Optimization Strategy of Customer Relationship Management based on Big Data Analysis

Baohua Zhang[1*], Sanbao Zhang[2], Chaojie Zhang[3]
[1]School of Business Administration Tongling University, Tongling Anhui, 244061, China
[2]School of Finance Tongling University, Tongling Anhui, 244061, China
[3]School of Mathematics and Computer Science Nanchang University, Nanchang Jiangxi, 330031, China
E-mail: 164108@tlu.edu.cn
[*]Corresponding author

This paper analyzes the basic customer management process and chooses the Apriori algorithm to build a CRM model based on data mining. In addition, this paper designs a customer relationship management system based on big data. The system is divided into three layers: data source, batch processing, and real-time processing. In the part of constructing the system architecture, this paper adopts the Hadoop platform. The batch processing layer, it has consisted of four parts, which include No SQL database, Oracle database, ETL architecture, and Hadoop platform. This paper gives the logical architecture design diagram. The real-time processing layer mainly includes a real-time decision engine and service bus. The key part of this layer is the real-time decision engine, in the design of which the Bayesian algorithm and product recommendation prediction model are used. Finally, this paper takes K company as an example to demonstrate the model and management system. After applying the analytical model and management system, the sales of K company keep increasing.

Povzetek: Analiziran je proces upravljanja odnosov s strankami in predlagan CRM model na osnovi algoritma Apriori in analize velikih podatkov. Sistem, zasnovan na Hadoop platformi, izboljša prodajo podjetja.

1   Introduction

Today's world is gradually stepping into a new economic era. Social productive forces are continuously developing, and our society is changing from the era of low production efficiency into the current surplus of various products with many more types and styles in our surroundings [1]. The variety and channels of products that customers can choose from are also significantly increased. The market is constantly cyclical, and the market may boom or bust. In the cycle of market economy development, the only thing that can be determined is that market competition will become increasingly fierce. Under such fierce competition, customer relationships have become increasingly obvious [2]. It is increasingly important for the company to consider the customer relationship as an extremely important factor for sustainable development in the future.

As is known to all, customers are becoming the main force of a company's economy in this society. To develop customers to their maximum value and increase sales, a company must pay attention to managing customer relations and continuously develop new products to satisfy customers. Thus, the company can reach the goal of maintaining its competitive advantage in the market [3]. The way to build a good relationship with customers is mainly three parts. Developing new customers, improving customer satisfaction, and conducting effective customer segmentation are three important things for companies to optimize customer relationship management. In the business process oriented by customer demand, companies must take customer demand as the starting point in all aspects of the product. The company should pay attention to the whole production process, from the first step, product design, to the last step, after-sales service. Only by continuously meeting the needs of customers can the company's own operational capabilities and business capabilities be improved. Focusing on the customer as the centre, innovate the company's marketing, sales, service and other aspects so that the company can more efficiently provide customers with satisfactory and thoughtful services.

Because the existing database does not have the perfect data mining function, it is not possible to find the rules implied in the massive data and the correlation between the data, it is difficult to evaluate and predict the future development situation based on the existing data, and there is no technology to mine the rules implied by a large number of data, so that a large number of data becomes "data garbage". The emergence of customer relationship management just solves the problem that a large number of customer data in the enterprise database cannot be utilized. CRM is the technology and method that enterprises use computer information technology to realize the investigation and understanding of customer information, analysis and service, and finally achieve the

retention of original customers and the discovery of new customers. Nowadays, the development of enterprises is customer-centric. Shopping malls are like battlefields. In a highly competitive environment, it is impossible to retain customers by improving product quality. Only by improving service can enterprises maximize their interests and win the final victory. How to improve service in such fierce business competition? How to easily achieve customer management and maintain good contact with existing customers, while attracting more new customers? How to reduce the total cost of the enterprise and maximize the profit? The emergence of customer relationship management system makes these problems easily solved. Association rules are one of the necessary algorithms to implement customer relationship management system. The application of association rules in CRM mainly includes classification and prediction, which can be summarized in many aspects such as customer group classification, customer profitability analysis, customer acquisition and retention, and customer satisfaction analysis.

Big data processing technologies currently provide more methods for each company's data needs. In this era, every company generates a large amount of data every year, which ends up building a huge database. Therefore, every company can use information data analysis to promote its business development. More and more companies realize the value of the big database and use it as an effective way to build good customer relationships, which also require information technology support. It can be said that the development of big data applications provides new ideas and new technical support for companies to implement customer relationship reform. This paper designs a new CRM model and system to help companies get new tools to optimize customer management. The application of current science and technology can quickly and effectively analyze customer demand preferences, which provides a new economic growth point for companies and promote the development of related technology as well. From the perspective of practical application, customer relationship management research is limited to theoretical research, not really applicable to the actual development needs of companies. This paper applies the theoretical model and design system to K company's customer relationship management example and analyzes the current research theories in detail.

2   Related work

In the middle and late 20th century, research on customer relation management gradually emerged, and the United States first began to explore the field of scientific management decision-making. In the 1980s, the United States first put forward the very important topic of "contact management" in the marketing field. It put forward the marketing conception of collecting customer information- relationship marketing. In the 1990s, based on relationship marketing, scholars put forward the concept of "Customer Care", which can be used by all enterprises and industries [4]. At that time, some scholars

proposed that the core of customer relationship management should be automation, which can be used in the entire business process. Processes can be improved in an automated way to form a systematic system. Some scholars in academia believe that using data warehouse technology in companies, comprehensively integrating and analyzing massive user data, subdividing customers through different classifications, and dividing customer preferences can help corporate decision-makers improve business strategy better [5]. This is the starting point of the combination of customer relationship theory and user data, but so far, it only exists in the theoretical stage.

From the mid-1990s to 2014, customer relationship management was gradually transformed into practical applications. After 2000, the system software for customer relationship management was developed, and customer relationships began to receive attention in the field of commercial applications. The main feature of this stage is technology [6]. At this time, the research focus planned to apply database technology to customer relationship management research. However, the technology is still in a relatively immature stage at this stage, and the methods of data acquisition and calculation are both immature.

The term big data first appeared in a research report submitted by the famous McKinsey Company in 2011. In this famous research report, researchers from McKinsey Company have very strategically analyzed that big data technology will change the economic development of the whole world. Nowadays, with the rapid development of computer technology, network technology and database technology, big data technology is applied to the whole process of customer relationship management. At this stage, many foreign scholars have carried out research on the integration of customer relationship management and big data technology [7]. Some scholars believe that accurate customer search and precision marketing through big data in e-commerce is a major trend in the future. The customer satisfaction obtained by data mining is integrated, and effective improvement measures are finally proposed, which can also enhance the competitiveness of enterprises [8].

Nowadays, big data has been widely used in the information management process all over the world. Related research from the perspective of information systems development. In 2001, someone designed a CRM framework with seven parts and pointed out that this framework should include seven parts, which are customer behaviour database, acquisition of the data mentioned above sources, application of the database, customer preference and choice, customer acquisition, relationship marketing, CRM evaluation criteria, etc. Some companies have already applied a complete set of customer relationship management systems in their workflow [9]. Its basic function is to collect customer information in real-time, process related customer management services, better maintain customer relationships and increase customer activity. Some companies have added support vector machine technology to the customer relationship management system they developed, which has broadened the application scope of the customer relationship management system, and finally

can better mine customer data and transmit big data to the server to realize customer input, automation of business processing, customer discovery, and customer return visits. Some scholars believe that big data is an important tool for enterprises to create value in the future. Suppose future enterprises do not keep up with the pace of big data and cannot effectively organize and discover information that can create value from the information left by customers. In that case, this company that relies only on traditional experience will be unable to maintain old customers and acquire new customers efficiently, so that they will be eliminated. Big data technology is no longer the future for us. Now using data can help companies better analyze the market, better increase market share, improve profit margins and form new economic growth points, which is an important direction that every company should pay attention to [10]. The methods, results and shortcomings of existing CRM models and systems are shown in Table 1.

However, the prediction of customer value in the above study is based on the assumption that customer behavior will not change, and does not dynamically measure the entire customer lifecycle. It is only based on the historical static indicator division, does not grasp the customer buying trend, does not consider the possibility of cross-selling opportunities customer growth and upgrading, will miss the potential and worth tracking quality customer groups.

3 Customer relationship management model based on big data

3.1 Customer relationship management process

According to the customer life cycle theory, we have sorted out the entire customer contact process with the company. We have set up some external channels because there are too many online and offline channels. Customer management ability becomes particularly important in this complex customer acquisition and maintenance system [11].

Table 1: Methods, results and shortcomings of existing CRM models and systems

| Existing CRM models and systems	Contact management	Customer RFM classification model	RBF neural network model
The methods adopted	Customer care with call centres to support data analysis	Clustering and SOM neural network technology	Establish customer evaluation index system
Achieved result	Customer relationship management sprouts	Customer value was evaluated	Categorize by category
Shortcoming	Only limited to the collection of customer information	There is no dynamic measurement of the entire customer lifecycle	There is no consideration of cross-selling opportunities for customer growth and upgrading
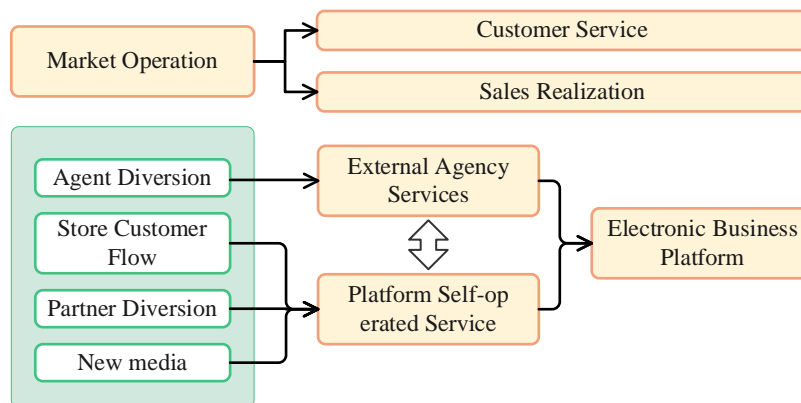
Figure 1: The flow chart of customer acquisition, service and sales realization

According to the flow chart (Figure 1) of customer-enterprise contact, we can divide customer relationship management into three aspects. The first is the customer information management system. The basis of customer information management is to collect customer information, including the customer's basic information, consumption information, consumption habits, credit value, etc.

The second part of customer relationship management is operation information management. This part integrates multiple modules, such as policy information and competitor information. Aspects of the situation to plan a scientific and reasonable management policy for the development of the enterprise. It can be seen that factors such as the number of customers, information, business and market competition can greatly affect the final results

of customer analysis and the direction of business operations.

The third part is sales information management. Sales information management mainly manages customer and sales-related information, mainly product information, sales activity information, sales channels, and after-sales management. Through sales analysis, companies can better understand customer preferences and market dynamics and formulate better directions and strategies for maintaining customer relationships and increasing market share. The three parts of sales information management are shown in Figure 2.

3.2 CRM model based on data mining

Before starting to dig and make plans, it is necessary to clarify the project objectives, be familiar with the business fields of relevant departments, have relevant knowledge background, understand the business content, determine the business object, and make feasibility analysis and evaluation of the project from the aspects of resource allocation, technology and economy. In the process of data mining, the preliminary data preparation work and the

model evaluation of mining results are very important. The analysis objectives set at the beginning of this mining task have a great guiding role in the evaluation of the mining results, and then the novelty and validity of the discovered knowledge model are evaluated by relevant experts in the field. After evaluation by experts and machines, it is necessary to remove redundant or meaningless patterns from these patterns. Sometimes some patterns cannot meet the actual requirements or reach the ideal effect. In order to obtain an effective knowledge pattern, it is necessary to return and repeat the previous processing steps to extract the knowledge pattern repeatedly until a meaningful knowledge pattern is found, so as to discover more effective and accurate knowledge.

Generally speaking, there are two types of data mining process models commonly used in academia, the process model summarized by Fayyad and the process model that follows the CRISP-DM standard. This paper adopts the Fayyad process model. Its process model mainly includes the following seven steps: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge representation [12]. The procedure is shown in Figure 3.
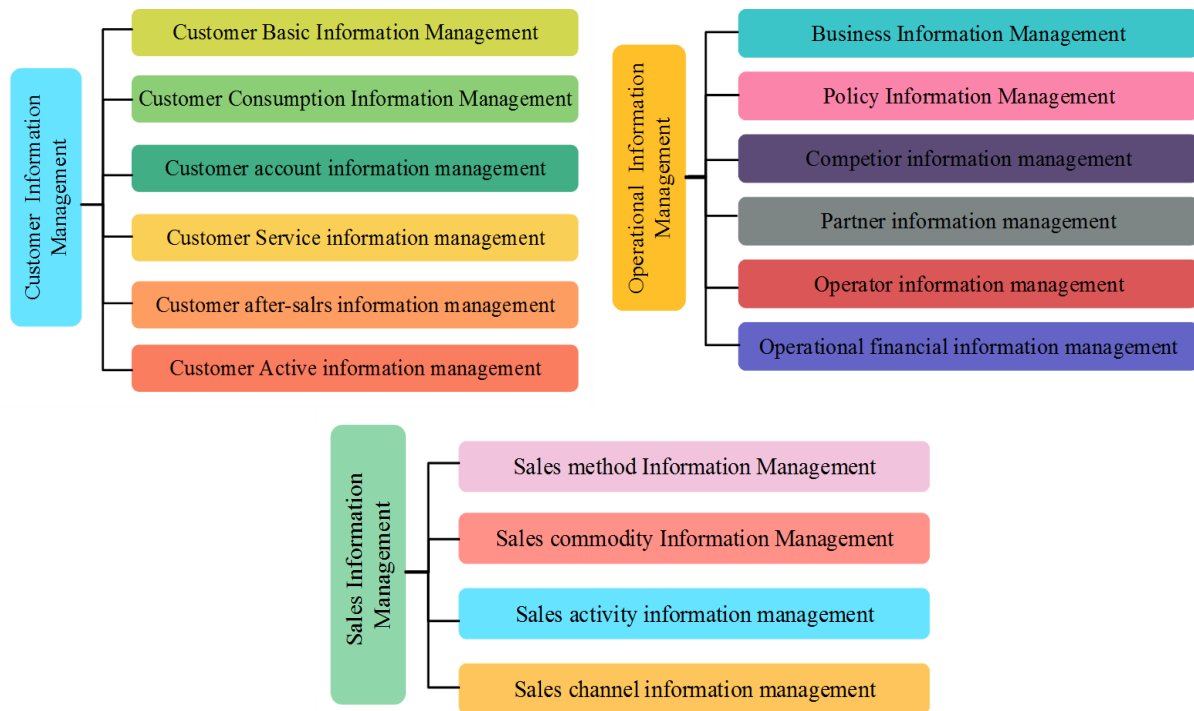

Figure 2: The three parts of sales information management
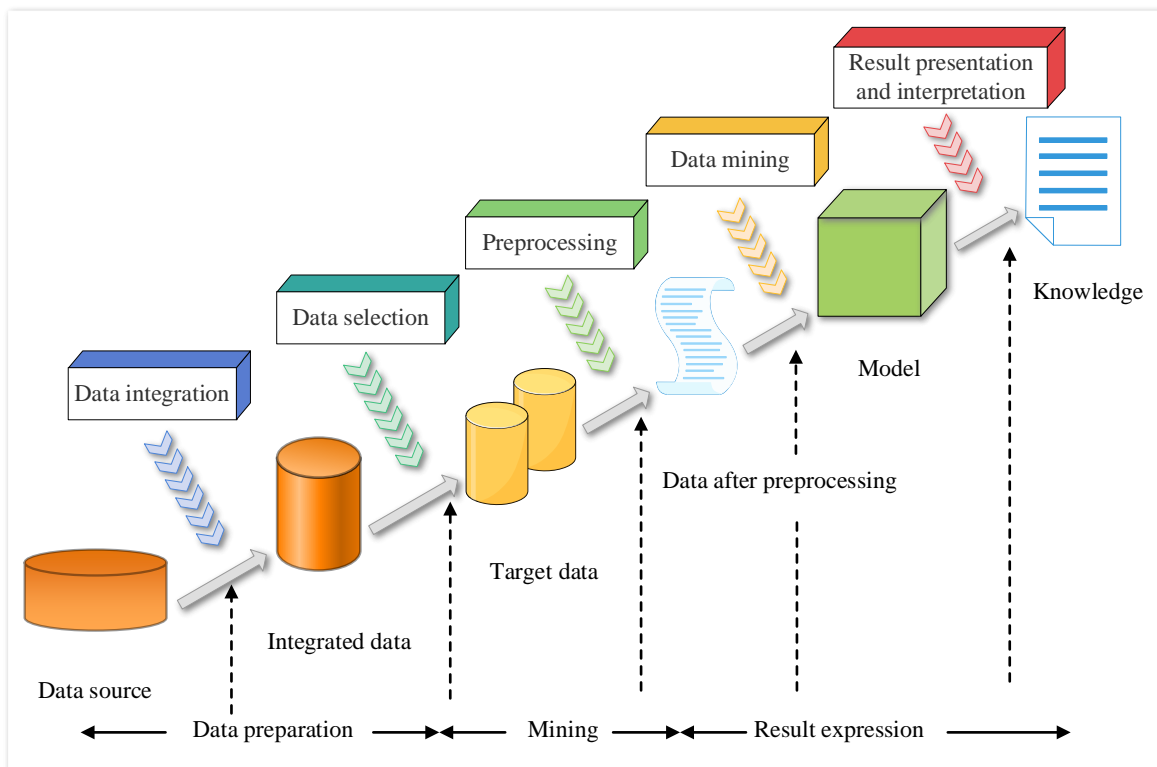
Figure 3: Fayyad process model

First, the noise data is filtered. After the filtering work is completed, use the Clementine platform to filter and analyze the data to ensure the effectiveness of data processing on each platform. After obtaining good data processing nodes, integrate the data content to realize the processing of logical business information.

The Apriori algorithm's basic idea is to first find all frequent itemsets in the original data set. The dominant thinking method of the algorithm is a recursive method based on the theory of frequency sets. It must meet the minimum requirements. Eliminate the association rules that do not meet the minimum confidence threshold, and the remaining association rules are strong association rules that satisfy both requirements. This algorithm needs to scan the transaction database many times, which takes a long time and needs to write a calculation program to complete the calculation. The algorithm runs for a long time when the program is executed, so the number of iterations of the algorithm cannot be too large, so further optimization of the program algorithm and related parameters may obtain more ideal results [13].

The algorithm uses a prior knowledge of the property of frequent itemsets: all non-empty subsets of frequent itemsets must also be frequent, and all supersets of infrequent itemsets are also infrequent itemsets. In this way, Apriori algorithm uses this prior knowledge, adopts the iterative method of layer-by-layer search, explores $(k+1)$ item sets with frequent k item sets to identify all frequent item sets higher than the set support in the target data set, and then computs conditional probability to construct strong association rules that meet the set confidence in frequent item sets.

This algorithm used the following two properties to reduce the search space.

Property 1: Any non-empty subset of a frequent itemset is frequent.

Property 2: Any superset of an infrequent itemset is infrequent.

On the basis of the above two properties, the Apriori algorithm generates all itemsets through the following process:

M1=. (frequentl-itemsets};

for (g=2J Ml- 1≠null ; l++) do begin Dl=apriori-gen (Ml--1) for every transaction u∈E do begin

du=subset(dl, t) ;

for every candidate D∈Dl do

C. count++ ;

Although the Apriori algorithm is simple and accurate, it has certain defects in efficiency. Therefore, a derivative algorithm can be used to make up for the shortcomings of Apriori. In this paper, we choose the Apriori_RD as the algorithm to make up for the shortcomings of Apriori. The Apriori_RD algorithm operates on the database bits based on the logical "&" and mines and analyzes frequent itemsets and strong association rules.

According to the operation process of the picking algorithm based on the Apriori property, the basis for the improvement of the Apriori_RD algorithm mainly includes the following three aspects:

(1) L subsets of an l-item are frequent as well

(2) If the value of l- is less than l+1, then l+1 does not exist.

(3) The value of every item set repetitions in the l-candidate item set Cl generated by Ll-1 self-connection is l*(l-1)/3.

During the execution process, set the minimum confidence data as 11% and set the maximum antecedent data as the relevant threshold of 2. The customer data clustering model is shown in Figure 4.

The calculating process is divided into three steps. The flowchart is shown in Figure 5.
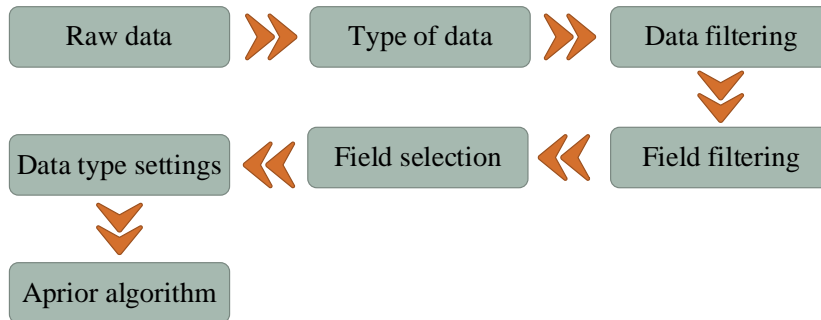

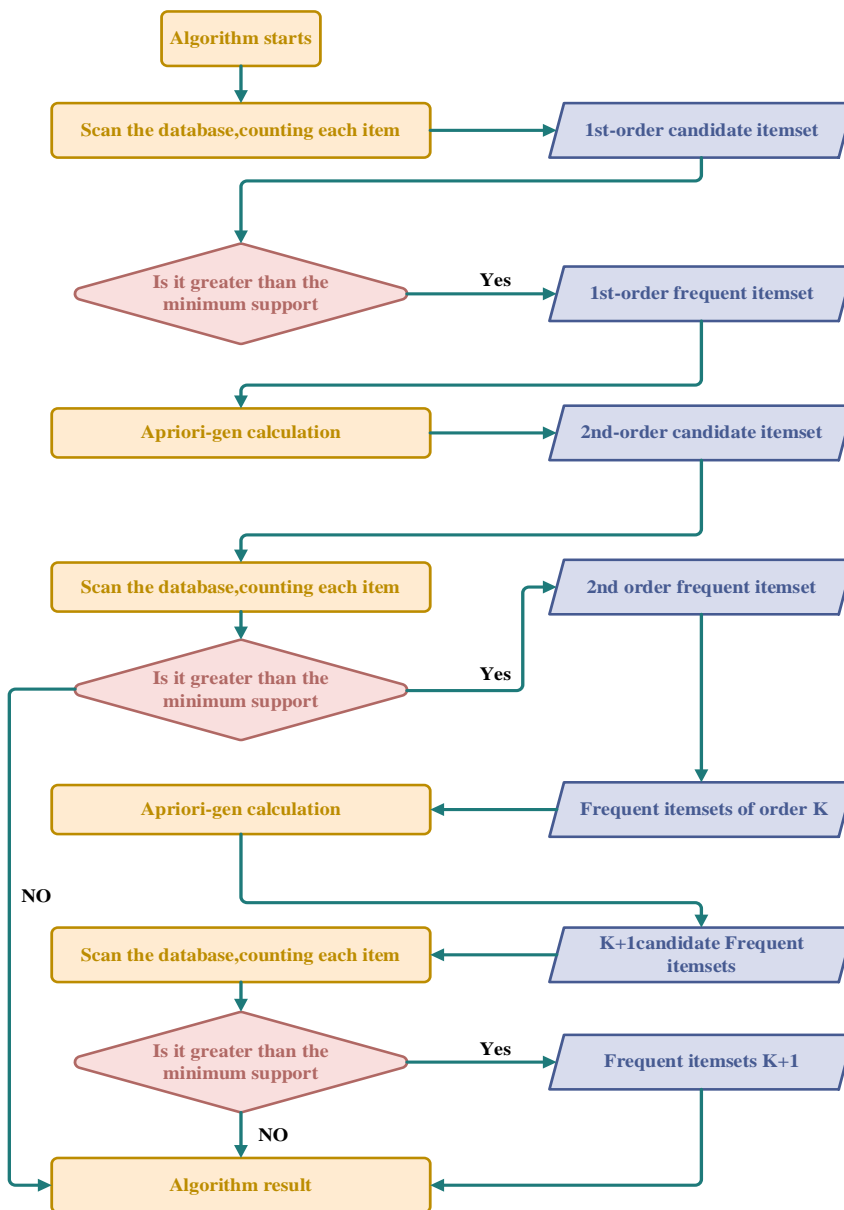Figure 4: Customer data clustering model based on apriori algorithm

Figure 5: Apriori algorithm flowchart

After mining all the data that can make use of the database, the next step would be to gain the corresponding association rules. The following formula can be used to calculate the obtained confidence.

$$Confidence(B => C) = Q(C \mid B)$$
$$= support\_um(B \cup C)/support\_um(B) \tag{1}$$

4 Customer relationship management system design

4.1 CRM system architecture

According to the results of customer segmentation, enterprise resources should be allocated reasonably, and the operation means and business processes of enterprises should be carried out around the customer, so as to bring more benefits to enterprises and customers' loyalty and satisfaction to enterprises. The integration of enterprise information system with existing computer technology provides the conditions for the successful implementation of CRM. CRM realizes the informationization of business processes, that is, the integration of sales, marketing and customer service. The integration of customer communication modules and the automation of communication means are realized. The generated information is integrated and processed to achieve customer intelligence and give decision-making guidance

for the business marketing skills and development strategies of enterprises.

The Hadoop platform used in this project adopts a distributed computing cluster of big data all-in-one machines based on an X86 server, which is composed of multiple nodes [14]. Any query or processing request for platform data is processed by multiple nodes. When system capacity or processing capacity becomes the bottleneck, nodes can be added by "building blocks", and each data block exists on multiple data server nodes, which ensures data reliability. This greatly reduces the coupling between the business layer and the data layer, improves the scalability and maintainability of the system, and improves the development efficiency of the system [15]. The system architecture has good scalability, which ensures the dynamic expansion of cloud platform services and the rapid launch of new services. The immutable nature of distributed data storage also enhances the credibility of the data, thereby increasing the platform's credibility. In the specific design, the big data is first cut into pieces. Then according to the concept of distributed decentralized storage, these pieces of data are stored in different clusters or computers of the Hadoop system respectively. When reading system data, this distributed storage structure can freely realize parallel extraction of related diced files from different machines or clusters. The logical structure of the project is shown in Figure 6.
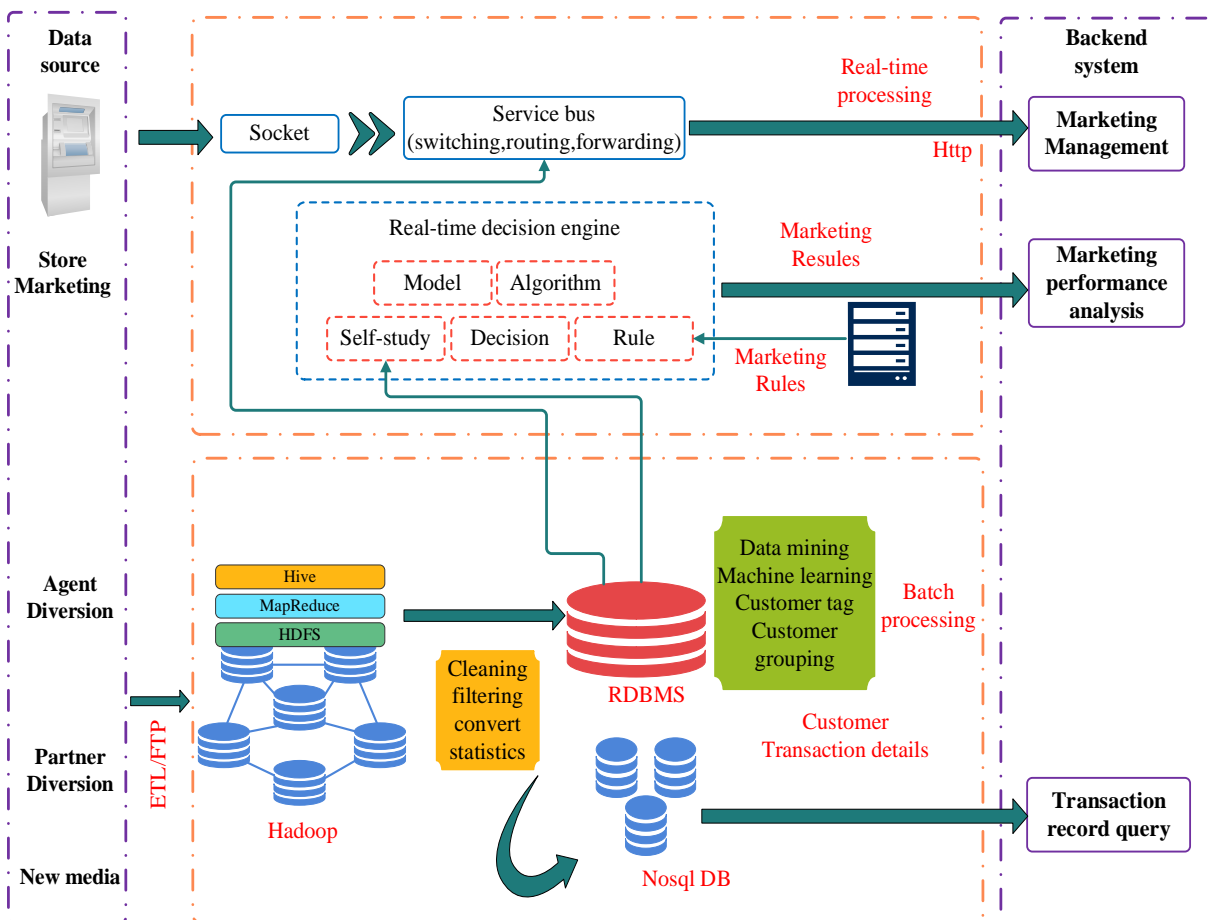

Figure 6: Logical architecture design

The data source layer is the provider of Hadoop platform data. There are multiple data sources in the system, and data is collected and stored in different ways. This requires complete or differential extraction from different data platforms, including various data source tables, databases, and data files of various forms. The data are distributed in different regions, run in different system environments, and the data formats produced are also different. The company's data sources mainly include the following components: customer information of stores, customer data brought by all media, external partners, etc., and data generated in the process of operating services. A lot of business data comes from suppliers, which can also be used to manage suppliers. The full data file can be exported from the private data analysis platform and loaded into the Hadoop platform to obtain the stored historical data. For daily data acquisition requirements, the business system is the main source of data. Through the combination of full complement and incremental, the source system provides corresponding data according to relevant interface specifications and downloads data through the data download platform. The data is transmitted in the form of text. Enterprises generate a large amount of enterprise data in the business development process, providing a reliable analysis basis for enterprise management and decision-making. Companies can use this data to conduct precision marketing or provide data for other companies or units to gain profits.

The batch processing layer is composed of the Hadoop platform, No SQL database (SequoiaDB), and Oracle database combined with ETL architecture. The batch processing layer architecture diagram is shown in Figure 7.

The linear expansion and low cost of the Hadoop platform solve the shortcomings of the traditional architecture (IOE) for building a data platform [16]. The introduction of the Hadoop platform has greatly improved the elastic space for data storage and processing performance expansion. The distributed file system (Hadoop Distributed File System, HDFS) in Hadoop can store a large amount of structured and unstructured data by means of distributed storage. The distributed computing model MapReduce simplifies the complex data

processing and computing process and is very suitable for analyzing and processing huge, complex and disordered data. The data in the distributed storage server is not whole but a piece of data, which can better protect the security of the data. In addition, it can help businesses and enterprises create large data centres without increasing costs by utilizing efficient processing power. The cooperation of these tools enables the Hadoop platform to exchange data with existing database systems and become a platform with comprehensive functions in the field of massive data computing, such as file systems, computing models, database systems, and management modules.

The Hadoop platform processes and stores data from business systems through Hive and Hbase components. The result set processed by the Hadoop platform is imported into the Oracle database for use by the real-time decision engine, and other result sets, such as the after-sales information thin and wide table, are imported into the NoSQL database.

NoSQL is a non-relational database that can store different types of structured and unstructured data [17]. It is very suitable for accessing massive data and massive customer information. The database is a dynamic database, and the underlying architecture is a distributed, highly integrated, and high-response-speed architecture, which is very suitable for relational and non-relational data with repeated records. The NoSQL database adopts a more flexible data model than the relational one to achieve real-time data analysis, good usability and scalability in a big data environment. Each data mirror of a NoSQL database is stored in a different location, ensuring data availability and non-loss, with high availability and flexibility.

By carrying out ETL architecture design, the main purpose is to provide a long-term foundation and architecture so that it can meet actual needs as business needs grow and change. In the ETL development process, the SAP data service tool is used for data extraction, conversion and integration process design. The data in the business system is extracted into the ODS and data warehouse, and the automatic data loading processing is realized in a graphical way in the design process, shown in Figure 8.
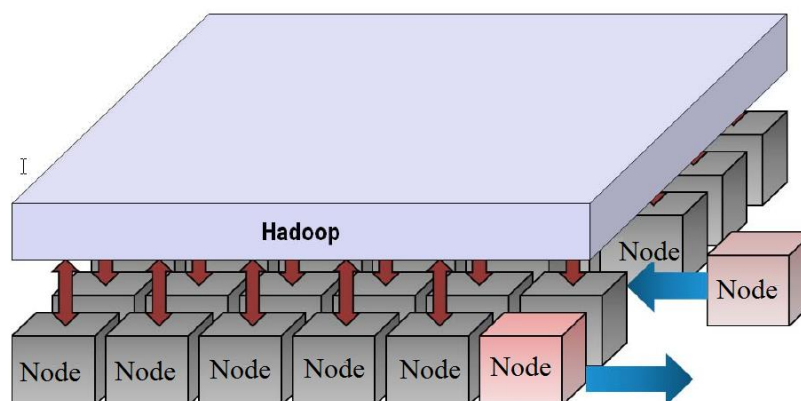

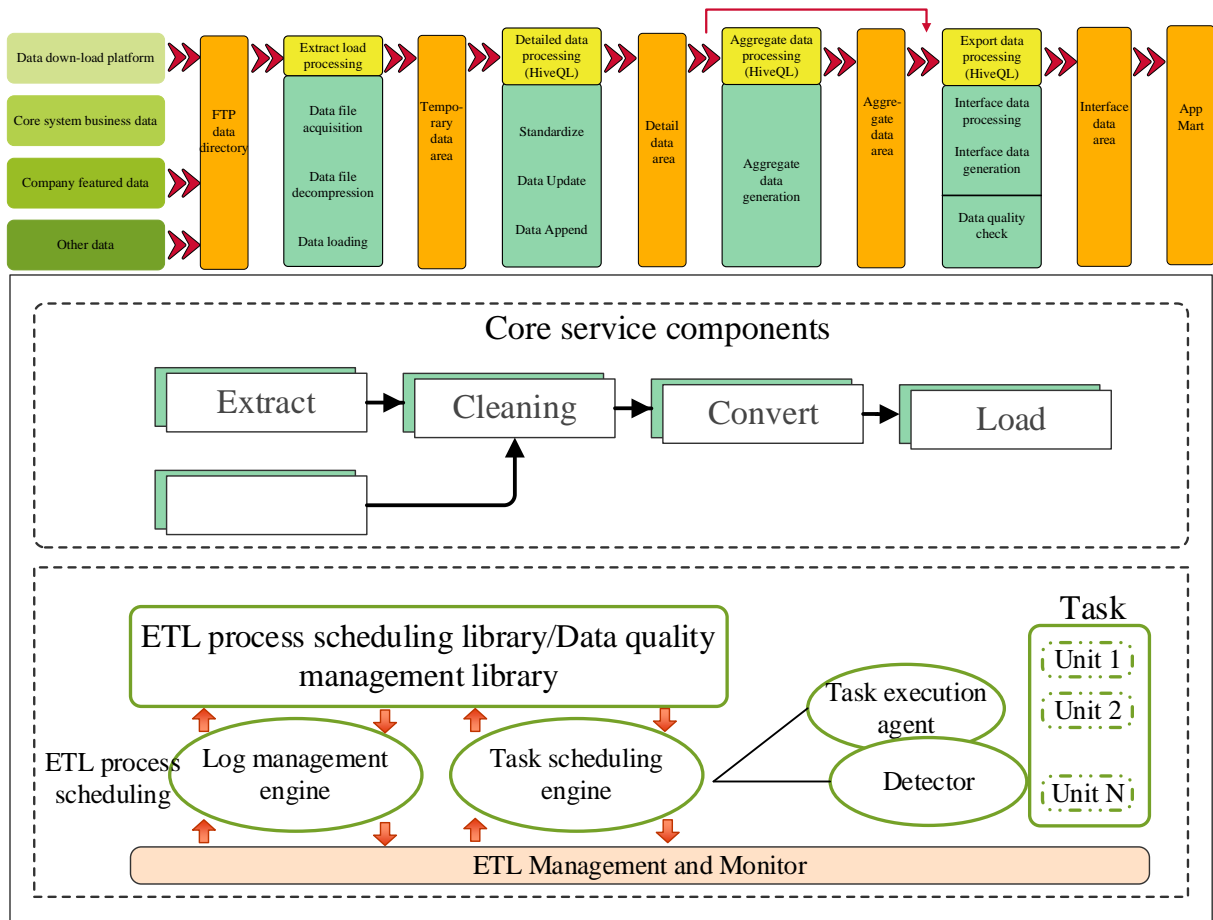Figure 7: Batch processing layer architecture

Figure 8: ETL technology architecture

The real-time processing layer mainly includes two parts, the real-time decision engine and the service bus. Real-time data processing is the default business logic of the data processing system and is also the core part of the data processing system. The key part of the project is the real-time decision engine, including modules such as models, algorithms, rules, decision-making, and self-learning [18]. Business logic is centrally located, interacting with all data sources to meet actual business needs. The real-time decision engine adopts the Bayesian algorithm and product recommendation prediction model. The driving factors of the Bayesian prediction model are derived from the wide table of customer information processed by the Hadoop platform.

The real-time decision-making engine receives customer information from various channels such as stores, external partners, and new media drainage, performs analysis and calculates recommended products based on customer information, product purchases, and historical marketing campaign results, and connects marketing channels through the service bus. Customers push product recommendation information, conduct comprehensive analysis and prediction on customer needs, and fully tap customer needs. On this basis, collect, sort out and analyze customers' individual needs and carry out demand forecasting. At the same time, it combines data visualization tools and decision optimization algorithm tools to provide users with decision support.

The real-time decision engine receives the user's purchase product information downloaded from the business system. Further, it optimizes the Bayesian prediction model through its self-learning function so that its recommendation and success rates are getting higher and higher. The model mainly predicts the response degree of target customers to a certain product or service based on basic customer information and historical transaction data to carry out targeted marketing, improve the marketing response rate and reduce marketing costs. At the same time, collect a large amount of data for product feedback, update and adjust products, provide products that are more in line with customer needs, realize the visualization of products and services, and meet the service needs of different customers. The design of the real-time decision engine is shown in Figure 9.
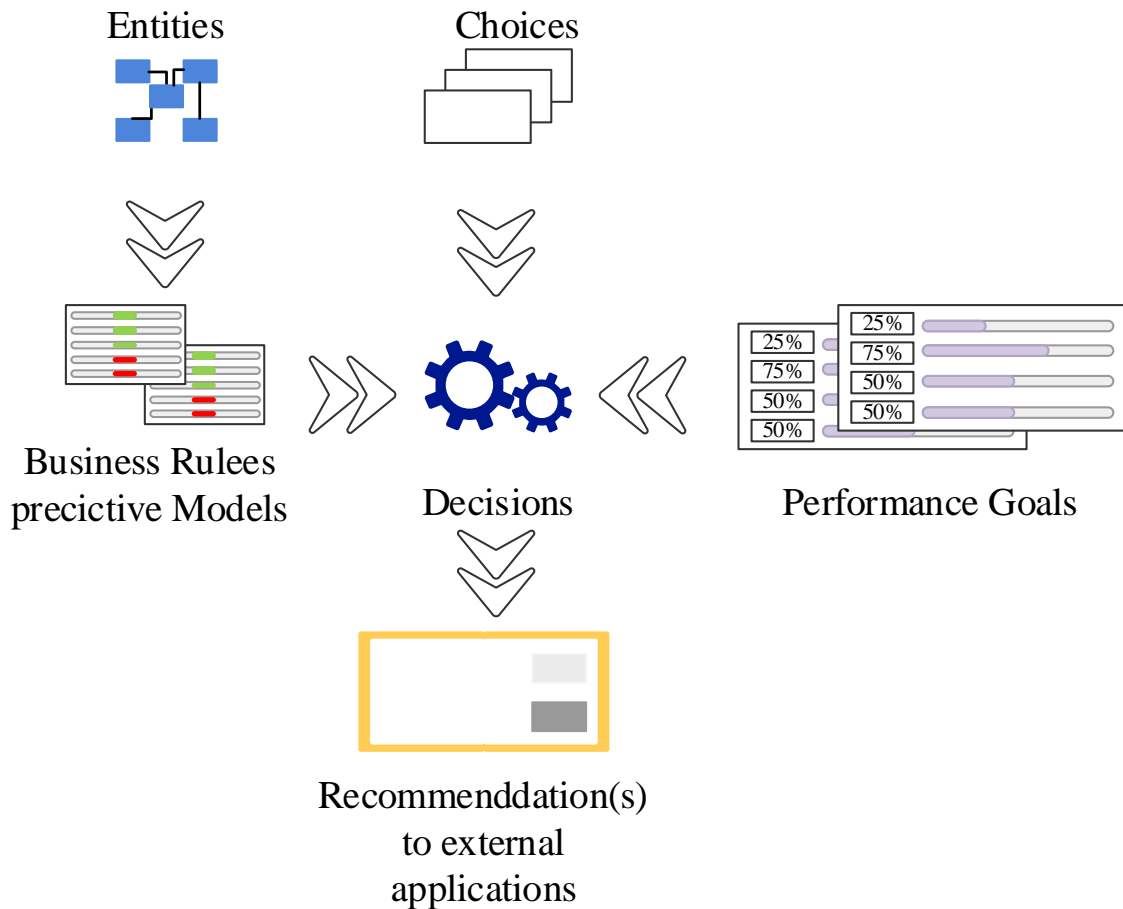
Figure 9: Schematic diagram of the real-time decision engine

4.2 Customer relationship management system application

In this part, we take Company K as an example. Company K is mainly engaged in the communication business. In the modern era of rapid development of communication technology, value-added services such as mobile data communication have brought high development opportunities to Company K. With the introduction of new network technologies, the introduction of new business requirements, and the construction of new operation management systems, these will become the inevitable trend of enterprise development. Under such a background, precision marketing emerges as a new and efficient marketing method. It continuously integrates advanced marketing concepts and, through a high degree of integration of information, helps enterprises to accurately locate customers and achieve precision marketing, which can not only reduce the cost of enterprises' marketing costs and improve the economic benefits of enterprises. Nowadays, the competition for value-added services of various operators is very fierce. In order to better enhance the operator's customer value and provide new growth points for operating income, Company K must improve its logical business and design

new business strategies and development paths, forming a new growth point.

In the process of applying the model to the instance of Company K, we found that the Apriori algorithm is very inefficient when searching for a large number of frequent itemsets during the running process. For example, for a data item with n data items, the data items owned by the frequent item set are 2o-1, so the huge amount of data causes the algorithm to run a very large amount of computation. After many experiments, it can be proved that the newly generated candidate item sets are infrequent. At the same time, the number of candidate item sets can be reduced as much as possible and verified. The number of itemsets further improves the number of candidates itemsets subsets. In addition, the Apriori algorithm adopts the depth-first search strategy during the running process. The depth-first search strategy can deeply search the candidate item set in the database. The algorithm scans the database repeatedly during the search process, which can consume a lot of algorithms running space and practice.

Regarding data preparation, Company K has data information on a number of value-added services. After screening, the value-added services left in this paper are shown in Table 2.

The Apriori algorithm analyzes the data to obtain the association rules between value-added services. The specific association rules are shown in Table 3.

From the analysis in the table, we can conclude that the correlation between mobile game data information and mobile video data information is the strongest, followed by mobile phone report data information and mobile phone reading data information. Because the data processed by the Aprior algorithm is of Boolean type, Boolean data is more suitable when using symbols such as true and false, 01, etc., which can focus on describing the data structure, thereby showing the relationship between the data. At the same time, it can realize the data analysis. Specific description. To better describe the relationship between the numerical values processed by the algorithm, numerical data is used to describe the related variables, and the algorithm is used to perform related data mining operations. Add the network graph node after the type node, and select the fields participating in the network graph construction. This article uses the thickness of the connecting line to reflect and demonstrate the strong and weak relationship between businesses. The resulting correlation diagram is shown in Figure 10.

Table 2: Business of Z company found by association rules

Company Z's major business
Mobile game ordering identification information
Weather forecast order identification information
Mobile video ordering identification information
Wireless music order identification information
Mobile news order identification information
Mobile securities order identification information
Mobile phone reading order identification information

Table 3: Association rules find the required value-added services

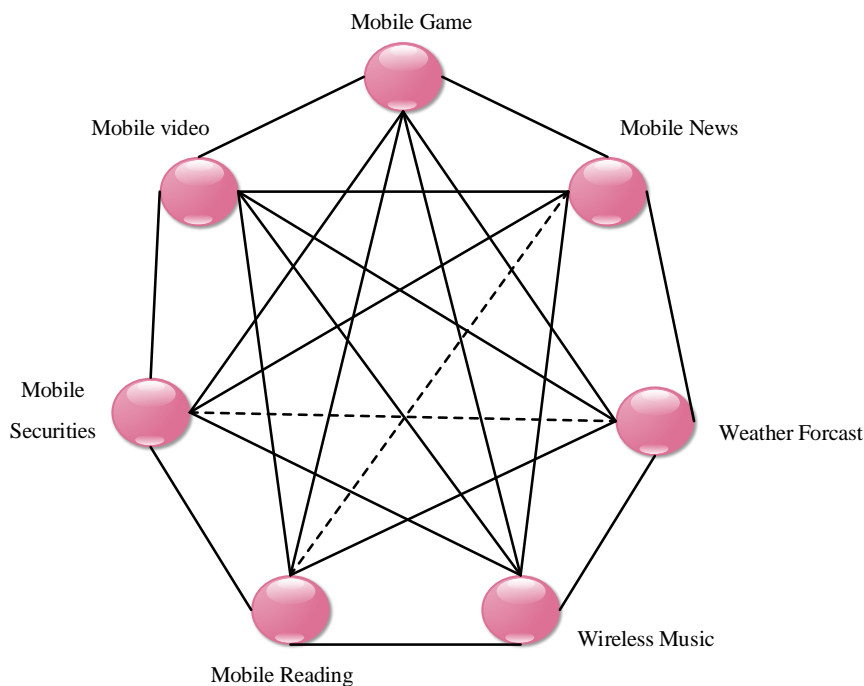| Latter	Former	Support (%)	Confidence (%)
Game	Video	3.21	24.34
Video	Game	6.2	12.34
Weather forecast	Wireless Music	4.3	12.32
Game	News	2.3	14.21
Weather forecast	News	2.5	10.22
Game	Reading	1.77	14.23
Video	Reading	1.77	13.54
News	Securities	0.53	22.08
Game	Securities	0.53	17.23
Video	Securities	0.53	13.22

Figure 10: Associated network of value-added services

The association network of value-added services can effectively display the associated relationship between a value-added service, enabling us to conduct data mining and analysis through an intuitive knowledge graph. The relevant right-side table can describe the relevant left-side business relationship so that the relationship between each mobile phone can be better represented to display the relationship between the data. At the same time, it can realize the specific description of the data. In the above figure, the value-added services with the largest number of user links are the subscription of mobile game value-added services and mobile video value-added services, with a value of 344. Secondly, the number of users of wireless music is 123. The analysis shows that the current marketing direction of K company should focus on these two aspects. For K company, it can not only carry out precise marketing through the analysis results but also increase the sales volume of this business, which is also very important for K company's mobile phone sales.

This paper takes K company as an example to apply the customer management model based on big data. It applies this model to the enterprise OA management system to improve its practicability. And the sales of company K within one year after applying the optimized customer management system are counted. It can be seen from Figure 11 that the sales of company K in 2022 will increase significantly compared with the sales in 2021, and the growth trend is obvious.

Good customer relationships are the reason why the company's performance continues to grow, and it is also the reason why the company is becoming more and more influential in the market. New customers have opened up new development space for the company and brought new growth points for the company. Stable customer resources have laid the foundation for the company's stable development. At the same time, the company is also actively exploring new technologies in technology research and development, actively adopting new technologies, and launching new products in a timely manner according to market demand to meet the needs of consumers. Management to enhance customer value. With the improvement of the company's informatization level, the company's operation efficiency has been continuously improved, which has promoted the improvement of the company's performance, making the company's operation enter a virtuous circle.
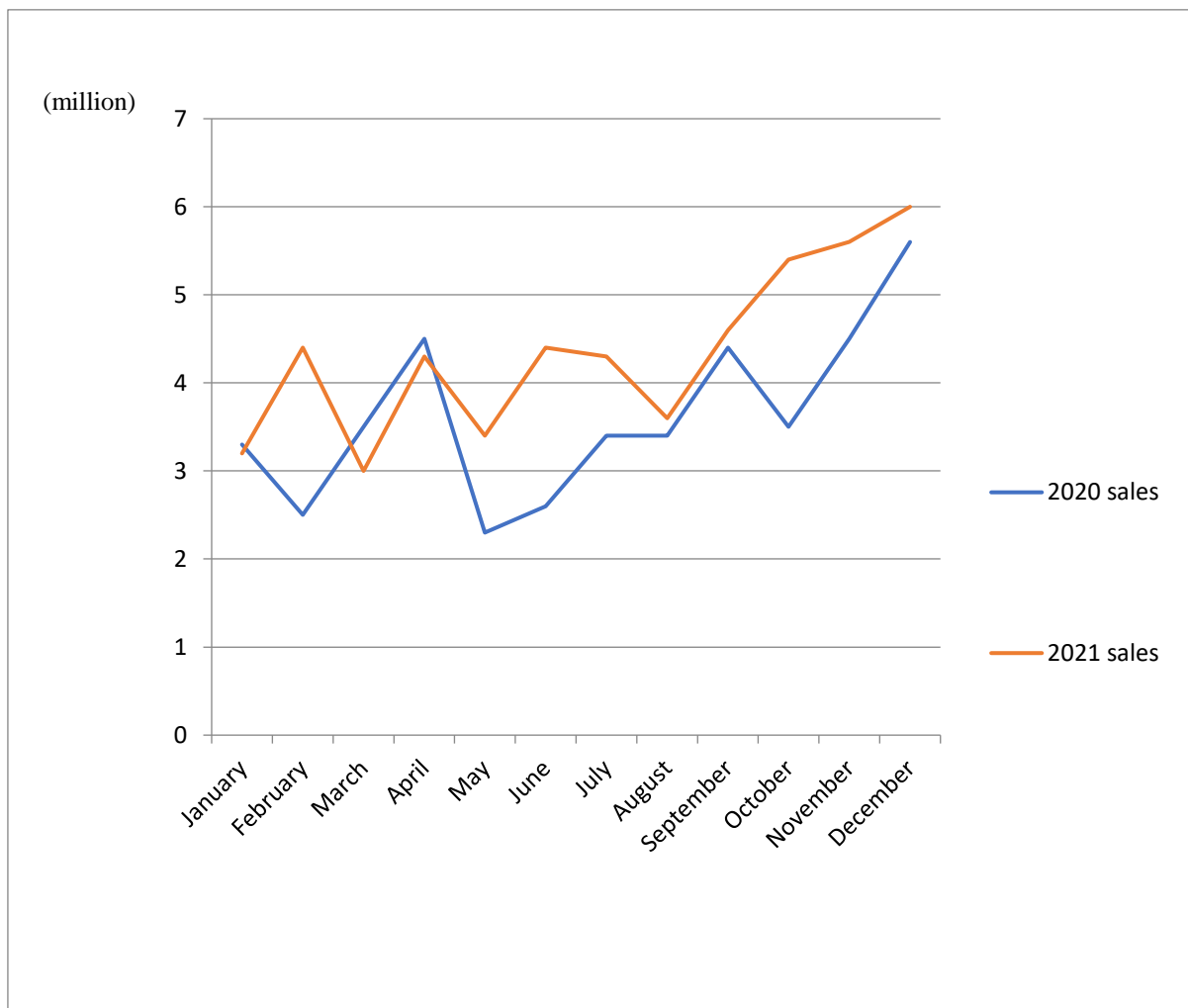

Figure 11: Company K's 2020-2021 sales comparison chart

5   Discussion

With the development of economy, the needs of customers are diversified, changeable and personalized, and the complex relationship between customers and enterprises is growing exponentially. The existing enterprise customer relationship management system has been unable to fully meet the management needs of the actual development of enterprises. In the era of the Internet and big data, how to intelligently and automatically dig out potential and valuable business models from the complex historical transaction information? It is an inevitable trend that the focus of enterprise customer relationship management shifts from operation to analysis and intelligent decision making. In recent years, the field of enterprise customer relationship management is undergoing revolutionary changes. The extensive application of mathematical models and the rapid development of computer technology have laid a good foundation for dealing with the complex customer relationship of enterprises. The emergence of data mining technology provides a new opportunity to improve the level of enterprise customer relationship management. Therefore, it is necessary to make full use of enterprise information resources, actively promote the transformation of enterprise management mode from product-centered to customer-centered, combine intelligent knowledge mining technology with enterprise customer relationship management system, improve the utilization of information resources to the advanced stage of knowledge innovation, and realize scientific enterprise management decision-making. It is of great practical significance for enterprises to stand out in the fierce market competition.

The research of this paper aims to realize the scientific and intelligent analysis method of customer relationship management data and information, and improve the level of enterprise customer relationship management. It intends to analyze the current situation and demand of enterprise customer relationship management, integrate the knowledge of logistics field, and comprehensively apply the intelligent technology of data mining. From the aspects of data mining process framework construction, association rule mining implementation, association rule analysis, customer relationship management improvement, etc., this paper conducts a systematic and comprehensive in-depth research on enterprise customer relationship management data knowledge mining, establishes a preliminary application framework of enterprise customer relationship management data mining method and implements its application. It effectively breaks the subjectivity and efficiency limitation of manual prediction of customer potential demand, and promotes the improvement of enterprise customer relationship management with the help of intelligent technology.

The research of this paper has great theoretical and practical significance for the research and application of enterprise customer relationship management theory and intelligent technology method of association rules mining.

This study initially establishes a set of systematic methods and technical routes for enterprise customer relationship management knowledge mining, and improves the scientificity of customer relationship management data and information analysis methods. The method system of customer relationship management knowledge mining process framework is described from the aspects of enterprise business goal definition, transaction data preprocessing, rule mining and result analysis.

The association rules mining process framework is effectively applied to the specific field of enterprise customer relationship management, and the excavated association rules can provide scientific and intelligent support for enterprise management decision-making, help enterprises find customer purchasing trend, accurately predict customer potential demand, guide cross-selling, promote the development of customer value, enhance enterprise competitiveness, and improve the scientific management level of enterprises.

Using the transaction data information effectively, a new customer value analysis method is put forward from another angle. Based on the analysis of the association rules mined by Apriori algorithm, this paper selects the purchasing pattern characteristics of customer groups that are more likely to purchase high-margin services and increase the net present value of customers' life cycle, and puts forward another method to identify the potential value brought by high-quality customers from the future purchasing trend.

6   Conclusion

In this paper, the basic process of customer management is analyzed, and the Aprior algorithm is chosen to build a CRM model based on data mining. In addition, this paper designs a customer relationship management system based on big data. The system is mainly divided into three layers, the data source layer, batch processing layer and real-time processing layer. In the part of constructing the system architecture, this paper adopts the Hadoop platform. In the batch processing layer, it is consisted of four parts, which include No SQL database, Oracle database, ETL architecture, and Hadoop platform. This paper gives the logical architecture design diagram. The real-time processing layer mainly includes a real-time decision engine and service bus. The key part of this layer is the real-time decision engine, in the design of which the Bayesian algorithm and product recommendation prediction model are used. Finally, this paper takes K company as an example to demonstrate the model and management system. After applying the analytical model and management system, the sales of K company keep increasing. The research scope of big data is very wide, and the variables of customer relationship management in the era of big data need to be explored from multiple perspectives and dimensions. This paper only analyzes several aspects of customer relationship management from the management perspective, and the scope and depth of the research are far from enough. It is hoped that

with the deepening of research in the future, more dimensional research can be done on strategies to improve customer management relationships.

The data sets in the customer relationship management system of enterprises are characterized by huge capacity, different abstraction layers and multi-dimensional structure. The Apriori algorithm adopted in this paper takes a long time to calculate and has limited application scope. The subsequent improvement of the algorithm can make the mining results more efficient and accurate.

Competing of interests

The authors declare no competing of interests.

Authorship contribution statement

Baohua Zhang: Writing-Original draft preparation, Conceptualization, Supervision.
Sanbao Zhang: Language review, Project administration
Chaojie Zhang: Methodology, Software, Validation

Data availability

On Request

Declarations

Not applicable

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] N. Capuano, L. Greco, P. Ritrovato, and M. Vento, "Sentiment analysis for customer relationship management: an incremental learning approach," *Applied intelligence*, 51: 3339–3352, 2021. https://doi.org/10.1007/s10489-020-01984-x

[2] B. Melović, B. Rondović, S. Mitrović-Veljković, S. B. Očovaj, and M. Dabić, "Electronic customer relationship management assimilation in Southeastern European companies—cluster analysis," *IEEE Trans Eng Manag*, 69(4): 1081–1100, 2020. https://doi.org/10.1109/TEM.2020.2972532

[3] M. Parast and D. Golmohammadi, "The impact of firm size and business strategy on response to service disruptions: evidence from the US domestic airline industry," *IEEE Trans Eng Manag*, 69(5): 1944–1957, 2020. https://doi.org/10.1109/TEM.2020.2994828

[4] R. Zhang, W. Chen, T.-C. Hsu, H. Yang, and Y.-C. Chung, "ANG: a combination of Apriori and graph computing techniques for frequent itemsets mining," *J Supercomput*, 75: 646–661, 2019. https://doi.org/10.1007/s11227-017-2049-z

[5] A. Kumar *et al.*, "Effects of HTL and ETL thicknesses on the performance of PQT-12/PCDTBT: PC 61 BM/ZnO QDs solar cells," *IEEE Photonics Technology Letters*, 32(12): 677–680, 2020. https://doi.org/10.1109/LPT.2020.2991536

[6] M. Zaharia *et al.*, "Apache spark: a unified engine for big data processing," *Commun ACM*, 59(11): 56–65, 2016. https://doi.org/10.1145/2934664

[7] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, 63(5): 3137–3147, 2016. https://doi.org/10.1109/TIE.2016.2519325

[8] S. Athey, "Beyond prediction: Using big data for policy problems," *Science (1979)*, 355(6324): 483–485, 2017. https://doi.org/10.1126/science.aal4321

[9] Y. Chen, J. D. E. Argentinis, and G. Weber, "IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research," *Clin Ther*, 38(4): 688–701, 2016. https://doi.org/10.1016/j.clinthera.2015.12.001

[10] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Netw*, 30(1): 44–51, 2016. https://doi.org/10.1109/MNET.2016.7389830

[11] X. Wang, Y. Zhang, V. C. M. Leung, N. Guizani, and T. Jiang, "D2D big data: Content deliveries over wireless device-to-device sharing in large-scale mobile networks," *IEEE Wirel Commun*, 25(1): 32–38, 2018. https://doi.org/10.1109/MWC.2018.1700215

[12]   Y. Chen and Y. Chi, "Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization," *IEEE Signal Process Mag*, 35(4): 14–31, 2018. https://doi.org/10.1109/MSP.2018.2821706

[13]   M. Asch *et al.*, "Big data and extreme-scale computing: Pathways to convergence-toward a shaping strategy for a future software and data ecosystem for scientific inquiry," *Int J High Perform Comput Appl*, 32(4): 435–479, 2018. https://doi.org/10.1177/1094342018778123

[14]   N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of big data and 5G wireless networks: opportunities, approaches, and challenges," *IEEE Wirel Commun*, 25(1): 12–18, 2018. https://doi.org/10.1109/MWC.2018.1700193

[15]   Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to cache: Machine learning for network edge caching in the big data era," *IEEE Wirel Commun*, 25(3): 28–35, 2018. https://doi.org/10.1109/MWC.2018.1700317

[16]   Y. Zhao, S. Bin, and G. Sun, "[Retracted] Research on Information Propagation Model in Social Network Based on BlockChain," *Discrete Dyn Nat Soc*, 2022(1): 7562848, 2022. https://doi.org/10.1155/2022/7562848

[17]   H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A.-X. Zhu, and W. Chen, "Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China," *Science of the total environment*, 625: 575–588, 2018. https://doi.org/10.1016/j.scitotenv.2017.12.256

[18]   M. Sattarian, J. Rezazadeh, R. Farahbakhsh, and A. Bagheri, "Indoor navigation systems based on data mining techniques in internet of things: a survey," *Wireless Networks*, 25: 1385–1402, 2019. https://doi.org/10.1007/s11276-018-1766-4