Market Segmentation and Tag Optimization in Customer Behavior Data Analysis: Current Status and Future Research Directions

Zixiang Yan
York University, Faculty of Science, Toronto, Ontario, Canada

Abstract: With the rapid collection of customer behavior data, a key factor for businesses in their digital transformation was obtaining accurate marketing and personalized recommendations by effectively segmenting the market and optimizing labels. Traditional methods face difficulties in dealing with multidimensional, dynamic and heterogeneous customer data. This paper systematically discusses current market segmentation techniques and tag optimization strategies based on literature reviews, analyzes their optimization paths in terms of model accuracy, real-time performance, semantic significance and more. Combining typical examples of application in industry, the paper addresses the potential shortcomings and challenges of existing methods in practical application and provides recommendations for future research in areas such as intelligent systems, semantics-based approaches, and dynamic tag generation.

Keywords: Customer Behavior Data; Market Segmentation; Tag Optimization; Ontology Methods; Multi-Layer Tags; Precision Marketing

1. Introduction

1.1 Research Background

With the rapid development of information technologies and digital transformation, customer behavior data is gradually becoming a key element for businesses to understand market dynamics and optimize services provided. By collecting and analyzing data such as the number of views, clicks, and transactions, companies can develop more competitive marketing strategies by analyzing customer characteristics and anticipating their potential needs. Whether it is personalized recommendations, creating a portrait of a user or providing differentiated services, Market Segmentation and label optimization functions are necessary. However, with increasingly complex aspects of customer data and frequent changes in behavior patterns, traditional classification methods face many difficulties in dealing with multidimensional and variable data (Qu, Cheng, & Xu, 2021)[2]. These include inefficient data processing, a semantically weakened labeling system, and insufficient model adaptability, which makes it difficult for its comprehensive application in real-world operations.

In recent years, various optimization strategies have been proposed to solve many problems, such as improving tag semantics based on the ontological model, building a multi-layered tag system, and the mechanism of dynamic tag generation. The aim of these strategies is to increase the rationality of consumer classification and the expressiveness of labels. These studies have contributed to the intelligent development of consumer behavior data analysis, and at the same time feasible technical support for future optimization, as well as various relevant studies have been collected.

1.2 Research Significance

This study focuses on "Market Segmentation and label optimization in consumer behavior data analysis" and aims to comprehensively explore basic research methods and practical results, as well as semantically expressive explore their adaptability and possibilities for improvement in various industries. Based on the detailed study of theoretical models, algorithmic strategies, semantic expression and system architecture, an optimization path is proposed that is more intelligent, semantically manageable and flexible. In the context of digital transformation of small and medium enterprises, we strive to create a reusable data analysis system, develop intelligent customer management tools, promote the widespread use of behavioral data in financial, retail, medical and other areas, enhance customer engagement or efficiency, and improve service personalization.

2. Current Research Status

2.1 Overview of Market Segmentation Methods

As for the analysis of customer behavior data, common methods of market segmentation mainly include the following: the first thing to mention is the division by demographic characteristics, which is used quite widely. In addition, there is a classification based on consumer behavior, such as the frequency of purchases and the amount spent that can be used for detailed separation. It is also more common to divide according to psychological factors such as lifestyle, interests, and preferences. If these segmented data are combined and analyzed, the company's reaction to the market can become more accurate and focused.

Cluster analysis: the clustering algorithm automatically separates customer groups according to characteristics similarity into multiple parameters. The k-means algorithm has the advantages of simple and fast computation and is suitable for use on large scale data and moderate dimension variables. Hierarchical grouping is more applicable when it is necessary to present a visual structure or the number of clusters is not defined in advance (Chen, 2023)[1].

Latent Class Analysis (LCA): this can be used for the interpretation of variables dataset such as multiple and categorical variable datasets. It finds the hidden categorical variable in the observed data to understand the structure of customers. LCA can be applied in many aspects like in social science, medical, and difficult modeling for customer behavior. On the contrary, LCA faces the risk of parameter instability with too small of a sample size, resulting in biased class assignment. High dimensionality data will increase the multicollinearity among variables, which makes the model complex, so it is easy to lead to over fitting, and the latent class structure is not easy to be converged.

Rule analysis: decision tree algorithms such as ID3, C4.5 or cart are good at shaping interpreted classification rules and the impact of different characteristics on customer behavior can be clearly demonstrated. The results are easy to understand, making this method extremely applicable in practical operating scenarios that require a high degree of interpretability, such as financial risk management and customer group management.

Segmentation: in the process of continuous collection of internet data, there is an increasing emphasis on analysis techniques based on behavioral sequences. This type of data encompasses various forms, such as click flows and access paths, and is essential in scenarios such as e-commerce platforms and mobile applications. By in-depth analysis of the user experience sequence or access path, it can effectively identify potential purchasing intentions or preferences and promote market segmentation, becoming more flexible and adaptable to dynamic demand changes.

2.2 Overview of Tag Optimization Methods

As a basic tool for creating portraits of users and implementing marketing strategies, the customer tagging system has the function of encoding the behavior, interests and attributes of customers, and the data becomes more structured and efficient. However, traditional labeling systems tend to have static features or be manually adjusted and problems such as high redundancy and update delay occur, limiting their effectiveness in complex scenarios.

In recent years, many optimization techniques have emerged to improve the expressiveness and efficiency of the application of labeling systems.

Semantic optimization: ontology is a way of expressing conceptual formalization. When hierarchical relationships between tags, ontological attributes and a network of connections are established, tags cease to be only a set of keywords, but become a unit of knowledge with a clear semantics and completed structure.

Tag systems: tags are divided into major tags, sub-tags, and contextual tags, and a tag system with a hierarchical structure is created. This method contributes to improving labels in terms of versatility and adaptability to scenarios and can be widely used in Task scenarios at many levels, such as e-commerce recommendation system and financial Client Management (Zhu, Huang, Jia, et al., 2013)[3].

Dynamic tag generation mechanisms: real-time behavior data controls automatic tag updates that in turn dynamically display customer characteristics. For example, when users frequently review a specific category of products, the system can automatically label them as "potential customers", and as this behavior stops, such labels will be lowered or even removed. This method is widely used in retail, e-commerce, advertising recommendation systems and other areas.

3. Optimization Strategy Analysis

The key to optimizing the market segmentation method and labeling systems is to improve their adaptability and extensibility to a complex environment of customer behavior data. Existing models often face problems such as insufficient accuracy, slow response, and poor semantic representation. The scientists proposed three optimization methods: improving accuracy and stability, dynamic adaptability, and improving real-time performance, as well as deepening semantic association while reducing excess tags.

3.1 Accuracy and Stability Optimization

In the traditional process of Market Segmentation, the quality of initial clustering has a significant impact on subsequent results. As a result, improving the accuracy of the initial classification and improving the model's resistance to data distortion were major advances aimed at improving overall system performance.

The hybrid clustering method can effectively improve initial clustering results. Such methods often first display the General Data Structure by hierarchical clustering, and then use K-means to achieve exact separation. As a result, the quality of the clustering is significantly improved, as well as the reduced likelihood of hitting the local optimal solution.

In addition, the multi-model voting mechanism can reduce the instability caused by sample deviation or sensitivity to the parameters of a single model. This process involves integrating the results of several clustering or classification models and forming a final solution based on voting results or weighted averages. This not only increases the resistance to interference and efficiency of system generalization, but also makes it more applicable in high-noise data processing scenarios.

In addition, accuracy optimization should also focus on the comprehensibility and performance of the model's output, thereby increasing the model's interpretability and transparency of the decision-making process. This is especially important in the customer's business strategy, as interpretive results make it easier to make more informed and accurate decisions.

3.2 Enhancing Real-Time and Dynamic Adaptability

Customer behavior data tends to be very dynamic, especially in areas where user behavior often changes, such as e-commerce and finance. If the tagging system is in a static state for a long time, this is likely to reveal the problem of lagging or obsolescence of information. The importance of timely and dynamic adaptation and updating of models and labeling systems to improve adaptability is obvious.

The gradual learning mechanism and sliding window technology offer a way to solve the problem of model upgrades. Gradual learning allows the model to dynamically adjust the parameters in the process of obtaining new data, while a sliding window limits the model's focus to fresh data, thus excluding interference caused by historical information. The combination of these two methods allows the application of online simulation and real-time updates.

On the other hand, the integration of data flow and rule mechanism architecture can be used to generate tags. The system performs a real-time analysis of customer behavior and generates its own labels in accordance with pre-established rules. For example, if a user browses a particular type of product more than three times within 30 minutes, the system labels them as "high-intent customers". This mechanism can support dynamic labeling and continuous adjustment of the label weight, which significantly increases the timeliness and accuracy of the labelling system.

At the same time, we need to avoid incorrect label generation too. In the dynamic label generation process, errors in labeling can happen because of data noise or wrongly set thresholds Reliability needs to be enhanced through multiple mechanisms:

Confidence Assessment: Give each label a confidence score from 0-1. The label gets triggered just when the score reaches or goes beyond 0.8.

Behavioral Logic Verification: Add time windows to exclude sporadic actions. Labels require a full path through behavior. Take "Prospective Purchase Intent" as an example, it's a label requiring a series of actions: "Browse - Add to Cart - Save to Wishlist".

Human-Machine Collaborative Review: Carry out manual sample checks for higher value customer labels. Continually rectifying bias with machine labeling-human review-model adjustment closed-loop.

3.3 Semantic Association and Tag Redundancy Removal

As time passes, the product labeling system for customers will gradually face situations such as different labeling accuracy, ambiguous semantics, and redundant and overlapping tags. Such problems will seriously affect the expression and efficiency of the system. To solve these problems, an extremely important part of label optimization was the introduction of semantic improvement mechanisms and strategies for removing redundancy.

The construction of the ontological model and map of knowledge provides shortcuts with a hierarchical structure, division into synonyms and possibilities of semantic reasoning and analysis, and gradually forms a semantic structure with inherent associations. Tags are no longer keywords that exist in one form, but are converted into the contents of a node with interlinked links. Such processing allows intelligent systems to better combine, interpret and optimize them in different contexts. In addition, TF-IDF and tag similarity matrix can be used to effectively measure and analyze the semantic proximity between tags. The former is used to find high-frequency labels with an indistinct weight, while the latter is used to compare the similarity characteristics between labels. This method helps to automatically remove duplicates and redundant information on labels to achieve real benefits from optimizing system performance and distributing label maintenance according to the goals of intensive resource management and efficiency improvement.

The combination of semantic association and elimination of label redundancy will significantly improve the interpretability and mobility of the labeling system, providing an important framework for the future development of customer needs analysis and intelligent recommendations.

4. Typical Application Cases

4.1 Application of Customer Segmentation and Tag Optimization in Smart Banking

The trend of digitalization and intelligence of financial services is becoming increasingly apparent, and customer demand for personalized banking services is growing. Traditional methods of static segmentation and label management are difficult to adapt to the needs of modern operations, which often makes customer portraits vague and the marketing effect weak. Smart banks have started using market segmentation and label optimization techniques based on customer behavior data to improve service delivery speed and customer satisfaction.

A large commercial bank uses the K-mean and Latent Class Analysis (LCA) algorithm to build a customer cluster segmentation model. It displays customer behavior characteristics in several parameters, such as transaction frequency, product preferences, and channel activity, and divides customers into different groups, such as "high-value active customers" and "multi-channel preferences". This method is based on data and effectively detects the heterogeneity of the consumer, which allows for a higher practical value of the stratification of the consumer production.

To optimize the labeling system, the bank has developed a multi-layered labeling structure that includes tag types such as basic attributes, behavior, and scenes, as well as an ontological model to improve labeling efficiency at the semantic level. Combined with data processing and rule processing mechanism, the system can support dynamic label adjustment. For example, when it is found that customers often make "frequent interbank transfers", appropriate labels such as" mobile usage preferences "and" highly active transactions" will be linked to support personalized recommendation strategies.

Practice after the system went online showed that response time to label updates decreased by 40%, that marketing transitions increased by 12%, and customer satisfaction increased by more than 20%. This case allowed optimization of grouping and labeling based on behavioral data, which showed great value in intelligent banking (Jiang, Chen, & Liu, 2025)[4].

4.2 Practice of Customer Segmentation and Tag Optimization in the Power Industry

With the gradual popularization of smart grids and the digitalization of energy management, energy companies are beginning to recognize the value of data on the dynamics of consumer electricity consumption. Demand for differentiated services and demands for precise management are constantly growing. Many energy companies have implemented consumer behavior analysis systems to try to improve operational efficiency and consumer satisfaction by market segmentation and labeling optimization.

When energy companies segment their customers, they rely on analyzing customer payment records and electricity load growth trends to separate customer types. Users are divided into four categories: high value, conventional, potential, and small value. The process is completed using the K-mean clustering algorithm using two metrics: "average payment sum" and "electricity consumption growth rate". This method naturally brings together users with similar consumer behavior, which not only improves the accuracy of hierarchical operations, but also makes business strategies more feasible. Two types of key indicators allow for a clearer classification of users, so that the entire operational process presents more efficient and accurate characteristics of the model.

As part of label optimization, the system implements a mechanism to model behavior characteristics to transform data such as consumer's usual electricity consumption time, payment time, and load fluctuations into quantified user behavior labels, which are then dynamically displayed on the visualization platform. Combined with the value model for the user, the label can be assigned automatically, so that businesses can easily identify core customers and provide them with priority services.

The results of the experiments show that the clustering model allows to significantly differentiate the characteristics of customer groups. The accuracy of the prediction of a group of customers with high costs is significantly improved. In load prediction, a model built using short-and long-term memory network (LSTM) outperforms traditional regression methods in terms of average absolute error (Mae) and mean square error (RMSE), and its total prediction error is within 2%. The system has wide application perspectives and is of important practical importance (Wei, Sun, Lu, et al., 2024)[5].

4.3 Application of Customer Segmentation and Tag Optimization in the Retail Industry

When companies classify customers, they often use behavioral metrics such as purchase frequency, average transaction volume, and rate of return, and also use the RFM model and K-Means clustering algorithm to divide customers into "high value active type", "medium potential interactive type", and "low risk type". and other species (Jian, 2025)[6]. This allows companies to clearly identify customers who are the main source of profit and provides a framework for implementing clear service and marketing strategies.

When creating a Labeling System, retail companies pay special attention to replacing static labels with dynamic ones. By means of a dynamic marking mechanism based on behavioral events and consumer value levels, multi-layered marking combinations can be applied. With the integration of the "customer profiling" system, the processing of the structure and visualization of the labeling system can be completed, and the process of content recommendation and distribution of advertising material is carried out in accordance with the automation of branding. This series of operations significantly improved user experience and conversion rate[7].

Experimental data shows that after retail companies optimize the modeling of customer behavior and labeling systems, the repeat purchase rate increases by 17%, the unit price for valuable customers increases by 12%, and the overall marketing conversion rate increases by about 15%. This example shows that in a data-driven retail scenario, customer grouping and label management gradually allowed businesses to move from ""from 'customers searching for goods' to 'goods being found by customers'" and increased customer attraction and brand loyalty (Jian, 2025)[6].

5. Challenges and Future Research Directions

Although Market Segmentation and label optimization techniques have made it possible to achieve certain results in many practical situations, there are still many challenges and limitations. These problems are largely manifested due to the large amount of heterogeneous data, the dynamic adaptability of the labeling system, and the interpretability of the model. These difficulties are still areas of research that require urgent study.

5.1 Data Heterogeneity and the Difficulty of Behavioral Data Fusion Remain Major Obstacles

Customer behavior data sources are complex and diverse, including many heterogeneous types, such as records of transactions, social activities, and movement paths. This significant structural difference places high demands on unified modeling and integration of functions. Synchronously collecting data from multiple sources, standardizing formats, and ensuring semantic consistency continue to face major challenges. Although there is currently no generally recognized processing system, the issue of efficient integration of cross-modal data with minimal losses has become a key issue in improving hierarchical strategies.

5.2 Dynamic Adaptability and Generalization Ability of Tag Systems Need to Be Improved

Most existing labeling systems are still based on static settings or periodic updates, with low refresh rates and no real-time interaction, making it difficult to adapt to rapid changes in customer behavior in a dynamic business environment. At the same time, the possibilities for the transfer and generalization of labels in different scenarios are weak, which is a significant obstacle for the implementation of labeling systems in different industries and on different platforms. Introduction of a reusable labeling system, which is dynamically evolving and has interdisciplinary capabilities, it is crucial to improve the efficiency and extensibility of the labeling system.

5.3 Lack of Model Interpretability Limits Practical Deployment

Although some market segmentation models show good results in terms of accuracy, their interpretability and visualization remain at a relatively low level. As a result, transparency is highly sought after when making business decisions, and this lack makes it difficult to align the model with real needs. Especially in highly regulated industries such as finance and healthcare, interpretability and verifiability are key factors, which also limits the application of these models.

5.4 Integration of Deep Learning and Behavioral Understanding Engines

To more accurately identify potential customer needs and behavior patterns, future research can be supplemented with deep learning systems that will help with label creation and classification. With the help of the graph neural network (GNN), Bert, and other structures to build models of understanding behavior, the system can extract the characteristics of customer preferences and apply understanding and adaptive updating of semantic tags. In the process of work, we continue to deepen the study and analysis of the characteristics of customer preferences[8].

5.5 Establishment of an Industry-Shared Tag Ontology System

In situations where multiple systems coexist in a real environment, the inconsistency of tag semantics often interferes with mutual analysis and data exchange. In the future, the creation of an inter industrial library of Ontology tags for sharing could be considered. Standardize naming conventions and create a unique semantic tag network to enable interoperability between tags, which can increase the extensibility and stability of the system that manages data in different application scenarios.

6. Conclusion

This article systematizes the basic methods of segmentation and labeling optimization mechanisms in the field of consumer behavior data analysis, as well as examples of their application in financial, power, retail and other industries. The article then discusses existing barriers to data integration, tag system adaptability and model interpretability, as well as considering future research directions, such as applying deep learning to facilitate tag creation, creating inter-domain tag ontologies, and improving the efficiency of real-time updates.

Overall, market segmentation and label optimization relate to the main challenges in customer data management and analysis, and there is still a lot of room for development in theoretical research and practical mastering. Future research may further focus on integrating intelligent algorithms, improving the ability to understand model semantics, and expanding the applicability of systems engineering to advance customer behavior analysis technology in a more reasonable, efficient, and interpretive direction, as well as providing technical support to companies to achieve the set goals. customer-oriented high quality development goals.

References

[1] Chen, T. (2023). Market segmentation

strategies for cross-border e-commerce based on data analysis. Chengdu University of Arts.
[2] Ou, J., Cheng, B., & Xu, Y. (2021). Segmentation of intercity railway passengers based on ticket-buying behavior data. Journal of Wuyi University (Natural Science Edition), 35(4), 52–59.
[3] Zhu, H., Huang, Z., Jia, D., Chen, G., & Chen, D. (2013). Research on tag ontology for mining multidimensional preferences of telecom customers. C-Enterprise Management, (1), 84–86.
[4] Jiang, Q., Chen, H., & Liu, W. (2025). Research on data analysis and customer behavior prediction of AI in smart banking operations. Focus on Artificial Intelligence, 3, 11–13.
[5] Wei, Q., Sun, J., Lu, X., Lu, Y., Li, N., & Li, Y. (2024). Electricity customer behavior analysis system based on electricity big data. Guangxi Electric Power, 47(4), 37–43.
[6] Jian, K. (2025). Enterprise marketing strategy research from the perspective of customer value. Modern Marketing (Upper Monthly), 5, 157–159.
[7] Zhang Lihong. Analysis of the Impact of Large Model-Based Customer Behavior Interaction and Countermeasures [J]. Changjiang Information & Communications, 2025,38(02):227-230.
[8] Zheng Chenrui. Innovation Analysis of Enterprise Marketing Strategy in the Era of Big Data [J]. Modern Business, 2025(02):40-43.