. Isabelde, T. D. Diez, M. D. A. Samad, and I. Ashraf, "A review of approaches for rapid data clustering: Challenges, opportunities, and future directions," vol. 12, pp. 138086-138120, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3461798.