The probabilistic reasoning of hierarchical diagnosis decision-making

Yawei Ge¹ and Xiqian Hou²

¹ Assessment Demonstration Research Center, PLA Academy of Military Science, Beijing 100091, China.
² PLA Army.93735, Tianjing 301700, China.

E-mail: vvrues11@163.com

Abstract. Aiming at the problems of information incompleteness, model complexity and conclusion inaccuracy existing in diagnosis decision-making reasoning under uncertain conditions, probability information and its unique attributes are introduced on the basis of ontology reasoning to realize the formal expression of uncertainty in various elements of diagnosis decision-making ontology, and the probability ontology is obtained. On this basis, Probabilistic Graphical Model (PGM)—Dynamical Uncertain Causality Graph (DUCG) is selected to implement complex uncertain causality in hierarchical diagnosis decision-making ontology and compact expression of effective probability reasoning. And the hierarchical diagnosis decision-making ontology based dynamic uncertain diagnosis decision-making causality diagram model is put forward, and its system structure is built. Combining with the ontology concepts, relations, attributes to build, simplify, split, and delete the causality diagram, the chain of reasoning, probability calculation and sorting under the uncertain conditions are realized.

1. Introduction
With the continuous development and expansion of high-tech in complex weapon systems, the performance of weapon systems has been greatly improved, and the ability to perform missions has been greatly enhanced, meanwhile, the complexity of weapon systems has been significantly increased, which is bound to bring many problems, such as the increase of support equipment, cumbersome support processes, and difficulties in obtaining information. The rapid development of sensor technology has led to differences and diversity of knowledge information obtained between different types of systems, and monitoring data information between different levels in the same complex system also shows ambiguity, multi-source and coupling. At the same time, due to the uncertainty of hierarchical diagnosis and maintenance decision-making environment, the hierarchical diagnosis ontology of complex weapon support systems is incomplete, which will directly affect the accuracy of diagnosis decision-making reasoning. The research of knowledge expression and reasoning methods for hierarchical diagnosis and decision-making under uncertain conditions is of great significance.

2. Ontology reasoning
Based on hierarchical diagnosis decision-making ontology model and knowledge base, the semantic web-based ontology reasoning machine can make logical reasoning combining ontology classes,
subclasses, class attributes, inter-class relations, and related instances. The specific hierarchical diagnosis decision-making ontology reasoning process is shown in figure 1.

![Diagram](image)

**Figure 1.** Schematic diagram of hierarchical diagnosis decision-making ontology reasoning process.

Based on logical inference based on Web Ontology Language (OWL) is Description Logic [1]. Hierarchical diagnosis decision-making ontology description logic includes four main components: Concept set, Relation set, Axiom set used for describing domain structure and its reasoning mechanism, and Axiomatic set used for describing individual and its reasoning mechanism. Among them, the domain structure axiom set includes two forms of axiom and theorem, while the individual axiom set includes concept assertions and relation assertions. At the same time, when the description logic system stores the term axioms and assertion axioms, it can also complete the reasoning process related, which is used to discover the hidden knowledge in the knowledge base. In terms of reasoning, ontology description logic is mainly based on its decidability. For a certain type of problems, if and only if there is an algorithm, it can be used to solve any problem in this type and give the correct answer in finite steps [2]. However, when the relevant information in the reasoning environment is incomplete, it will lead to incompleteness of reasoning and be more difficult to guarantee the accuracy of reasoning. Therefore, it is necessary to combine the incompleteness of knowledge expression in the process of hierarchical diagnosis decision-making to find an effective reasoning method applicable to the knowledge ontology of hierarchical diagnostic decision-making under uncertain conditions.

### 3. Probability ontology and DUCG

Based on ontology to expand the knowledge of uncertainty, the classes that represent probability are defined, and the probability is introduced to quantify and process the uncertain and incomplete information in the hierarchical diagnostic decision-making ontology. Then, the unknown or hidden states can be reasonably expressed valid conclusions can be obtained.

Definition 1 (Probability Ontology): Probability ontology is the formal representation of uncertainty in each element of the ontology with the introduction of probability information and its unique attributes based on ontology. The uncertain knowledge in the probability ontology can be formalized as \( \text{Prob}(\text{Predicate}(\text{Subject}, \text{Value})) \), where \( \text{Prob} \) is the probability information of the uncertain knowledge and its value is between 0 and 1; \( \text{Predicate} \) is the assertion axiom contained in the knowledge instance; \( \text{Subject}, \text{Value} \) is the entity and the corresponding attribute value [3].

Considering probability ontology, probabilistic graphical model can be used to represent and process the probability distribution of hierarchical diagnosis decision-making ontology [4]. Probabilistic graphical model is an effective method for solving the complexity and incompleteness of complex system diagnosis modeling and inference, and providing a unified framework for the
representations of complex dependencies among system multivariate and probability inference operations [5]. Frequently-used probabilistic graphical models include Bayesian Networks (BNs), Markov Networks, and probability relationship models. As a classical model, BNs can use Directed Acyclic Graph (DAG) and Conditional Probability-distribution Table (CPT) to quantify the uncertainty of relationships between variables, however, it is only applicable to single-valued cases and not to multi-valued cases. At the same time, confirming a CPT requires a large amount of parameters, and its inference calculation is also an NP-hard problem, which the calculation amount increases exponentially with the parameter scale. This is in contradiction with the facts of small sample data and high experimental costs due to complex structures, diverse functions, and high complexity of complex system. Therefore, a probabilistic graphical model for complex uncertain causality and compact representation of effective probability inference is required.

The dynamic uncertain causality graph (DUCG) [6] was proposed by Professor Zhang Qin in 2012. It is based on the Dynamic Causality Diagram (DCD) [7,8]. Adding conditional connection events (conditional action events) and default events to form a new intelligent system framework, the purpose is to concisely and intuitively express the uncertain causality knowledge, to provide an effective method for probability reasoning, and to make the reasoning results interpretable. The advantages are: ① Explicit and simple representation of complex uncertain cause-effect relationships with graphic symbols; ② Separately construct DUCG sub-modules of different domain knowledge, and on the basis of ensuring that the definitions of same variable are unchanged, fuse each sub-module to facilitate the DUCG knowledge base ③ Add conditional connection events (applicable to single assignments) and conditional events (applicable to multiple assignments), which can significantly reduce the scale of observation evidence problems, and simplify the cause-effect diagram to get a higher level of reasoning efficiency.

DUCG is a directed graph, which can be represented by a four-tuple \(\{X,B,D,G\} \). At the same time, DUCG contains a series of nodes and directed arcs. The schematic diagram and the specific meanings of related variable symbols are shown in figure 2 and table 1.

![Figure 2. The schematic diagram of DUCG.](image)

**Table 1.** Variable definitions in DUCG.

| Variable symbol | Variable definitions |
|-----------------|----------------------|
| \(X_n\)         | It is used to represent the data or effect variables that can be monitored in DUCG, which are generally represented by circular symbols, where \(n\) is the variable number. |
| \(B_i\)         | It is used to represent the dependent variable in DUCG, which is generally represented by a rectangular symbol, where \(i\) is its variable number. |
| \(G_n\)         | It is used to represent the logic gate variables in DUCG. It can logically combine the states of the input variables into one state, thereby simplifying the expression of DUCG, where \(n\) is the variable number; |
| \(D_o\)         | It is used to indicate the default event in DUCG, which \(X_n\) represents the unknown reason. The specific symbol is shown in the figure. |
The variable $V_i \in \{X, B, D, G\}$, its corresponding state event is recorded as $V_{i,j}$, which $j$ represents the state $V_i$ is in, and can also be recorded as $V_i$.

$r_{a,j}$

It is used to indicate the causality between parent $V_i$ and child variables $X_n$,

$$r_a = \sum_j r_{a,j},$$

where ";" is used to separate the parent and child variables.

It is used to indicate the directed arc in DUCG, that is, the causal event of weights between variables, where $F_{a,k,i,j} = \left(\frac{r_{a,j}}{r_a}\right)A_{a,k,i,j}$, $r_{a,j}/r_a$ is the weight coefficient bound to.

$A_{a,k,i,j}$

It is used to define the event of this random action $X_{n,k}$ caused by independence $V_{i,j}$, where $k$ represents the state $X_n$ is in, and can also be recorded as $A_{a,k,i,j}$.

$a_{a,j}$

It is used to quantify the uncertainty in the action event, where

$$a_{a,k,i,j} = \text{Pr}\{A_{a,k,i,j}\}.$$

For the complex system, its hierarchical diagnosis decision-making ontology has a lot of knowledge and complex relationships, and there must be a large number of single and multiple assignments. Therefore, this article mainly discusses the probability ontology and applies to both single and multiple assignments M-DUCG fusion inference model. The basic idea of M-DUCG is to express the causality relationship of uncertainty between knowledge with relatively independent random events. Let $V_{i}(V \in \{X, B, D, G\})$ represents parent variable of variable $X_n$, and expand the event into a series of "sums of products" according to the meaning of the causality relationships of knowledge, which is:

$$X_{n,k} = \sum_i \sum_k \left(\frac{r_{a,i,j}}{r_a}\right)A_{a,k,i,j}V_{i,j}$$  \hspace{1cm} (1)

Where, $j$ represents the state $V_i$ is in, and the meanings of other symbols are shown in the table. An intuitive explanation of the meaning of logical events is shown in figure 3.

**Figure 3.** Expansion diagram of DUCG causality under multiple assignments.
In M-DUCG, the parent variable is related due to the existence of the weighting system \( \left( r_{a,j}, r_s \right) \), the parameter \( a_{n,k,i,j} \) and \( r_{a,j} \) are given independently. In this way, M-DUCG skillfully solves the contradiction between “the state of the sub-variable in the compact expression is related by mutual exclusion” and “independent expression of causality relationships between each parent variable and child variable” [9], which makes it possible to achieve a concise expression of CPT in the case of multiple assignments.

In M-DUCG, the meaning of the parameter \( a_{n,k,i,j} \) is different from the conditional probability in probability theory, that is \( a_{n,k,i,j} \neq \Pr \left\{ X_{n,k} \mid V_{i,j} \right\} \), where the causality parameter is used by DUCG to express and quantify the uncertain causality, not the traditional CPT conditional probability values. DUCG satisfies the automatic normalization of variable states and the self-dependence of chained reasoning [10], so that operators need only pay attentions to significant diagnosis events and give corresponding causality parameters in the process of making diagnosis decisions. This means that a certain of parameters in DUCG can be missing without affecting the accuracy of diagnosis reasoning. In this way, DUCG can achieve causality expressions of incomplete knowledge, and then meet the requirements for adequacy and separability of uncertain knowledge expressions [10]. For incomplete historical fault data and online observation data of complex systems, causality modeling of multi-valued uncertain knowledge can still be performed, and the hierarchical diagnosis decision-making reasoning process under incomplete conditions can be realized.

4. Hierarchical diagnostic decision-making OntoDUCG architecture

Based on the analysis of a series of uncertain factors such as hierarchical diagnosis decision-making knowledge expression, knowledge fusion, etc., the probability ontology and DUCG are combined to propose a dynamic uncertainty causality diagram model based on the hierarchical diagnosis decision-making ontology (OntoDUCG). Through building the architecture, the extension of hierarchical diagnosis decision ontology to a dynamic uncertainty diagnosis decision-making causality graph model is achieved.

For the problem of diagnostic reasoning, we need to solve the posterior probability \( \Pr \left\{ H_{k,j} \mid E \right\} \) of a certain fault hypothesis based on specific evidence, where \( H_{k,j} \) represents the hypothetical event or event expression, \( k \) represents the combination of variables in \( H_{k,j} \), \( j \) represents the state combination of variables, and \( E \) represents all the observed set of data, \( E \equiv \bigcap_{i} V_{i,h} \) \( V_{i,h} \subseteq \{ X, B \} \).

Combining with the posterior probability, the expression of the conditional state probability \( h_{k,j}^{*} \) of the hypothetical event under the given evidence \( H_{k,j} \) is given:

\[
h_{k,j}^{*} = \frac{\Pr \left\{ H_{k,j} \mid E \right\}}{\Pr \left\{ E \right\}} = \frac{\Pr \left\{ H_{k,j} \bigcap V_{i,h} \right\}}{\Pr \left\{ \bigcap_{i} V_{i,h} \right\}} \tag{2}
\]

Combining hierarchical diagnosis decision-making ontology with DUCG, a dynamic uncertain causality diagram model based on diagnosis decision-making ontology is proposed for hierarchical diagnosis decision-making reasoning under uncertain conditions. The main reasoning process is divided into three stages:

1. Ontology construction stage.
2. Logic operation stage. The corresponding \( H_{k,j} \mid E \) and \( E \) of hierarchical diagnosis decision-making ontology is expanded according to equation (2) into a "sum of products" expression composed of causality, dependent variable, and random action events leading to failure, and then is combined with logical operations ("and ", " Or ", " not ", " mutually exclusive ", etc.) to achieve inner product
deduplication within the expression and not including between products. In the expression, each "product" process is a causality chain, and the "sum" process is the expansion of all causality chains.

(3) Probability expansion and calculation. Add the ontology instances involved in the causality chain expression obtained in the logical operation to the probability information, add the corresponding parameter values and prior probabilities, and use formula (2) to complete the probability ranking of the fault events in the hierarchical diagnosis decision-making ontology calculation.

The knowledge flow of the specific hierarchical diagnosis decision-making reasoning process is shown in figure 4.

![Figure 4](image)

**Figure 4.** Schematic diagram of the knowledge flow of the hierarchical diagnosis decision process OntoDUCG inference process.

Therefore, we give a schematic diagram of the hierarchical diagnosis decision-making process OntoDUCG system, as shown in figure 5.
As shown in the figure, on the basis of constructing a hierarchical diagnosis decision-making ontology semantic model, four ontology knowledge processing levels from hardware equipment to intelligent decision-making are realized, including the data information involved in the diagnosis decision-making the entire level from system components, process status, diagnosis mode, to maintenance decision-making. The subjective reliability of the diagnosis decision-making ontology at each level is gradually given the in a probabilistic form through DCUG reasoning, and a complete hierarchical diagnosis decision-making DUCG model is constructed. At the same time, under the conditions that the hierarchical diagnosis decision-making semantic model and its related description logic are compatible with each other, the concept, relationship, attribute construction, simplification, splitting, and deletion of the causality diagram are combined with the ontology model, and the obtained uncertain causality diagram is made. The chained reasoning is performed to obtain the logical "sum of products" expansion for the convenience of OntoDUCG-based probability calculation and ranking.

5. Probability reasoning of OntoDUCG

According to the hierarchical diagnosis decision-making OntoDUCG system structure, based on the "knowledge reduction and modularization" uncertainty knowledge expression strategy, the logical reasoning and probability calculation process of OntoDUCG is researched.

5.1. Chained reasoning process of OntoDUCG

The important link of OntoDUCG chained reasoning is how to construct, simplify, and split the causality diagram, so that the domain experts only need to focus on the knowledge of the events with diagnosis significance according to the "self-dependency" of chained reasoning. For incomplete hierarchical diagnosis decision-making knowledge, the use of OntoDUCG chained reasoning does not affect the accuracy of diagnosis decision-making reasoning. The hierarchical diagnosis decision-making OntoDUCG chained reasoning process can be divided into the following sub-processes:

(I) Division of fault state interval

In the process of diagnosis decision-making, it is necessary to correctly distinguish the abnormal working states and normal working states of target equipment. Sometimes, the fault condition can be
directly determined by the abnormal working state of the equipment. However, in the actual process, most failures states need to collect, process and analyse system state data to obtain the relationship between the fault symptoms and faults, and then determine the specific situation of fault state.

In the OntoDUCG model, the division of state intervals is essentially an ontology mapping from the equipment operating state space to the fault state space. In the process of hierarchical diagnosis decision-making, the division of equipment fault status intervals can be completed according to the knowledge device status, fault mode, equipment component, and the interconnection relationships between them in each core domain.

(II) Causality diagram construction, simplification, and split

(1) Construction: According to the division of equipment status intervals in the OntoDUCG model, select the variables \( X_i \) or their sets that can constitute a complete failure mode, and determine their corresponding parent variable sets; use the action variables to connect the parent variables to the child variables \( X'_i \), where logic gates are used to express the parts with complex logical relationships; when there is a sub-variable \( X_i \) for unknown reasons or not being concerned, a default variable \( D_i \) is introduced as the parent variable.

(2) Simplification: The purpose of simplification is to eliminate impossible and meaningless causality relationships and variables in the causality diagram based on the evidence obtained. Reference [6] gives 10 simplified rules. By excluding irrelevant, unnecessary, and unnoticed parts during the simplification process, the problem areas of OntoDUCG model in the process of diagnosis decision-making reasoning can be timely paid attention. The scale of the causality diagram is reduced to the greatest extent, and the computational complexity is greatly reduced under the condition of incomplete information, but the accuracy of diagnostic reasoning is not affected.

(3) Split: NC Rasmussen and others proposed the PRA / PSA (probabilistic risk / safety assessment) method in the "WASH-1400" report, and it is widely used in the field of large and complex systems such as aerospace case evaluation. Among them, according to the PRA / PSA method, in the continuous and stable operations of the system, when the system state changes from a normal state to a fault state, compared with the occurrence probability of only one initial event or no initial event, multiple initial events occur, and the probability is a high-order small value, which confirms the correctness of the assumption that DUCG in the abnormal state can be split without affecting the accuracy of diagnosis decision-making reasoning. Specifically, that is, by assuming different initial events, a large and complex DUCG can be divided into a set of local diagnosis causality diagrams, which are comprehensive and exhaustive with each other. Any sub-diagram is meaningful if and only if the sub-diagram can explain the anomaly evidence obtained, and only if the initial event in a valid sub-diagram can be a candidate cause for anomaly observation. In the diagnosis decision-making reasoning process, all meaningless branches are excluded from consideration. Obviously, splitting can speed up the reasoning process.

(III) Weighted logical reasoning

By performing weighted event expansion and weighted logical operations on the observed knowledge evidence, a meaningful hypothesis space \( sub-DUCG_g \) \((g = 1, 2, \ldots)\) can be obtained.

Evidence \( E \equiv \bigcap_i E_i = \bigcap_j V_{i, j} \) is defined as complete evidence, which \( E_i \equiv V_{i, j} \) is the observed evidence contained in the sub-DUCG. Define abnormal evidence \( E' \equiv \bigcap_i E'_i \) as incomplete evidence and normal evidence as \( E^* \), \( E = E' \cup E^* \). Let \( H_{k, j} \) denote a candidate hypothesis, a possible root cause of the evidence, then the hypothesis space can be defined as \( S_{H_g} = \{H_{k, j} | H_{k, j} \in sub-DUCG_g \} \).

As the event unfolds, the original static logic cycle is destroyed, and logical operations such as AND, OR, XOR, NOT, absorption, exclusion, and addition are implemented. The final logical expression can be obtained by removing the products of exclusive events containing logical expressions and attracting the inclusive events in the expression of the product, and expressing it in the
form of "sum of products" of independent events. The logical expression operation can reduce the redundant probability calculation and further reduce the calculation cost.

5.2. Probability calculation process of OntoDUCG

Based on the construction, simplification and splitting of the causality diagram of the hierarchical diagnosis decision-making in the previous section, and the knowledge logic expansion in the form of "sum of products", we can perform probability calculations based on the hierarchical diagnosis decision-making OntoDUCG reasoning. At present, most qualitative or deductive expert systems usually only include abnormal process information into diagnosis decision-making reasoning, without considering the normal state or the unobserved abnormal state in the diagnosis process. Therefore, in OntoDUCG, normal observations are incorporated into the model as negative evidence for the hypothesis. First, an approximate state probability \( h_{k,j}' \) of \( H_{k,j} \) obtained by using incomplete evidence \( E' \) is used to modify it based on supplementary normal evidence \( E'' \). From this, the accurate state probability \( h_{k,j} \) in the case of complete evidence can be obtained. The calculation process is:

\[
h_{k,j}' = \frac{\Pr\{H_{k,j}|E'\}}{\Pr\{E'\}}
\]

\[
h_{k,j} = \frac{\Pr\{H_{k,j}|E\}}{\Pr\{E\}} = \frac{\Pr\{H_{k,j}\cap V_{i,k}\}}{\Pr\{V_{i,k}\}}
\]

\[
h_{k,j} = h_{k,j}' \cdot \sigma_{k,j} \tag{5}
\]

\[
\sigma_{k,j} = \frac{\Pr\{E''|H_{k,j}E'\}}{\Pr\{E''|E'\}} \tag{6}
\]

Among them, \( \sigma_{k,j} \) represents the correction coefficients of the exact state probability and the approximate state probability.

According to the definition of state probability, we can get \( \sum_{H_{k,j}\in S_{k}} h_{k,j} = 1 \). When only one hypothetical event \( H_{k,j} \) is considered, then can be obtained without calculation \( h_{k,j}' = 1 \). If more than one meaningful \( sub-DUCG_g \) are considered, the local state probability \( h_{k,j}' \) need to be modified to obtain the global state probability \( h_{k,j} \) based on the weights associated with the prior probability in each evidence containing the hypothesis. The probability can be used to quantify the reliability value of whether the hypothetical event is the root cause of the current fault. Therefore, the most possible root cause of the failure is the maximal posterior hypothesis, which can also be called the root cause that can best explain the observed symptoms, that is \( \arg\max_{H_{k,j}}(h_{k,j}') \), where \( H_{k,j} \in S_H \), \( S_H \equiv \bigcup_{g \subseteq \bar{g}} \bigcup_{k \subseteq k_{g}} \) is a complete hypothesis space that brings together all the root causes in meaningful \( sub-DUCG_g \). At the same time, because the hypothetical events involved in logical calculations are limited to the hypothetical events contained in \( S_H \), the influence of the parameter accuracy on the calculation results is low. Therefore, OntoDUCG has a high robustness to the parameter accuracy.

6. Conclusions

Through the analysis of the above-mentioned OntoDUCG probability inference method for hierarchical diagnosis decision-making, the proposed OntoDUCG model has the characteristics of high efficiency, accuracy, which does not depend on the accuracy of parameters. The logical reasoning process guarantees that the accuracy of the reasoning results can still be guaranteed in the case of
incomplete historical data for system diagnosis decision-making and incomplete online data, thereby reducing the scale of causality analysis and the calculation of the hierarchical diagnosis decision-making process for complex systems complexity and computational costs.

References
[1] D.L. Mcquinness. OWL Web Ontology Language Overview [EB/OL] http://www.23.org/TR/2004/REC-owl-features-20040210 2010.06.
[2] R L Lu 1995 Artificial Intelligence Beijing: Science Press
[3] Zhongli Ding and Yun Peng 2004 A Probabilistic Extension to Ontology Language OWL Proceedings of the 37th Hawaii International Conference on System Sciences
[4] D Koller and N Friedman 2009 Probabilistic Graphical Models: Principles and Techniques Cambridge: MIT Press
[5] P Larrañaga and S Moral 2011 Applied Soft Computing 11(2) 1511-28
[6] Q Zhang 2012 Journal of Computer Science and Technology 27(1) 1-23
[7] Q Zhang, C L Dong and Y Cui et al 2014 Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: statistics base, matrix, and application IEEE Transactions on Neural Networks and Learning Systems 25(4)
[8] Zhang Qin 1994 Probabilistic reasoning based on dynamic causality trees/diagrams Reliability Engineering and System Safety 46
[9] A Pfeffer 2001 Sufficiency, separability and temporal probabilistic models In Proc. the 17th Conf. Uncertainty in Artificial Intelligence 08
[10] C L Dong, Q Zhang 2014 Acta Automatica Sinica 40(12) 2766-81