

Postprint: Research on Causal Analysis Methods for Civil Aviation Emergencies Based on Bayesian Networks

Authors: Wang Hong, Guo Xiaodan, Zhu Han

Date: 2018-05-18T00:00:00+00:00

Abstract

To address the issue of ineffective evaluation and correlation analysis of causal relationships in civil aviation emergency events, this paper proposes a causal relationship analysis method for civil aviation emergency events based on Bayesian networks. By introducing Bayesian theory into the domain ontology for civil aviation emergency management, the conversion of concepts, relations, and instances from the domain ontology to Bayesian networks is first realized through rule design. Subsequently, the Bayesian network knowledge synthesis algorithm E-IPFP is adopted to construct the conditional probability tables of Bayesian network nodes, and a message passing mechanism is employed to calculate the probabilistic relationships between parent and child nodes, thereby obtaining the probability distribution of causal relationships in civil aviation emergency events. Utilizing the domain ontology for civil aviation emergency management and cases from the World Civil Aviation Accident Investigation Tracking Report as experimental data, the analysis of causal relationships among civil aviation emergency events is presented, providing methodological support for big data-based correlation analysis and inference of emergency events.

Full Text

Preamble

Title: Research on Causal Analysis Method for Civil Aviation Emergencies Based on Bayesian Networks

Authors: Wang Hong, Guo Xiaodan, Zhu Han

Affiliation: School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

Abstract: To address the challenge of effectively evaluating and analyzing causal relationships in civil aviation emergencies, this paper proposes a causal analysis method based on Bayesian networks. By introducing Bayesian theory into the domain ontology of civil aviation emergency management, the method first implements the transformation of concepts, relations, and instances from the domain ontology into a Bayesian network through rule design. It then employs the Bayesian network knowledge synthesis algorithm E-IPFP to construct conditional probability tables for the network nodes, and calculates probabilistic relationships between parent and child nodes via a message passing mechanism to obtain the probability distribution of causal relationships in civil aviation emergencies. Using the civil aviation emergency management domain ontology and cases from the World Civil Aviation Accident Investigation Tracking Report as experimental data, the paper presents an analysis of causal relationships between civil aviation emergencies, providing methodological support for big-data-driven correlation analysis and reasoning of emergency events.

Keywords: civil aviation emergencies; domain ontology; Bayesian networks; causality; conditional probability

0 Introduction

Bayesian networks (BN), first proposed by Judea Pearl in 1985, are graphical models for probabilistic reasoning that serve as powerful tools for uncertain knowledge inference and data analysis. Civil aviation emergency management encompasses a series of planned and organized processes for responding to civil aviation emergencies, with the primary objective of effectively preventing and handling various emergencies to minimize negative impacts, casualties, and property losses. Analysis of causal relationships between events plays a crucial role in the prevention and response phases of civil aviation emergency management. Current research on domain ontology-based civil aviation emergency management primarily focuses on ontology construction, rule-based emergency plan reasoning using SWRL and description logic, and similar case retrieval, while lacking investigation into causal relationship analysis and reasoning based on domain ontologies.

This paper integrates the E-IPFP algorithm with the civil aviation emergency management domain ontology, designing a set of rules to transform the domain ontology into a Bayesian network, thereby achieving the conversion between ontology and Bayesian network. Building upon evidence node probabilities of emergency causal relationships, the method employs the E-IPFP algorithm combined with a message passing mechanism to further calculate posterior probabilities of nodes, thereby obtaining the probabilistic distribution of causal relationships in emergencies. This approach aims to solve the problem of correlation analysis and reasoning for civil aviation emergency causal relationships based on domain ontologies, providing better data support for emergency prevention

and response.

1 Research Approach

The analytical framework for civil aviation emergency causal relationships based on Bayesian networks is illustrated in [Figure 1: see original paper]. The process consists of three main phases:

- a) **Data Preprocessing:** The civil aviation emergency management domain ontology is parsed into an RDF graph, and probability attributes are assigned to nodes in the RDF graph based on expert analysis from the World Civil Aviation Accident Investigation Tracking Information Report to generate node probability files.
 - b) **Bayesian Network Conversion:** Rules are designed to transform the domain ontology RDF graph into a Bayesian network (BN), and the Bayesian knowledge synthesis algorithm E-IPFP is employed to achieve probability synthesis, generating the domain ontology Bayesian network BN_O.
 - c) **Event Causal Relationship Analysis:** Bayesian network inference based on message passing mechanisms is used to propagate probability values of event causal relationships, obtaining the probabilistic distribution of event causality.
-

2 BN Construction Method Based on Domain Ontology

A Bayesian network comprises a directed acyclic graph (DAG) and a set of conditional probability tables (CPT). The DAG is represented as $G = (V, E)$, where V denotes the set of nodes (random variables) and E represents the set of directed edges indicating relationships between variables. The CPT collection is denoted as $P = \{P_i\}$, where each node corresponds to a conditional probability table that quantifies the degree of relationship between a node variable and its parent nodes. If node V_i directly influences node V_{i+1} , i.e., $V_i \rightarrow V_{i+1}$, then a directed edge is established between them, with the connection strength (i.e., connection weight) represented by conditional probability $P(V_{i+1}|V_i)$ stored in the conditional probability table.

2.1 Definition of Conversion Rules

A domain ontology describes concepts and relationships between them, including classes, relations, and instances. The domain ontology RDF consists of a series of triples (subject, predicate, object), where subjects and objects represent

concepts C in the ontology, predicates represent relations R , forming the domain ontology RDF graph shown in [Figure 2: see original paper]. The following rules transform the domain ontology RDF graph into a Bayesian network:

Rule 1: For the concept set $\{C_1, C_2, \dots, C_n\}$ in the ontology RDF graph, each concept C_i corresponds to a node V_i in the Bayesian network, with state attributes a_i representing the probability values of True and False states for the concept.

Rule 2: For the relation set $\{R_1, R_2, \dots, R_m\}$ in the ontology, if concept C_i and concept C_{i+1} have relation R_x and C_i directly influences C_{i+1} , then the influence relationship between the two concepts is mapped to a directed edge E_x between nodes V_i and V_{i+1} in the Bayesian network, strictly adhering to hierarchical relationships where upper-level concepts serve as parent nodes of lower-level concepts.

Rule 3: For instance I_j in the ontology, it corresponds to node V_{I_j} in the Bayesian network, with state attributes a_{I_j} representing the probability values of True and False states for the instance.

The mapping process is illustrated in [Figure 2: see original paper].

2.2 BN Node Conditional Probability Table Construction Based on E-IPFP

Bayesian knowledge synthesis algorithms integrate given probabilistic knowledge into Bayesian networks. The fundamental approach is as follows:

- a) Input the joint probability $P^0(X)$ and treat the probabilities in the BN as constraint conditions $\{D_1, D_2, \dots, D_m\}$.
- b) Determine convergence under new constraint conditions. If converged, calculate the current joint probability distribution $P^k(X)$. If not converged, add α parameters to both the constraint set and probabilities, then iterate until convergence.
- c) The iterative calculation continues until convergence is achieved, where X represents random variables in the Bayesian network, n represents the number of variables, $P(V_i|\pi(V_i))$ denotes node conditional probabilities, α is a parameter for constraint conditions, $P^k(X)$ represents the current joint distribution, $Q(Y)$ represents the child node joint probability distribution, $\pi(X)$ is the parent node set of X , D represents new constraint conditions, and m is the number of constraint sets.

2.2.1 E-IPFP Algorithm The E-IPFP algorithm is employed for knowledge synthesis. Let S , R , and G represent the cause, result, and severity level of events as class concepts in the domain ontology, with E representing the event instance set. Based on the mapping rules, the following E-IPFP-based Bayesian knowledge synthesis algorithm performs the conversion:

- (a) Apply Rule 1 to assign state attributes True and False to concept nodes C_i in the RDF graph. Event cause nodes C_i^S are transformed into BN nodes V_i^S , event result nodes C_i^R into BN nodes V_i^R , and event severity nodes C_i^G into BN nodes V_i^G .
- (b) Apply Rule 2 to transform relationships between event instances, causes, and results in the emergency ontology, mapping them to directed edges in the network while strictly maintaining hierarchical relationships where upper-level concepts serve as parent nodes of lower-level concepts.
- (c) Apply Rule 3 to transform event instance nodes I_j in the RDF graph, assign state attribute values, and map them to BN nodes V_{I_j} . The cause and result attributes determine parent-child node relationships, while three data attributes—casualties, economic losses, and aircraft damage degree—determine the connection relationship from node I_j to event severity G_i .

The transformed domain ontology Bayesian network is shown in [Figure 3: see original paper].

2.2.2 BN Node Conditional Probability Table Construction Constructing node conditional probability tables involves converting True and False state values from node probability files in the RDF graph into conditional probability tables in the BN:

- a) When node probability attribute values in the ontology RDF graph are not empty, i.e., when concept C_i directly influences C_{i+1} , they correspond to nodes V_i and V_{i+1} in the BN. The conditional probability of V_{i+1} is $P(a_{i+1}|a_i)$, where a_{i+1} represents $\text{True}(C_{i+1})$ and a_i represents $\text{True}(C_i)$. In the probability file, this is represented as:

```
<owl:Probability rdf:ID="P(ai+1|ai)">
  <hasProposition>ai+1</hasProposition>
  <hasCondition>ai</hasCondition>
  <hasValue>0.63</hasValue>
</owl:Probability>
```

This is converted to conditional probability table format in the Bayesian network. If V_i is the parent node of V_{i+1} , the conditional probability of node V_{i+1} is shown in .

- b) When node probability attribute values in the ontology RDF graph are empty, default values are applied during conversion, i.e., initial state values are set to 0.5 (True=0.5, False=0.5).

2.3 Bayesian Inference Based on Message Passing

Message passing is a process for solving posterior probabilities based on each node's conditional probability and probability values transmitted from adjacent nodes in the node set. Nodes with state True=1 in the network are defined as

evidence nodes U , while others are non-evidence nodes, as shown in [Figure 4: see original paper]. Messages from child node V_i to parent node V_j are called ρ messages, while messages returned from parent to child are called λ messages. Posterior probabilities of all nodes are calculated based on given evidence.

The message passing-based inference algorithm proceeds as follows:

Input: Bayesian network BN_O (network structure and conditional probability table set), evidence node set $U = \{U_1, U_2, \dots, U_n\}$.

Output: Conditional probability distributions of non-evidence nodes V_i' : $P(V_i')$.

Set initial message transmission values for evidence nodes $V_i \in U$:

If non-evidence node V_i' has parent nodes, it receives messages from them and calculates $\rho(V_i')$ using formula (2). If V_i' has child nodes with set $S(V_i')$, formula (3) calculates $\lambda(V_i')$.

If V_i' has both parent and child nodes, calculate $\rho(V_i')$ and $\lambda(V_i')$ separately, then compute the posterior probability $P(V_i'|U)$ using formula (4).

Repeat until no message passing occurs in the entire network, then output the posterior probabilities of nodes.

3 Experiments and Results

The experiment employs the civil aviation emergency management domain ontology and cases from the World Civil Aviation Accident Investigation Tracking Information Report between 2014-2016, setting state attribute values for each node in the domain ontology RDF. In the emergency concept class, the initial probability values for the occurrence or existence of aviation-related emergencies and non-aviation emergencies are set to 0.5, generating node probability files for the ontology RDF.

The core concepts in the civil aviation emergency management domain ontology include emergency plans, emergency resources, emergency drills, emergencies, emergency response, resource reporting, and aftermath handling. The RDF graph of emergencies and their causal relationships is shown in [Figure 5: see original paper].

Where emergency U includes aviation emergencies U_1 and non-aviation emergencies U_2 ; subclasses related to U_1 and U_2 events include S_i ; event results related to aviation in emergencies are shown in ; event severity levels are divided into four categories: general G_1 , relatively serious G_2 , major G_3 , and extremely serious G_4 .

3.1 Domain Ontology to Bayesian Network Conversion

The ontology data and probability data serve as inputs, and the method described in Section 2.2.1 is applied to incorporate probabilities as algorithm inputs, forming a complete Bayesian network BN_O. Emergency cause node S_i corresponds to BN node V_i^S (as parent nodes), event result node R_i corresponds to leaf node V_i^R ; event severity G_i corresponds to leaf node V_i^G ; event instance E_i corresponds to node V_{E_i} in the Bayesian network.

3.2 Bayesian Inference Based on Message Passing

The algorithm from Section 2.3 is used to set the evidence node set in BN_O, using node conditional probabilities from the set as inputs for the message passing inference algorithm to reason about posterior probabilities of nodes.

Let nodes V_{E1} and V_{E2} represent non-aviation emergencies (e.g., 7.17 MH17 incident where a Malaysia Airlines aircraft was shot down by a missile, 8.24 VD8387 Henan Airlines crash on August 24), while V_{E3} and V_{E4} represent aviation emergencies (e.g., 11.4 QF-667 EgyptAir cockpit fire incident). Nodes $V_{S_{i,1}}$ (pilot operation error), $V_{S_{7,1}}$ (air traffic control command error), and $V_{S_{n,2}}$ (weather conditions) represent event causes.

The message passing-based inference implementation:

Input: Bayesian network BN_O (network structure and conditional probability table set), evidence node V_{E2} with state True=1.

Output: Conditional probability distributions of non-evidence nodes $V_{S_{i,1}}$, $V_{S_{7,1}}$, $V_{S_{n,2}}$: $P(V_{S_{i,1}})$, $P(V_{S_{7,1}})$, $P(V_{S_{n,2}})$.

Set initial message transmission values for evidence node V_{E2} :

The conditional probability tables for non-evidence nodes $V_{S_{i,1}}$, $V_{S_{7,1}}$, $V_{S_{n,2}}$ are set as follows:

Through formula (3), calculate V_{E2} 's λ value. For nodes $V_{S_{i,1}}$, $V_{S_{7,1}}$, $V_{S_{n,2}}$, after receiving messages from child nodes, output the posterior probabilities using formula (6) to compute $P(V_{S_{i,1}})$, $P(V_{S_{7,1}})$, $P(V_{S_{n,2}})$. Similarly, calculate the posterior probability of node V_{E2} .

Repeat until no message passing occurs throughout the network, then output the posterior probabilities. [Figure 6: see original paper] shows partial probability distribution calculation results related to event causal relationships.

3.3 Effectiveness Analysis

Based on the probability distribution of emergency causal relationships in the domain ontology shown in [Figure 6: see original paper] and the conditional probability tables of each node, the probability distribution of event causes can be derived from calculations with node V_{E_i} as evidence nodes. The conditional probability table for event node V_{E2} is shown in .

presents the probability transfer relationships between the emergency event represented by node V_{E2} and its three cause parent nodes $V_{S_{i1}}$, $V_{S_{71}}$, $V_{S_{n2}}$:

- a) When all three cause nodes $V_{S_{i1}}$, $V_{S_{71}}$, $V_{S_{n2}}$ are in state T (all causes present), the probability of V_{E2} event occurrence is 1.
- b) When two of the three cause nodes exist simultaneously, the event occurrence probability varies with the correlation degree between cause 1 and cause 2. In , the probability of 0.86 shows stronger correlation with $V_{S_{n2}}$, while 0.84 correlates more with $V_{S_{71}}$. Notably, the probability of 0.3 demonstrates stronger correlation with $V_{S_{i1}}$, reflecting significant differences in how various causes impact the event.
- c) When only one of the three cause nodes exists, the maximum probability of event occurrence is 0.7 (even higher than in some multiple-cause scenarios), while the minimum is 0.14, further demonstrating differential impacts of various causes on aviation safety events.
- d) When all three cause nodes $V_{S_{i1}}$, $V_{S_{71}}$, $V_{S_{n2}}$ are in state F (no causes present), the probability of V_{E2} event occurrence is 0.

These results enable further reasoning about probabilistic relationships between different events and multiple causes.

4 Conclusion

The causal analysis method for civil aviation emergencies based on Bayesian networks combines the semantic knowledge representation capability of domain ontologies with the probabilistic reasoning power of Bayesian networks. Through designed conversion rules from domain ontology to Bayesian network, the method enables ontology-based BN construction. Using expert evaluations from the World Civil Aviation Accident Investigation Tracking Information Report to set probability attributes for evidence nodes in the Bayesian network, and integrating the E-IPFP probability fusion algorithm with message passing mechanisms, the method completes the calculation of posterior probability values for variables in emergency causal relationship nodes. While obtaining causal relationship distributions for different categories of emergencies, the approach effectively solves the problem of analyzing civil aviation emergency causal relationships based on domain ontologies, establishing a solid data foundation for further ontology-based causal reasoning. Future research will focus on large-scale domain ontology RDF graph to Bayesian network conversion and distributed inference.

References

- [1] CCAR396-R3, Regulations on Civil Aviation Safety Information Management [S].
- [2] Gan Honghua. Research on Computational Analysis of Event Causality [D]. Hangzhou: Zhejiang University, 2008.
- [3] Zhong Jun, Yu Long, Tian Shengwei, et al. Extraction of Causal Relations for Uyghur Emergency Events Based on a Two-layer Model [J]. *Acta Automatica Sinica*, 2014, 40(4): 771-779.
- [4] Zhou Wentao, Wang Hong, Wang Jing. Research on Semantic Model Construction Method for Civil Aviation Emergency Decision-making [J]. *Application Research of Computers*, 2013, 30(1): 195-198.
- [5] Li Lin, Wang Hong, Fu Yu, et al. Research on Semantic Retrieval Method for Civil Aviation Emergency Cases [J]. *Computer Engineering and Design*, 2011, 32(3): 1130-1133.
- [6] Wang Hong, Yang Xuan, Wang Jing, et al. Research on Ontology-based Knowledge Representation and Reasoning Method for Civil Aviation Emergency Decision-making [J]. *Computer Engineering and Science*, 2011, 33(4): 129-133.
- [7] Wu Xianguo, Ding Baojun, Zhang Limao, et al. Research on Subway Construction Risk Management Based on Bayesian Networks [J]. *China Safety Science Journal*, 2014, 24(1): 84.
- [8] Xie Hongwei, Yan Ting, Che Jinqiang. Application of Ontology-driven Bayesian Network Model in Medical Diagnosis [J]. *Journal of Taiyuan University of Technology*, 2016, 47(3): 389-393.
- [9] Elster C, Wübbeler G. Bayesian inference using a noninformative prior for linear Gaussian random coefficient regression with inhomogeneous within-class variances [J]. *Computational Statistics*, 2017, 32: 1-19.
- [10] Wood F, Willem V D M J, Mansinghka V. A New Approach to Probabilistic Programming Inference [J]. *Computer Science*, 2015: 1024-1032.
- [11] Cai Bingwan, Shi Yuqiang, Li Minghui, et al. Research on Ontology-based Bayesian Network Knowledge Reasoning [J]. *Machinery Design and Manufacture*, 2016(1): 84-87.
- [12] Zhang Shiyong. Research on Uncertain Reasoning in Bayesian Networks [D]. Hefei: University of Science and Technology of China, 2013.
- [13] Liu Y, Chen S, Li S, et al. Bayes-SWRL: a probabilistic extension of SWRL [C]// Proc of the 9th International Conference on Computational Intelligence and Security. Washington DC: IEEE Computer Society, 2013: 702-706.
- [14] <http://safety.caac.gov.cn/index/initpage.act> [EB/OL]. 2017.

[15] Zhao Jianzhe, Li Kai. An Improved Local Inference Algorithm for Multi-module Bayesian Networks [J]. Journal of Northeastern University: Natural Science Edition, 2015, 36(9): 1251-1255.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.