Complex Product System Risk Assessment Model Based on Bayesian Algorithm Optimized By Task Decomposition

Tao Cui
Administration office, Zhongyuan University of Technology, Zhengzhou 450000, Henan, China

Abstract
In view of accuracy problem of standard Bayesian algorithm in risk assessment model, this paper proposes a kind of optimized one based on task decomposition for complex product system risk assessment. Firstly, chaotic mapping is used to introduce the chaotic states into the optimal variables in artificial fish swarm algorithm, and the ergodicity range of chaotic motion is extended to the range of optimal variables value. Then the searching is optimized by chaotic variables. The task of Bayesian network is decomposed by the improved artificial fish algorithm accompanied by the construction of evaluation function to improve the accuracy. The simulation experiment shows the improved artificial fish swarm algorithm has better convergence and higher accuracy in complex product system risk assessment.

Key words: Risk Assessment, Bayesian Network, Complex Product System, Artificial Fish, Chaotic Mapping, Task Decomposition.

1. INTRODUCTION
The research and development of complex product system has the characteristics of high risk and low success rate. Except for the risks of research and development of general products, the complex product systems have many other risks caused by their own characteristics so that the success rate is far less than the ordinary large-scale manufacturing product development (James, 2013). In the past project development process, the risk analysis and assessment of each stage is a weak link, especially in the research and development stage. Therefore, in the current situation of rapidly increasing complex product systems, it has strong practical significance for this paper to conduct the research on the risk of complex product system.

Domestic risk management of complex product development is still in the initial stage. It is lack of effective risk identification, risk assessment and risk decision-making method for the cost risk and the progress risk in the process of complex product development. The risk awareness is not well rooted in each stage of project management. Feng et al. established the relationship model among three kinds of risks, and developed a multi-layer technology risk simulation model based on design structure matrix from the parallel iteration of complex product development process. The model modified the simulation deviation of the only considered cost risk and schedule risk in the traditional simulation (Lucy, 2013). Through the study of technical maturity assessment method in the weapons and equipment acquisition project by British and American armed forces, Hou Yan proposed technical risk assessment method based on the quality function deployment and technical utilization level, which provided a new idea for theoretical stage risk assessment of the equipment development project. However, this method is not based on historical statistics or scientific analysis model meaning that it is sensitive to subjective factors, so the results of risk assessment will be different due to the different human knowledge level (Mile, 2013). Xu et al. discussed the concept and connotation of the technical risk in the project of weapon equipment development, and put forward a kind of quantitative and qualitative method which is based on the fuzzy evaluation of risk factors according to the complexity and diversity of technical risks (Hu and Zhao, 2013). Liu Yangqing put forward a method to evaluate the risk of weapon equipment development project and a procedure of evaluating this project risk.

A method based on historical data was also proposed to adjust the correlation coefficient between the progress and the cost of weapon equipment development and to derive the conditional distribution form of progress about cost when they follow the normal distribution (Berger and Ghosh, 2013). Tian Xinguang established the progress risk analysis model on the basis of multivariate risk probability and gave the progress risk assessment method on the condition of normal distribution or exponential distribution, setting the totally delayed time limit of each risk factor as the delayed time of the project (Aas, 2013). Qiu Wanguo built a random network schedule risk analysis model using Monte Carlo simulation technology to simulate and analyze the progress of the whole project, and finally solved the project schedule risk (Kadam, 2013). Huang et al. sampled randomly the cost data in the stage of modeling, and built a cost model and estimated the cost in terms of random sample data obtained from traditional cost factor (Migration and Lenk, 2012). Wei Gaole presented the
level risk assessment method of weapon system development based on rough set theory, established a system of cost risk knowledge system, and derived the weights of critical subsystems and each subsystem by related reduction and calculation steps, which provided relevant data for the fee risk control (McNeil and Wendin, 2012). Du Rong et al. introduced the earned value method in the risk control of the ship construction. Through the calculation of the cost error, the cost risk is determined and then the cost of the ship construction is further controlled (Rosenberg and Schuermann, 2012).

In view of the defects of risk evaluation model based on Bayesian algorithm, this paper put forward a model based on task decomposition optimized Bayesian algorithm for complex product system risk assessment. The simulation experiment is conducted to prove the effectiveness of this improved strategy.

2. RISK ASSESSMENT MODEL BASED ON BAYESIAN ALGORITHM

Bayesian algorithm is a kind of classification method among the data mining analysis. Bayesian algorithm is used to predict the unknown speculation based on current data. Supposing there are cases \(C_1, C_2, \ldots, C_N\) among the data set.

The predicted equation for certain structure is followed.

\[
P(x_{N+1} | D, S^b) = \prod_{i=1}^{S} \prod_{j=1}^{D} \frac{N_{ij}^b + N_{ij}}{N_{ij} + N_{ij}^b} P(\theta_i | D, S^b) d\theta_i
\]

And the following is the equation for uncertain structure.

\[
P(x_{N+1} | D, S^b) = \sum_{y} P(x_{N+1} | D, S^b) P(S^b | D)
\]

It is very difficult to calculate the uncertain structure because of the solving of all the possible structures. Therefore, a selective average mode is used commonly.

Simply speaking, if a training sample data is given, we can study from the sample data set and classify the new data into a category without hesitation. It is seen that the abovementioned Bayesian theory is consistent with this task.

\[
s_{\text{MAP}} = \arg\max_{x \in S} P(x_i | x_1, x_2, \ldots, x_n)
\]

It can be expressed with Bayesian equation,

\[
s_{\text{MAP}} = \arg\max_{x \in S} \frac{P(x_1, x_2, \ldots, x_n | s_i)P(s_i)}{P(x_1, x_2, \ldots, x_n)}
\]

\[
= \arg\max_{x \in S} P(x_1, x_2, \ldots, x_n | s_i)
\]

According to the training data, the value of equation (4) is calculated. Actually, each of \(P(s_i)\) is easy to estimate by calculating the frequency of each objective value \(s_i\) in the training data set. However, if there is not a large enough training data set, it is impossible to estimate different \(P(x_1, x_2, \ldots, x_n | s_i)P(s_i)\) cases by this method. The number of these cases may be equal to the product of number of possible cases and possible objective value, which requires considerable training set. In order to obtain a proper estimation, each case must appear several times in instance space.

Bayesian algorithm is based on a simple assumption that if the objective value is given, then the probability of \(x_1, x_2, \ldots, x_n\) is equal to product of the probability of each properties.

\[
P(x_1, x_2, \ldots, x_n | s_i) = \prod_i P(x_i | s_i)
\]

Adding it to equation (4), then the method in Bayesian algorithm is known.

\[
s_{\text{NB}} = \arg\max_{x \in S} P(s_i) \prod_i P(x_i | s_i)
\]

\(s_{\text{NB}}\) is the objective value of Bayesian algorithm. In Bayesian algorithm, the number of different \(P(x_i | s_i)\) estimated from training data set is only the product of number of different objective values and the number of attribute value, which requires less calculation than the prediction of \(P(x_1, x_2, \ldots, x_n | s_i)\) .
To overcome the zero probability problem caused by narrow range of training samples, this paper chooses m-estimate to optimize it so as to make accurate classification. M-estimate can be defined as,

\[ P(m_i | n_i) = \frac{v_i + up}{v + u} \]  

(7)

The number of examples in class \( n_i \) is denoted as \( v \); the number of sample cases with value \( m_i \) is denoted as \( v_i \); the equivalent parameter of sample is expressed as \( u \); the designated parameter by tester in expression is \( p \); if training set is zero, then \( P(m_i | n_i) = p \). Therefore, the prior probability of attribute \( m_i \), recorded in class \( n_i \), is expressed as \( P \). The equivalent number can determine the relationship between probability \( P \) and \( v_i / v \).

In the risk assessment model based on Bayesian algorithm, if the capital return is \( y_i \), \( \mu \) the conditional mean, \( \epsilon_i \) residual term, \( Z \) disturbance term, \( \psi_{i-1} \) the information of all time including \( t-1 \), \( \sigma \) the conditional heteroscedasticity of \( \epsilon \) at \( \psi_{i-1} \), taking \( p = 1 \) and \( q = 1 \), then the model can be described as follows.

\[ y_i = \mu + \epsilon_i \]  

(8)

\[ \epsilon_i = \sigma_i z_i \mid \psi_{i-1} \sim N(0, \sigma_i^2) \]  

(9)

\[ \sigma_i^2 = \alpha_i + \epsilon_{i-1} + \beta \sigma_{i-1} \]  

(10)

\( \alpha_i \) reflects the influence of earlier disturbance on future conditional variance; \( \beta \) reflects the influence of earlier conditional variance on future conditional variance; \( (\alpha_i + \beta) \) reflects the continuity of economic fluctuation. The larger this value is, the stronger the continuity will be.

However, there are a certain of limits when using Bayesian network model, namely the image it constructs must be loop-free and static. The loop-free means the model has no loop and only has mono-lateral relationship, but in reality, the relationship is mutual and hard to be one-way. Static state means the Bayesian network is static without consideration of time, yet the casual relationship is time-related.

3. IMPROVED ARTIFICIAL FISH SWARM ALGORITHM FOR THE TASK DECOMPOSITION OPTIMIZATION OF BAYESIAN ALGORITHM

3.1. Convergence Optimization of Artificial Fish Swarm Algorithm

This paper introduces the idea of chaotic searching into the initialization of fish swarm. The chaotic searching has the characteristic of ergodicity that chaotic motion can experience all the state without repetition following its own law and easily approach the optimal region. This can be used to improve the convergence speed of basic fish swarm algorithm when the fish swarm size is large. Generally speaking, the basic idea is to introduce the chaotic state into the optimization parameters with chaotic mapping and extend the range of ergodicity to the value range of optimization parameters for the following searching optimization.

Logistic mapping is a kind of simple and widely used chaotic system, with following definition.

\[ Z_{i+1} = \mu Z_i (1-Z_i), Z_i \in (0,1) \]  

(11)

Here, \( Z_i \) is the real number series and \( \mu \) is the control parameter of system.

Taking the maximum value of complex function for example, the mathematic model of this optimization problem is described as,

\[ \max f(x) = f(x_1, x_2, ..., x_n), x_i \in [a_i, b_i], i = 1, 2, ..., n \]  

(12)

The steps of chaotic system initialization in artificial fish swarm algorithm are,

1. The maximum iteration number of known chaotic variables is set as \( M \), then corresponding initial chaos sequence \( Z^0 = (Z_1^0, Z_2^0, ..., Z_n^0) \) is obtained where the range of \( Z^0 \) is \((0,1)\) and the fixed point of Logistic mapping is passed.

2. According to \( x_i^0 = a_i + (b_i - a_i) \times Z_i^0 \), \( x^0 = (x_1^0, x_2^0, ..., x_n^0) \) is obtained, \( f^0 = f(x^0) \), and the iteration number \( k = 0 \).
(3) According to Logistic mapping \( Z_{k+1} = \mu Z_k (1-Z_k) \), continuous iteration yield chaotic variables sequence \( Z_i^{k+1} \). Then the function variable sequence \( x_i^{k+1} \) and function fitness \( f(x_i^{k+1}) \) is calculated from \( x_i^{k+1} = a_i + (b_i - a_i) \times Z_i^{k+1} \).

(4) If \( k < M \) and \( f(x_i^{k+1}) \geq f^0 \), then \( x_i^0 = x_i^{k+1} \) and \( f^0 = f(x_i^{k+1}) \). If \( f(x_i^{k+1}) < f^0 \), \( x_i^0 \) and \( f^0 \) keep constant, \( k = k + 1 \). \( Z_i^k = Z_i^{k+1} \) and turn to step (2) until the maximum iteration number is reached namely the chaotic searching ends.

(5) After the searching finishes, a new chaos sequence \( x_i^0 \) is obtained. This sequence is the new initial state of artificial fish swarm by chaotic initialization. The position of fish now is very close to optimal area.

The initial position of fish swarm by chaotic idea doesn’t change the random property of artificial fish swarm algorithm, but also improves the ergodicity of the searching and convergence speed when the fish swarm scale is large. This initialization makes the fish swarm firstly searches the optimization in a larger range and improves the convergence performance.

In the standard artificial fish swarm algorithm, the key problem is to adjust the relationship between global searching and local searching ability. To make the adjustment effective, this paper introduces the adaptive strategy to optimize it. Firstly, the optimal fitness value change rate \( K \) and change variance is defined.

\[
k = \frac{f(t) - f(t-n)}{f(t-n)} \quad (13)
\]

\[
\sigma = D(f(t), f(t-n), f(t-2n)) \quad (14)
\]

Here \( f(t) \) is the optimal fitness value of population at generation \( t \); \( f(t-n) \) is the optimal fitness value of population at \( (t-n) \); \( f(t-2n) \) is the optimal fitness value of population at \( (t-2n) \). Then \( K \) represents the relative change rate of optimal fitness value within \( n \) generation. \( D \) is the change variance among \( 2n \) generation.

The game theory is set at the reference whether it is required to change parameters so as to adaptively adjust the parameters, with following expression.

\[
\begin{cases}
\text{step} = f(\text{step}), \delta = f(\delta) & K \leq \theta, \sigma \leq \phi \\
\text{step} = \text{step}, \delta = \delta & K > \theta, \sigma > \phi
\end{cases} \quad (15)
\]

\( f(\text{step}) \) is the adjustment of step length according to the performance of fish swarm algorithm, usually taking the principle of “big then small”. According to the section optimization theory mentioned above, an early bigger step length is conducive to optimal region searching. A late smaller step length can improve searching accuracy and avoid local convergence; \( f(\delta) \) represents the corresponding adjustment to congestion degree; \( \theta, \phi \) is the evaluation coefficient and dependent on different problems.

3.2. Bayesian Algorithm Based on Improved Artificial Fish Swarm Algorithm

Artificial fish experiences the outside environment and the state of other fish to make better its own state, that is to say, the optimization process of individual variable \( X \), in artificial fish swarm algorithm is an adaptive behavior of artificial fish.

To obtain the optimal Bayesian network structure, each artificial fish variable is randomly initialized as a kind of Bayesian network structure \( \text{dag} \). Then the artificial fish swarm can be represented as,

\[
dags = \{\text{dag}1, \text{dag}2, ..., \text{dag}i, ..., \text{dag}n\} \quad (16)
\]

Where, \( dags \) is equal to \( X \), \( \text{dag}i \) is \( Xi \) and \( \text{dag} \) is the variable to be optimized.

\[
X = [M \cdot M \cdot \text{double}][M \cdot M \cdot \text{double}]...[M \cdot M \cdot \text{double}] \quad (17)
\]

The number of \( [M \cdot M \cdot \text{double}] \) is the scale of artificial fish swarm.

If the scale of artificial fish swarm is \( N \), then it is a \( 1 \times N \) dimension cellular array with \( N \) possible Bayesian network structure \( dags \). Here \( N \) dimension cellular array refers to the \( M \times M \) directional and circle-free image.
Parameter initialization includes the size of artificial fish population $N$, the visible range of each individual fish $\text{Visual}$, maximum step length $\text{step}$, possible maximum time of each random motion $t$, iteration number $k$ and Bayesian network node number $M$.

![Bayesian network structure state obtained from random initialization of artificial fish](image)

**Figure 1.** Bayesian network structure state obtained from random initialization of artificial fish

Bulletin board is used for recording the optimal Bayesian network including structure, nodes order and evaluation values. Bayesian evaluation function is the measure of network structure, namely the fitness value of each artificial fish. Our goal is to obtain the maximum posteriori probability $P(B_s \mid D)$ under the condition of data set $D$.

$$P(B_s \mid D) = P(D \mid B_s) / P(D)$$ \hspace{1cm} (18)

And $P(D)$ is independent of optimized network structure, thus $P(B_s \mid D)$ can be calculated directly. The node of Bayesian network structure is $x_i$, its state number $r_i$, its father node number $q_j$. $N_{ij}$ is the $j$ state number of father node of $i$ th node $X_i$ in the data set; $N_{ijk}$ is the intersection of $k$ state of state $r_i$ of node $X_i$ and $j$ state of corresponding node. The relationship between $N_{ijk}$ and $N_{ij}$ is,

$$N_{ij} = \sum_{k=1}^{s} N_{ijk}$$ \hspace{1cm} (19)

$P(B_s)$ is the prior probability of structure. If it is not given firstly, then it is evenly distributed. The higher value the Bayesian evaluation function gives, the better the fitness between network structure state and data set will be.

$$P(B_s, D) = P(B_s) \prod_{i=1}^{n} \prod_{j=1}^{s} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{s} N_{ijk}!$$ \hspace{1cm} (20)

To acquire the maximum value of Bayesian evaluation, the value of each node should be the maximum.

$$\text{score}(i, pa_i) = \sum_{j=1}^{s} \left( \log \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} + \sum_{k=1}^{s} \log \left( N_{ijk}! \right) \right)$$ \hspace{1cm} (21)

The state and data in the bulletin board each time will be the optimal Bayesian network structure in current system. When the algorithm ends, the data in bulletin board will be the optimal solution of whole algorithm.

**4. SIMULATION EXPERIMENT**

To verify the effectiveness of the proposed algorithm, this paper conducted the simulation experiment. Firstly, typical function is used to test the performance of improved artificial fish swarm algorithm with following expression.

$$F = \frac{\sin^2 \sqrt{x^2 + y^2} - 0.5}{(1 + 0.001(x^2 + y^2))^2} - 0.5$$ \hspace{1cm} (22)
The maximum iteration number is set as 100, \( N = 10 \), \( T = 60 \), \( \delta = 0.618 \), \( visual = 8 \) and \( step = 0.6 \). The optimization searching curve of the test function is plotted.

**Figure 2.** The initial state of artificial fish distribution

![Figure 2](image)

**Figure 3.** The state distribution of artificial fish after 100 iterations

![Figure 3](image)

Then the improved artificial fish algorithm is used to optimize the Bayesian network and construct the risk assessment model for complex product system. 10 practical examples are sampled for the project risk evaluation comparing with the real value with error analysis plotted in Figure 4.

**Figure 4.** Error complex product system risk assessment analysis

![Figure 4](image)

From the simulation experiment, the proposed artificial fish swarm algorithm has better convergence and higher accuracy in complex product system risk assessment analysis based on task decomposition optimized Bayesian algorithm.
5. CONCLUSIONS

With the pavement of technology, complex products are widely used for military use and civil use. The research and development of complex product is paid more and more attention. The development process is commonly decomposed to several independent subsystems in parallel or serial. However, the more or less connection between each other increases the risk of complex product development. In view of the defect of traditional risk assessment model based on Bayesian algorithm, this paper put forward an improved one by task decomposition optimization. The simulation experiment shows this proposed model has better convergence and higher accuracy.

Acknowledgements

This work was supported by National Natural Science Foundation, China (No. 51475290,51075261) and Henan Science and Technology Department Soft Science Project (No. 152400410589).

REFERENCES