



OPEN Application of improved FP-tree algorithm in rural accurate information service

Yanfeng Jin, Jianwei Zhao[✉], Xuzhao Li & Jianmin Wu

The imbalance between supply and demand of rural information is a prominent problem faced by rural informatization. The lack of an effective supply of rural information is the biggest cause of this problem. The innovation of rural accurate information service mode has important practical significance in improving the contradiction between the supply and demand of rural information. This study analyzes the advantages and disadvantages of the existing rural information service modes at home and abroad. Based on the existing problems of rural information service model in China, a rural accurate information service model based on an improved FP Tree algorithm is proposed. The effectiveness of the mode is verified by simulation experiments, and the specific implementation path is given. The results show that the main problems of the rural information service model in China are lack of pertinence, low information quality of farmers, high information cost, and imperfect long-term mechanism; the rural accurate information service model based on improved FP Tree algorithm overcomes the above problems in terms of organization mode, service content, communication mode, benefit distribution, and technical support. Under the same data scale and tolerance, the improved FP Tree algorithm is better than the traditional Apriori algorithm.

Keywords FP-tree algorithm, Information service model, Precision service, Data mining

Rural information services are of importance to the development of all aspects of rural areas. They can assist in agricultural production decision-making, promote the popularization of agricultural technologies, expand the sales channels of agricultural products, enhance the farmers' living standards, and promote rural social governance. They have become a key factor in promoting the all-round development of rural areas^{1,2}. The content of rural information services is also very extensive, mainly including agricultural production, rural economy, social life, and public services³. These information services are provided through various channels to meet farmers' demands, and ultimately achieve the goals of improving rural social welfare.

To enhance the level of rural information services, many countries have carried out specific practices and exploratory research, achieving remarkable accomplishments and accumulating a considerable amount of experience. Through the literature review, we also find that, in the process of rural information services, there are still some problems. The most prominent one is the mismatch between the supply and demand of rural information services⁴. Specifically, the supply of information in rural areas is not what the farmers need, while the information that farmers truly need is lacking in supply, which has greatly affected the efficiency and effectiveness of rural information services.

The demand for rural information services refers to the needs of farmers for all kinds of information in the process of production, life, and social development. Their demands have the characteristics of diversity, timeliness, practicality, and easy accessibility^{3,5}. Meeting farmers' information needs is the fundamental purpose of rural information services; Rural information service supply refers to the sum of all information services provided by information providers, such as the government, enterprises, social organizations individuals, etc. It covers all the links of information production, dissemination, distribution, and application.

The reasons for the mismatch between the supply and demand of rural information services are multi-faceted. Among them, the lack of an accurate rural information service model is generally regarded as a crucial one. The rural information service model is a comprehensive service system in which information services are provided by information providers based on the farmer's needs. It serves to meet farmers' information needs in multiple fields, thereby promoting the all-around development of rural areas. Therefore, given the mismatch between the supply and demand of rural information services, it is urgent to develop an accurate rural information service model that can precisely achieve the balance of "supply=demand" for rural information services, that

Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China. [✉]email: jianwei_zhao8599@163.com

is, provide farmers with the information they truly need. At present, there is no unified definition of accurate information service in academics. However, many scholars believe that an accurate information service system is an information service model that uses big data analysis technology to identify the images of users, and based on their different needs, delivers personalized information to them through a modern network.

Previously, scholars from various countries have also researched the rural information services model^{4,6–11}, and put forward some ideas dedicated to coordinating the mismatch between the supply and demand of rural information services. However, most of them still take “self” as the center, only considering what the service provider “can offer” rather than “what the farmers need”. The supply of information services is still mainly a one-way push, which fails to meet the farmers’ personalized needs and does not take into account the interactivity between information supply and demand. However, in reality, farmers’ demand for information services is constantly increasing, and the trend of personalized demands is becoming more and more obvious. They are also more cautious about accepting information.

In particular, literature that is truly oriented towards the farmers’ information service demands, dedicated to addressing the imbalance between the supply and demand of rural information services, and utilizes modern information technology to study the accurate information service model in rural areas remains sparse. **To fill in the gap**, this paper utilizes big data mining and analysis techniques to propose a rural accurate information service model that balances the information supply and demand in rural areas. It analyzes and mines the personalized information demands and behavioral tendencies of farmers, and provides accurate information services to them. On the whole, the effectiveness of information service supply in rural areas has been enhanced.

The accurate information service model proposed in this paper, under the collaboration of the government, enterprises, and third parties, utilizes big data mining technology to analyze the personalized farmers’ needs and integrates the collected agricultural technology information, agricultural production information, market information, etc. An innovative model that precisely delivers to farmers through a combination of online and offline channels. This model consists of three steps: the first is the collection and processing of relevant agricultural-related data; the second is to process and analyze the collected data.; the third is to predict the farmer’s information needs and provide them with accurate and personalized rural information services.

The remainder of the paper proceeds as follows. We review the related literature in Section “[Literature review](#)”, and then we provide the information organization and communication model design in Section “[Information organization and communication mode design](#)”. In Section “[Construction of rural accurate information service model](#)”, we construct the rural accurate information service model. In Section “[Conclusion](#)” we conclude the paper.

Literature review

There are three streams of literature related to the research in this paper, and we will review them separately. The first stream is concerned with the rural information service model. The rural information service model is the external form of rural information service supply, reflecting the interactive relationship between the information provider and farmers. Through the reform and innovation of the rural information service model, the enthusiasm of all parties involved can be mobilized and exerted, and the mismatch between the supply and demand of rural information services can be improved. It is precisely because of its significance that many scholars are dedicated to researching rural information service models. For example, Liu et al⁸ proposed a susceptible exposed infected recovered (SEIR) information dissemination model in the era of artificial intelligence and proved that this model is more effective than the traditional ones. Zhang et al¹¹ reviewed and analyzed seven Agricultural information dissemination models based on Information and Communication Technologies (ICTs) in China, and provided some typical successful cases. Harbanskova and Gertsberg⁶ developed an information service model for the rural settlement system, considering five influencing factors (society, technology, economy, environment, and politics). After being tested for a period of time, it was affirmed by the certification department. Jin et al⁴ developed a probit discriminant model of influencing factors of rural information demand using big data technology. Kukar et al⁷ developed an information management and decision support model applied to rural farms. This model can predict the simulated scenarios of users and can better understand the dependencies among models. Muangprathub et al.⁹ predicted the soil fertility level through the analysis of soil information in rural areas of India. This study provides a reference for the rural information service supply model from the perspective of technical implementation. Based on data mining and the K-means algorithm, Shang and Shang¹⁰. Analyzed agricultural fertilization behavior and its influencing factors. Wang and Wang¹². Analyzed an uncertainty model in the integration path of rural tourism information service. Wang¹³. Built an image evaluation and optimization model utilizing deep learning technology to enhance the information service of rural tourism. Li and Zhou¹⁴. Developed an intelligent monitoring and analysis model based on federated learning to optimize the information services regarding rural sports. Ifty, R. A.¹⁵ offered an federated Learning-Driven Agricultural Innovations model to improve the accuracy of rural information services.

This research has a certain guiding significance for agricultural production and also provides ideas for the rural information service supply model. From the literature above, it can be found that the current rural information service models are diverse in types, while the accurate service that can truly meet the diverse needs of farmers is rarely studied by scholars. Most models still remain confined to the research on the information supply side, lacking in-depth research and investigation on the information demand side, and not even attempting to address the mismatch between the supply and demand of rural information services. In reality, the problem of mismatch between information supply and demand in rural area is still quite prominent. Therefore, it is very necessary to study a rural information service model that can better match the supply and demand of rural information service.

The second stream is on the farmer’s demand for rural information services. Scholars’ research on the farmer’s information demands began with how to effectively deliver information such as agricultural

information techniques and pest control to them. With the continuous expansion of information demands in rural areas, these types of information have been unable to meet farmers' actual needs. Researchers have begun to expand the research scope of information that farmers are concerned about, from the initial information on agricultural technology and pest control to that on agricultural production materials, and then to market information on agricultural products. For example, Bopape et al.¹⁶ Found that the information required by users of public libraries and information services in rural areas is centered around specific categories, and suggested understanding other services provided by public libraries. From time to time, we conduct in-depth research on user needs to enhance the effectiveness of information supply and meet the constantly changing needs of users. Folitse et al.¹⁷ evaluated the information needs and sources in rural areas. And concluded that some of the main contractual factors that farmers face when obtaining information include the lack of the skills to obtain information, insufficient information resources, the absence of information centers, and the inappropriate timing of agricultural programs being broadcast on radio stations. Chen and Lu⁵. Adopted the correlation analysis method to explore the relationship between farmers' information demands and channel preferences. The results show that individual characteristic factors, social factors, and family factors have different degrees of influence on farmers' information needs and access channels. Nikam et al.² examined the farmers' information demands in two different regions mainly relying on rain-fed and irrigated crops, as well as the ways to obtain information from different sources, and concluded that the farmers' information demands in the two regions are slightly different. Cui et al.¹⁸ applied the Urban Network Analysis Toolbox (UNA) toolkit and demand intensity measurement methods to establish a theoretical framework to examine the supply of and demand for rural public service facilities and information services Under the background of China's rural revitalization strategy. Shukla et al.³ studied the differences in information demands of farmers with different population structures and regions, by interviewing 90 respondents from the Sitapur region of Uttar Pradesh, India, to formulate targeted information dissemination strategies. With the advancement of rural informatization, the information provided in rural areas is no longer sufficient to meet the basic needs of production and living. Because of this, researchers shifted their attention to the fields of culture and entertainment, and even to the farmers' information awareness and behavior.

The third stream is about the supply of rural information services. Improving the supply quality of rural information services is an important way to solve the mismatch between the supply and demand of rural information services. Currently, scholars have conducted in-depth research on aspects such as technology and approaches to information services, attempting to find methods to improve the quality of rural information services. There is also much literature on this aspect. This article only lists some representative ones. Amir et al.¹ took the Lodland area of South Punjab Province, Pakistan as an example to study the interpersonal service information dissemination channels preferred by farmers. The conclusion of the research indicates that farmers prefer to have face-to-face communication with sales representatives of marketing companies regarding agricultural product service information. Chen et al.¹⁹ took the information and communication technology (ICT) platform as an example to study the role of rural information service supply in promoting rural e-government. The research results show that the information asymmetry and persistent tensions between the government and villagers have hindered the effectiveness of rural governance to varying degrees and run through all stages of e-government. Saridewi and Annisa²⁰. Studied the impact of the popularity of the Internet on the supply of information services in rural areas and pointed out that smartphones are one of the most extensive and effective channels for the supply of information services. Rabiu²¹. Studied the potential of Selective Dissemination of Information (SDI) as a strategic information supply tool in meeting the information needs of rural poor communities. The conclusion shows that customized and accessible SDI initiatives have a positive impact on community development. Nowfal et al.²². Studied the impact of agricultural cooperatives, as the main providers of service information, on strengthening farmers' access to market information. Studies show that agricultural cooperatives play an important role in providing service information.

From the literature above, it can be known that the rural information service supply system refers to an entity composed of the government, information enterprises, farmers, information products, and information communication media. The current achievements in the supply of rural information services mainly focus on issues such as the service providers, government roles, information supply channels, as well as supply effectiveness, but most of which are at a micro level. Some scholars have improved and expanded the current rural information service supply model with the government playing a leading role, all of which are aimed at addressing the existing problems. The suggestions put forward within their theoretical framework are not highly operational.

Information organization and communication mode design

Organization of rural information

The rural fine information service model constructed in this paper adopts a government-led, market-participated, and third-party-followed collaborative organization system with farmers as the information service object. Therefore, it can be divided into the following two modes: collaborative subject + service platform + farmer, collaborative subject + service platform + technician + farmer. These two models are influenced by the communication channels. If you choose an online channel, you will use the model of collaborative agent + service platform + farmer. If you choose an offline channel, you will use the model of collaborative agent + service platform + technician + farmer. As shown in Fig. 1.

Collaborative subject + service platform + farmers. This model is coordinated by the government, the market, and third parties, and uses information service platforms to send rural information to farmers in an online form. From the perspective of the subject, since the government subject in the collaborative subject still plays a leading role under this model, the government's own supervisory function is played; the participation of the market subject is conducive to the market-oriented operation of information services and promotes the maximization

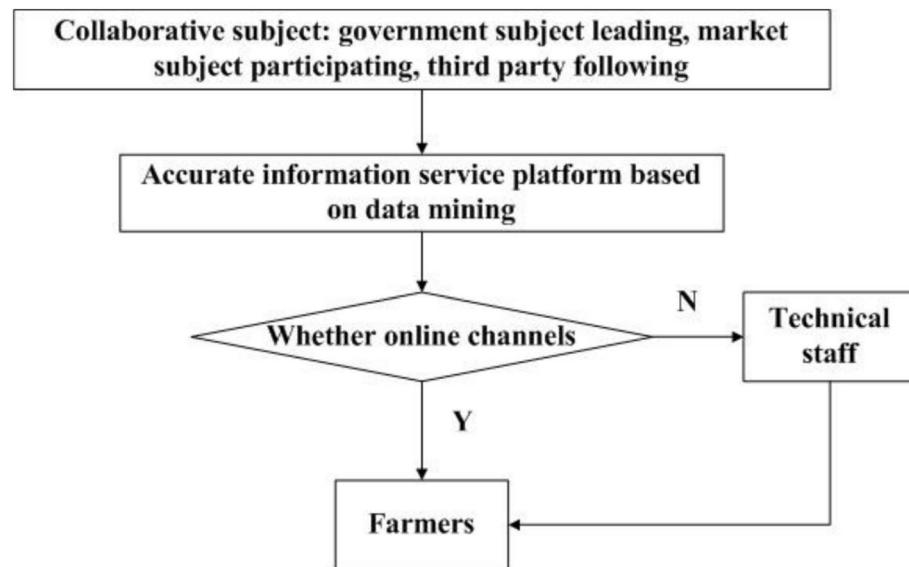


Fig. 1. Accurate information service organization model.

of the interests of all parties; The follow-up of third-party entities is conducive to the improvement of service levels. Therefore, this organizational model has greater influence and execution. At the same time, the service platform provides support for farmers' accurate information services, and analyzes the individual needs of farmers through data mining to ensure the effectiveness of this model organization.

Collaborative subject + service platform + technician + farmer. This model is similar to the above model and has the same advantages and features. The only difference is the participation of technical personnel under the offline communication channel. The technical personnel through the selection of offline channels ensure the effective delivery of information services. Of course, the quality and ability of technical personnel will also have a certain impact on the efficiency of the service organization model.

Rural information dissemination model

The information dissemination mode in the rural precise information service model directly determines the communication between service providers and farmers, but the information dissemination model is not directly related to the information service providers. The front-end information collection and processing, information service platform, and information The data mining function in the service platform provides an important basis for the selection of the dissemination mode of information services. As shown in Fig. 2.

Accurate push considering the differences in rural information environment. Due to the different advantages and disadvantages of the rural information environment, the dissemination of information services needs to consider regional differences, and selectively push different types of information to farmers in different regions.

Consider the precise push of farmers' individual needs. In the promotion of information services, we must fully consider the impact of different factors, such as pushing different information service content for farmers of different ages; the form and quality of rural information services should take into account the farmers' cultural level, and push the matching information services; The duality of peasants' occupations makes the information services that they care about quite different, so it is necessary to distinguish between the pushes of information services.

Consider the dynamic and accurate push of farmers' information feedback. From the perspective of time, the information needs of farmers are obviously affected by seasonal factors. At the same time, as the information environment changes, and the age and experience of farmers increase, the information needs of farmers will continue to change. From the perspective of information service application effects, farmers need to provide timely feedback on information service results and satisfaction. Therefore, the service subject should timely adjust the information service content according to the needs of farmers, and carry out dynamic and accurate push.

Construction of rural accurate information service model

Through research, it is found that although the current domestic and foreign rural information service models are diversified, they still cannot meet the individualized needs of farmers. From the perspective of the service provider, it is more self-centered and considers "what can be provided" rather than "what the farmer needs"; from the perspective of the farmer demand side, the farmer's demand for information exists objectively, and personalized demand is increasing The more obvious, and at the same time take a more cautious attitude towards information; from the perspective of rural information dissemination channels, operators provide basic network information platforms, which need to be further tapped and improved in terms of functions; from government policies, although Under the rural revitalization strategy, the trend of rural informatization is irreversible, but

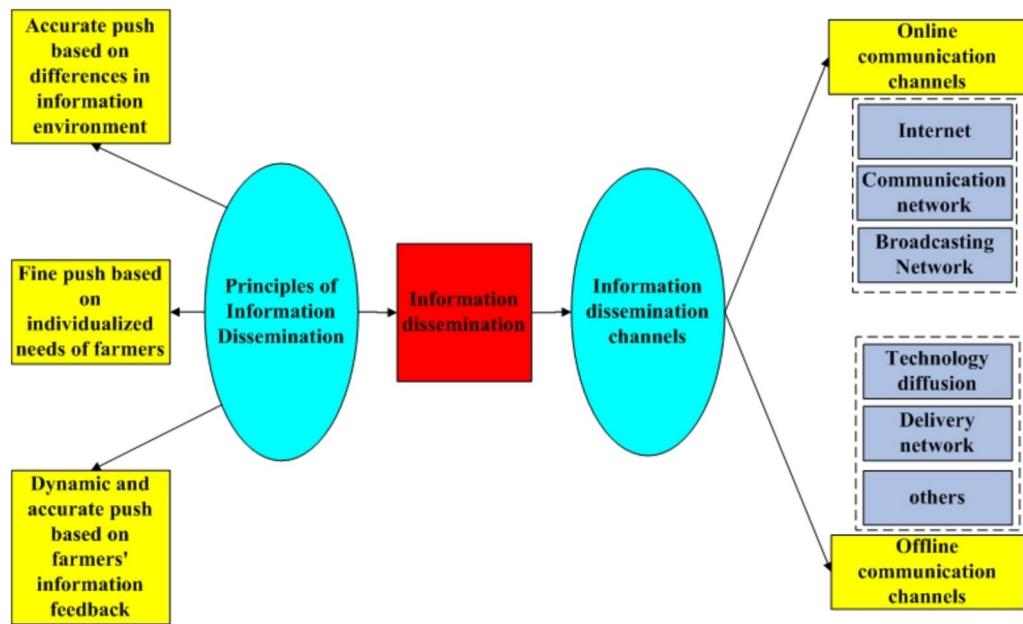


Fig. 2. Information dissemination mode of precise information service.

there is still a lack of targeted policies for path guidance in the construction of digital villages and smart villages. Based on the above problems, considering the individual needs of farmers for information and their willingness to adopt information, based on the idea of data mining, the rural information service model innovation is proposed^{23,24}.

Model building

The rural precision information service model refers to the use of data mining technology to analyze the personalized needs of farmers under the collaborative cooperation of the government, enterprises and third parties, and accurately transfer the collected agricultural technology information, agricultural production information, market information, etc. to farmers. The first is the collection of relevant agricultural-related data, information processing, processing and analysis, guided by the needs of farmers, to predict and provide personalized rural information services to farmers. The collection and processing of data is the basis of the work and the most difficult part. Different types of data from multiple sources need to be considered. Information processing and analysis mainly use data mining tools to identify and predict different needs of farmers and establish different service categories²⁵. The information service owner provides accurate and effective personalized services to farmers and obtains farmers' feedback in a timely manner. As shown in Fig. 3.

From the perspective of the effectiveness of the service model, accuracy is the biggest advantage of the model, and the design of data mining algorithms is the greatest technical guarantee to achieve the accuracy of rural information service supply. In the following, the technical support guarantee under this mode will be discussed from the process of rural information data collection, data preprocessing, algorithm design, and data mining.

Data collection and preprocessing

Data collection of farmers' information needs

The study's respondents are villagers from administrative villages. The data mainly come from two sources: First, county-level statistical yearbooks; second, survey data, which primarily include interviews with relevant personnel and questionnaire data collected from the sample. To facilitate data collection and ensure a uniform sample distribution, while also covering three types of terrain in terms of geographic space (plains, hilly areas, and mountains), this study selected 15 towns under a central China province and 30 administrative villages within them, totaling 1,543 respondents as the study sample.

Based on the research on farmers' information needs and farmers' willingness to accept information, the data table is designed. See Table 1 for details. Data attributes are mainly divided into two categories: First, the basic information attributes of farmers, such as name, gender, age, etc.; Second, the types of farmers' information needs, such as life, production, market, education, and entertainment and so on.

Each row represents the type of information needs of each farmer, corresponding to a transaction, and each column corresponds to an item. Let $I = \{i_1, i_2, \dots, i_d\}$ be the set of all items in the farmer's demand information, while $T = \{t_1, t_2, \dots, t_N\}$ is the set of all transactions. Each transaction t_i contains a subset of I .

Data preprocessing

Information collection is a very important work. After data collection, we need to introduce the concept of association rules, and then construct FP Tree algorithm through association analysis to preprocess the information data.

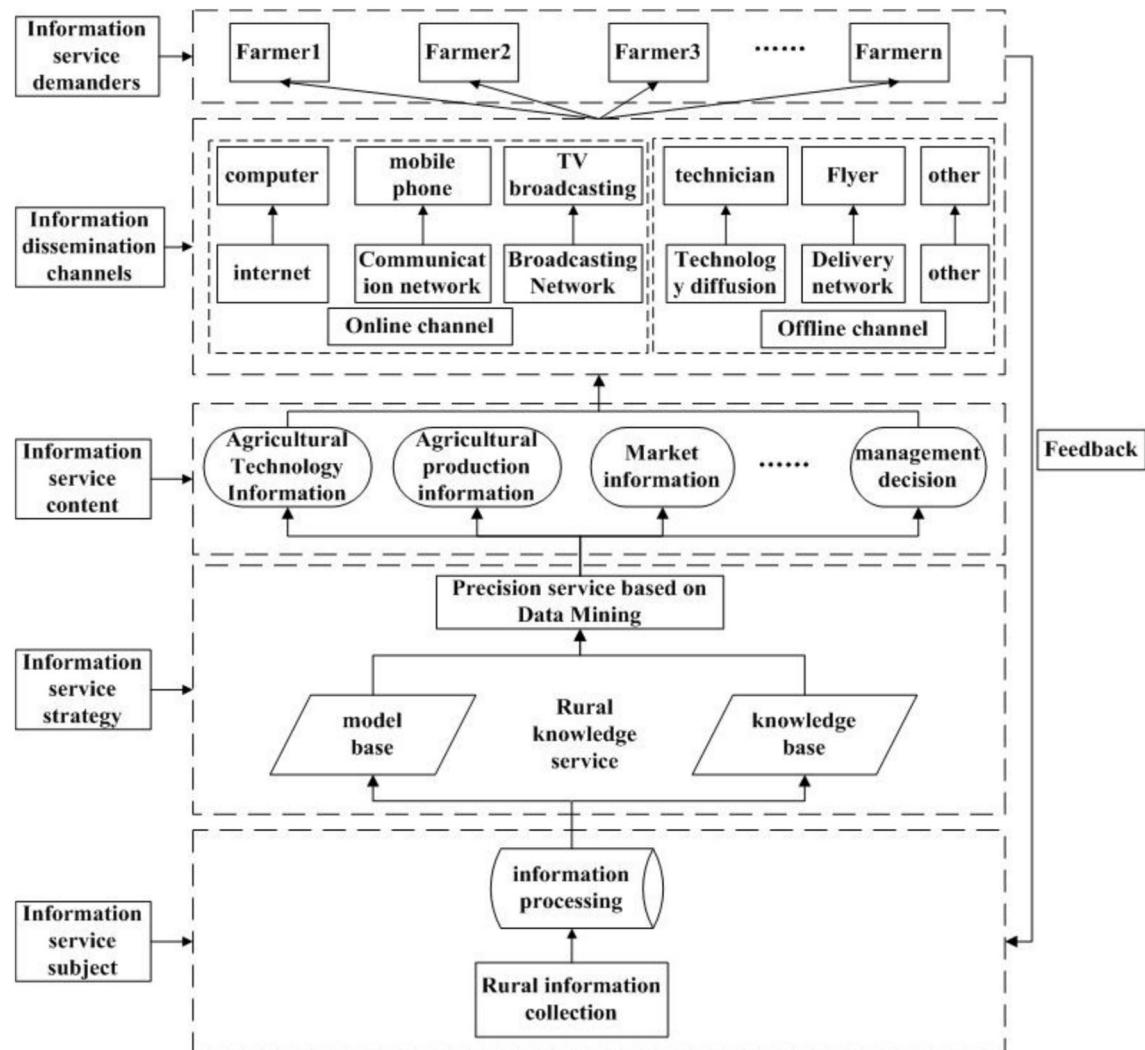


Figure 3. Rural precision information service model based on data mining.

Basic information					Information category and intention			
					Rural technical information	Agricultural production information	Market information	Management decision information
ID	Name	Sex	Age	Area	Y/N	Y/N	Y/N	Y/N
1	ZHANG	M	41	Town	Y	N	N	N
2	LI	F	53	village	Y	N	Y	Y
3	WANG	M	30	Town	Y	Y	Y	Y
4	ZHOU	F	34	Town	Y	N	N	Y
5	ZHAO	M	23	village	Y	Y	Y	Y
6	MA	F	35	village	Y	N	Y	N
7	ZHENG	M	45	village	N	Y	Y	Y

Table 1. Contents of data collection.

Association rules reflect the interdependence and association between one thing and other things. Association rules are one of the main technologies of data mining. In essence, association rules are used to find frequent patterns, associations, correlations or causal structures between item sets or object sets in transaction data or other information carriers. Shopping basket analysis is a good example of association rules, such as promotion classification and cross marketing. They all discover association rules from data, identify and inform the relationship between certain human behaviors and projects, and finally describe specific local patterns. At the same time, association rules are also applicable to other fields, such as bioinformatics, medical diagnosis, web mining and scientific data analysis²⁶.

The definition of association rule mining is as follows: for a binary data table, $I = \{i_1, i_2, \dots, i_d\}$ is a set of items, $T = \{t_1, t_2, \dots, t_N\}$ is the set of all transactions, each transaction t_i contains a set of items is a subset of I . In association analysis, a set containing 0 or more items is called an item set. The so-called association rules refer to expressions of the form $X \rightarrow Y$, where X and Y are disjoint itemsets, X is called the leader, and Y is called the successor of the rule. In association rules, the two attribute values of support and confidence are usually used to directly describe the nature of the association rule. The support determination rule can be used for the frequency of a given data set, and the confidence determines that Y is in a transaction containing X . The frequency of occurrence in. The two measures of support (s) and confidence (c) are defined as follows:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}, c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

In the process of mining association rules, if you do not consider the support and confidence thresholds, you will find an infinite number of association rules from the database, but in our actual life, we need association rules with practical significance to reflect the implicit data Rules, so you need to set a minimum value, that is, the minimum support and minimum confidence. Association rules can be divided into two stages: The first stage must find all frequent item sets from the data set, and the second stage generates strong association rules from these frequent item sets²⁷.

The key of association mining is the mining of frequent itemsets. The purpose is to find all frequent itemsets. When the frequency of occurrence of an item is “frequent” relative to other items, we call it frequent itemsets. After frequent itemsets are identified, the corresponding association rules can be directly pushed out. Compared with the first stage, the second stage is easier. It uses the frequent itemsets obtained in the previous step to generate rules. If the confidence of a rule meets the minimum confidence, it is called a strong association rule. There are many algorithms that can extract frequent item sets²⁸. The two typical strategies adopted by these algorithms are: One is to reduce the candidate set combination search space through an effective pruning strategy (Apriori technology), and one is to use compressed data representation To promote the core computing of the program (FP tree technology).

FP-tree is a prefix tree, which is composed of frequent item header table and item prefix tree, arranged in descending order of support, the higher the support, the closer the frequent items are to the root node, so that more frequent items can share the prefix. The FP-growth algorithm compresses the information in the transaction database by building an FP-tree (Frequent Pattern tree), thereby generating frequent item sets more efficiently.

The initial FP-tree root node is null, meaningless. Afterwards, the transactional database is continuously scanned to create an FP-tree. For each transaction, a node of an item is created. The special support is greater than the support threshold. When the same node is encountered, a support is added. After FP-tree is constructed in this way, frequent item sets can be mined to build a conditional pattern library. (The authors declare that all experiments in this paper were approved by the E-commerce Vocational Education Group of Hebei Province, China, and the experimental methods applied conformed to the regulations of the group on scientific experimental methods).

Analysis of improved algorithm based on FP-tree

Algorithm improvement

In order to increase the mining speed of the Apriori algorithm based on FP-tree, research is conducted from the direction of improvement to further reduce the amount of scanned data. In this paper, the following methods are used to improve and optimize: First, change the partitioning method of the algorithm and use the tail meta-partitioning method for partitioning, which can get a smaller amount of data. Second, based on the nature of the Apriori algorithm, this paper dynamically deletes redundant data in the sub-data set that is smaller than the number of dimensions of the current iteration. The dynamic reduction of the data further reduces the data in the sub-data set¹². Finally, scan the sub-data set for quick statistics, quickly count the number of support of the candidate set, and determine whether the candidate set is a frequent item set, so as to achieve rapid mining.

The initial Apriori algorithm is an algorithm that uses the iterative idea of layer-by-layer search to mine frequent itemsets with association rules. Its specific operation process is mainly divided into the following three steps:

Generate candidate sets. Randomly select two frequent item sets, where each item in the k item set is arranged in order, and connect to generate a candidate set;

Candidate set cutting technique. The anti-monotonicity of frequent item sets is used. If the candidate $K + 1$ item candidate set contains any infrequent items or item sets, then the infrequent item sets can be directly determined, otherwise the support number statistics of the database traversal need to be further Verify whether it is frequent;

Support statistics. Scan the data set and accumulate the number of occurrences of the item candidate set in the data set. Finally, $K + 1$ frequent itemsets are generated according to the given minimum support number threshold.

The support number calculation follows a theorem when determining frequent itemsets: If the length of a transaction in the database is K , then this transaction cannot contain any frequent itemsets with a number of items greater than K .

Corollary: If the length of a transaction is less than, the transaction data can be ignored in the support count statistics of the $K + 1$ candidate set.

The specific improvement methods based on FP-tree structure are as follows:

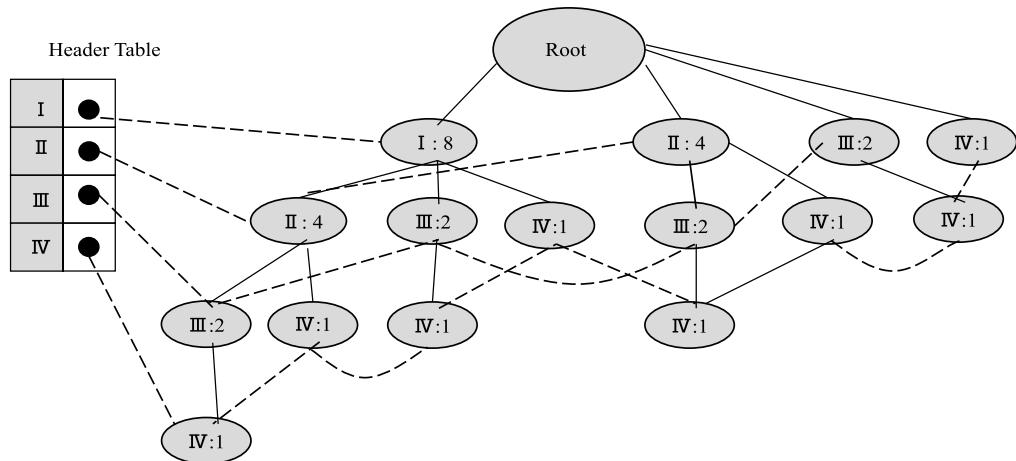


Fig. 4. Schematic diagram of FP-tree structure.

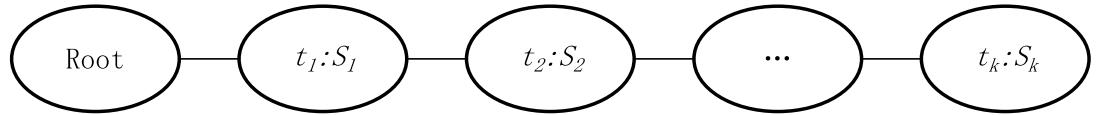


Fig. 5. FP-tree branch.

The FP-tree structure is a tree data structure that compresses data, where each node corresponds to an item element, each node is identified by an item element, the number of supports through the node, the item element chain, and the parent node pointer 4 Domain composition. In addition, in order to facilitate the operation of the tree, a header table is also needed to record the two fields of the item meta identifier and the item chain link. In FP-tree, the solid line is the parent node pointer, and the dotted line is the item chain.

Tail meta partition: Search the FP-tree through the tail meta of the candidate set, divide the original data set to generate several sub-data sets, so as to achieve the purpose of reducing the amount of data. There is only one copy of each generated sub-data set in memory, and there is no need to generate it again after generation. The specific process of tail element partitioning is shown in Fig. 4.

If a tail item t_k exists in the candidate set, the sub data set generation process of t_k is as follows: Find the node-head of the item element chain in the header table Header Table of the item element t_{k1} , and use the node-head to find the first node $T1$ whose item-item identifier item-name is t_k in the FP-tree. Assuming that the general form of the branch of the FP-tree prefix where node t_{k1} is located is shown in Fig. 5, the information of all nodes on the branch can be obtained by traversing the branch, and the item-name of each node on the branch is summarized into a vector representation. The general representation of the branch vector is: $k = [t_1, t_2 \dots t_k]$.

The t_{k1} branch vector can be obtained as k_1 , and record the support count of node t_{k1} as S_1 . Find the node t_{k2} at the next position t_k through the node-link of node t_{k1} , and generate the vector k_2 of the branch where the node t_{k2} is located in the same way, record the support number as S_2 . Traverse the next node according to the node-link of the current node until the last node t_{ki} . Get the branch vector k_i , support number S_i , stop traversing. Summarizing all the branch vectors, the sub-data set of this element t_k can be obtained as: $M_k = \{k_1, k_2, \dots k_i\}$. The corresponding support data set is $S_k = \{S_1, S_2, \dots S_i\}$. The S_k data set records the number of supports corresponding to each branch in the sub-data set, which is used for the following quick support number statistics.

By tail element partitioning, the data to be scanned for frequent item sets with the same tail element are grouped into the same sub-data set, thereby avoiding the scanning of irrelevant data. In the case of a large number of frequent item sets, the amount of data is reduced. More, the algorithm mining speed is also improved.

Dynamic data reduction: The iterative generation of candidate sets recurs from low to high dimensions. When the number of transactions is less than the number of dimensions of the iteration, it is known that the frequent itemsets of the current number of dimensions must not be included in this transaction, then this transaction can be ignored in the support statistics, and this transaction will be deleted from the sub-data set to achieve reduced scanning. The purpose of the data volume. The description of the dynamic reduction of the sub-data set data is as follows:

On the basis that the sub-data set has been generated, if you want to count the support number of a certain candidate set $C_k = [t_1 \ t_2 \dots t_k]$, then select the last item t_k of C_k , find the sub-data set corresponding to t_k , such as the sub-data set M_k in Eq. (2). Among them, if the dimension of the current iteration is j , then delete the transaction with the number of transactions less than j in M_k , get a new M_k with a smaller amount of data.

The improved method of dynamic data reduction is to reduce the data based on the reduced sub-data set in the previous layer. When the number of iterations is higher, the more redundant data can be deleted. Therefore, the more obvious the mining speed of the algorithm is improved.

Fast statistics support number: Traverse new M_k , compare the vector of each branch with C_k , if k_i contains C_k , then this branch is considered to contain candidate set C_k , the cumulative number (3) supporting number in the data set S_k corresponding to this branch, if there are integers n branches containing candidate set C_k ,

Then the support number of the final candidate set C_k is: $S = \sum_{i=1}^n S_i$. If S is greater than the minimum support number, the candidate set C_k is a frequent item set.

The improved algorithm proposed in this paper is improved through the above ideas and methods, which improves the algorithm mining speed, especially for data sets containing a large number of high-dimensional number frequent item sets, the amount of scanned data will be greatly reduced, and the algorithm mining speed can be significantly improved²⁹.

Improved algorithm experiment simulation analysis

In this section, the self-built database collected and processed using questionnaires is analyzed using the original Apriori algorithm and the improved Apriori algorithm based on FP-Tree, and then compared from two aspects: The running time of the two algorithms on different scale data under the same conditions; Comparison of the running time of the two algorithms under different tolerances. In order to reduce the experimental error, the experimental results are averaged after multiple experiments for comparative analysis.

Comparison of running time at different data scales. While keeping the support and the number of cluster points unchanged, the data set is divided into six subsets of different sizes, and the different subsets are tested separately to see the time required to run the data, as shown in Fig. 6.

It can be seen from Fig. 4 that when the amount of data is small, the running time is more spent on parallel work scheduling. Therefore, the running time of the Apriori algorithm and the FP-Tree algorithm is basically the same, and there is no obvious advantage of the algorithm. When the amount of data gradually increases, the difference between the Apriori algorithm and the FP-Tree algorithm is gradually reflected. The running time of the FP-Tree algorithm is much shorter than that of the Apriori algorithm. When the amount of data continues to increase, the Apriori algorithm will first cause problems such as insufficient memory in the FP-Tree algorithm, which causes the algorithm to fail. If the memory reaches a relative value and remains constant, then the data processed by the FP-Tree algorithm will be several times higher than the Apriori algorithm. Therefore, the experimental results are basically consistent with the proposed advantages, indicating that the improved algorithm is significantly better than the original algorithm.

The running time under different tolerances is compared as follows: This experiment only changes the preset support degree of the algorithm, and explores the efficiency of the two algorithms under different support degrees. With the same data size and the same execution in a distributed environment, the number of cluster nodes is 3. The experimental results are shown in Fig. 7.

It can be seen from the figure that with the increase of the minimum support, the overall running time of the algorithm is continuously decreasing, but the overall decreasing trend is gradually gradual. This is because as the minimum support increases, the resulting frequent itemsets will decrease, and the algorithm calculation time will also decrease. Overall, when the data set is kept at a fixed size, no matter how the tolerance of the algorithm is changed, the running time of the FP-Tree algorithm will always be less than the Apriori algorithm, which also verifies the superiority of the algorithm.

Simulation experiment

The database used in this study is a self-constructed database composed of county annual data and survey data. It employs high-performance hardware configuration and a stable operating system environment to ensure the

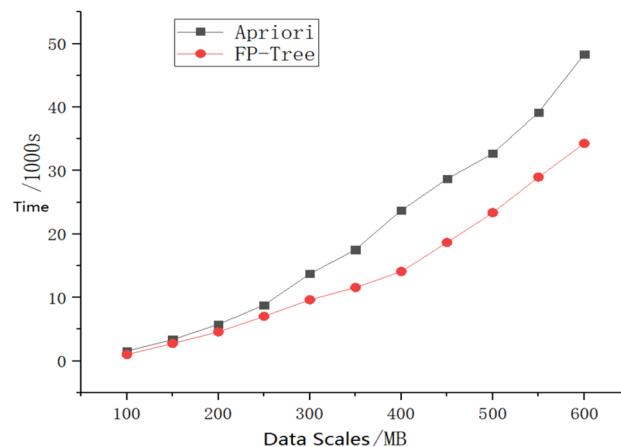


Fig 6. Comparison of running time of the two algorithms under different data scales.

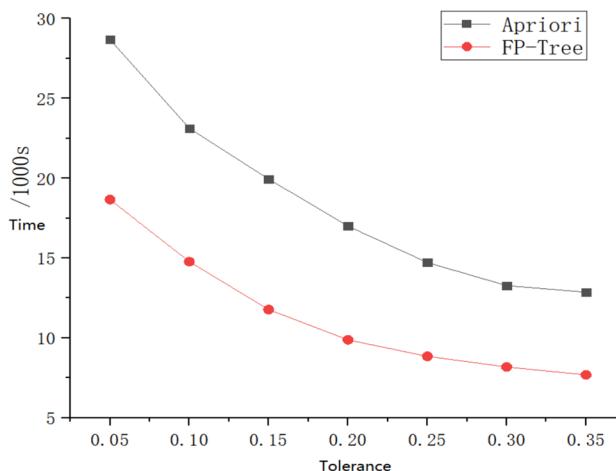


Fig 7. Comparison of the running time of the two algorithms under different tolerances.

ID	Items	Frequent items
1	b,c,a,f,g,h,m,k	f,c,a,m,k,b
2	a,b,f,m,e	f,a,m,b
3	c,f,w,g,k	f,k
4	d,a,b,n,k	a,b,k
5	f,a,d,k,m	f,m,k,a
6	j,a,m,k,e	a,m,k
7	a,f,g,h,k,m,c	a,f,k,m,c

Table 2. Farmers' demand for different types of information.

Rules	Confidence	Support
$1 \rightarrow 4^*6$	$2/5 = 40\% (\{1, 4, 6\}\text{Frequency}/\{1\}\text{Frequency})$	$2/7 = 29\%$
$4 \rightarrow 1^*6$	$3/5 = 60\% (\{1, 4, 6\}\text{Frequency}/\{4\}\text{Frequency})$	$3/7 = 43\%$
$6 \rightarrow 1^*4$	$4/5 = 80\% (\{1, 4, 6\}\text{Frequency}/\{6\}\text{Frequency})$	$4/7 = 57\%$
$1^*4 \rightarrow 6$	$3/4 = 75\% (\{1, 4, 6\}\text{Frequency}/\{1, 4\}\text{Frequency})$	$3/7 = 43\%$
$1^*6 \rightarrow 4$	$3/4 = 75\% (\{1, 4, 6\}\text{Frequency}/\{1, 6\}\text{Frequency})$	$3/7 = 43\%$
$4^*6 \rightarrow 1$	$3/4 = 75\% (\{1, 4, 6\}\text{Frequency}/\{4, 6\}\text{Frequency})$	$3/7 = 43\%$

Table 3. Association rules between types of information.

accuracy and repeatability of the experiments. The computer used for the experiments is equipped with an Intel Core i7-14700KF processor, 32 GB DDR5 6000 MHz high-frequency memory, and a 1 TB NVMe M.2 solid-state drive. The operating system is 64-bit Windows 11, and the software environment relied upon during the experiments is IBM SPSS Modeler 18.0. This setup ensures the objectivity and accuracy of algorithm efficiency testing during hardware and software operation.

For the types of farmers' information needs, the improved Apriori algorithm is used to mine association rules between different types of information, and the effectiveness of the rules is determined based on support and credibility.

According to the principle of the algorithm, the data records in the constructed farmers' information demand data set are searched for frequent itemsets. Table 2 shows the types of information required by different farmers.

Table 2 shows the transactional database used for farmers' information needs. Among them, a, b, ..., m represent the types of information required by farmers. First, scan the transactional database to calculate the support of various information requirements in each row of records; then arrange in descending order of support, only retain frequent item sets, and remove items below the support threshold. The support threshold is taken in this paper. Is 3, so as to get $\langle (f:4), (a:4), (b:3), (c:3), (m:3), (k:3) \rangle$, the third column of Table 2 is sorted result.

According to the frequent item set $\{f,a,b\}$ in the results of the association rule algorithm, its non-empty true subset is $\{f\}$, $\{a\}$, $\{b\}$, $\{f,a\}$, $\{a,b\}$, $\{f,b\}$. So 6 association rules were obtained, and the corresponding confidence and support were calculated. The specific rules are shown in Table 3.

Therefore, based on the improved Apriori algorithm based on FP-tree, the established rural information demand data is established for association rule mining, which can effectively mine the association between different types of information, and decide whether to adopt the rule based on confidence and support. And at the same time predict the types of information that farmers may need in the future based on the types of information the farmers need. Realized accurate information services based on data mining results.

Result analysis

Based on the improved Apriori algorithm based on FP-tree, the basis for farmers' personalized information needs is established. According to the actual demand of farmers for information, the association rule analysis method was used to find out the association relationship among all agricultural-related information categories, and establish association rules based on confidence.

The establishment of the above rules has provided technical support for the precise model of rural information services, greatly improved the accuracy of farmers' information, and truly realized the provision of information services according to the information environment and on demand. At the same time, considering that the time complexity of the original association rule algorithm is too high, which may lead to a decline in data mining efficiency in the case of large amounts of data, this paper optimizes and improves apriori in advance to effectively avoid this problem.

The accuracy of the results reached over 90% when compared with the interview data from the respondents. This not only significantly improved the efficiency of matching information supply and demand, but also ensured the precision of the information provided.

Conclusion

Firstly, the types of foreign rural information service models are analyzed, and a comparative analysis is carried out. It is found that there is a large gap in the information service models of different countries. This diversity provides a reference space for my country's rural information service. Secondly, it sorts out and summarizes the domestic rural information service models, and finds out the problems in China's rural information services, mainly including the lack of targeted rural information services, the peasants' own deficiencies, the high cost of rural information services and the lack of rural information services Long-term mechanism, etc. This part is not only a supplement to the current situation of rural information service supply but also a continuation study. According to farmers' information needs and willingness to adopt information, an accurate information service model with feedback mechanism based on data mining is proposed. It mainly uses data mining methods to predict the needs of different farmers and provide differentiated information based on different needs. At the same time, when the farmers receive the information, they can provide feedback and realize two-way communication, and the information providers can dynamically adjust the types and methods of information services. Through the accurate information service model, the value of rural data can be brought into full play to realize the effective use of information.

Data availability

The datasets generated and/or analysed during the current study are not publicly available. Due that this article uses a self-built database, which is based on research data from local governments and cannot be made public, but are available from the corresponding author on reasonable request.

Received: 11 July 2025; Accepted: 11 November 2025

Published online: 24 November 2025

References

1. Amir, A., Akhter, P., Cheema, A. N. & Asif, A. Facilitating agricultural knowledge dissemination: An exploration of farmers' preferred interpersonal communication channels and styles. *J. Res. Soc. Dev. Sustain.* **3**, 49–57 (2024).
2. Nikam, V., Ashok, A. & Pal, S. Farmers' information needs, access and its impact: Evidence from different cotton producing regions in the Maharashtra state of India. *Agr. Syst.* **196**, 103317. <https://doi.org/10.1016/j.agrys.2021.103317> (2022).
3. Shukla, G., Ansari, M. N. & Lal, S. P. Assessment of agricultural information needs of farmers: Triangulating reliability of standardized information need index. *Guj J Ext Educ.* <https://doi.org/10.56572/gjoe.2024.37.2.0005> (2024).
4. Jin, Y., Li, G. & Wu, J. Research on the evaluation model of rural information demand based on big data. *Wirel. Commun. Mob. Com.* **14**, 8861207. <https://doi.org/10.1155/2020/8861207> (2020).
5. Chen, Y. & Lu, Y. Factors influencing the information needs and information access channels of farmers: An empirical study in Guangdong China. *J. Inf. Sci.* **46**, 3–22 (2019).
6. Harbiankova, A. & Gertsberg, L. Information model for sustainable rural development. *Energies* **15**, 4009. <https://doi.org/10.3390/en15114009> (2022).
7. Kukar, M., Vrancar, P. & Kosir, D. AgroDSS: A decision support system for agriculture and farming. *Comput. Electron. Agric.* **161**, 260–271 (2019).
8. Liu, J., Yu, Y., Ji, P. & Liu, D. Information dissemination model in rural live broadcasting under blockchain in the era of artificial intelligence. *Comput. Intell. Neurosci.* **11**, 4590578 (2022).
9. Muangprathub, J. et al. IoT and agriculture data analysis for smart farm. *Comput. Electron. Agric.* **156**, 467–474 (2019).
10. Shang, J. & Shang, M. Agricultural fertilizer application behaviour and its influencing factors based on data mining and K-means algorithm. *Rev. Fac. Ing.* **32**, 627–634 (2017).
11. Zhang, Y., Wang, L. & Duan, Y. Agricultural information dissemination using ICTs: A review and analysis of information dissemination models in China. *Info. Proc. Agri* <https://doi.org/10.1016/j.inpa.2015.11.002> (2016).
12. Wallace, L. G. & Sheetz, S. D. The adoption of software measures: A technology acceptance model (TAM) perspective. *Inf. Manag.* **51**, 249–259 (2014).
13. Wang, X. The analysis of rural tourism image optimization under the internet of things and deep learning. *Sci. Rep.* **14**, 29898. <https://doi.org/10.1038/s41598-024-81868-z> (2024).

14. Li, Q. & Zhou, Q. Intelligent monitoring and analysis of sports data in the context of rural revitalization based on federated learning. In *International Conference on Neuromorphic Computing (ICNC)*.<https://doi.org/10.1109/ICNC64304.2024.10987783> (Chongqing, China, 2024).
15. Ifty, R. A. Project AgriSage Tech: Federated learning-driven agricultural innovations for precision agriculture. In *IEEE 12th Region 10 Humanitarian Technology Conference (R10-HTC)*.<https://doi.org/10.1109/R10-HTC59322.2024.10778905> (Kuala Lumpur, Malaysia, 2024).
16. Bopape, S. et al. Identifying the information needs of public library and information services users in Limpopo province. *S. Afr. J. Libr. Inf. Sci.* **83**(1), 10 (2017).
17. Folitse, B. Y. et al. Poultry farmers' information needs and sources in selected rural communities in the greater accraregion, ghana. *Int. Inf. Libr. Rev.* **50**, 1–12 (2018).
18. Cui, X., Ma, L. & Zhang, W. Do the supply of and demand for rural public service facilities match? Assessment based on the perspective of rural residents. *Sustain. Cities Soc.* **82**, 103905. <https://doi.org/10.1016/j.scs.2022.103905> (2022).
19. Chen, X., Cheng, X., Zhang, T. & Guo, H. How do information and communication technology platforms shape rural e-governance: The case of Zhao-lou Village on the WeCounty platform. *Inf. Syst. J.* **35**, 545–576 (2025).
20. Saridewi, L. P. & Annisa, L. Use of information and communication technology at the farmer level in kebumen regency, indonesia. *Golden Rat. Soc. Sci. Educ.* **5**<https://goldenratio.id/index.php/grsse/index> (2025).
21. Rabiu, A. M. Selective dissemination of information (SDI) as a strategic information dissemination and therapeutic for poverty alleviation. *Aj. Inf. Sci. Technol.* **9**, 34–41. <https://doi.org/10.11648/j.ajist.20250901.14> (2025).
22. Nowfal, S. H. et al. The role of agricultural cooperatives in enhancing credit access, market information, and smart farming among rural farmers. *Res. World Agric. Econ.* **6**, 654–672. <https://doi.org/10.36956/rwae.v6i1.1536> (2025).
23. Abebe, A. & Mammo, C. Y. Factors affecting the use of information and communication technologies for cereal marketing in Ethiopia. *J. Agric. Food Inf.* **20**, 59–70 (2019).
24. Jin, Y., Li, G. & Zhang, H. Evaluation of regional rural information environment based on Fuzzy method in the era of the internet of things. *IEEE Access.* **6**, 78530–78541 (2018).
25. Chiachio, M. et al. Plausible Petri nets as self-adaptive expert systems: A tool for infrastructure asset monitoring. *Comput. Aided Civ. Infrastruct. Eng.* **34**, 281–298 (2019).
26. Gupta, D., Richharia, B. & Borah, P. A fuzzy twin support vector machine based on information entropy for class imbalance learning. *Neural Comput. Appl.* **31**, 7153–7164 (2019).
27. Ievoli, C. et al. Information and communication infrastructures and new business models in rural areas: The case of molise region in Italy. *Eur. Countrys.* **11**, 475–496 (2019).
28. Lessanibahri, S., Gastaldi, L. & Fernández, C. G. A novel pruning algorithm for mining long and maximum length frequent itemsets. *Expert. Syst. Appl.* **142**, 113004 (2020).
29. Yang, A. et al. Research on feature extraction of tumor image based on convolutional neural network. *IEEE Access.* **7**, 24204–24213 (2019).

Author contributions

Yanfeng Jin: Writing – original draft, formal analysis. Jianwei Zhao: Writing – Review and formal analysis. Xu-zhao Li and Jianmin Wu: Writing – review and editing.

Funding

The funding was provided by Shijiazhuang Posts and Telecommunications Technical College Doctor Pre-research Foundation (Grant No. G99993069).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

Neither the manuscript nor any parts of its content are currently under consideration or published in another journal.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent to publish

The participant has consented to the submission of the manuscript to the journal.

Additional information

Correspondence and requests for materials should be addressed to J.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025