Design of Mining Rules and Prediction System for the First Hit Ratio of Artillery

Na Zhou¹, *, Yonghao Li², Qianya Guo⁴ and Xinan Liu²

¹North China Vehicle Research Institute, Beijing 100072, China
²Beijing Institute of Special Vehicles, Beijing 100072, China

*Corresponding author e-mail: znjojo@126.com

Abstract. In order to realize the data-driven cost-effective prediction of the first hit rate of artillery, this paper applies data mining, minHash and other core algorithms and ideas to design a set of artillery first hit ratio correlation with data analysis, storage, analysis, mining and other functions. Rule mining and forecasting system. In this paper, the system architecture, core algorithm and system software of this system are designed. The actual case test of the first hit rate prediction is carried out based on the test data of a certain type of artillery. The analysis of the test results shows that the function of the system is normal and the operation is stable. The average deviation of the predicted test results is 3.08%, which meets the index requirement of the predicted deviation of the first hit ratio of the artillery <5%.

1. Introduction

The first hit rate of the artillery (FRHP) is one of the main performances of the fire control system, directly reflecting the health status of the fire control system, which is related to the survival of the artillery on the battlefield. At this stage, the method of assessing the first hit rate of the fire control system is to conduct a live ammunition shooting experiment, shooting a certain number of bombs to varying degrees, and hitting the hit rate with the hit frequency. The method is: according to the ammunition spread error and different distances provided by the shot table. The required FRHP determines the upper bound of the aiming error of the fire control system. With the aid of the shot table, it is possible to determine or estimate the maximum distance of the FRHP that guarantees that the fire control system is not less than 50%, thereby obtaining the predicted value of the first hit rate. However, there are some problems with this estimation method: the traditional prediction method can only be evaluated for the system, that is, if the subsystems of the vehicle are normal, it cannot be evaluated for the bicycle, and it is necessary to carry out a large number of accurate hit rates. The live ammunition shooting experiment wastes manpower and material resources [1].

In the process of system testing, vehicle task operation, etc., a large amount of information data from sensors, control systems and off-board detection systems of various subsystems in the vehicle were recorded. In the past, the vehicle detection level was limited and the multi-type data processing capability was lacking. The impact of this information is only part of the data through simple data analysis to achieve performance testing and fault analysis and positioning work, most of the data cannot be fully utilized. Mainly reflected in the following aspects: 1. Various devices continue to generate a large amount of data, but it is difficult to conduct data analysis. 2. Equipment generates various data, its
types are complex and diverse, which increases the difficulty of unified analysis; 3. Vehicle fault diagnosis system based on equipment data is imperfect; 4. The association rules between data are not clear, and most cases are carried out in an independent manner. Data analysis; 5. The depth of data mining is difficult, and it is difficult to extract useful knowledge from massive data. In view of the above problems, this paper studies the "Team Artillery First Hit Ratio Association Rule Mining and Prediction System", and takes a real vehicle shooting test record data as the analysis object, and designs a set of data analysis, storage, analysis and mining functions. The software system platform uses the basic ideas and basic methods of data mining to improve and optimize the algorithm, finds the relationship between the information to form knowledge, and uses this knowledge to predict the artillery hit rate.

2. Artillery first hit ratio association rule mining and prediction system
The DM (Data Mining) core algorithm was used in the mine artillery first hit ratio association rule mining and prediction system. Finally, a set of data processing logic including data import, data integration, data analysis, data mining, and auxiliary decision making was established. After preliminary data verification, it can be proved that these core algorithms play a big role in assisting decision-making, and the prediction algorithm has high reliability.

The data analysis process of this system is shown in Figure 1.

(1) Setting the information attention interval: manually segmenting the real field of the information parameter to obtain multiple attention intervals for each information.
(2) Raw data import, ETL: Data cleaning.
(3) Information and hit rate association rule search: Apriorism algorithm is used in data mining process;
(4) Hit rate prediction: Using data mining results, the hit rate prediction algorithm based on MinHash algorithm and BP neural network is adopted.

3. System software design

3.1. System software architecture design

Figure 1. Data analysis process diagram.

Figure 2. Software system architecture diagram.
The system consists of two parts: the database and the client. The database is responsible for the data management function. The client is responsible for providing the user with an interactive interface, implementing system logic, and accessing the database.

Database: Responsible for the entry, management and retrieval of system data. The database operating system is SQL Server 2008 R2, and the 3NF rules are strictly adhered to during the data table construction process. The data sheet and view summary design is detailed in the database design section;

Database access module: implements a stable access interface to the database, and opens the driver call interface to the upper module. The database module is at the lowest level of the client, and other modules cannot be referenced to avoid circular references.

Entity definition layer (module): The client description of the table structure in the database.

Database logic layer (module): The logic for database access, including: data insertion, deletion, query, update and other basic operations.

Functional logic layer (module): system form template, main or core function logic implementation;

User level reference: All user interface, whose entry is driven in the function menu of the main form.

3.2. Functional logic layer core algorithm design

3.2.1. Apriorism based hit rate association rule algorithm. Since the algorithm supports hit ratio prediction, the function is to find the association rule of the association combination of the attention interval combination of the information parameter and the artillery hit rate.

The algorithm is designed as follows:

Let \( X \) be any associated combination event of information parameters, \( Y \) be the artillery firing miss event, \( D \) be the set of all sample things, and the degree of association between \( X \) and \( Y \) can be described by support and confidence. The support of \( X \Rightarrow Y \) is \( \text{Support}(X \Rightarrow Y) \), as in equation (1); the confidence \( \text{Confidence}(X \Rightarrow Y) \), as in equation (2).

\[
\text{Support}(X \Rightarrow Y) = \frac{\text{Count}(X \cap Y)}{|D|} \quad (1)
\]

\[
\text{Confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \Rightarrow Y)}{\text{Support}(X)} \quad (2)
\]

Find all the information classification and combination events that satisfy \( X \Rightarrow Y \) minimum support and minimum confidence, that is, find the association rule between the information attention combination and the artillery firing hit rate, wherein the support \( \text{Support}(X \Rightarrow Y) \) reflects the associated combination event \( X \) caused the artillery firing miss. Frequentness; Confidence \( \text{Confidence}(X \Rightarrow Y) \) reflects the degree of credibility of the associated combination event \( X \) causing an artifact missed event.

3.2.2. MinHash-based hit rate prediction algorithm. The MinHash algorithm is used to quickly estimate the similarity of two sets. It is originally used to detect duplicate web pages in search engines, and can also be applied to large-scale clustering problems.

The MinHash algorithm is an algorithm based on the concept of Jaccard coefficients.

The Jaccard coefficient is an algorithm used to calculate similarity. If there are sets \( A, B \), then \( J(A, B) = \frac{|A \cup B|}{|A \cap B|} \). The Jaccard coefficient of the set \( A \) and \( B \) is equal to the ratio of the number of elements jointly owned by \( A \) and \( B \) to the elements jointly owned by \( A \) and \( B \), and the numerical interval is \([0, 1]\).

The MinHash algorithm can be derived as follows: Let \( h(x) \) be a hash function that maps \( x \) to an integer. \( \text{Hmin}(S) \) has the smallest hash worth element after the element in set \( S \) has been hashed by \( h(x) \).

Then for the set \( A, B \), \( \text{hmin} \ A \cap B = \text{hmin}(B) \), the condition is that the element with the smallest hash value in \( A \cup B \) is also in \( A \cap B \). So,
That is, the similarity between the set A and the set B is the probability that the minimum hash values of the sets A and B after the hash is equal.

3.3. Database and its related access interface design
The database is responsible for the entry, management and retrieval of system data. There are three types of database storage, the first one is transaction type data, that is, the fire control system information data; the second is big data warehouse data, which stores the conversion of transaction data for analysis and mining; the third is the expert database. Data, which is based on the data warehouse and manages the associated combination knowledge obtained through data mining analysis. The system has designed the data table, data view, data entity design and database access interface.

(1) The data table includes the vehicle information table (Table 1), the vehicle type table (Table 2), the data sample table (Table 3), the associated combination table, the associated combination sub-table, the associated combination sample, the test vehicle table, and the test data record. 13 data tables such as table, test item information table, parameter partition information table, import rule information table, and data sample temporary table.

(2) The view includes the hit ratio and the associated combined view (Table 4), the associated combined view corresponding to the vehicle (Table 5), the corresponding interval view of the vehicle, the associated combination analysis result view, the associated interval correlation view, and the associated combination correlation 7 views, such as the sex view and the associated composite attribute result view.

(3) The data includes 13 entities such as vehicles (Table 6), vehicle models (Table 7), and data sample information (Table 8).

(4) The database access interface is a function for querying the database, including log, user management, import rules, vehicle model, vehicle information, test project management, attention interval management, association portfolio management, data sample library, associated combination sample library, 13 access interfaces such as association analysis, display column, and underlying access [2].

3.4. Software function and interface design
The system functions are as follows:

(1) System management: maintain stable operation of the system; system management functions include system settings

(2) Basic information management: Through the maintenance and management of users, vehicles and observation parameters, the system can work normally, including user management, import rules, vehicle model, vehicle information, and test items.

(3) Data import to achieve data cleaning function: According to the configured parameters, the binary stream can be cleaned as valid information;

(4) Focus on interval management: By segmenting the parameter interval, the attention interval is specified, and the parameter status is monitored based on this;

(5) Association management: Combine the attention intervals into association combinations, establish a related portfolio knowledge base through long-term observation, analysis and diagnosis, and use the system to manage and maintain the knowledge base, through the combination of the attention interval and the associated combination data. Monitoring, which can quickly discover the associations hidden in the data, and then predict the hit rate;

(6) Data analysis function: Forming a data analysis tool with the function of drilling down and scrolling, enabling users to perform autonomous analysis on massive data; analysis tool functions include test-focus interval analysis, test-association combination analysis, attention interval-test analysis, association combination - test analysis.
(7) Data mining function: Using mature data mining algorithm, mining and exploring data on the basis of a large amount of sample data, finding the attention interval, paying attention to the internal relationship between the interval and the association combination, helping users to establish more scientific and reasonable The analysis system, the final mission rate prediction is more accurate.

The hit rate prediction mining function includes 2-order attention interval association mining, 3-order attention interval association mining, 2-order association combination mining, and 3-order association combination mining.

Figure 3. 2-order attention interval association mining interface diagram.

Figure 4. 3-order attention interval association mining interface diagram.

Figure 5. 2-order association combination association mining interface diagram.

Figure 6. 3-order association combination association mining interface diagram.
(8) Hit rate prediction: Through the analysis of the attention interval and association combination, the hit rate is predicted based on the expert database;

![Hit rate prediction interface diagram.](image)

**Figure 7.** Hit rate prediction interface diagram.

4. Test case

The test data comes from two sets of 6 batches of live ammunition shooting test data recorded on the same day by a certain type of weaponry and artillery. The data includes the hit rate of each batch of shots, and 18 kinds of information of the fire control system during the test period, including the horizontal angle of advance, the vertical angle of advance, the length of the gun mirror, the vertical mirror solver, Gun link solver, gun trunnion solver, gyroscope group angle, gyroscope group speed, turret gyro angle, turret gyro speed, motor code wheel horizontal angle, motor code plate vertical angle, etc. The live shots were shot 7 times per batch, and the hit rate for each batch is 71.43%, 57.14%, 85.71%, 71.43%, 57.14%, 57.14%.

In the prediction test, 18 kinds of information data of one batch are selected as the predicted analysis data, and the batch data and the hit rate data before the shooting time are used as the training data, thereby obtaining the predicted hit ratio and actuality of the batch. The deviation of the hit rate. In this prediction test, since the test was carried out in one day, and the artillery and related systems were not adjusted during the test, the state of the artillery remained basically unchanged. Therefore, the test data was selected to select any batch as the forecast data, and other batches were used as the forecast data. The training data scheme was tested, and the hit rate prediction bias of 6 trials was obtained, and the average deviation was calculated. When the average deviation was < 5%, the prediction scheme considered effective.

| Test | Batch | Actual hit rate | Predicted hit rate | Deviation (real-pre) |
|------|-------|----------------|-------------------|----------------------|
| Test one |
| 1 | 71.43 | 72.41 | -0.98 |
| 2 | 57.14 | 58.18 | -1.04 |
| Test 2 |
| 1 | 85.71 | 71.43 | 14.28 |
| 2 | 71.43 | 72.41 | -0.98 |
| 3 | 57.14 | 57.52 | -0.38 |
| 4 | 57.14 | 57.98 | -0.84 |

From the results of the hit rate prediction case, the results of the 6 predictions are satisfactory 5 times, and the deviation is 14.28%, the average deviation is 3.08%, which meets the requirements of the forecast technical indicators (|mean deviation|<5%)  

5. Conclusion

This paper applies data mining, minHash and other core algorithms and ideas to design a set of artillery first hit ratio association rules mining and prediction system with data analysis, storage, analysis, mining and other functions. In this paper, the system architecture, core algorithm and system software of this system are designed. The actual case test of the first hit rate prediction is carried out based on the test
The analysis of the test results shows that the function of the system is normal and the operation is stable. The average deviation of the predicted test results is 3.08%, which meets the index requirement of the predicted deviation of the first hit ratio of the artillery <5%.

References

[1] Li Xiaolong, Liu Jianying, Wang Qinzhao. Simulation Research on the First Hit Ratio of Tanks. Fire Control & Command Control, 12 (31) (2006) 66-68.

[2] Zhang Changquan. Comprehensive Test and Detection System for Armored Vehicles. Fire Control & Command Control, 3 (5) (1997) 34-40.

[3] Zhu Jingfu, Zhao Bijun, Wang Qinzhao. Modern Tank Fire Control System. Beijing: National Defence Industry Press, 30 (1) (2003) 325-329.