

Stanisław DUER¹
Konrad ZAJKOWSKI²

PROBABILISTYCZNA METODA W TWORZENIU BAZY WIEDZY EKSPERTOWEJ

W chwili obecnej przy tworzeniu baz wiedzy ekspertowej stosuje się różne podejścia w celu ich pozyskiwania. W literaturze brak jest sformalizowanego podejścia do problemu metod stosowanych przy budowie baz wiedzy. Jedną z najstarszych metod tworzenia wiedzy specjalistycznej jest metoda probabilistyczna. W artykule zawarto wiedzę dotyczącą opisu probabilistycznej metody wyznaczania informacji obsługowej. W pracy przedstawiono także podstawy analityczne tej metody oraz jej właściwości i uwarunkowania.

Uzyskane efekty z wpływu danej metody na profilaktykę obiektu poddano analizie jakościowej.

PROBABILISTIC METHOD IN CREATING EXPERT KNOWLEDGE BASE

At present, the creation of expert knowledge bases uses different approaches to their recovery. In literature, there is no formalized description of the expert's approach to the problem of the methods used in the construction of knowledge bases. One of the oldest methods of creating expertise is a probabilistic method. The article includes the description of knowledge of the probabilistic method to determine the maintenance of information. The paper presents the analytical basis for this method and its properties and conditions. The effects of the impact of the method for preventing an object subjected to qualitative analysis.

1. WSTĘP

Opracowywanie programów wspomagających proces organizacji obsługi technicznych na bazie informacji diagnostycznej jest znacznie utrudnione. Podstawą takiego stanu jest brak ogólnych metod formalnego zapisu wiedzy diagnostycznej, wiedzy specjalisty (eksperta) i innej oraz brakiem możliwości pełnej algorytmizacji procesu efektywnego wnioskowania i podejmowania decyzji.

¹ Koszalin University of Technology, Department of Mechanical Engineering, Raclawicka 15-17
75-620 Koszalin, Poland, tel. 0943478262, e-mail: stanislaw.duer@tu.koszalin.pl

² tel. 0943478426; konrad.zajkowski@tu.koszalin.pl;

Ogólnymi metodami poszukiwania rozwiązań w takich złożonych warunkach zajmuje się *sztuczna inteligencja*. Jest to dziedzina informatyki i innych dotycząca metod i technik wnioskowania symbolicznego-podejmowania decyzji przez komputer oraz symbolicznej reprezentacji wiedzy stosowanej podczas takiego wnioskowania, gdzie *reprezentacja wiedzy* oznacza ogólny formalizm zapisywania, gromadzenia i jej przechowywania. Wśród stosowanych metod i technik sztucznej inteligencji występują w niej: sztuczne sieci neuronowe, systemy ekspertowe, algorytmy genetyczne i ewolucyjne itp. oraz możliwe kombinacje między nimi tworzące nowe jakościowe rozwiązania, które są nazwane hybrydowymi systemami lub układami inteligentnymi. W praktyce spotykane rozwiązania wykorzystujące metody sztucznej inteligencji są nazwane systemami inteligentnymi. W inteligentnych systemach wspomagających proces podejmowania decyzji w procesach obróbkowych szczególne znaczenie może mieć połączenie w pewnym stopniu metod sieci neuronowych, logiki rozmytej oraz algorytmów ewolucyjnych w postaci inteligentnych systemów hybrydowych.

2. ISTOTA METODY PROBABILISTYCZNEJ PODEJMOWANIA DECYZJI EKSPERTOWEJ

Znaczną grupę metod wnioskowania w systemach ekspertowych stanowią metody, w których w procesie wnioskowania zastosowano podejście probabilistyczne w rozwiązywaniu problemów ekspertowych. W opisie probabilistycznym metod wnioskowania ekspertowego zakłada się, że na różnych poziomach wnioskowania ekspertowego występują fakty. Fakty są losowo scharakteryzowane prawdopodobieństwami o znanych współczynnikach wagowych. Dokładne zatem stwierdzenie, że fakt (V_i) jest losowy oznacza także, że prawdziwość tego faktu jest także zdarzeniem losowym, natomiast prawdopodobieństwo tego faktu wynosi:

$$P(V_i) = P(V_i = 1) \quad (1)$$

Zależność (1) oznacza prawdopodobieństwo zdarzenia ($V_i = 1$). Innymi słowy ($V_i \in \{0,1\}$) są wartościami zmiennej losowej (V_i) z rozkładem prawdopodobieństwa (2) i (3):

$$P(V_i = 1) = P(V_i) \quad (2)$$

oraz

$$P(V_i = 0) = P(\sim V_i) = 1 - P(V_i = 1) \quad (3)$$

Jeśli ekspert w czasie zestawiania (tworzenia) bazy wiedzy ekspertowej podaje zbiór możliwych faktów ($V_{i,j}$), z których jeden i tylko jeden jest prawdziwy. To wówczas fakt losowy (V_i) scharakteryzowany jest zbiorem prawdopodobieństw zestawionych odpowiednio dla tego faktu (4):

$$V_i = \{V_{i,1}, V_{i,2}, \dots, V_{i,j}\} \text{ oraz } P_{V_{i,j}} = (P_{V_{i,1}}, P_{V_{i,2}}, \dots, P_{V_{i,j}}) \quad (4)$$

gdzie:

$$P_{V_{i,j}} = P(V_{i,j}) = P(V_{i,j} = 1) \text{ dla } \sum_{j=1}^{i=1} P_{V_{i,j}} = 1 \quad (5)$$

Można zatem stwierdzić, że podobnie tak jak fakty również poprzez prawdopodobieństwo scharakteryzować można zdania losowe. Reguła wnioskowania (zdanie losowe) w metodzie probabilistycznej w postaci ogólnej przedstawia zależność:

$$RW = \langle \alpha, F, P_V \rangle, P_{V_{i,j}} = (P_{V_{i,1}}, P_{V_{i,2}}, \dots, P_{V_{i,j}}) \quad (6)$$

Istota opisywanej metody przedstawiono w przykładzie. Jeśli w opisie diagnostycznym badanego obiektu-sygnалу diagnostycznego (zdarzenia) występują fakty przedstawione jako: (F, F_c, F_b) wyrażone poprzez:

$$C = F(\alpha), V = F_c(\alpha), F_b(\alpha) \quad (7)$$

gdzie: $C = (C_1, C_2, \dots, C_s)$, $V = (V_1, V_2, \dots, V_k)$ oraz dane są prawdopodobieństwa (P_{C_i}) dla $i = \overline{1, S}$ oraz P_{V_i} dla $i = \overline{1, K}$ dla których zachodzą zależności:

$$\begin{aligned} C_{i,s} &= \{C_{i,1}, C_{i,2}, \dots, C_{i,s}\} \\ P_{C_{i,s}} &= (P_{C_{i,1}}, P_{C_{i,2}}, \dots, P_{C_{i,s}}) \\ P_{C_{i,s}} &= P(C_{i,s}) \end{aligned} \quad (8)$$

oraz $\sum_{s=1}^{i=1} P_{C_{i,s}} = 1$. Zadanie eksperta, w realizowanym procesie wnioskowania ekspertowego

(diagnostycznego) polega na wypracowaniu decyzji $B = \{B_1, B_2, \dots, B_r\}$, która następnie musi wygenerować (utworzyć) zbiór decyzji (prawdopodobieństw) wynikowych-pośrednich, które przedstawiono w postaci zależności:

$$P_b = (P_{b_1}, P_{b_2}, \dots, P_{b_r}) \quad (9)$$

$$P_{b_j} = P(B_j) = P(b_j = 1) \text{ dla } \sum_{j=1}^r P_{b_j} \quad (10)$$

Proces wnioskowania staje się bardziej skomplikowany, gdy dla każdego zaistniałego faktu zachodzi możliwy mu zbiór symptomów ($S_i = K_i = 2$) oznacza to, że formułowane są pojedyncze fakty (V_i) i zdania wejściowe (C_i) z odpowiednimi prawdopodobieństwami ($P_{V_1}, P_{V_2}, \dots, P_{V_k}$) oraz ($P_{C_1}, P_{C_2}, \dots, P_{C_s}$). Należy wówczas wyznaczyć wartość wielkości (P_b), która jest prawdopodobieństwem prawdziwości zdania (B) sformułowanego przez eksperta (jako zdania głównego). Wcześniej zbiór (B) oznaczał, że dla podanego przez eksperta faktów (F_b) okazały się możliwe różne zdania (B_j), a system ekspertowy powinien podać-zaprezentować te wszystkie możliwości rozwiązania (wyniki) danego zadania wraz z ich prawdopodobieństwami-współczynnikami prawdziwości. Problem ten często jest rozwiązywany poprzez wykorzystanie prawdopodobieństw warunkowych.

Jeżeli występuje przypadek taki, że w analizie ekspertowej wyróżniono pojedyncze zdania losowe (C_i) dla którego podane są prawdopodobieństwa (P_{C_i}), ekspert (moduł wnioskujący) zakłada wówczas możliwość prawdziwości implikacji w postaci:

$$H_{i-1} \wedge C_i \Rightarrow H_i \quad (11)$$

oraz

$$\sim (H_{i-j} \wedge C_i) \Rightarrow H_i \text{ dla } i > 1 \quad (12)$$

Analogicznie można stwierdzić, że dla zadania (C_i) istnieje fakt (H_i) dla, którego jest słuszna zależność ($C_i \Rightarrow H_i$), stąd dla każdego i -tego faktu (V_i) występują rozwiązania tego problemu ($K_i = 4$), czyli istnieją cztery możliwości, z których wystarczy rozpatrywać dwie poprzednio wymienione implikacje, bo dwie pozostałe są ich negacjami. W rezultacie każdy i -ty fakt (V_i) jest scharakteryzowany parą prawdopodobieństw oznaczających prawdziwość obu wymienionych implikacji, czyli parą prawdopodobieństw warunkowych:

$$\begin{aligned} P(H_1 / C) &= P_{V_{11}} \\ P(H_1 / \sim C_i) &= P_{V_{12}} \\ P(H_1 / C_i) &= P_{V_{i1}} \\ P(H_1 / \sim C_i) &= P_{V_{i2}} \end{aligned} \quad (13)$$

Przy założeniu, że (C_1, C_2, \dots, C_s) są niezależne, to znaczy istnieje zależność (14)

$$P(C_1 \wedge C_2 \wedge \dots \wedge C_s) = P_{c_1} \cdot P_{c_2} \cdot \dots \cdot P_{c_s} \quad (14)$$

stąd otrzymano zależność (4.33)

$$P(H_1) = P_{h_1} = P_{c_1} \cdot P_{V_{11}} + (1 - P_{c_1}) \cdot P_{V_{12}} \quad (15)$$

i ogólnie można stwierdzić, że istnieje zależność w następującej postaci:

$$P(H_1) = P_{h_i} = P_{c_i} \cdot P_{h-1} P_{V1i} + (1 - P_{c_i} \cdot P_{h-1} P_{V1i}) \cdot P_{C_{i2}} \quad i = \overline{2, k} \quad (16)$$

Zależność (16) umożliwia wyznaczenie kolejnych wartości prawdopodobieństw (P_{h_i}), aż do wartości ($P_{h_k} = P_b$), stąd można otrzymać (wyznaczyć) ogólną postać zależności (P_b). Wielkości (C_i i H_i) mogą być scharakteryzowane zbiorami możliwych zdań i odpowiednimi dla nich prawdopodobieństwami. Dla uproszczenia zapisu można przyjąć zasadę, że występuje tu tylko jedno zdanie wejściowe $C_1 = \{C_{11}, C_{12}, \dots, C_{1s}\}$, któremu przypisano pewien możliwy dla niego zbiór (zestaw) prawdopodobieństw:

$$P_{C_1} = (P_{C_{11}}, P_{C_{12}}, \dots, P_{C_{1s}}) \quad (17)$$

oraz dla zbioru możliwych faktów w postaci:

$$H_1 = \{H_{11}, H_{12}, \dots, H_{1n}\} \quad (18)$$

istnieje zbiór prawdopodobieństw dla faktów (zależność (19)):

$$P_{h_1} = (P_{h_{11}}, P_{h_{12}}, \dots, P_{h_{1n}}) \quad i = \overline{1, k} \quad (19)$$

wówczas na podstawie zależności (19) utworzono zbiór prawdopodobieństw uogólnionych takich, że:

$$P_{C_1}^T = M_1 P_{C_1}^T, P_{h_1}^T = M_1 P_{h_1}^T \quad (20)$$

gdzie: (P_{C_1}) i (P_{h_1}) traktowane są jako macierze jednowierszowe, a macierze wynikowe [M] są macierzami prawdopodobieństw warunkowych:

$$M_1 = [P(H_{1,j} / C_{1,l})] \quad j = \overline{1, N}, \quad l = \overline{1, s} \quad (21)$$

$$M_i = [P(H_{i,j} / C_{i,l})] \quad (22)$$

ostatecznie wyznaczono zależność:

$$P_{h_1}^T = P_{b_1}^T = M_1 P_{C_1}^T, M_2 P_{C_2}^T, \dots, M_k P_{C_k}^T = M P_{C_i}^T \quad (23)$$

W zaprezentowanym opisie metody probabilistycznej przedstawiono, że wnioskowanie ekspertowe w niej ma charakter losowy wtedy, gdy zarówno przesłanka (objaw) jak i wniosek (reguła) są zdaniami losowymi. Znając zatem prawdopodobieństwa, które są znane dla różnych możliwych objawów, wyznacza się prawdopodobieństwa dla różnych możliwych wniosków (reguła). Jeżeli w wyniku obserwacji możliwe jest określenie prawdziwości zdania ($C_{1,l}$) to znaczy, że ($P_{C_{1,l}} = 1$) i wówczas jest zależność:

$$P_{b,j} = m_{j,l}; j = \overline{1, N_k} \quad (24)$$

gdzie: $m_{j,l}$ – jest elementem w j -tym wierszu i l -tej kolumnie macierzy (a), czyli prawdopodobieństwem prawdziwości implikacji $C_{j,l} \Rightarrow B_j$ lub inaczej prawdopodobieństwem warunkowym $P(B_j / C_{j,l})$

Z podanych przez system ekspertowy możliwości (wyników) (B_j) użytkownik wybiera zdanie (regułę) najbardziej prawdopodobną, dla której wartość prawdopodobieństwa (P_{bj}) jest największa. W opisanym metodzie probabilistycznej przedstawiono, że podstawą tej metody wnioskowania są objawy oraz możliwe zdania (reguły) z określonym prawdopodobieństwem z przedziału (0,1). Problem wnioskowań w tym wypadku polega na przyporządkowaniu znanym objawom (C_i) najbardziej prawdopodobnym zdarzeniom (regułom) z jak największym prawdopodobieństwem (P_{ci}).

3. BADANIE JAKOŚCI ODNAWIANIA OBIEKTU TECHNICZNEGO DLA PROBABILISTYCZNEGO SPOSOBU BUDOWANIA BAZY WIEDZY EKSPERTOWEJ

Badanie jakości odnawiania obiektu w systemie obsługiwanym zorganizowanym na podstawie informacji obsługowej przeprowadzono dla trzech następujących sposobów wnioskowania ekspertowego:

1. Dedukcyjnego-model I.
2. Probabilistycznego-model II.
3. Indukcyjnego-model III.

Isotą badania jakościowego procesu profilaktyki obiektu technicznego jest badaniem dwuetapowym. W pierwszym etapie jest wykonana odnowa (profilaktyka) obiektu przy danym sposobie wnioskowania ekspertowego. Dopiero w drugim etapie badania jest oceniana jakość tej profilaktyki. W badaniu przyjęto założenie, każdy z tych trzech sposobów wnioskowania zapewnia odnowienie pełne obiektu. Wykonaną profilaktykę obiektu przeprowadzono przy wykorzystaniu wektora testowego informacji obsługowej wyznaczonej dla metody probabilistycznej wnioskowania ekspertowego (model II), który ma następującą postać zależności (25):

$$\{M_E(e_{i,j})_{II}\} = [M_E(e_{1,1})_{II}, \dots, M_E(e_{i,j})_{II}, \dots, M_E(e_{I,J})_{II}] \quad (25)$$

gdzie: $M_E(e_{i,j})_{II}$ – obsługowa informacja obsługowe j -tego elementu w i -tym poziomie struktury obsługowej dla modelu II.

Po wykonanej profilaktyce obiekt techniczny, w tym rodzaju badania poddawany jest diagnozowaniu kontrolnemu z wykorzystaniem programu DIAG. Uzyskane wyniki diagnozowania obiektu dla modelu II zaprezentowano w postaci (Rys. 1).

Definicja: Funkcja jakościowa oceny procesu profilaktyki obiektu technicznego (F_r) w systemie obsługiwanym jest to wielkość, która określa zbiór w pełni odnowionych elementów funkcjonalnych obiektu technicznego, w stosunku do zbioru wszystkich ele-

mentów obiektu, które poddano procesowi obsługi. Wielkość tą przedstawiono w postaci zależności:

$$F_r = \frac{n - k}{n} \cdot 100\% \quad (26)$$

gdzie: n - zbiór wszystkich elementów obiektu poddanych procesowi obsługi-zbiór elementów wyznaczających strukturę usługową obiektu $\{e_{i,j}\}$, k - zbiór w pełni odnowionych (przy rozpoznanych stanach zdadności) elementów podstawowych obiektu.

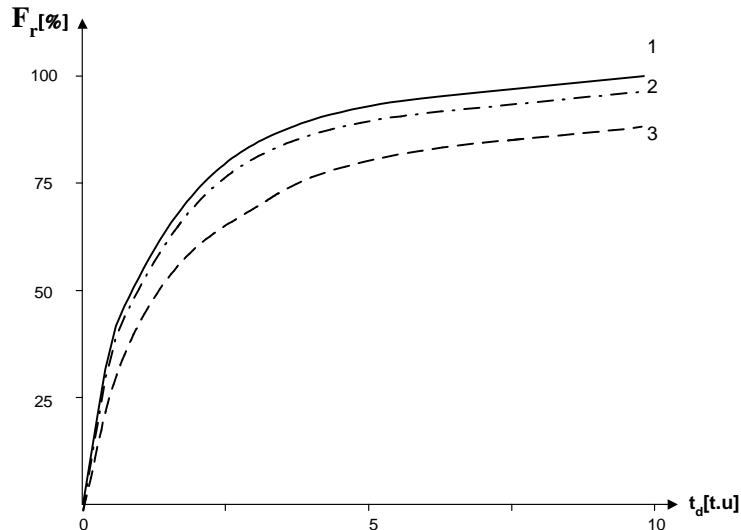
Uzyskane wyniki z badań w eksperymencie poddano analizie i opracowaniu przy wykorzystaniu zależności (26). Wykreślone wielkości przedstawiono graficznie na (Rys. 2).

Model	Data	Metrics	Deviation	Variances	N-metrics	Distribution	Probabilities	ZG Probabilities	WG Weight	Classification
p	2									
2	2	2	2	2	2	2				
2										
2	2	2	0							
2										
2	2									
2	1	2	2	1						

Rys. 1. Postać kontrolnej informacji obsługowej z wykorzystaniem programu DIAG po wykonanej profilaktyce (model II)

4. WNIOSKI

Z analizy jakości odnawiania obiektu w systemie obsługi zorganizowanym na podstawie testowych informacji obsługowej dla (modeli I; II i III) (Rys. 2) wynika, że najbardziej wiarygodną ekspertową informację obsługową uzyskano dla (modelu I) = 98%, nie co gorsza jakościowo jest informacja obsługowa dla (modelu III) = 96%. Natomiast dość znacznie jakościowo odbiega od tych dwóch model informacja dla (modelu II) = 94%. Na podstawie przeprowadzonych badań trzech sposobów wnioskowania ekspertowego: dedukcyjnego, probabilistycznego i indukcyjnego największe efekty w zakresie podejmowanych decyzji posiada wnioskowanie dedukcyjne.



Rys. 2. Wykres funkcji jakościowej odnawiania elementów obiektu technicznego (F_r) przez system obsługowy, gdzie: 1 – wykres dla wektora obsługowej informacji testującej (model I); 2 – wykres dla wektora obsługowej informacji testującej (model III); 3 – wykres dla wektora obsługowej informacji testującej (model II); $t_p(e_{i,j})$ – jest czasem odnawiania j -tych elementu podstawowego w i -tych zespołach funkcjonalnych obiektu

BIBLIOGRAFIA

- [1] Będkowski L., Dąbrowski T.: Podstawy eksploatacji cz. 2. Wyd. WAT, Warszawa 2006, str. 187.
- [2] Duer S.: System ekspertowy wykorzystujący trójwartościową informację diagnostyczną wspomagający obsługiwane złożonego obiektu technicznego. ZEM, Z. 4(152) VOL. 42, 2007, str. 195-208.
- [3] Duer S.: An algorithm for the diagnosis of reparable technical objects utilizing artificial neural Network. ZEM, Vol. 43, No. 1(53) 2008, pp. 101-113.
- [4] Duer S.: Determination of a diagnostic information of a reparable technical object on the basis of a functional and diagnostic analysis on example of a car engine. ZEM, Vol. 43, No. 4(156) 2008, pp. 85-94.
- [5] Duer S.: Artificial Neural Network-based technique for operation process control of a technical object. Defence Science Journal, DESIDOC, Vol. 59, No. 3, May 2009, pp. 305-313.
- [6] Duer S.: System ekspertowy, ze sztuczną siecią neuronową obsługujący zestaw stacji radiolokacyjnej. VII Krajowa Konferencja Inżynierii Wiedzy i Systemów Ekspertowych, Politechnika Wroclawska, Instytut Informatyki, 23–25 czerwiec, Wrocław, 2009, str. 377-388.
- [7] Nakagawa T.: Maintenance Theory of Reliability. Springer – Verlag London Limited 2005, p. 264.