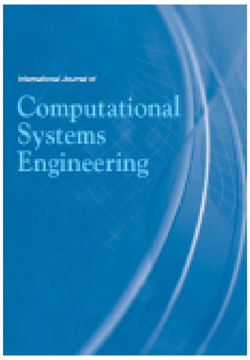


International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

Application value of data mining technology in ultra dense heterogeneous wireless networks

Yuming Zhong, Leyou Chen

DOI: 10.1504/IJCSYSE.2025.10059939

Article History:

Received:	13 July 2023
Last revised:	10 August 2023
Accepted:	19 August 2023
Published online:	18 June 2025

Application value of data mining technology in ultra dense heterogeneous wireless networks

Yuming Zhong*

Educational Technology and Information Center,
Guangzhou Panyu Polytechnic,
Guangzhou, 511483, China
Email: zhongyuming134@163.com
*Corresponding author

Leyou Chen

School of Law,
South China Normal University,
Guangzhou, 510006, China
Email: pc235@163.com

Abstract: In the era of the internet, a large amount of data is constantly generated, which has led to the emergence of network data mining technology. To improve access network security and user network experience, data mining technology is applied to ultra dense heterogeneous wireless networks. A switching algorithm based on user personalised preferences is proposed and a network security prediction module based on data mining is designed. Experimental data shows that when the number of networks is 10,000, the computational time cost based on the multi-attribute vertical switching algorithm is 3.45 ms. The switching algorithm based on user consumption preferences has a computational time cost of 0.97 ms, saving approximately 71.9% of the time. When the number of users exceeds 200, the throughput of the predictive network security switching algorithm based on data mining exceeds that of the analytic hierarchy process switching algorithm. The blocking rate is lower, which can better achieve balanced network selection and improve user network experience.

Keywords: data mining; ultra dense isomerism; wireless network; user preferences; network security.

Reference to this paper should be made as follows: Zhong, Y. and Chen, L. (2025) 'Application value of data mining technology in ultra dense heterogeneous wireless networks', *Int. J. Computational Systems Engineering*, Vol. 9, No. 10, pp.10–19.

Biographical notes: Yuming Zhong holds a Bachelor's degree and is an Intermediate Laboratory Technician. He graduated from the Computer Science and Technology Program at South China University of Technology and currently works at the Education Technology and Information Center of Panyu Vocational and Technical College in Guangzhou. He is engaged in research on computer technology, network information security, and management technology.

Leyou Chen is a Master's degree candidate, Asset Preservation Manager, graduated from South China Normal University with a major in Civil and Commercial Law, and engaged in financial technology research and non-performing asset recovery and preservation work.

1 Introduction

With the progress of technology and human development, communication technology has undergone tremendous changes (Kim, 2021). There are many constraints on wired signals, such as laying requirements, construction difficulty, and communication costs. This problem has been solved through the joint efforts of researchers, and wireless communication technology has emerged (Yates and Islam, 2022). Mobile communication technology has developed from 1G to 5G, spanning the analogue era, digital era, mobile internet era, and even the internet of things. In the future, the number of online users will increase sharply. Mobile communication technology has developed from 1G to 5G, spanning the analogue era, digital era, mobile internet era, and even the internet of things. In the future, the number of online users will increase sharply. The increase in user throughput requirements promotes the emergence and application of ultra dense heterogeneous networks (Zhang et al., 2022). Among them, 5G ultra dense heterogeneous wireless network (UDHWN) technology is gradually entering the commercial market, making the networking forms more diverse. However, there are differences in the scope and capacity of different wireless access technologies. UDHWN architectures are also rapidly developing, providing high throughput and low-cost services to end users. At the same time, UDHWNs face many challenges while providing services, such as increasing demand from network users (Huerta et al., 2023; Franceschini et al., 2022). With the development of network technology and the increase in data volume, data mining (DM) technology has emerged. Data pre-processing is an important step in DM, providing a foundation for subsequent DM. A basic method of DM is classification, which is mainly used to divide data types, usually including supervised learning and unsupervised learning classifiers. Clustering in DM is mainly used to calculate the similarity between different datasets. Common clustering methods include hierarchical clustering and density clustering. Association mining in DM is a method of mining relationships between different data, represented by the Apriori algorithm. DM technology can be used in e-commerce to predict user needs; for financial risk control, it can be used to predict credit risks, etc.; for healthcare, it can be used to analyse case data and treatment plans. UDHWNs can provide more wireless access options, but current switching algorithms have insufficient user personalised preferences and security considerations. This study is based on DM technology, with the goal of improving user experience. From the perspectives of mining user personalised consumption preferences and predicting network security, corresponding solutions are proposed. Therefore, DM technology has been applied to UDHWNs to improve user network experience and security. The study is divided into four parts. Section 2 is a summary of research related to DM technology. Section 3 is the application design of DM technology in UDHWNs. Section 4 is the application effect verification of DM technology in UDHWNs. Section 5 is a summary of the entire article.

2 Related works

In recent years, DM has attracted great attention from the information industry. For a large amount of data, it needs to be transformed into useful information. The Liu et al. (2021) applied DM to the smart grid. DM technology is applied to predict power load and guide decision-making in power enterprises. Test findings demonstrate that the predicted results of this method have a high similarity to the actual situation. Liu and Zhou (2021) proposed an intelligent manufacturing system evaluation model based on DM. Based on the characteristics of intelligent manufacturing systems, the data contains cross sets, training sets, and testing sets. Neural networks are applied to training. The results showed that the highest accuracy of the training group was 69%. Guo (2021) proposed a web information DM method based on agent crawlers. The architecture of web information search and DM has been constructed. The subject crawler technology is combined to achieve web information search and DM. Test findings demonstrate that this method can obtain various information resources of web pages for network monitoring and management. Hezarkhani (2021) proposed a DM technique based on decision numbers. The number of samples from the Pakham porphyry system was used to separate outlier. DM algorithms are constructed to achieve anomaly prediction of unknown points. Experimental data shows that this method has a smaller sample size for error estimation, which is superior to Bayesian and other methods. Gu (2022) proposed a DM-based model for vocational education model reform, which constructs effective indicators for vocational education model reform. By combining Simhash algorithm and data information gain algorithm, the cleaning and selection of indicator features are achieved. DM methods can achieve DM of effectiveness indicators. Test findings demonstrate that the accuracy of the model can reach 97.2%.

With the growth of the Internet, a large amount of data is constantly generated and accumulated. The mining of useful information from these data has become an important issue. Network big DM technology has emerged, which is one of the current popular technologies. The Shastri and Pandit (2021) proposes a new reshaping technique that combines K-anonymity and K-means algorithms. Privacy protection DM technology is used to achieve privacy protection. According to the test findings, this method can effectively protect data privacy and maintain data diversity. Gong (2021a, 2021b) proposed research on DM based on artificial intelligence inference. Based on the terminology extraction method, a method for deploying a large number of expert system technologies was designed. A framework construction based on main sequence induction was implemented. According to the test findings, the accuracy of this method is high, which can provide more feasible strategy estimation. Wang and Liang (2021) proposed big DM for social wireless network users based on the Python platform. DM of social network users is implemented based on hierarchical clustering calculation. The Python platform is used to perform simulations. The experimental results indicate that this method is effective. Jiang et al. (2020) proposed an anomaly network DM model depending on deep training learning. The enhanced local linear embedding algorithm is applied to extract features. The target network clustering was completed using the K-means algorithm. The parameters are optimised by improving particle swarm optimisation. According to the test findings, this method has good clustering effect, providing support for anomaly network DM. Gong (2021a, 2021b) proposed a low latency deterministic network design depending on DM association. A DM algorithm based on page association rules has constructed. The time probability relationship of continuous requests is used for session recognition. The results indicate that this method can reasonably organise cache space, improve cache hit rate, and reduce network latency.

In summary, many researchers have conducted various studies on DM technologies in different fields, especially network DM. However, the effectiveness of these network DM methods still needs to be improved. Therefore, DM technology is applied in UDHWNs. It is expected to improve users' online experience and network security.

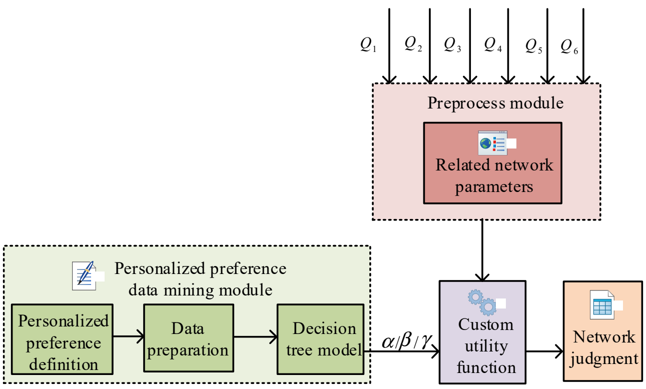
3 Application of DM technology in UDHWN

This chapter applies DM technology to UDHWN. A switching algorithm based on user personalised preferences and a network security prediction module based on DM are proposed to improve network security prediction performance and enhance user network experience.

3.1 Design of switching algorithm based on user personalised preferences

Although traditional information mining algorithms reduce unnecessary switching times, they cannot meet the needs of diverse networks (Saritas et al., 2022). Nowadays, information collection has become convenient. The mining of user historical information and user personalisation tend to enhance the user experience. The switching algorithm process based on user personalised preferences is shown in Figure 1.

Figure 1 Switching algorithm process based on user personalised preferences (see online version for colours)



The five parameters involved in threshold judgment directly affect the quality of business services. The threshold screening criteria are shown in equation (1).

$$Q_n \geq \varepsilon_i^k \tag{1}$$

In equation (1), the network parameters are represented by Q_n. The business type is represented by k. The threshold parameter is represented by ε. The value of i is [1, 2, 3, 4, 5]. The normalisation process of network parameters is shown in equation (2).

$$\begin{cases} C1_{ij}^k = \dfrac{c_{ij}^k - \min c_j^k}{\max c_j^k - \min c_j^k} \\ C2_{ij}^k = \dfrac{\max c_j^k - c_{ij}^k}{\max c_j^k - \min c_j^k} \end{cases} \tag{2}$$

In equation (2), the normalised benefit type parameter is $C1_{ij}^k$. The normalised cost parameter is $C2_{ij}^k$. The j^{th} parameter of the i^{th} network in k-service is c_{ij}^k. The maximum and minimum values of the j in the k business are represented by $\max c_j^k$ and $\min c_j^k$, respectively. The evaluation value of service quality, cost-effectiveness, and price utility value are the key to a custom utility function. The evaluation of network service quality is shown in equation (3).

$$X = P \cdot \left(W_1^k\right)^T = (x_1)_{m\times 1} \tag{3}$$

In equation (3), the network service quality evaluation vector is X. The evaluation matrix is P. The feature vector is W_1^k. The number of candidate networks is m. The network service evaluation value is x_1. The ratio of service quality evaluation to price is cost-effectiveness. In a heterogeneous network environment, users have personalised preferences for consumption. The user's preference for price and performance in online services constitutes their personalised preferences. User personalised preferences are divided into three categories. The first type of personalised preference is the pursuit of ultimate network performance, regardless of consumer price. The second type is to achieve good network performance through a certain consumption cost, which is a comprehensive choice between the two. The third type is to hope for lower consumer prices and lower network performance requirements. The abstract relationship of three types of personalised preferences can be represented by equation (4).

$$\alpha + \beta + \gamma = 1 \tag{4}$$

In equation (4), the preferences for pursuing extreme network performance, higher network performance, and network price sensitivity are represented by α, β, and γ, respectively. The values of all three are [0, 1]. After completing the data preparation work, a decision tree model is designed to calculate the weight of user personalised consumption preferences. The decision tree model is shown in Figure 2.

The potential preferences in a past switch can be reflected through the decision tree. After several switches, the preference weights for the past few days can be obtained accordingly. The network detected by the terminal decision is E. The objective network performance rating threshold is ϑ_1. The new set formed by network nodes above the threshold is I. The potential number of times users pursue ultimate network performance is v. The constraints of each set and variable are shown in equation (5).

$$E = \left\{N_c^f \,\middle|\, f = 1, 2, \ldots, m\right\} \tag{5}$$

In equation (5), the objective network performance rating number is f. The network identifier is c. The elements quantity in set I is shown in equation (6).

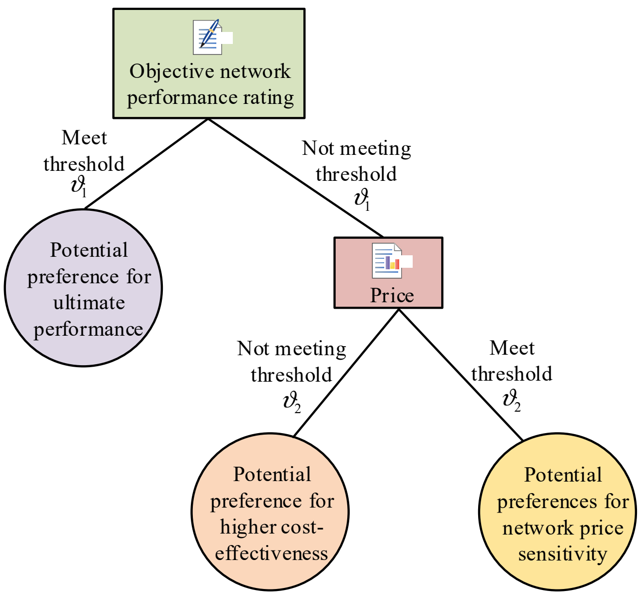
$$|I| = \lceil \mu \cdot |E| \rceil, I \subset E \tag{6}$$

In equation (6), the elements in set I is $|I|$. The elements in set E is $|E|$. The extreme rate of network performance is μ.

The objective network performance score values corresponding to set I are shown in equation (7).

$$I = \{ N_c^f \mid f = 1, 2, \ldots, |I| \} \tag{7}$$

Figure 2 Decision tree model (see online version for colours)



The remaining networks that do not meet the threshold are represented by set M. The network price threshold is represented by ϑ_2. The networks in set M that are less than or equal to the threshold ϑ_2 are represented by set L, indicating the set of low-priced networks. The number of times a user's network price sensitive potential preferences are expressed as p. The constraint conditions that each set and variable satisfy are shown in equation (8).

$$\begin{cases} |M| = m - |I| \\ M = \{ N_c^g \mid g = 1, 2, \ldots, |M| \} \end{cases} \tag{8}$$

In equation (8), the descending order of network prices is g. The elements in set M is expressed as $|M|$. The calculation of the elements in set L is shown in equation (9).

$$|L| = \lceil \delta \cdot |M| \rceil, \; L \subset E \tag{9}$$

In equation (9), the elements in set L is expressed as $|L|$. The price sensitivity is δ. The number of potential preferences for users seeking higher network cost-effectiveness is q. The above values are obtained through decision tree model analysis. The constraint relationships satisfied by each variable are shown in equation (10).

$$\begin{cases} v + p + q = z \\ \alpha = \dfrac{v}{z} \\ \beta = \dfrac{q}{z} \\ \gamma = \dfrac{p}{z} \end{cases} \tag{10}$$

In equation (10), the total number of user preference switches is expressed as z. The custom utility function constructed by network parameters and user personalised consumption preferences is shown in equation (11).

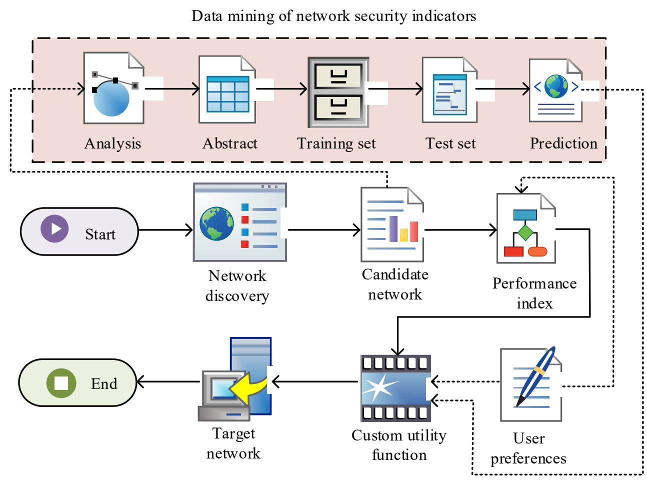
$$y = \alpha x_1 + \beta x_2 + \gamma x_3 \tag{11}$$

In equation (11), the network parameters are represented by x_1, x_2, and x_3. The comprehensive utility value of the user candidate network is represented by y. When making network decisions, the network with the highest comprehensive utility is used as the target network.

3.2 Design of network security prediction module based on DM

The popularisation of UDHWN not only brings convenience to users, but also increases the risk of network access (Jayasri and Aruna, 2022). Figure 3 is the predictive network security switching algorithm based on DM.

Figure 3 Process of predictive network security switching algorithm based on DM (see online version for colours)



After the terminal is connected to the network, candidate networks are screened through the pre-processing module. The personalised consumption weights obtained from DM are used to calculate network performance indicators. Then, the DM module for network security indicators is used for prediction. Finally, a custom utility function is constructed based on various parameters to achieve network switching. By utilising historical data from base stations and access points, a DM module for network security indicators is constructed to predict network security indicators and further enhance user experience.

The first characteristic of pseudo base stations and pseudo access points is their high transmission power, ensuring the reception of signals detected by the user end. The second characteristic is the forgery of identification codes, specifically manifested as IP addresses. The third characteristic is the movement of criminals and intermittent activation (Kaur and Dharni, 2022). Based on the above specific, the attributes of the predicted network security indicator dataset are constructed. Accept strength as one of the attributes of the candidate network, denoted as f_1. By comparing candidate networks, the existence of the same identifier is determined and used as another attribute, represented as f_2. According to the third characteristic, the number of device service locations is taken as the third attribute, denoted as f_3. The fluctuation of the daily service duration of the device, as the fourth attribute, is represented as f_4. The rectangular deviation of the device's start-up service time within the past φ days is used as a measurement indicator. The fourth attribute calculation formula is shown in equation (12).

$$\begin{cases} f_4 = \dfrac{\sum_{i=1}^{\varphi}\left(V_i - \overline{V}\right)^2}{\varphi} \\ \overline{V} = \dfrac{\sum_{i=1}^{\varphi} V_i}{\varphi} \end{cases} \tag{12}$$

In equation (12), the set of daily service hours for the last φ days of the device is represented as V. The average service duration is expressed as $\overline{V}$. The abstract target dataset attribute set is shown in equation (13).

$$F = \{f_1, f_2, f_3, f_4\} \tag{13}$$

In equation (13), the attribute set of the target dataset is represented as F. The labels are assigned to the collected attribute values and legal attribute values to form the training set. The label is the result of online reporting, denoted as ς. $\varsigma = 1$ indicates that the network has been reported. $\varsigma = 0$ indicates that the network is legal. The labels of pseudo base stations and pseudo access points need to be obtained through user feedback and relevant professionals. The labels of legitimate base stations and access points can be obtained through the operator's database. The same number of different label networks are used to construct the training set. The training set formula is shown in equation (14).

$$D_{ts} = \left\{\left(f_{ij}, \varsigma_i\right)\right\}_{i=1, j=1}^{\psi, 4} \tag{14}$$

In equation (14), the training set is represented as D_{ts}. The number of samples in the training set is expressed as ψ. The j^{th} attribute value of the i^{th} sample in the training sample is represented as f_{ij}. The i^{th} sample label in the sample is represented as ς_1. The constraint conditions for the sample labels in the training set are shown in equation (15).

$$\begin{cases} \left(\varsigma_i\right)_{i=1}^{\frac{\psi}{2}} = 1 \\ \left(\varsigma_i\right)_{i=\left(\frac{\psi}{2}+1\right)}^{\psi} = 0 \end{cases} \tag{15}$$

By collecting relevant data from candidate networks, a test set is generated for future network security indicator prediction. The test set formula is shown in equation (16).

$$T_{ts} = \left\{\overline{f}_{i',j}\right\}_{i'=1, j=1}^{m, 4} \tag{16}$$

In equation (16), the test set is represented as T_{is}. The candidate networks represents m. The j^{th} attribute value of the i'^{th} sample is represented as $\overline{f}_{i',j}$. The K-nearest neighbour (KNN) classification algorithm and training set are used to predict the test set. The process of KNN algorithm is shown in Figure 4.

The KNN is a typical example of lazy learning, suitable for situations where data updates are frequent and data representativeness is high (Chen, 2022). The given training set and test set data are normalised. The sample distance calculation is shown in equation (17).

$$d_{i'i} = \sqrt{\sum_{j=1}^{4}\left(f_{ij} - \overline{f}_{i',j}\right)^2} \tag{17}$$

In equation (17), the distance between the i'^{th} test sample and the i^{th} training sample is expressed as $d_{i'i}$. The distance between the i'^{th} test sample and all training samples is expressed as D_{dis}. The KNN set of the test sample is shown in equation (18).

$$D_{knn} = \left\{\left(f_{ij}^k, \varsigma_i^k\right)\right\}_{k=1}^{K}, i \in \{1, 2, \ldots, \psi\} \tag{18}$$

In equation (18), D_{knn} is the KNN set of the test sample. The prediction calculation of network security indicators for test samples is shown in equation (19).

$$\xi = \frac{K - \left(\sum_{k=1}^{K} \varsigma_i^k\right)}{K}, \xi \in [0, 1] \tag{19}$$

In equation (19), ξ is the network security prediction index of the test sample, with a value range of [0, 1]. The number of neighbours in the test sample is K. By repeating the test m times, the network security indicators of all test samples are obtained. The quantitative formula for preference degree and the corresponding highest preference network performance parameters are shown in equation (20).

$$\lambda = \max\{\alpha, \beta, \gamma\} \tag{20}$$

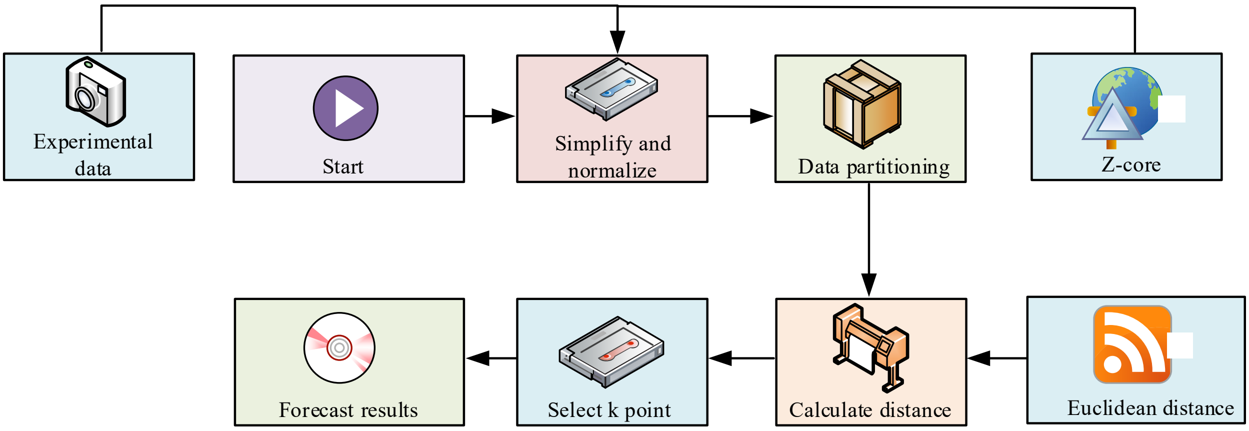
In equation (20), λ represents the qualitative analysis of the user's main consumption preferences. x_0 is the corresponding network performance parameter. It includes three types of network performance parameters with user preferences. The new utility function is shown in equation (21).

$$y = \lambda x_0 + (1 - \lambda)\xi \tag{21}$$

In equation (21), y is the comprehensive utility value of the user on the candidate network. The set of comprehensive utility values and the target network are shown in equation (22).

$$\begin{cases} T = \{y_{i'}\}_{i'=1}^{m} \\ R = \max\{T\} \end{cases} \tag{22}$$

Figure 4 KNN classification algorithm process (see online version for colours)



In equation (22), T is the set of comprehensive utility values for the candidate network. R is the target network.

4 The application effect of DM technology in UDHWNs

This chapter first introduces heterogeneous wireless network environments. The MATLAB platform is used to validate switching algorithms based on user personalised preferences. Then, pseudo access points and pseudo base stations are added to heterogeneous wireless network environments. Based on this experimental environment, the effectiveness of the network security prediction module based on DM is verified.

4.1 Effect of switching algorithm based on user personalised preferences

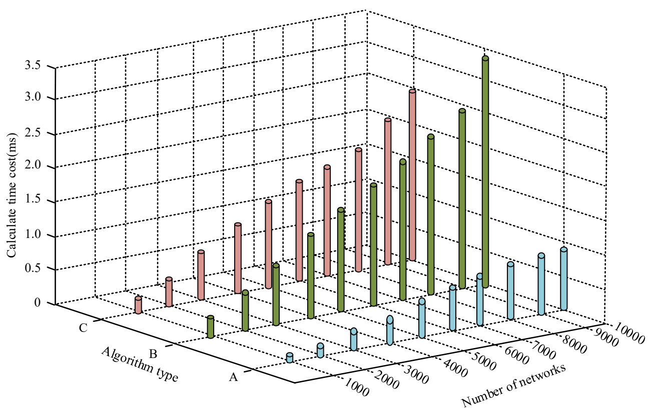
The experiment uses a combination of six access technologies to form a heterogeneous wireless network environment. It includes three WLAN access points, one 4G base station and one 5G microcell. MATLAB is combined for simulation. Table 1 is candidate network parameter values.

Table 1 Candidate network parameter values

Network	*RSS (dBm)*	*Bandwidth (Mbps)*	*Time delay (ms)*	*Shake (ms)*	*Packet loss rate ($\times 10^{-6}$)*	*Cost (¥)*
WLAN1	–60	35	50	20	3	8
WLAN2	–80	15	140	80	30	4
WLAN3	–55	30	65	35	15	5
WLAN4	–50	35	65	45	15	9
WLAN5	–40	15	70	50	20	5
WLAN6	–35	30	80	40	30	3
4G	–65	5	20	5	20	9
5G microcell	–40	100	4	2	1	15

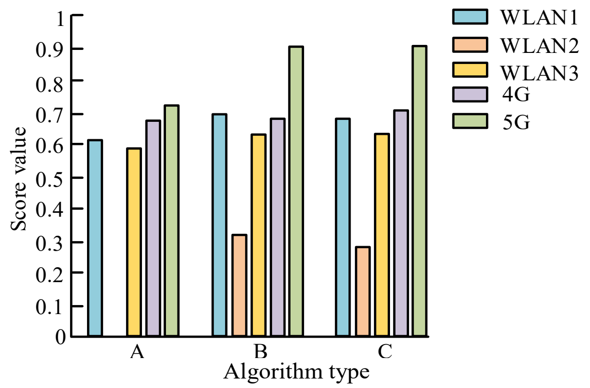
In Table 1, the channel bandwidth represents the maximum data transmission rate that the channel can achieve, which is different from the signal bandwidth, which represents the signal frequency components that can be observed in the signal spectrum. The channel bandwidth of wireless networks is a key factor determining their performance. Choosing the correct frequency band and channel bandwidth can greatly affect the speed and reliability of the network. The 2.4 GHz frequency band provides a longer range and better obstacle penetration ability, but other devices operating on this frequency band are more crowded. The 5 GHz frequency band provides less interference but has a shorter range. WLAN coverage radius is 50 metres, 4G coverage radius is 700 metres, and 5G microcell coverage radius is 200 metres. To better validate the effectiveness of the switching algorithm based on user consumption preferences (marked as A), the multi-attribute vertical switching algorithm (marked as B) and the neural network-based vertical switching algorithm (marked as C) are used as comparative methods. The calculation time cost of each algorithm is shown in Figure 5.

Figure 5 Calculation time cost of each algorithm (see online version for colours)

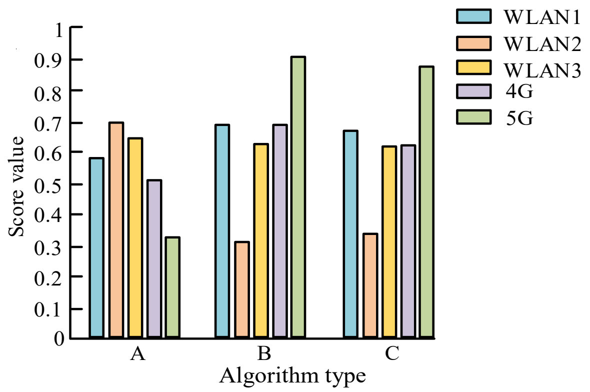


In Figure 5, as the number of networks increases, the computational time cost of each algorithm also gradually increases. The time cost of switching algorithms based on user consumption preferences is relatively low. As the number of networks increases, the computational time cost difference between algorithms gradually increases. When the number of networks is 10,000, the computational time cost of method B is 3.45 ms, while method A is 0.97 ms, saving approximately 71.9% of the time. The switching algorithm based on user consumption preferences can save more computing time. The scoring results of each algorithm on the network for two different users are shown in Figure 6.

Figure 6 The scoring results of each algorithm on the network for two different users, (a) user 1 network score value (b) user 2 network score value (see online version for colours)



(a)

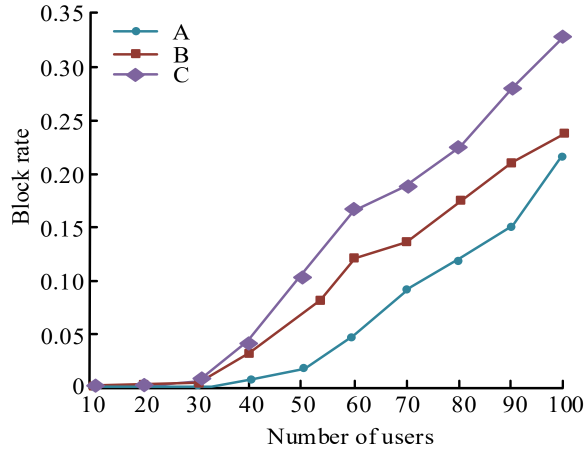


(b)

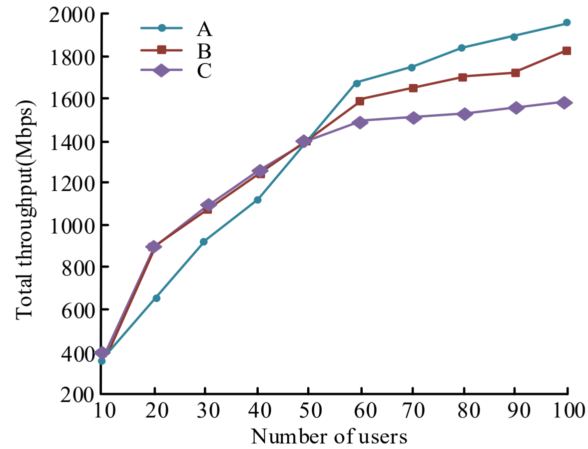
Figure 6(a) shows the network score of user 1. User 1's business is real-time. The pursuit of ultimate network performance is 0.7, the pursuit of high cost-effectiveness is 0.2, and the sensitivity of network price is 0.1. The switching algorithm based on user consumption preferences does not score WLAN 2. The reason is that WLAN 2 has a long latency and does not meet real-time business requirements. It is already excluded before scoring. The highest score for 5G networks is due to users' pursuit of extreme network performance requirements. Figure 6(b) shows the network score of user 2. User 1's business is non-real-time. The pursuit of ultimate network performance is 0.15, the pursuit of high cost-effectiveness is 0.15, and the network price sensitivity is 0.7. The highest score of the multi attribute vertical handover algorithm and the neural network-based vertical handover algorithm is still the 5G microcell. However, users have a high sensitivity to online prices and tend towards cheaper consumer prices. The highest scoring switching algorithm based on user consumption preferences is WLAN 2, which is the cheapest.

Experimental data shows that the scoring effect of switching algorithms based on user consumption preferences is more in line with user needs, and the scoring is more rigorous. The switching blocking rate and throughput of each algorithm are shown in Figure 7.

Figure 7 Algorithm's switching blocking rate and throughput, (a) comparison of blocking rates (b) throughput comparison (see online version for colours)



(a)



(b)

Figure 7(a) shows the comparison of switching blocking rates among various algorithms. As the number of users increases, the blocking rate of each algorithm gradually increases. The multi-attribute vertical switching algorithm and the neural network-based vertical switching algorithm start blocking when the number of users exceeds 30. The switching algorithm based on user consumption preferences is blocked when the number of users exceeds 40. The overall blocking rate curve is the lowest. Experimental data shows that switching algorithms based on user consumption preferences make user access to the network more balanced. It improves the rationality of network resource allocation to reduce blocking rates. Figure 7(b) shows the relationship between the total throughput of each algorithm and the users quantity. When the users are small, the throughput of each algorithm increases faster. The throughput of multi-attribute vertical switching algorithms and neural network-based vertical switching algorithms is greater than that of switching algorithms based on user consumption preferences. The reason is that the first two algorithms do not consider user preferences and tend to favour networks with larger bandwidth. After the users reaches 50, the total throughput of the switching algorithm based on user consumption preferences is greater than the other two algorithms, improving network utilisation.

4.2 Effectiveness of network security prediction module based on DM

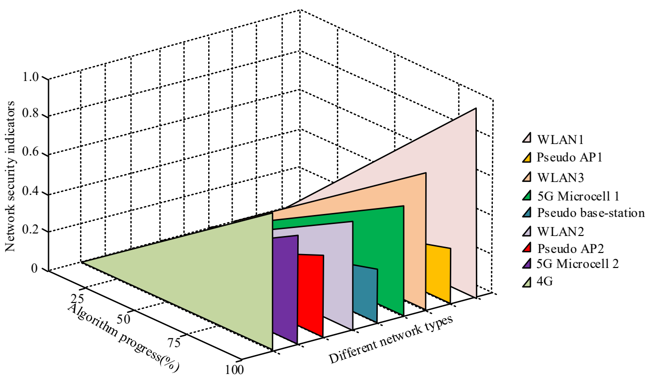
In a heterogeneous wireless network environment formed by the combination of three access technologies, pseudo access points and pseudo base stations are added. Table 2 is relevant parameters of each network.

Table 2 Relevant parameters of each network

Network	*RSS (dBm)*	*Bandwidth (Mbps)*	*Time delay (ms)*	*Shake (ms)*	*Packet loss rate ($\times 10^{-6}$)*	*Cost (¥)*
WLAN1	–60	35	50	20	3	8
WLAN2	–80	15	140	80	30	4
WLAN3	–55	30	65	35	15	5
4G	–65	5	20	5	20	9
5G microcell 1	–40	100	4	2	1	15
5G microcell 2	–20	95	4	2	2	15
Pseudo AP1	–15	30	20	10	10	2
Pseudo AP2	–10	30	20	10	8	5
Pseudo base-station	–15	20	10	6	10	2

The simulation environment includes three WLAN access points, one 4G base station, two 5G microcell, two pseudo access points and one pseudo base station. WLAN coverage radius is 200 metres, 4G coverage radius is 700 metres, and 5G microcell and pseudo base station coverage radius is 150 metres. The KNN is used to predict the aforementioned network. The simulation prediction results are shown in Figure 8.

Figure 8 Network simulation prediction results (see online version for colours)



In Figure 8, the KNN classification algorithm has set a total of six legitimate networks for classification, while the remaining three are set as dangerous networks, which is the same as the simulation scenario settings. This indicates that the KNN algorithm can successfully distinguish network properties. To better validate the effectiveness of the DM-based predictive network security switching algorithm (labelled as A), the analytic hierarchy process switching algorithm (labelled as B) and the service-based neural network switching algorithm (labelled as C) are used as comparative methods. The number of users accessing dangerous networks under each algorithm is shown in Figure 9.

Figure 9 Number of users accessing dangerous networks under each algorithm (see online version for colours)

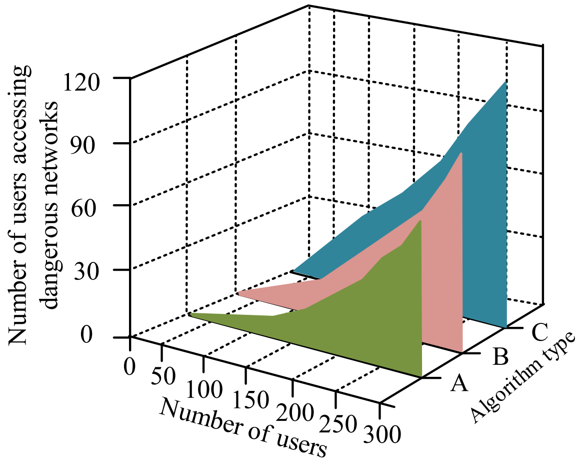
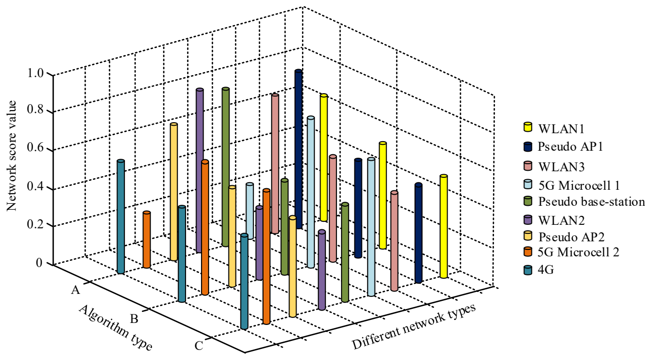
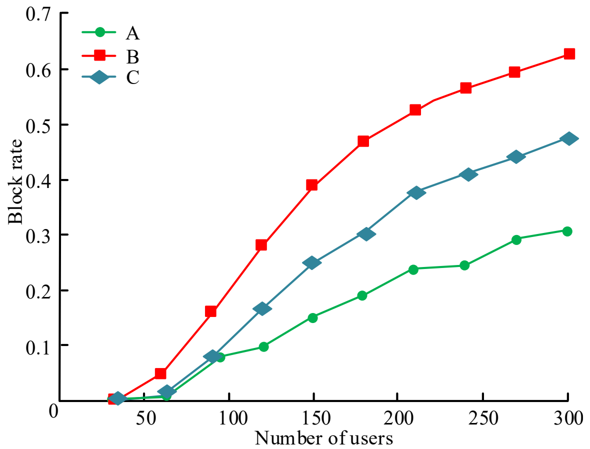


Figure 10 Scoring results of three algorithms on networks (see online version for colours)

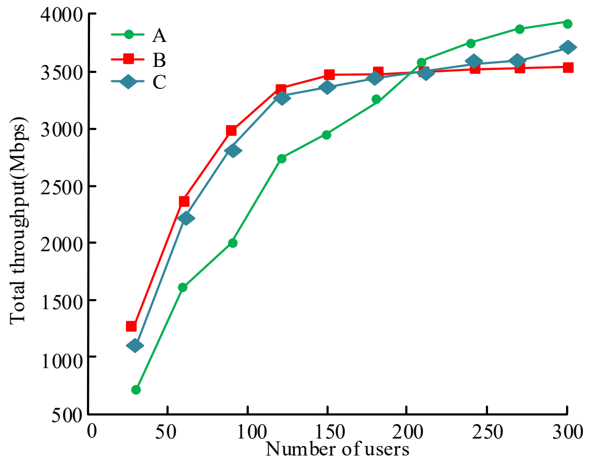


In Figure 9, the number of users accessing dangerous networks for the three algorithms increases with the total number of users. Among them, the service-based neural network switching algorithm has the highest number of access dangerous users. The predictive network security switching algorithm based on DM has the fewest number of access dangerous users. When the users quantity is 3,000, the number of dangerous users accessed by the service-based neural network switching algorithm is 117. The dangerous users quantity accessing the analytic hierarchy process switching algorithm is 93. The dangerous users quantity accessing the predictive network security switching algorithm based on DM is 74. Compared with the previous two, it has decreased by 36.8% and 20.4%, respectively. The user's current business is non-real-time and the price sensitive type is 0.9. The scoring results of the three algorithms on the network are shown in Figure 10.

Figure 11 Blocking rate and throughput of three algorithms, (a) comparison of blocking rates (b) throughput comparison (see online version for colours)



(a)



(b)

In Figure 10, the user is price sensitive, resulting in low scores for two 5G microcell. The prices of pseudo access point 1, pseudo base station, and WLAN 2 are cheap, so the score is higher. Security indicators have been added to the predictive network security switching algorithm based on DM. Therefore, the security performance prediction of WLAN 2 is higher than that of pseudo access point 1 and pseudo base station. The other two algorithms do not consider user preferences and network security indicators, and the highest score is still 5G microcell 1. Experimental data shows that the predictive network security switching algorithm based on DM is more practical, which enhances the users' network experience. The blocking rate and throughput of the three algorithms are shown in Figure 11.

In Figure 11(a), the blocking rates increase with the number of users. Among them, the analytic hierarchy process switching has the highest blocking rate and the fastest growth rate. The prediction network security switching algorithm depending on DM has the smallest blocking rate and the slowest growth. In Figure 11(b), when the users quantity is less than 150, the throughput growth rate of each algorithm is relatively fast. The throughput of the predictive network security switching algorithm based on DM is smaller than the other two algorithms. When the users quantity exceeds 200, the throughput of the DM-based predictive network security switching algorithm exceeds that of the other two algorithms. Low blocking rate makes it easier to achieve balanced network selection.

5 Conclusions

To improve user network experience and network security, DM technology is applied in UDHWNs. A switching algorithm based on user personalised preferences was designed using a decision tree model. A predictive network security switching algorithm based on DM has been proposed. Experimental data shows that when the number of networks is 10,000, the computational time cost based on the multi-attribute vertical switching algorithm is 3.45 ms. The switching algorithm based on user consumption preferences has a computational time cost of 0.97 ms, saving approximately 71.9% of the time. The switching algorithm based on user consumption preferences can save more computing time. The KNN algorithm has setup a total of six legitimate networks, while the remaining three are setup as dangerous networks, which is the same as the simulation scenario settings. This indicates that the KNN can successfully distinguish network properties. When the users quantity is less than 150, the throughput growth rate of each algorithm is faster. The throughput of the predictive network security switching algorithm based on DM is smaller than the other two algorithms. When the users quantity exceeds 200, the throughput of the DM-based predictive network security switching algorithm exceeds that of the other two algorithms. Low blocking rate makes it easier to achieve balanced network selection. The throughput of switching algorithms can be further improved by increasing clock frequency, data channel bandwidth, and multithreading processing. The research results show that the handoff algorithm based on user consumption preference can enable users to access a network that is both consistent with user consumption preference and relatively safe, reduce the constrictivity, improve the total network throughput, and thus improve the user experience. The limitation of this study is that it only considers low-speed mobility in UDHWNs, without considering high-speed or even ultra high-speed mobility. In future research work, further consideration can be given to how vehicle terminals or high-speed trains can improve user experience in high-speed or ultra high-speed mobility.

Acknowledgements

The research is supported by Basic and Applied Basic Research Project of Guangzhou Municipal Science and Technology Bureau in 2021 – Research on Data Mining and Behavioral Portrait Application of Vocational College Students Based on Big Data Technology (202102080252).

References

Chen, Z. (2022) 'Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm', *Journal of Computational and Cognitive Engineering*, Vol. 1, No. 3, pp.103–108.

Franceschini, R., Rosi, A., Catani, F. and Casagli, N. (2022) 'Exploring a landslide inventory created by automated web data mining: the case of Italy', *Landslides*, Vol. 19, No. 4, pp.841–853.

Gong, J. (2021a) 'Design and analysis of low delay deterministic network based on data mining association analysis', *Journal of Web Engineering*, Vol. 20, No. 2, pp.513–532.

Gong, J. (2021b) 'Network data mining based on artificial intelligence inference engine', *Microprocessors and Microsystems*, April, Vol. 82, pp.103794–103798.

Gu, J. (2022) 'An effectiveness model of vocational education mode reform based on data mining', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 32, No. 1, pp.111–127.

Guo, H. (2021) 'Research on web data mining based on topic crawler', *Journal of Web Engineering*, Vol. 20, No. 4, pp.1193–1205.

Hezarkhani, B. (2021) 'Applying data mining techniques for forecasting geochemical anomalies', *Geosystem Engineering*, Vol. 24, No. 3, pp.122–136.

Huerta, C.M., Atahua, A.S., Guerrero, J.V. and Andrade, A.L. (2023) 'Data mining: application of digital marketing in education', *Advances in Mobile Learning Educational Research*, Vol. 3, No. 1, pp.621–629.

Jayasri, N.P. and Aruna, R. (2022) 'Big data analytics in health care by data mining and classification techniques', *ICT Express*, Vol. 8, No. 2, pp.250–257.

Jiang, X., Zhang, H., Xu, J., Wu, W. and Xie, X. (2020) 'Abnormal network data mining model based on deep training learning', *International Journal of Internet Protocol Technology*, Vol. 13, No. 4, pp.228–236.

Kaur, J. and Dharni, K. (2022) 'Application and performance of data mining techniques in stock market: a review', *Intelligent Systems in Accounting, Finance and Management*, Vol. 29, No. 4, pp.219–241.

Kim, M. (2021) 'A data mining framework for financial prediction', *Expert Systems with Applications*, July, Vol. 173, pp.114651–114658.

Liu, X. and Zhou, Q. (2021) 'Intelligent manufacturing system based on data mining algorithm', *International Journal of Grid and Utility Computing*, Vol. 12, No. 4, pp.396–405.

Liu, Y., Wang, G., Zhang, Y., Dong, W., Guo, W., Wang, Y. and Zeng, Z. (2021) 'Power data mining in smart grid environment', *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 2, pp.3169–3175.

Saritas, M.T., Börekci, C. and Demirel, S. (2022) 'Quality assurance in distance education through data mining', *International Journal of Technology in Education and Science*, Vol. 6, No. 3, pp.443–457.

Shastri, M.D. and Pandit, A.A. (2021) 'Remodeling: improved privacy preserving data mining (PPDM)', *International Journal of Information Technology*, Vol. 13, No. 1, pp.131–137.

Wang, K. and Liang, X. (2021) 'Social wireless network user big data mining based on Python platform and hierarchical clustering computing', *International Journal of Networking and Virtual Organisations*, Vol. 25, No. 1, pp.62–82.

Yates, D. and Islam, M.Z. (2022) 'Data mining on smartphones: an introduction and survey', *ACM Computing Surveys*, Vol. 55, No. 5, pp.1–38.

Zhang, M., Fan, J., Sharma, A. and Kukkar, A. (2022) 'Data mining applications in university information management system development', *Journal of Intelligent Systems*, Vol. 31, No. 1, pp.207–220.