

基于 D-S 理论的因果图区间分析方法*

蒋 杰, 王洪春

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】针对实际应用中事件发生概率具有不确定性的特点,研究解决获取事件发生概率精确值的问题。【方法】运用 D-S 理论提出了因果图区间分析方法,避免了获取概率的精确值,使用似然函数和信任函数分别作为因果图中基本事件区间概率的上限和下限。针对出现异常现象的事件先确定出可能导致其发生的基本事件,再分别求出这些基本事件的后验概率,然后利用相对优势度对区间数比较大小。【结果】区间数最大的基本事件就是故障源。【结论】用区间数来表示事件发生的概率值,解决了事件发生概率精确值的不确定性问题。

关键词:因果图;D-S 理论;后验概率;相对优势度;区间分析

中图分类号:TP181

文献标志码:A

文章编号:1672-6693(2018)01-0088-05

故障诊断研究主要包括判断故障是否发生、故障发生概率的大小以及确定故障的发生位置和发生时间。故障诊断的方法有很多,如有向图、故障树、贝叶斯网络、因果图等,这些方法都具有定性和定量分析的特点。因果图理论是一种基于概率论的知识表达推理方法^[1],能够对复杂系统进行故障诊断。在工程实际中会遇到各种不确定因素,使得事件发生的精确概率值难以获得,因此引入 D-S 证据理论。在因果图中研究用区间数代替事件发生的概率值,扩大了因果图的应用范围,具有非常重要的实际意义。因果图理论在近些年来已得到了一些应用,其中文献[2-3]研究了因果图在化工过程中的应用,文献[4]将因果图应用到了医疗领域,文献[5]对因果图在核电领域的应用进行了研究。目前,大量的文献^[6-7]在因果图中引入模糊数学的方法,用模糊数来描述基本事件和连接事件的发生概率,避免了人们在获取事件发生概率的精确值时遇到的难题,但对于复杂的系统,要建立正确的模糊规则和隶属函数是十分困难的。

因果图区间分析是基于 D-S 理论,再利用区间分析四则运算规则进行区间数值运算,解决因果图量化计算的问题,从而获得事件发生的故障区间概率。与传统的因果图分析方法相比,区间分析方法可以不用知道不确定变量的概率分布。

区间概率理论^[8]是用于量化表示专家系统证据的理论。区间概率理论是对经典概率理论的一个推广,可以用来描述一般的不确定性,是描述不确定性的一个统一的概念。区间概率的实质是把区间数的概率用于概率论中,就是说任意概率都可以用区间数表示,然后用区间数表示的概率来处理系统的模糊性和不完整性。

1 因果图模型

因果图理论是一种可以用来表达不确定性知识的推理模型。这种图形的知识表示方式,可以使得需要表达的因果关系非常直观、自然、明显。因果图是在信度网基础上发展起来的推理方法^[1],是一个有向图模型。在因果图中是有可能出现循环的情况,出现这种情况的时候需要用到解环规则^[9]进行逻辑解环。图 1 是一个典型的因果图模型,其中圆圈表示中间事件,方框表示基本事件,门节点表示逻辑门事件,有向弧表示连接事件。

采用因果图进行推理的主要目的是求解某事件在已知证据下的后验概率^[10]。在得到后验概率的过程中一般要经过如下 4 个步骤:

* 收稿日期:2016-11-10 修回日期:2017-11-30 网络出版时间:2018-01-18 15:21

资助项目:国家社会科学基金(No.13BTJ008);重庆师范大学研究生科研创新基金(No.YKC17002)

第一作者简介:蒋杰,女,研究方向为因果图、不确定性推理,E-mail: 1016724527@qq.com;通信作者:王洪春,教授,E-mail: wanghc@cqnu.edu.cn

网络出版地址: <http://kns.cnki.net/kcms/detail/50.1165.N.20180118.1521.026.html>

- 1) 求出各个中间(节点)事件的一阶割集表达式;
- 2) 求出各个中间(节点)事件的最终割集表达式,在这一步中对于因果图出现环路的情况要进行解环;
- 3) 求出各个中间(节点)事件的不变化割集表达式:设 $X = \bigcup_{i=1}^m C_i$, 其中 $C_i = \bigcap_{j=1}^{n_i} V_{ij}$, C_i 是一个割集, 则节点事件的不变化割集表达式可表示为 $X = C_1 + C_2 \bar{C}_1 + C_3 \bar{C}_1 \bar{C}_2 + \dots + C_m \bar{C}_1 \bar{C}_2 \dots \bar{C}_{m-1}$, 为互斥或操作符;
- 4) 通过观测事件的异常现象, 计算基本事件也就是故障源在这些异常现象发生情况下的后验概率。

在利用因果图进行推理时, 很多时候各个事件的概率都被看作是精确值, 由此得到的结果也是精确的, 这样通过直接比较大小就可以确定故障源。然而在实际系统中存在的不确定性以及数据的不精确性等因素使得事件发生概率的精确值难以得到。针对这些模糊性、不确定性问题, 本文运用区间概率理论来处理, 避免了获取精确知识的难度。

2 区间运算法则

定义 若 $X = [\underline{x}, \bar{x}] = \{x \in R \mid \underline{x} \leq x \leq \bar{x}\}$, 就称其为有界闭区间。用 $I(R)$ 表示 R 上所有有界区间所构成的集合。

在 $I(R)$ 上定义区间的四则运算, 对于任意的区间 $X = [\underline{x}, \bar{x}], Y = [\underline{y}, \bar{y}] \in I(R)$, 四则运算定义为:

$$X + Y = [\underline{x} + \underline{y}, \bar{x} + \bar{y}], X - Y = [\underline{x} - \bar{y}, \bar{x} - \underline{y}],$$

$$X \cdot Y = [\min(\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}), \max(\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y})], X/Y = [\underline{x}, \bar{x}] \cdot [1/\bar{y}, 1/\underline{y}].$$

其中 $0 \notin [\underline{y}, \bar{y}]$ 。

在区间分析理论中实数 1 可以用区间数 $[1, 1]$ 来表示。

3 D-S 理论

D-S 理论是处理不确定性问题的一种数学方法, 可以处理随机不确定性, 因此特别适合数据较少的情况。这一理论的优点在于可以综合专家们给出的不同意见。通常情况下各位专家给出的基本概率函数值是不同的, 而 D-S 理论可以合理地融合这些基本概率函数, 为因果图分析输入数据的获取提供一种途径。

定义 1 识别框架是一组被研究的问题的集合用 Θ 表示, 称 Θ 为识别框架。

定义 2 设 m 是 2^Θ 到 $[0, 1]$ 的函数, 并满足 $m(\emptyset) = 0, \sum_{A \subseteq \Theta} m(A) = 1$, 称 m 是 2^Θ 上的概率分配函数, $m(A)$ 称为 A 的基本概率函数。其中 2^Θ 表示 Θ 的幂集。

定义 3 信任函数 B_{Bel} 是 2^Θ 到 $[0, 1]$ 的函数, 且 $B_{\text{Bel}}(A) = \sum_{B \subseteq A} m(B)$ 对所有的 $A \subseteq \Theta$, 其中 2^Θ 表示 Θ 的所有子集。 $B_{\text{Bel}}(A)$ 表示对 A 为真的信任程度。

定义 4 似然函数 P_{Pl} 是 2^Θ 到 $[0, 1]$ 的函数, 且 $P_{\text{Pl}}(A) = 1 - B_{\text{Bel}}(\bar{A}) = \sum_{B \cap A \neq \emptyset} m(B)$, 其中 $A \subseteq \Theta$ 。 $P_{\text{Pl}}(A)$ 表示对 A 为非假的信任程度。

根据 D-S 理论的合成公式有:

$$m_1 \oplus m_2(C) = \begin{cases} 0, C = \emptyset, \\ \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{K}, C \neq \emptyset. \end{cases}$$

其中, $m_1(A)$ 表示根据证据 A 的基本概率函数; $m_2(B)$ 表示根据证据 B 的基本概率函数, K 为正交化系数, 且 $K = 1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$ 。

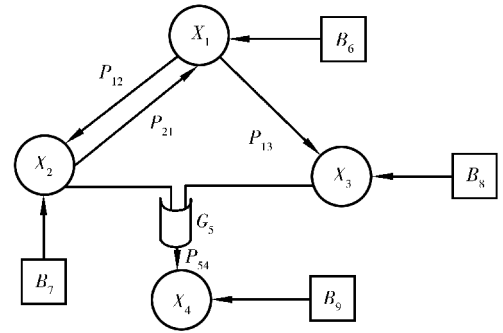


图 1 经典因果图模型

Fig. 1 A classical causality diagram model

用信任区间 $[B_{\text{Bel}}(A), P_{\text{Pl}}(A)]$ 描述对 A 的确认程度, 由此获得因果图中基本事件和连接事件发生的区间概率。再根据文献[11]的相对优势度, 提出比较区间数大小的方法, 最后选择区间数最大的基本事件作为故障源。

比较区间数大小的具体步骤如下:

第 1 步, 得到区间数的极大区间数用 M 来表示, M 以区间数所有的左右端点中的次最大和最大作为左右端点。

第 2 步, 对区间数进行两两比较得到相对优势度 $p_{ij} = P(A_i > A_j) = \frac{d(A_j, M)}{d(A_i, M) + d(A_j, M)}$, 若 $p_{ij} > 0.5$ 称 $A_i > A_j$; $p_{ij} = 0.5$ 称 $A_i = A_j$; $p_{ij} < 0.5$ 称 $A_i < A_j$, 其中 $0 \leq p_{ij} \leq 1$, $d(A_i, A_j)$ 指的是区间数之间的距离, 令 $A_i = (\underline{a}, \bar{a})$, $A_j = (\underline{b}, \bar{b})$, 则有 $d(A_i, A_j) = \frac{\sqrt{(\underline{a} - \underline{b})^2 + (\bar{a} - \bar{b})^2}}{2}$ 。在这个过程中每比较一次就可以把小的区间数删除掉, 也就是说这个较小的区间数不再做之后的比较, 这会使得计算量大大降低, 提高了计算效率。

第 3 步, 通过相对优势度比较大小后, 最大的区间数所对应的基本事件就是要找的故障源。

用区间来描述事件的发生概率和模糊数在因果图中的应用的实质是一样的, 但是这种方法比通过模糊数来获取事件的发生概率更容易, 因为只需要取得基本事件概率, 这是由专家给出的一个概率, 鉴于每位专家对知识的认识角度方法不一样, 所以得到的结果又会有差异, 此时通过 D-S 理论的合成公式对多位专家给出的概率进行融合就能得到一个更准确地结果。而模糊数中隶属函数的获需要大量的数据这个过程是比较复杂的, 而且对于数据缺乏的系统是难以实现的。

4 算例分析

为了验证诊断推理算法的有效性, 下面以图 1 的因果图作为例子来说明因果图和证据理论相结合进行故障诊断的过程。

在因果图推理中基本事件作为故障源, 在故障现象发生的情况下来求故障源发生概率的大小是本文研究的目的, 但在实际应用中通常精确的概率值是很难获得的。为了解决这样的一个问题, 可以用 D-S 证据理论来获取事件发生的区间概率。不同领域专家可以针对导致异常现象出现的各个故障源给出基本概率函数为: $m_1(B_6, B_7, B_8, B_9, \Theta) = \{0.2, 0.25, 0.1, 0.4, 0.05\}$, $m_2(B_6, B_7, B_8, B_9, \Theta) = \{0.3, 0.25, 0.05, 0.25, 0.15\}$, 然后利用 D-S 证据理论推理算法对这两组数据进行融合, 首先计算正交化系数 K 的值:

$$K = 1 - [m_1(B_6)m_2(B_7) + m_1(B_6)m_2(B_8) + m_1(B_6)m_2(B_9) + m_1(B_7)m_2(B_6) + m_1(B_7)m_2(B_8) + m_1(B_7)m_2(B_9) + m_1(B_8)m_2(B_6) + m_1(B_8)m_2(B_7) + m_1(B_8)m_2(B_9) + m_1(B_9)m_2(B_6) + m_1(B_9)m_2(B_7) + m_1(B_9)m_2(B_8)] = 1 - 0.58 = 0.42.$$

融合后的结果如下:

$$m_1 \oplus m_2(B_6) = \frac{1}{K} [m_1(B_6)m_2(B_6) + m_1(B_6)m_2(\Theta) + m_2(B_6)m_1(\Theta)] = \frac{0.12}{0.42} = 0.2857,$$

$$m_1 \oplus m_2(B_7) = \frac{1}{K} [m_1(B_7)m_2(B_7) + m_1(B_7)m_2(\Theta) + m_2(B_7)m_1(\Theta)] = \frac{0.105}{0.42} = 0.2679,$$

$$m_1 \oplus m_2(B_8) = \frac{1}{K} [m_1(B_8)m_2(B_8) + m_1(B_8)m_2(\Theta) + m_2(B_8)m_1(\Theta)] = \frac{0.0225}{0.42} = 0.05357,$$

$$m_1 \oplus m_2(B_9) = \frac{1}{K} [m_1(B_9)m_2(B_9) + m_1(B_9)m_2(\Theta) + m_2(B_9)m_1(\Theta)] = \frac{0.175}{0.42} = 0.4107,$$

$$m_1 \oplus m_2(\Theta) = \frac{1}{K} m_1(\Theta)m_2(\Theta) = \frac{0.0075}{0.42} = 0.01786,$$

$$B_{\text{Bel}}(B_6) = m_1 \oplus m_2(B_6) = 0.2857, B_{\text{Bel}}(B_7) = m_1 \oplus m_2(B_7) = 0.2679,$$

$$B_{\text{Bel}}(B_8) = m_1 \oplus m_2(B_8) = 0.05357, B_{\text{Bel}}(B_9) = m_1 \oplus m_2(B_9) = 0.4107,$$

$$P_{\text{Pl}}(B_6) = 0.3036, P_{\text{Pl}}(B_7) = 0.2858, P_{\text{Pl}}(B_8) = 0.0715, P_{\text{Pl}}(B_9) = 0.4286.$$

最后得到基本事件的故障区间概率如表 1 所示。

对于连接事件概率的确定也用相同的方法来获得,不同领域专家再根据父节点事件和子节点事件之间的因果关联强度给出连接事件的基本概率函数为: $m_1(P_{12}, P_{13}, P_{21}, P_{54}, \Theta) = (0.4, 0.3, 0.1, 0.1, 0.1)$, $m_2(P_{12}, P_{13}, P_{21}, P_{54}, \Theta) = (0.2, 0.2, 0.05, 0.5, 0.05)$ 。

接下来对这两组数据进行融合然后得到概率区间,方法和上述基本事件的融合方法一样在这里就不再重复叙述。连接事件概率融合后的结果如表 2 所示。

表 1 基本事件故障区间概率

Tab. 1 Basic events failure interval probability

基本事件名称	区间概率
B_6	[0.285 7, 0.303 6]
B_7	[0.267 9, 0.285 8]
B_8	[0.053 6, 0.071 5]
B_9	[0.410 7, 0.428 6]

表 2 连接事件区间概率

Tab. 2 linkage events interval probability

连接事件名称	区间概率
P_{12}	[0.353 0, 0.367 7]
P_{13}	[0.279 4, 0.294 1]
P_{21}	[0.044 1, 0.058 8]
P_{54}	[0.308 8, 0.323 5]

在对图 1 进行推理时应该注意的是 G_5 表示逻辑或门。

首先,求出各个中间事件的一阶割集表达式:

$$X_1 = B_6 \cup P_{21} X_2, X_2 = B_7 \cup P_{12} X_1, X_3 = B_8 \cup P_{13} X_1, X_4 = B_9 \cup P_{54} (X_2 \cup X_3)。$$

然后,求出各个中间事件的最终割集表达式:

$$X_1 = B_6 \cup P_{21} B_7, X_2 = B_7 \cup P_{12} B_6, X_3 = B_8 \cup P_{13} B_6 \cup P_{13} P_{21} B_7, \\ X_4 = B_9 \cup P_{54} B_7 \cup P_{54} P_{12} B_6 \cup P_{54} B_8 \cup P_{54} P_{13} P_{21} B_7 \cup P_{54} P_{13} B_6。$$

其次,求出各个中间事件不交化割集的表达式:

$$X_1 = B_6 + P_{21} B_7 \bar{B}_6, X_2 = B_7 + P_{12} B_6 \bar{B}_7, X_3 = B_8 + P_{13} B_6 \bar{B}_8 + P_{13} P_{12} B_7 \bar{B}_8 \bar{B}_6, \\ X_4 = B_9 + P_{54} B_7 \bar{B}_9 + P_{54} P_{12} B_6 \bar{B}_9 \bar{B}_7 + \bar{B}_9 \bar{B}_7 \bar{B}_6 P_{54} P_{12} B_8 + \bar{B}_9 \bar{B}_7 P_{54} \bar{P}_{12} B_8 + P_{54} P_{13} B_6 \bar{B}_9 \bar{B}_7 \bar{B}_8 \bar{P}_{12}。$$

再次,假设事件 X_3 发生异常,通过事件 X_3 不交化割集的表达式可以看出导致它异常的可能故障源有 B_6, B_7, B_8 。为了便于描述这 3 个后验区间概率大小,分别用 A_1, A_2, A_3 来表示 3 个区间数即: $A_1 = P_r(B_6 | X_3) = \frac{P_r(B_6 X_3)}{P_r(X_3)} = [0.162 0, 0.215 8], A_2 = P_r(B_7 | X_3) = \frac{P_r(B_7 X_3)}{P_r(X_3)} = [0.065 85, 0.097 36], A_3 = P_r(B_8 | X_3) = \frac{P_r(B_8 X_3)}{P_r(X_3)} = [0.097 06, 0.145 2]$ 。

最后,根据上文中提出的区间数比较大小的主要步骤对这 3 个区间数进行大小比较,首先得到区间数 A_1, A_2, A_3 的极大区间数,则有 $M = [0.162 0, 0.215 8], p_{12}(A_1 > A_2) = 1 > 0.5$, 所以有 $A_1 > A_2$ 成立; $p_{13}(A_1 > A_3) = 1$, 所以有 $A_1 > A_3$ 成立,由此可知 A_1 已是最大区间数,也就是说 B_6 是导致事件 X_3 发生异常的主要原因。同理可以根据上述步骤对所观察到的异常现象做同样的分析来进行故障诊断确定故障源。

5 结束语

本文提出的基于 D-S 理论的因果图区间分析方法,很容易得到事件发生的区间概率,解决了对于因果图传统的分析方法中不确定性不容易量化的问题,从而避免了获取精确知识的难度,通过 D-S 理论对因果图进行区间分析的过程中,最重要的问题是基本概率函数的获取它是由不同领域专家来给定的,为了得到更准确的结果需要先对这些专家给定的数据进行融合,再进行后面的分析,特别是对于数据缺乏的情况有明显的优势。

参考文献:

[1] ZHANG Q. Probabilistic based on dynamic causality trees/ diagrams[J]. Reliability Engineering and System Safety,

- 1994, 46: 209-220.
- [2] 杨佳婧, 张勤, 朱群雄. 动态不确定因果图在化工过程故障诊断中的应用[J]. 智能系统学报, 2014, 9(2): 154-160.
YANG J J, ZHANG Q, ZHU Q X. Application of dynamic uncertain causality graph to fault diagnosis in chemical processes[J]. CAAI Transactions on Intelligent Systems, 2014, 9(2): 154-160.
- [3] 曲彦光, 张勤, 朱群雄. 动态不确定因果图在化工过程故障诊断中的应用[J]. 智能系统学报, 2015, 10(3): 354-360.
QU Y G, ZHANG Q, ZHU Q X. Application of dynamic uncertain causality graph to fault diagnosis In chemical processes[J]. CAAI Transactions on Intelligent Systems, 2015, 10(3): 354-360.
- [4] DONG C L, WANG Y J, ZHANG Q, et al. The methodology of dynamic uncertain causality graph for intelligent diagnosis of vertigo[J]. Computer Methods and Programs in Biomedicine, 2014, 113(1): 162-174.
- [5] ZHANG Q, DONG C L, CUI Y, et al. Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: statistics base, matrix, and application[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(4): 645-663.
- [6] 王洪春, 张勤. 基于模糊因果图的故障诊断[J]. 微电子学与计算机, 2005, 22(6): 109-112.
WANG H C, ZHANG Q. Fault diagnosis based on the fuzzy causality diagram[J]. Microelectronics & Computer, 2005, 22(6): 109-112.
- [7] 梁新元, 石庆喜. 三角模糊数的模糊因果图研究[J]. 仪器仪表学报, 2006, 27(6): 2495-2498.
LIANG X Y, SHI Q X. Fuzzy causality diagram with triangular fuzzy number[J]. Chinese Journal of Scientific Instrument, 2006, 27(6): 2495-2498.
- [8] CUI W, BLOCKLEY D I. Interval probability theory for evidential support[J]. International Journal of Intelligent System, 1990(5): 183-192.
- [9] 王洪春, 石庆喜, 张勤. 基于因果图的一种推理算法[J]. 微电子学与计算机, 2005, 22(5): 1-7.
WANG H C, SHI Q X, ZHANG Q. A reasoning algorithm based on causality diagram[J]. Microelectronics & Computer, 2005, 22(5): 1-7.
- [10] 梁新元, 张勤. 因果图在故障分析中的应用研究[J]. 计算机工程与应用, 2004, 40(19): 185-188.
LIANG X Y, ZHANG Q. The application of causality diagram in fault analysis[J]. Computer Engineering and Applications, 2004, 40(19): 185-188.
- [11] 徐改丽, 吕跃进. 不确定性的属性决策中区间数排序的一种新方法[J]. 统计与决策, 2008, 10(19): 154-157.
XU G L, LÜ Y J. A new method for the ranking of the number of Uncertain attributes in interval number[J]. Statistics and Decision Making, 2008, 10(19): 154-157.

Interval Analysis Method of Causality Diagram Based on D-S Theory

JIANG Jie, WANG Hongchun

(College of Mathematics, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] According to the actual application of event occurrence probability has the characteristics of uncertainty, In order to solve the problem of obtaining the accurate values of event occurrence probability. [Methods] The D-S theory is used to put forward the method of causality diagram of interval analysis. Applying this method avoids obtaining the accurate values of difficulty of the event occurrence probability, the likelihood function and the belief function in the D-S theory are used as upper and lower boundary on the interval probability of fault occurred in basic events in causality diagram analysis. When abnormal phenomenon is occurred, firstly, we should determine the basic events that may lead to its occurrence, and then calculate the posterior probability of basic events. [Findings] Interval numbers are compared based on the relative superiority degree. Basic event of the most interval number is fault source. [Conclusions] Using interval number to denote the probability of event value, solves the problem of uncertainty event probability.

Keywords: causality diagram; D-S theory; posterior probability; relative superiority degree; interval analysis

(责任编辑 许 甲)