Hierarchical Prediction of Infrared interference Environment based on Mutual Information and Random Forest

Xin Wu¹,a, Youli Wu¹,b*, Bian Chen¹,c, Yuepeng Gan¹,d, Yuxuan Cai¹,e

¹School of Aeronautics Engineering College, Air Force Engineering University, Xi’an, Shanxí, 710038, China
ªe-mail: 1097390663@qq.com, ªe-mail: 1916454941@163.com, ªe-mail: clsh118118@163.com, ªe-mail: 1251704557@qq.com
ªe-mail: kgdgcxy@afeu.cn

Abstract. In order to realize the classification prediction of environment in the process of combat, we set up the classification prediction model. First, we divided data set into blocks because of the big scale, wide range of influencing factors and strong nonlinear relationship among them. The missile hit rate is reflected in the miss distance, clustering each block data to analysis the correlation factors that affect the miss distance. Second, according to the mutual information theory, the strong correlation factors affecting the miss distance are identified from three aspects which are throwing strategies, maneuver mode and the relative situation by calculating the mutual information matrix. Third, the strong correlation factors as the sample attributes of node splitting in the decision tree, a hierarchical prediction model of infrared interference environment is constructed. This method can better identify the strong correlation factors that affect the miss distance, and avoid the adverse impact on the classification prediction due to the difference of correlation factors. The results show that the method is suitable for the processing of miss distance data. The prediction accuracy of infrared interference environment classification is about 98%, and the generalization ability of the model is strong.

1. Introduction

With the continuous upgrading and improvement of infrared countermeasure equipment in countries, using infrared decoys has become the norm of infrared countermeasure with high cost-effectiveness ratio in the process of close combat. Due to infrared decoys, the originally complex air combat environment becomes more complex, so in order to improve the hit rate and reduce miss distance, how to construct a hierarchical prediction model of infrared interference environment and assist pilots to maximize the operational effectiveness of missiles has become a research hotspot in this field.

In reference [1], the author used BP neural network to reasonably distribute the weight of each index to establish an interference classification model, but the generalization ability of the model was poor and it was easy to converge to the local optimum. In reference [2], the author proposed an environmental quantitative modeling method based on missile seeker and guidance control system. At the same time, various machine learning algorithms, such as support vector machine, Bayesian classifier and so on[3-4], were combined with the environmental classification to obtain the hierarchical prediction model under each algorithm. However, infrared environment is complex with large dimensions, these methods had slow convergence speed and poor generalization ability in dealing with this problem.

In this paper, the massive data are obtained from the infrared anti-jamming simulation platform, then the mutual information theory is used to get the correlation factors between miss distance (MD) and
different interference environments, and the test results are taken as the sample attribute of decision tree node splitting. A hierarchical prediction method of infrared interference environment based on mutual information and random forest is proposed.

2. Principles and Methods

Mutual information is the amount of information about another random variable in one random variable. The accuracy rate is reflected in the miss distance, and the characteristics of miss distance and environmental factors just meet the characteristics of mutual information. So the mutual information theory is chosen to deal with this problem. The process of modeling is shown in figure 1.

![Figure 1. The process of modeling](image)

(1) The data is obtained by simulation platform [5]. Based on these data, m groups of these data set, called $V_D$, and interference environmental data set, called $Y_D$, contained l related factors are constructed. The miss distance of $i$ block in $V_D$ can be expressed as $V_i(\alpha, \beta, \gamma)$ ($\alpha, \beta, \gamma$ are the influence factors). According to $V_i(\alpha, \beta, \gamma)$, the subspace in different blocks is clustered, $n$ cluster blocks are obtained, which are $G_k (k=1,2,\ldots,n)$.

(2) Calculating mutual information matrix between $V_i(\alpha, \beta, \gamma)$ and the correlation factors $Y_k$ in $G_k$, then averaging. Finally, obtaining the strong correlation factors of $G_k$ according to average value.

(3) Obtaining training samples $S_k$ based on $G_k$. The strong correlation factors are randomly selected as attribute subsets, and selecting optimal attributes to generate a decision tree. Finally, a hierarchical prediction model of interference environment ($F_k$) is obtained.

2.1 Construction of multi-dimensional evaluation index

Because there are differences in the hitting accuracy under different conditions, different jamming modes are clustered under jamming conditions. Based on the existing evaluation indexes, the multi-dimensional evaluation indexes of miss distance are as follows:

$$V_i = [\alpha_1, \alpha_2, \ldots, \alpha_n; \beta_1, \beta_2, \ldots, \beta_n; \gamma_1, \gamma_2, \ldots, \gamma_n]$$  \hspace{1cm} (1)

Where: $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the throwing strategies; $\beta_1, \beta_2, \ldots, \beta_n$ are the maneuvering modes; $\gamma_1, \gamma_2, \ldots, \gamma_n$ are the relative situations. The contents of index are shown in table 1.

2.2 Cluster analysis of miss distance feature subspace

Clustering the multi-dimensional evaluation index is helpful to analyze the influence factors under different conditions. Aggregating the data with similar feature vectors ($V_i$) into the same class is an
important method to implement this problem. The indexes are composed of throwing strategies \((V_α)\), maneuvering modes \((V_β)\) and relative situation \((V_γ)\), that is, \(V_c = [V_α, V_β, V_γ]\). Among them, the throwing strategy is combined by 9 bait throwing parameters; maneuver modes contain pitch down, zooming, left / right maneuver; according to the conclusion of reference [6], entry angle and missile-target distance have a significant influence on jamming effect and miss distance, so selected them as the main parameters of the relative situation.

| Table 1. The content of multi-dimensional evaluation index of miss distance |
|--------------------------|--------------------------|--------------------------|
| Strategy | Maneuver | Relative situation |
| \(α_1\) - Throwing number | \(β_1\) - Pitch down | \(γ_1\) - Approach angle |
| \(α_2\) - Per time number | \(β_2\) - Zooming | \(γ_2\) - Distance |
| \(α_3\) - Per group number | \(β_3\) - Left motor | |
| \(α_4\) - Throwing timing | \(β_4\) - Right motor | |
| \(α_5\) - Throwing interval | | |
| \(α_6\) - Throwing external | | |
| \(α_7\) - Throwing velocity | | |
| \(α_8\) - Horizontal angle | | |
| \(α_9\) - Vertical angle | | |

\(V_D\) is divided into three subspaces, which are \(L_1, L_2, L_3\), where \(L_j\) corresponds to throwing strategies, \(L_2\) corresponds to maneuvering modes, and \(L_3\) corresponds to relative situation. Finally, the fuzzy C-means algorithm is used to cluster the data. The specific steps are as follows:

(1) \(U\) is a membership matrix with the random number between \([0,1]\), by initializing \(U\) to make it meet the constraint condition. \(u_{ij}\) represents the membership degree of each sample point belonging to each class, and the formula is as follows:

\[
u_{ij} = \frac{1}{\sum_{k=1}^{c} d_{ik}^{2/(m-1)}}
\]

Where \(d\) is the distance function from the sample point to the cluster center.

(2) Calculating clustering centers of the three subspaces:

\[
c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}
\]

(3) Calculating \(U\) and the value function:

\[
J(U,c_1,...,c_r) = \sum_{i=1}^{n} J_i = \sum_{j=1}^{n} \sum_{k=1}^{c} u_{ij}^m d_{ij}^2
\]

(4) Calculating the new \(U\) and return to step (2). Because of the different data dimensions, the distance function is calculated by the standard Euclidean distance.

The three subspaces of \(V_D\) are clustered into \(r\), \(s\) and \(t\) classes, and the membership degrees of each data sample in the subspace can be expressed as follows:

\[
U_{r,s,t} = \begin{bmatrix}
u_{α_1} & \ldots & u_{α_s} & \ldots & u_{α_r}
\vdots & \ddots & \vdots & \ddots & \vdots
u_{β_1} & \ldots & u_{β_s} & \ldots & u_{β_r}
\vdots & \ddots & \vdots & \ddots & \vdots
u_{γ_1} & \ldots & u_{γ_s} & \ldots & u_{γ_r}
\end{bmatrix}
\]

Where: \(u_{α,s} \in [0,1], u_{β, r} \in [0,1], u_{γ, k} \in [0,1]\) and \(\sum_{i=1}^{s} u_{α,s} = 1, \sum_{j=1}^{r} u_{β, r} = 1, \sum_{k=1}^{t} u_{γ, k} = 1\).
In the whole sample space, there are \( n = r \times s \times t \) class, that is, there are \( n \) jammed modes of miss distance, which can be expressed as \( G_j (j = 1,2, \ldots, n) \). The \( U_{\text{max}} \) formed by the maximum membership degree sample data expresses as:

\[
U_{\text{max}} = \begin{bmatrix} u_{a,\text{max}} \\ u_{b,\text{max}} \\ u_{c,\text{max}} \end{bmatrix} = \begin{bmatrix} \max(a_{j1}, a_{j2}, \ldots, a_{jn}) \\ \max(b_{j1}, b_{j2}, \ldots, b_{jn}) \\ \max(c_{j1}, c_{j2}, \ldots, c_{jn}) \end{bmatrix}
\]  

(6)

2.3. Mutual information analysis

Miss distance is regarded as explanatory variable \( X \), and each influencing factor is regarded as condition variable \( Y \). The value of mutual information between \( Y \) and \( X \) reflects the influence of miss distance. The greater the mutual information value, the higher the correlation between them. In order to make the mutual information easier to interpret, the values are transformed into probability distribution intervals and expressed as:

\[
I(X,Y) = \sum_{i=1}^{N_X} \sum_{u=1}^{M_Y} \frac{M_{iu}}{M} \ln \frac{M_{iu}}{M_{i.} M_{..u}}
\]

(7)

Where: \( N_X \) is the number of intervals in \( X \); \( N_Y \) is the number of intervals in \( Y \), and \( M = N_X + N_Y \); \( M_i \) is the number of \( X \) falling in the \( i \) interval; \( P(Y_u) \) is the probability of \( Y \) falling in the \( u \) interval; \( M_{uv} \) is the number of \( Y \) falls in the \( u \) interval while \( X \) falls in the \( v \) interval.

Assuming that \( G_k (k = 1,2, \ldots, n) \) constitutes an explanatory variable data set, \( X_D = (X_1, X_2, \ldots, X_p) \), and the conditional variable data set is \( Y_D = (Y_1, Y_2, \ldots, Y_l) \), the mutual information between the two data sets can be expressed as:

\[
I(x_j, y_j) = \begin{bmatrix} I(X_1, Y_1) & \ldots & I(X_1, Y_l) \\ \vdots & \ddots & \vdots \\ I(X_p, Y_1) & \ldots & I(X_p, Y_l) \end{bmatrix}
\]

(8)

Where: \( k = 1,2, \ldots, n; X_j \in X_D; Y_j \in Y_D \).

The average mutual information between \( Y_j \) and \( X \) is defined as:

\[
\bar{I}(X, Y_j) = \frac{1}{p} \sum_{j=1}^{l} I(X, Y_j), j = 1, \ldots, l
\]

(9)

Figure 2. Hierarchical prediction modeling of based on random forest algorithm
2.4. Hierarchical prediction model

After analyzing the strong correlation factors by the previous step, the random forest algorithm is used to construct prediction model as shown in figure 2.

Using the self-help method to extract training samples from the original data set, the extracted samples contain the three data categories mentioned above, and each set of training subsets will generate a decision tree. About 36.8% of the samples will not be selected and they can be used as the test set to estimate the error of the classification of the decision tree.

Each randomly selected training subset will contain $M$ attributes screened by mutual information. According to the ranking results, the strong correlation factors are randomly selected as $m$ attributes. According to the Gini, the optimal attributes are selected to branch the decision tree, where $m < M$, so that each set of training subsets will generate a decision tree. Each decision tree forms random forest.

3. Results and Discussion

3.1. Mutual information analysis result

Regarding the miss distance data as the explanatory variable ($X$), and the interference environmental factors as the condition variable ($Y$), the mutual information matrix between the miss distance and the interference factors is established. Calculating the average mutual information between $X$ and $Y$ by (9), descends the order and selects the interference factors of the first 6 as strong correlation factors. The selected results are shown in Table 2.

| Index | Connected factors | $I_{11}$ | $I_{12}$ | $I_{10}$ | $I_{1}$ | $I_{9}$ | $I_{2}$ |
|-------|-------------------|---------|---------|---------|-------|-------|---------|
| Mean mutual | 0.4916 | 0.1993 | 0.1756 | 0.028 | 0.0193 | 0.0154 |

As can be seen from table 2, the top six correlation factors are: projectile-target distance ($I_{11}$), target maneuver mode ($I_{12}$), entry angle ($I_{10}$), bait throwing time ($I_{1}$), bait vertical throwing angle ($I_{9}$), number of each throw ($I_{2}$).

3.2. Hierarchical prediction

The results of the prediction model are obtained as shown in figure 3.
From figure 3, there are 19331 first level data in the interference environment data set. However, 19140 data are classified into first level, 183 data are divided into the second level, 8 data are divided into the third level, and the proportion of correct classification reaches 99%. There are 24047 second level data in the data set, of which 23789 data are divided into the second level data, and 0 data are divided into the third levels, and the proportion of correctly classified data reaches 99%. There are 6622 third level data, of which 6476 data are divided into the third level, of which 131 data are divided into the first and 15 data are divided into the second, and the proportion of correctly classified data reaches 98%.

Taking the unextracted data set as the test set, the test results are shown in figure 4.

![Test result](image)

(a) Test result  (b) ROC curve

Figure 4. Model test result

From 4(a), it is not difficult to see that when the test set is classified by the established model, the proportion of correctly divided into the first level is 98%, the proportion of correctly divided into the second level is 99%, and divided into the third level is more than 99%. Thus, it can be seen that the prediction model has strong generalization ability and has a good result in dealing with the classification of the miss distance.

At the same time, by observing the ROC curve of the model, the distance between the curve and the pure opportunity line (diagonal) is fast to the maximum value, which shows that the algorithm has a strong ability to identify the miss distance data, and also shows the accuracy of the model.

4. Conclusion

In view of the characteristics of large scale of interference environment data, wide range of influencing factors and strong coupling relationship among influencing factors in the process of infrared countermeasure, this paper studies the hierarchical prediction model of interference environment and draws the following conclusions:

(1) According to the average mutual information value, it is concluded that the interference factors that affect the miss distance are projectile-target distance, target maneuver mode, entry angle, bait throwing time, bait vertical throwing angle and the number of each throw.

(2) Taking the analysis result of mutual information theory as the optimal attribute of node division of decision tree, and letting it be selected randomly, a hierarchical prediction model of disturbance environment based on mutual information and random forest is established.

(3) The established is tested, and the results show that the correct classification accuracy of the model is about 98%, and the distance between ROC and the pure opportunity line reaches the maximum.

The model designed in this paper realizes the data-driven hierarchical prediction of interference
environment, and provides a new idea for the hierarchical prediction of infrared interference environment.

References
[1] Bo D. Research on simulation and grading method of infrared imaging interference for target recognition [D]. Huazhong University of Science and Technology. 2019:43-56.
[2] Niu D Q, Wu Y L, Xu Y and Xu R. Quantification modeling for environmental complexity under point source infrared decoy interference [J]. Infrared and Laser Engineering, 2020,49(02):211-219.
[3] Ge C J, Lu Z F, Hong Z H, Ma C, and et al. Anti-interference capability evaluation based on SVM regression and multi-kernel boosting[J], Aerospace Shanghai, 2019,36(05):94-98.
[4] Yang K, Li S Y, Zhang K, and Niu S S. Research on anti-jamming recognition method of aerial infrared target based on naive bayes classifier[J], Flight Control & Detection, 2019,2(04):62-70.
[5] Xu Y, Fang Y W, Li Y W, et al. Association rule mining for the infrared countermeasure by the pf-growth algorithm [C]//Proceedings of the 37th China Control Conference, 2018:8043-8048.
[6] Xu Y, Fang Y W, Wu Y L, et al. LOS-rate of seeker jump characteristic modelling under IR flare countermeasure [J]. Infrared and Laser Engineering, 2018, 48(5): 0504001.
[7] Wu D. Research on recommendation algorithm based on improved random forest [D]. Shenyang Ligong University,2017:35-50.
[8] Li Y, Qi L and Nie P Y. Random forest algorithm for large-scale data[J]. Statistics and Information Forum, 2020,35(06):24-33.
[9] Yang Z L, Shi W, Chen H X and Wang C P. Air quality forecasting with mutual information and random forests based on big data[J]. Environmental Engineering, 2019,37(03):180-185.
[10] Zhang W, Nie J, Teng S H. Fuzzy decision tree based on mutual information and incremental learning[J].Journal of Jiangxi Normal University (Natural Science), 2014,38(01):89-94.
[11] Niu D Q, Wu Y L, Xu Y, et al. Modeling of anti-jamming effectiveness of infrared air-to-air missile [J/OL]. Beijing University of Aeronautics and Astronautics:1-12[2020-10-21]. https://doi.org/10.13700/j.bh.1001-5965.2020.0334