

KRYTYCZNIE O WZORZE BAYESA W DIAGNOSTYCE OBIEKTÓW TECHNICZNYCH

Wzór Bayesa, a zwłaszcza wynikające z niego prawdopodobieństwa warunkowe są źródłem dość niepokojącego stereotypu panującego w diagnostyce – konieczności podziału diagnozowania na kontrolę zdatności i lokalizowanie niezdatności. Niepokój wynika przede wszystkim z faktu powstawania na tym tle wielu niejednoznaczności i strat, a także spełnienia nie przystających do rzeczywistości założeń. W referacie przedstawia się inne – niż występujące w literaturze przedmiotu – wyprowadzenie rzeczzonego wzoru i jego dyskusję w aspekcie diagnozowania. Podano rozwiązanie alternatywne oparte o związki prawdopodobieństw apriorycznych ze strukturą obiektu.

CRITICALLY ABOUT BAYES' FORMULA IN DIAGNOSTICS OF TECHNICAL OBJECT

This study expresses peculiar kind of protest against diagnostic tests division into two stages: recognition and fault finding. There is an opinion, that basic (free from wrong and incomplete diagnosis) way of object testing should take into consideration complete collection (skip Bayes' formula) its different reliability states, i.e. collection as a result of both components combination structure and two unquestionable rules of diagnostics, according to whom – the component is being considered faulted if all its input signals are admissible however output signal – inadmissible, and – component's output signal is inadmissible if at least one of its input signal is inadmissible too.

Wartości poszczególnych prawdopodobieństw wystąpienia każdego ze stanów obiektu przy dwuwartościowej ocenie stanów jego elementów składowych określone są wyrazami wielomianu powstałego z iloczynu sum:

$$\prod_{i=1}^{Card E} (p_i + q_i), \quad (1)$$

gdzie: $E = \{e_1, e_2, e_3, \dots, e_n\}$ – zbiór elementów obiektu,

p_i – prawdopodobieństwo zdatności elementu e_i ,

q_i – prawdopodobieństwo niezdatności elementu e_i .

Przyporządkowanie konkretnego jednomianu konkretnemu stanowi jest jednoznaczne z określeniem występujących w nim czynników - liczba tych czynników w każdym jednomianie jest równa liczbie elementów obiektu. Każdorazowo

$$(p_i + q_i) = 1, \quad (2)$$

więc zgodnie z zasadą neutralności „jedyńki” w mnożeniu

$$\prod_{i=1}^{Card E} (p_i + q_i) = 1. \quad (3)$$

Dla obiektu trójelementowego iloczyn sum wartości prawdopodobieństw przedstawia równanie (4):

$$(p_1 + q_1)(p_2 + q_2)(p_3 + q_3) = \begin{array}{l} \boxed{p_1 p_2 p_3} + 1 \\ + \boxed{q_1 p_2 p_3} + 2 \\ + \boxed{p_1 q_2 p_3} + 3 \\ + \boxed{p_1 p_2 q_3} + 4 \\ + q_1 q_2 p_3 + 5 \\ + q_1 p_2 q_3 + 6 \\ + p_1 q_2 q_3 + 7 \\ + q_1 q_2 q_3 + 8 \end{array} \quad (4)$$

Warto zauważyć, że przy dwuwartościowej ocenie stanów elementów liczba możliwych stanów obiektu trójelementowego wynosi 8, dla obiektu dziesięcioelementowego już - 1024, natomiast dla obiektu tysiącelementowego - wielokrotnie przekracza liczbę atomów występujących we wszechświecie [1]. Ta szczególna magia liczb stała się jednym z powodów, dla którego postanowiono diagnozowanie złożonych obiektów technicznych podzielić na dwa etapy. Zgodnie z polską normą (PN-93/N-50191) nazywane są one: rozpoznaniem niezdatności i lokalizowaniem niezdatności, lecz poprzez szeroko dostępną literaturę przedmiotu znane głównie pod nazwami: kontrola zdadności i lokalizacja uszkodzeń. Przeprowadzenie pierwszego etapu pozwala odpowiedzieć na pytanie, czy obiekt jest zdolny do wykonania swoich funkcji, natomiast drugiego - który z elementów obiektu powoduje jego obecny stan niezdatności. Jest przy tym oczywiste, że drugi etap rozpoczyna się po uzyskaniu negatywnego wyniku pierwszego etapu.

Oprócz „magii liczb” istotnym argumentem na uzasadnienie dokonanego podziału stało się przyjęcie założenia, że obiekt diagnozowania ma szeregową strukturę niezawodnościową i znikomą wartość prawdopodobieństwa wystąpienia jakiegokolwiek niezdatności mnogiej.

Dla obiektu trójelementowego prawdopodobieństwo wystąpienia stanu zdadności określone jest jednomianem - 1, natomiast prawdopodobieństw wystąpienia stanów wynikających z niezdatności pojedynczych elementów - jednomianami: 2,3 i 4, zamieszczonymi w podwójnych ramkach zależności (4).

Ponieważ

$$q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3 < 1, \quad (5)$$

postanowiono stosować wartości względne poszczególnych składników:

$$q_1' = \frac{q_1 p_2 p_3}{q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3}, \quad (6)$$

$$q_2' = \frac{p_1 q_2 p_3}{q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3}, \quad (7)$$

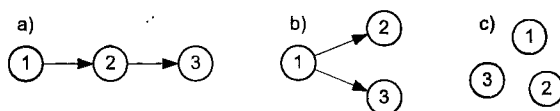
$$q_3 = \frac{p_1 p_2 q_3}{q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3} \quad (8)$$

Uogólnieniem zależności (6÷8), odnoszącym się do obiektu o dowolnej liczbie "n" elementów, jest zależność:

$$q_i = \frac{q_i \prod_{\substack{k=1 \\ k \neq i}}^n p_k}{\sum_{j=1}^n q_j \prod_{\substack{k=1 \\ k \neq j}}^n p_k} \quad (9)$$

Jest ona słynnym wzorem Bayesa wyrażającym w tej postaci wartość prawdopodobieństwa warunkowego wystąpienia niezdatności elementu e_i . Wzór ten bywa również nazywany wzorem na prawdopodobieństwo *a posteriori*.

Najwymowniejszym i najbardziej przekonującym przykładem, potwierdzającym „zasadność” wprowadzenia szeregowej struktury niezawodnościowej, była i zwykle jest analiza diagnostyczna obiektu, który podobnie jak ta struktura jest łańcuchem szeregowo połączonych elementów. (Dla obiektu trójelementowego może być nią struktura przedstawiona na rys. 1.a). Rozpoznanie niezdatności polega wtedy na sprawdzeniu sygnału wyjściowego ostatniego elementu tego łańcucha (elementu e_3 z rys. 1.a), natomiast lokalizowanie niezdatności – na analizie wyników sprawdzeń sygnałów na wyjściach pozostałych elementów (elementów e_1 i e_2 z rys. 1.a). Wartym zauważenia jest fakt, że pierwsza z czynności pozwala jedynie stwierdzić niezdatność bądź zdatność całego obiektu, natomiast druga – na wskazanie tylko jednego niezdatnego elementu.



Rys. 1. Przykładowe struktury funkcjonalne obiektu trójelementowego: a) łańcuch szeregowo połączonych elementów, b) struktura rozgałęziona, c) brak jakichkolwiek połączeń między elementami

Niestety, takiej samej jednoznaczności wyników badań nie można już odnieść do każdego innego obiektu. Dotyczy to zwłaszcza obiektu o rozgałęzionej strukturze połączeń elementów (np. takiego jak na rys. 1.b), czy obiektu, w którym występuje brak jakichkolwiek połączeń między elementami (np. takiego jak na rys. 1.c). Należy przy tym wyraźnie zaznaczyć, że takie obiekty też przecież mogą mieć szeregową strukturę niezawodnościową!

W świetle powyższego pojawia się także wiele wątpliwości co do zasadności użycia zależności (9). Z jej analizy wynika bowiem:

- obojętność położenia elementów w szeregowej strukturze niezawodnościowej;
- nieistotność wartości prawdopodobieństwa występowania stanu zdatności obiektu;
- brak związku struktury funkcjonalnej z szeregową strukturą niezawodnościową;
- nieistotność ukierunkowanego wzajemnego oddziaływania elementów, oraz
- w przypadku jednakowych wartości prawdopodobieństw q_i (lub p_i) – możliwość

wyrażenia prawdopodobieństw warunkowych q_i odwrotnością liczebności elementów obiektu: $1/n$.

W wielu publikacjach (np. [1, 5]) proponuje się zamiast zależności (9) stosować jej uproszczoną postać:

$$q_i \cong \frac{q_i}{\sum_{j=1}^n q_j}; n = \text{Card}E. \quad (10)$$

Oprócz minimalizacji ilości obliczeń, „zasadniczą korzyścią” jej wykorzystania staje się możliwość wyrażenia wartości prawdopodobieństw q_k , $k \in \{1, 2, 3, \dots, \text{Card}E\}$ dowolnymi dodatnimi liczbami rzeczywistymi. Wtedy przecież także

$$0 \leq q_k \leq 1. \quad (11)$$

Nieprawdaż?

Naturalną likwidację piętrzących się nieścisłości i trudności wynikających z dwuetapowego badania obiektu można uzyskać poprzez „wkomponowanie” charakterystyk probabilistycznych elementów w strukturę ich wzajemnych połączeń. Ważnym przyczynkiem do tego „wkomponowania” może stać się odpowiedź na pytanie:

Jaka jest wartość prawdopodobieństwa wskazania niezdatności elementów obiektu z rys.1.a, jeśli wartość prawdopodobieństwa „q” wystąpienia stanu niezdatności każdego z nich jest jednakowa i bliska jedności?

Kierując się pragmatyką dwuetapowych badań diagnostycznych należałoby powiedzieć, że wartość tego prawdopodobieństwa w odniesieniu do każdego elementu wynosi $1/3$, natomiast zdając się na intuicję – w odniesieniu do elementów e_2 i e_3 , – że jest bliska zeru, a elementu e_1 , – że bliska jedności.

Aby teoretycznie potwierdzić intuicyjne przypuszczenia wystarczy przyjąć, że

po pierwsze: - element uznaje się za niezdatny jeśli wszystkie jego sygnały wejściowe są dopuszczalne natomiast sygnał wyjściowy – niedopuszczalny;

po drugie: - sygnał wyjściowy elementu jest niedopuszczalny jeśli co najmniej jeden z jego sygnałów wejściowych jest niedopuszczalny.

Nietrudno zatem zrozumieć, że niezdatność pierwszego elementu łańcucha z rys. 1.a przesłania wszelkie stany techniczne jego pozostałych elementów a wartość prawdopodobieństwa jej wystąpienia musi być w konsekwencji równa sumie jednomianów 2, 5, 6 i 8 wielomianu (4):

$$\begin{aligned} r_1 &= q_1 p_2 p_3 + q_1 p_2 q_3 + q_1 q_2 p_3 + q_1 q_2 q_3 = \\ &= q_1 (p_2 p_3 + p_2 q_3 + q_2 p_3 + q_2 q_3) = q_1 (p_2 (p_3 + q_3) + \\ &\quad + q_2 (p_3 + q_3)) = q_1 (p_2 + q_2) = q_1. \end{aligned} \quad (12)$$

Warunkiem wykrycia niezdatności elementu e_2 , przesłaniającej stany techniczne elementu e_3 , jest zdatność elementu e_1 , więc wartość prawdopodobieństwa wystąpienia

tej niezdatności jest równa sumie jednomianów 4 i 7 wielomianu (4):

$$r_2 = p_1 q_2 p_3 + p_1 q_2 q_3 = p_1 q_2 (p_3 + q_3) = p_1 q_2. \quad (13)$$

Wykrycie niezdatności elementu e_3 wymaga zdatności elementów e_1 i e_2 , zatem prawdopodobieństwo jej wystąpienia jest równe jednomianowi 4. wielomianu (4):

$$r_3 = p_1 p_2 q_3. \quad (14)$$

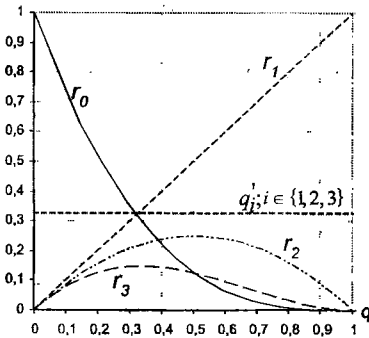
Orzeczenie stanu zdatności obiektu jest jednoznaczne ze zdatnością wszystkich jego elementów: e_1 , e_2 i e_3 , tak więc prawdopodobieństwo jego wystąpienia jest równe jednomianowi 1. wielomianu (4):

$$r_0 = p_1 p_2 p_3. \quad (15)$$

Suma wyznaczonych wartości r_i (równa jedności)

$$\sum_{i=0}^3 r_i = p_1 p_2 p_3 + q_1 + p_1 q_2 + p_1 p_2 q_3 = 1 \quad (16)$$

poddaje pod wątpliwość potrzebę wyznaczania prawdopodobieństw względnych q_i , tym bardziej, że te ostatnie są niezmiennie przy zmianie ich wartości bezwzględnych q_i (rys. 2.). Ponadto, jak zrozumieć obligatoryjną kolejność dwuetapowych badań, jeśli r_0 może być mniejsze od r_1 ?



Rys.2. Wartości prawdopodobieństw wystąpienia rozróżnialnych stanów niezawodnościowych obiektu z rys. 1.a przy q jednakowym dla każdego elementu.

Liczebność rozróżnialnych stanów technicznych poleca się wyznaczać poprzez wykorzystanie górnotrójkatnej macierzy osiągalności $D(G)$, natomiast ich charakterystyki probabilistyczne – poprzez zastąpienie zer i iksów tablicy diagnostycznej $M(O)$ prawdopodobieństwami p_i i q_i , (tablicy wyznaczonej na podstawie wspomnianej macierzy). Stosowne algorytmy przedstawiono między innymi w [2, 3]. Przykład ich zastosowania w odniesieniu do obiektu z rys. 1.b przedstawiają zależności (17).

$$M(O_b) \left\{ \begin{array}{l} D(G_b) = \begin{array}{c|ccc} & S_1 & S_2 & S_3 \\ \hline m_0 & 0 & 0 & 0 \\ m_1 & x & 1 & 1 \\ m_2 & 0 & x & 0 \\ m_3 & 0 & 0 & x \\ m_4 & 0 & x & x \end{array} \begin{array}{l} r_0 = p_1 p_2 p_3 = p_1 p_2 p_3 \\ r_1 = q_1 = 1 - p_1 \\ r_2 = p_1 q_2 p_3 = p_1 p_3 - p_1 p_2 p_3 \\ r_3 = p_1 p_2 q_3 = p_1 p_2 - p_1 p_2 p_3 \\ r_{2,3} = p_1 q_2 q_3 = p_1 - p_1 p_2 - p_1 p_3 + p_1 p_2 p_3 \end{array} \end{array} \right. \quad (17)$$

gdzie:

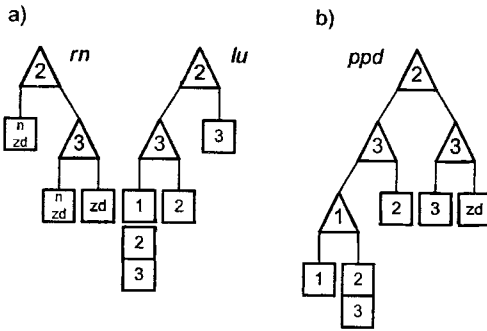
$s_i \in S$; $i \in \{1, 2, 3\}$ - sprawdzenie sygnału wyjściowego i -tego elementu,

$m_i \in M$; $i \in \{0, 1, 2, 3, 4\}$ - stan niezawodnościowy,

0 - pozytywny wynik sprawdzenia,

1, x - negatywne wyniki sprawdzeń.

Macierz $D(G)$ bezpośrednio może służyć badaniom wg programu stałego, lub pośrednio – opracowaniu badań wg któregoś z programów warunkowych [1, 2, 3] (np. takich jak na rys. 3 dla obiektu z rys. 1.b).

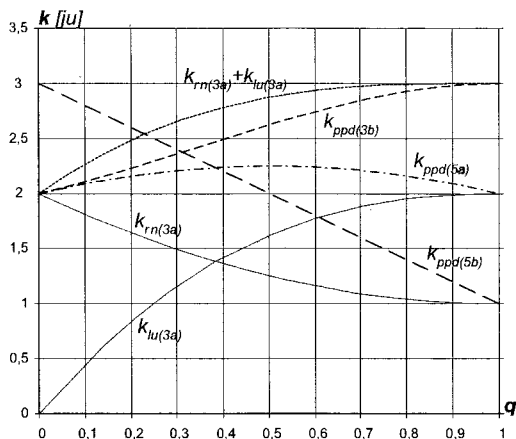


Rys. 3. Warunkowe programy diagnostowania obiektu z rys. 1.b, a) program dwuetapowy, gdzie: rn i lu – rozpoznanie i lokalizowanie niezdatności, b) pełny program diagnostowania (ppd). Trójkątami oznaczono sprawdzenia; kwadratami – stany techniczne; lewymi i prawymi gałęziami wychodzącymi z trójkątów – odpowiednio negatywne i pozytywne wyniki sprawdzeń.

Zasadniczą wadą dwuetapowego diagnostowania może być konieczność powtórzenia sprawdzeń rozpoznania w lokalizowaniu. Ponadto lokalizowanie niezdatności w tym diagnostowaniu może doprowadzić do niejednoznaczności diagnoz. Dla obiektu z rys. 1.b wiara w istnienie co najwyżej pojedynczych niezdatności nie wymusi przecież wykonania sprawdzenia sygnału wyjściowego elementu e_1 , aby upewnić się czy właśnie ten element jest niezdatny, czy oby nie przypadkiem elementy: e_2 i e_3 . Fałszywa diagnoza może przyczynić się do wymiany elementu zdatnego na zdatny, potem – po powtórnym diagnostowaniu i ponownej odnowie – do podejrzeń co do naruszenia zasad obsługi, niepoprawnie wykonanej odnowy lub niezdatności aparatury kontrolnej. Ostateczne stwierdzenie zdatności odnawianego elementu prowadzi zwykle do lokalizowania niezdatności opartego o tzw. „intuicję inżynierską”. Dla obiektów o bardzo złożonej strukturze zadanie to, wykonywane w taki sposób, staje się praktycznie rzecz biorąc niewykonalne.

Wady dwuetapowych badań diagnostycznych są przede wszystkim źródłem generacji strat. Na rys. 4. przedstawiono zmienności oczekiwanych kosztów realizacji zaprezentowanych w niniejszej pracy programów diagnostowania obiektu z rys. 1.b. Związek danego kosztu z określonym programem określono numerem rysunku, na którym został ten program przedstawiony. Okazuje się, że sumaryczny koszt: $k_{rn(3a)} + k_{lu(3b)}$, będący kosztem dwuetapowego diagnostowania, jest dla każdego q większy od kosztów pełnych programów diagnostowania: $k_{ppd(3b)}$ i $k_{ppd(3a)}$. Jednoznacznie

negatywnie nie można ocenić programu badań z rys. 5.b. Co prawda, dla małych q jego koszty są duże w porównaniu z kosztami innych programów, to jednakże już dla znacznie większych – zdecydowanie mniejsze. Należy przypuszczać, że dla obiektu charakteryzującego się dużą awaryjnością (nabywaną wraz z upływem czasu) program ten powinien stanowić podstawową formę jego badań.



Rys. 4. Zmienność oczekiwanych kosztów realizacji programów diagnostowania obiektu z rys. 1.b, gdzie: q – jednakowe dla wszystkich elementów prawdopodobieństwo niezdatności.

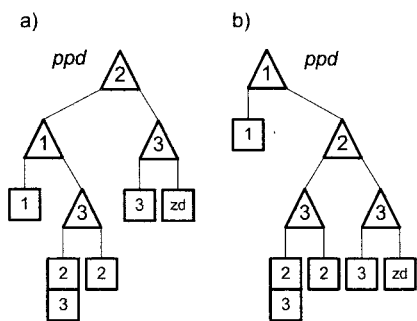
$$k_{rm(3a)} = 1(q + q^2(1-q) + q(1-q)^2) + 2(q(1-q)^2) + 2(1-q)^3 \quad (18)$$

$$k_{lu(3a)} = 2q + 2q^2(1-q) + 2q(1-q)^2 + 1q(1-q)^2 \quad (19)$$

$$k_{ppd(3b)} = 3q + 3q^2(1-q) + 2q(1-q)^2 + 2q(1-q)^2 + 2(1-q)^3 \quad (20)$$

$$k_{ppd(5a)} = 2q + 3q^2(1-q) + 3q(1-q)^2 + 2q(1-q)^2 + 2(1-q)^3 \quad (21)$$

$$k_{ppd(5b)} = 1q + 3q^2(1-q) + 3q(1-q)^2 + 3q(1-q)^2 + 3(1-q)^3 \quad (22)$$



Rys. 5. Modyfikacje pełnych programów diagnostowania (ppd) obiektu z rys. 1.b, a) dla $q < 0,6$ i b) dla $q > 0,6$

Dodatkowo, za połączeniem rozpoznania z lokalizowaniem niezdatności w pełnych programach diagnostycznych przemawia możliwość ich modyfikacji (optymalizacji) (patrz rys. 5) w funkcji zmian prawdopodobieństwa niezdatności elementów obiektu.

Choć pełne programy diagnostowania umożliwiają wskazanie stanów wynikających z niezdatności dwóch lub też większej liczby elementów, to należy jednakże zauważyć,

że elementy te muszą funkcjonować niezależnie jeden od drugiego (tak jak elementy e_2 i e_3 w obiekcie z rys. 1.b lub elementy: e_1 , e_2 i e_3 w obiekcie z rys. 1.c). Nierozróżnialne stany niezdatności (wynikające ze wzajemnego oddziaływania elementów, np. elementów: e_1 i e_2 w obiektach z rysunków 1.a i 1.b) mogą być wskazane np. w wyniku diagnozowania wielokrokowego [4].

LITERATURA

- [1] Rozwadowski T.: Diagnostyka techniczna obiektów złożonych (wydanie drugie uzupełnione i poprawione), WAT, Warszawa 1983,.
- [2] Szczepański P.: Określanie i zastosowanie prawdopodobieństw występowania rozróżnialnych stanów wadliwego funkcjonowania obiektu, Biuletyn WAT, nr 8/2001, (str. 25÷40).
- [3] Szczepański P.: Funkcjonalna struktura niezawodnościowa obiektu na przykładzie diagnozowania szeregowo połączonych elementów, Diagnostyka, 28/2003, (str. 53÷62).
- [4] Szczepański P.: O wielokrokowości procesu diagnozowania złożonego obiektu technicznego, referaty i komunikaty IX Konferencji Naukowo-Technicznej nt.: Diagnostyka maszyn roboczych i pojazdów, Bydgoszcz, wrzesień 1994.
- [5] Żółtowski B.: Podstawy diagnostyki maszyn, Wydawnictwo Uczelniane ATR, Bydgoszcz 1996, str. 336.