Research on Prediction Method of First Hit Ratio of Artillery Based on Data Driven

Na Zhou¹, *, Yonghao Li², Xinan Liu¹ and Qianya Guo¹

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

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

Abstract. This paper introduces the traditional firing rate (FRHP) assessment method of artillery and its insufficiency, and proposes a data-driven prediction method for the first hit rate of artillery. This method utilizes the control information, sensor information and other data recorded in the previous vehicle training test, and performs a small number of live ammunition shooting test result data to conduct the association of the first hit ratio of the artillery, and based on the association rules mining knowledge to predict the first hit rate of the artillery. This method is applied in the prediction test of the first hit rate of a certain type of weaponry. The test results show that the average deviation of the predicted test results is 3.08%, which meets the index requirement of the forecasting deviation of the first hit ratio of the artillery <5%, and the method is effective.

1. Introduction

The shooting hit rate of the weapon system is directly related to the survivability of weapons and equipment on the battlefield. Therefore, in the development of weapons as a whole, the major weapons-producing countries have placed an important role in improving and developing advanced fire control systems. In recent years, with the continuous improvement of the modernization level of weapons and equipment, more and more electronic devices in the fire control system, more and more diversified performance detection methods, and the amount of information and information types of the system have also increased. The degree of interaction is also getting higher and higher, which brings many difficulties to the fault detection, health management and design of the fire control system, which makes the cause of the single gun artillery hit rate unable to be accurately located, thus affecting the evaluation and prediction of the artillery hit rate.

During the work of the artillery system test, a large amount of information data from the sensors, control systems, and detection systems of each subsystem was recorded. In the past, it was affected by the limitation of detection level and the lack of multi-type data processing capability. Through simple data analysis to achieve performance testing and fault analysis positioning, most of the data cannot be fully utilized. This topic takes the live-fire shooting test record data as the research object, and based on data mining to find the knowledge relationship between the information to form knowledge, and use these knowledge and machine learning algorithms to predict the first hit rate of the artillery.
2. Traditional artillery first hit rate (FRHP) assessment method and its shortcomings

The first hit rate is a measure of the main performance indicators of the fire control system [1]. Due to the influence of various errors, it is impossible to make every bullet, but only how likely it is - that is, a probability hit. FRHP is a function of range $R$, fire control system height $\Delta x$, azimuth aiming deviation $\Delta z$, ammunition dispersion error $W_{T1}$ (high and low), $W_{T2}$ (azimuth) (for convenience, the aiming error of the human or target tracking system is ignored here and other Small error source), i.e.

$$F_{FRHP} = P(R, \Delta x, \Delta z, W_{T1}, W_{T2})$$ (1)

$\Delta x$, $\Delta z$, $W_{T1}$ and $W_{T2}$ all obey the two-dimensional independent Gaussian distribution, and their density functions are respectively

$$f_1(\Delta x, \Delta z) \triangleq \frac{1}{e^{x_1^2+\Delta z^2}} \exp \left(-\frac{(\Delta x - M_{x1})^2}{2e_{x1}^2}\right)$$ (2)

$$f_2(W_{T1}, W_{T2}) \triangleq \frac{1}{e^{y_2^2}} \exp \left(-\frac{(W_{T1} - M_{x1})^2}{2e_{y1}^2}\right)$$ (3)

In the shooting theory, the common intermediate deviation M

$$P(M_x - M < X < M_x + M) = \frac{1}{e^{\sqrt{2\pi}M_{x-M}}} \exp \left(-\frac{(x-M)^2}{2e^2}\right) dx$$ (4)

Easy to prove: $M = 0.6745e$

In the case where $x_1$, $e_{x2}$, $e_{y1}$, and $e_{y2}$ are known, FRHP is completely dependent on $R$, and FRHP is measured by the distance $R^*$ with a hit probability of 50% [2].

The above definition is only to facilitate the comparison of the performance of different systems or to make theoretical calculations. At present, 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 instead of the hit rate. The method is: determining the upper bound of the aiming error of the fire control system according to the ammunition dispersion error provided by the shooting table and the required FRHP at different distances. 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. Both the theoretical projections in 1.1 and the live-fire assessment methods for FRHP assessments and predictions have their shortcomings:

1. Theoretical calculations can only be evaluated for the system, that is, if the subsystems of the weapon equipment are normal, it is impossible to evaluate the individual artillery. Once the vehicle fails or the abnormality causes the fire control shooting elements to be out of the system tolerance, the first hit rate of the artillery cannot be accurately predicted.

2. The traditional theoretical calculation method and the hit rate are the information of the shooting elements, but as the level of modernization of weapons and equipment continues to increase, more and more electronic devices in the fire control system, the degree of interaction between the subsystems is also coming The higher the number, the more knowledge or patterns that have an impact on the hit rate may not be discovered, and the accuracy of the artillery’s first hit rate prediction cannot be guaranteed.

3. Live ammunition assessment method to get the accurate hit rate of a single artillery, a large number of live ammunition shooting experiments are needed, which wastes manpower and material resources.
3. Data-driven artillery first hit rate prediction method

Based on the data-driven prediction method of the first hit rate of the artillery, the recorded data of the previous training test, and a small number of experiments without real-shot shooting can be used to predict the first hit rate of the single artillery. The data-driven artillery first hit rate prediction method uses the core algorithms and ideas of data cleaning, transformation, loading (ETL), data mining (DM) and machine learning (ML), and has been targeted. With the corresponding improvement and optimization, a set of algorithms including data integration, data analysis, data mining, and auxiliary decision making were completed to complete the data processing logic. The forecasting process consists of three phases: raw data import and cleanup, hit rate correlation rule lookup, and hit rate prediction.

3.1. Raw data import and cleaning

The process of importing and cleaning raw data is shown in Figure 1.

![Figure 1. Raw data import and cleaning flow chart.](image)

3.1.1. Raw data conversion. The original data is binary stream data, and only the information recorded in the original data is only part of the information involved in the analysis of the prediction. Therefore, the cleaning flag should be configured to clean the useful data, and the cleaned out useful data needs to be defined by parameter flags and sampled. The data is defined and converted to information available to the system.

3.1.2. Data import. During the data import process, the same sample data is matched to the sample import rule, and the data is parsed according to the configuration, and finally the data is written into the ODS (Operational Data Store) data storage area. The ODS data storage area is a data storage area that is isolated from the data warehouse and is called an operational data storage area. Operational data, also
referred to as transaction data, refers to the generation of data in a trading system, typically such as sales data in a sales system, receipt data in a logistics system, and the like. A typical feature of such systems is the time concept and detail that focuses on the data. This system needs to import data, although it is in the same file, it still has a certain timing relationship. Therefore, it conforms to the basic characteristics of ODS data. During the data import process, the cleaning (E) and conversion (T) processes in the ETL process are emphasized.

3.1.3. Clustering operation. After all the data is imported into the temporary storage area, it needs to be clustered and written to the data warehouse. When the data is clustered, the concept of time series disappears, the individual concept disappears, and the overall concept is powerful. For example, in the supermarket sales data, after the aggregation operation of the fruit sales, the individual customers are not required to purchase the fruit quantity, but the supermarket overall fruit sales are emphasized.

A) One cluster operation

Before applying the aggregation operation, the system has carried out corresponding research on the basic structure of the data. Generally, the data in the same file is the information data collected in a short period of time. In many cases, these are in the same file or adjacent files. A parameter keeps the same value all the time, which makes the clustering operation have the premise of the application. Under this precondition, when the system is in the aggregation operation, the specific value of the parameter is used as the grouping condition, and the frequency of occurrence of the value is counted and recorded in the data warehouse.

The above process greatly reduced the amount of data storage. In one trial, a total of 3,719 transaction-type data was finally integrated into 48 data records into the data warehouse, and the compression ratio was as high as 98.7%. So, this is a very effective way to reduce the data analysis base. These data are closely related to the launch time, so during the data import process, the time format and the fire time can be specified, and the time concept is automatically established during the system cleaning process without manual intervention [1].

B) Secondary clustering operation

In order to further data compression, the system uses a quadratic clustering operation. In order to define the characteristics of the influence of information parameters on the hit rate, the system designed the parameter data interval division function. That is, the parameters are cut in the whole real number field as a number of intervals without gaps. These intervals cover the entire real field of $-\infty$ to $+\infty$, as shown in Figure 3.

![Figure 2. Schematic diagram of the interval of the real field of the parameter.](image)

Then, any imported data must belong to a certain interval. With this feature, in the process of secondary clustering, the specific value is no longer concerned, only the segment to which the data belongs, there is a record, corresponding the number of segment samples is increased by 1. After data compression using such a quadratic clustering method, at most $m*n$ records are generated in each test vehicle in one test task, $m$ is the number of parameters, and $n$ is the number of parameter average intervals.
3.2. Hit rate association rule lookup

The hit rate association rule search is implemented by the Apriori-based hit rate association rule algorithm, and its function is to find the association rule knowledge of the association combination of the interest 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).

\[
\text{Support}(X \Rightarrow Y) = \frac{\text{Count}(X \cap Y)}{|D|} \quad (5)
\]
\[
\text{Confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \Rightarrow Y)}{\text{Support}(X)} \quad (6)
\]

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 \( X \) that caused the artillery firing miss event.

In the algorithm, the value of \( \text{Support}(X) \) is set to the denominator of the associated combination frequency, the \( \text{Support}(X \Rightarrow Y) \) value is the numerator of the associated combination frequency, and \( \text{Confidence}(X \Rightarrow Y) \) is set to the frequency of the association combination.

Observing the selected records, we can see that the two interest intervals always appear at the same time, so when an abnormality occurs, the tendency can be analyzed to see if another abnormality occurs. For the conclusion of the excavation, the number of samples determines the reliability of the conclusion, so a reliable and reliable sample library is very important [2].

The Apriori algorithm has shortcomings, and the amount of data in the mining process is very large. Therefore, only the 2nd and 3rd order mining is performed in this method, and deeper mining is not realized. Such a setting not only improves the efficiency of the algorithm, but also makes the setting reasonable.

3.3. Hit rate prediction

The hit rate prediction is implemented by BP neural network-based algorithm. First, the data is divided into multiple batches such as training subjects or types. Each batch data is used as a data sample. The data samples are divided into training data samples and predicted data samples. There are no training data samples \( A_i (i = 0, 1, \ldots, n) \), the hit probability \( \hat{R}_i (i = 0, 1, \ldots, n) \), and the hit rate of test sample \( B \) is \( \hat{R}_B \).

In this algorithm, BP neural network three-layer neural network model is used to construct and realize \( \hat{R}_B \) calculation. The three-layer neural network is shown in the figure below.
The neural network algorithm is designed as follows:

1. Let input $x=(x_1, x_2, \ldots, x_n)$ be the similarity $Pr(i, B)$ between $A_i \ (i = 0, 1, \ldots, n)$ and $B$, which is obtained by the MinHash algorithm. Let the set of associated combinations in the associated combination expert library contained in the training sample $A_i$ be $\{p_{1i}, p_{2i}, \ldots\}$, and the set of associated combinations in the associated combination expert library included in the test sample $B$ be $\{p_{1B}, p_{2B}, \ldots\}$, then $Pr(i, B)$ can be described as the number of the same elements in $p(A_i)$ and $p(B)$. The ratio to the total number of elements.

$$Pr(i, B) = |p(A_i) \cap p(B)| / |p(A_i) \cup p(B)|$$  \hspace{1cm} (7)

2. Expectation $d = (d_1, d_2, \ldots, d_n)$ is the true hit rate of the training sample. After multiple parameter adjustment tests, the connection weight $\omega_{ih}$ of the output layer and the intermediate layer, the connection weight $\omega_{ho}$ of the output layer and the output layer, and the threshold $b_h$ of each neuron of the hidden layer. The threshold $b_o$ of each neuron in the output layer is set to 0.3, 0.42, 0.35, -0.27, the maximum number of learning times $= 50$, and the calculated accuracy value $\epsilon = 0.05$.

4. Artillery first hit rate prediction test

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 given in Table 1.
Table 1. Basic information table of shooting test

| Test  | Batch | Time             | Bomb amount | Hit rate  |
|-------|-------|------------------|-------------|-----------|
| Test one | 1 | 9:40:03-9:58:11 | 7           | 71.43%    |
|       | 2 | 11:04:28-12:09:27 | 7           | 57.14%    |
| Test 2 | 1 | 12:46:50-14:20:59 | 7           | 85.71%    |
|       | 2 | 15:23:14-15:38:24 | 7           | 71.43%    |
|       | 3 | 16:13:32-17:57:25 | 7           | 57.14%    |
|       | 4 | 18:51:49-18:58:27 | 7           | 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 [3].

Table 2. Hit Ratio Forecast Results Statistics

| 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
In order to solve the problem of predicting the first hit rate of single artillery, this paper proposes a data-driven prediction method for the first hit rate of artillery. This method utilizes the control information, sensor information and other data recorded in the previous vehicle training test, and a small number of live ammunition target test results data, using the core algorithms and ideas of ETL, DM and ML, and targeted the corresponding The improvement and optimization finally established a set of algorithms including data integration, data analysis, data mining, and auxiliary decision-making to complete the data processing logic, realize the association of the first hit ratio of the artillery, and mine the knowledge based on the association rules to achieve the first hit rate of the artillery. Prediction. This method is applied in the prediction test of the first hit rate of a certain type of weaponry. The test results show that the average deviation of the predicted test results is 3.08%, which meets the index requirement of the forecasting deviation of the first hit ratio of the artillery <5%, and the method is effective.

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 Qinzha. Modern Tank Fire Control System. Beijing: National Defence Industry Press, 30 (1) (2003) 325 - 329.