

Andrzej Fortuna

Instytut Automatyki Politechniki Śląskiej

## MIKROKOMPUTEROWE WSPOMAGANIE ANALIZY PRZYCZYN ZDARZEŃ LOSOWYCH. \*

Streszczenie. W pracy opisano pakiet programów służących do wspomaganie analizy przyczyn zdarzeń, poprzez analizę ich częstości (intensywności). Przedstawiono model matematyczny systemu składającego się ze zbioru obiektów, oraz powiązanych z tym zbiorem zdarzeń o określonych przyczynach oraz zaproponowano dwa algorytmy przetwarzania danych. Oprogramowanie zrealizowano w oparciu o technikę relacyjnych baz danych.

### 1. Wstęp

W wielu dziedzinach badań występuje problem poszukiwania przyczyn zaistniałych zdarzeń, szczególnie w odniesieniu do awarii, uszkodzeń, itp. Charakterystyczną cechą problemu jest w dużym stopniu jednolitość algorytmów przetwarzania danych, natomiast cechą odróżniającą są odmienne struktury danych, wynikające z odmienności opisu problemu.

W artykule przedstawiono propozycję uniwersalnego narzędzia informatycznego, zbudowanego w oparciu o zaprezentowany w pracy model matematyczny, które wspomaga przeprowadzenie analizy i które jest niezależne od cech i atrybutów badanych obiektów i zdarzeń.

Opis problemu tworzony jest przez użytkownika oprogramowania, w którym wymaga się jedynie jednorodności analizowanych obiektów w sensie ich opisu (czyli struktur danych opisujących te obiekty) i powiązanych z nimi zdarzeń. Przyjęto założenie, że w skład opisu pojedynczego zdarzenia wchodzi zawsze:

- przyczyna jego wystąpienia,
- element obiektu, z którym powiązane jest zdarzenie.

Tak zdefiniowany minimalny opis zdarzenia określa obszar zastosowań oprogramowania do zadań, w których każde zdarzenie jest determinowane określoną przyczyną, a stopień pewności takiego twierdzenia wynosi 1. Analiza częstości (intensywności) występowania tego typu zdarzeń prowadzona post factum dotyczyć może analizy zdarzeń w powiązaniu z cechami opisu obiektów i zdarzeń.

Przykładami tego typu analizy może być np.:

\* Praca finansowana w ramach programu RRI-14

- analiza przyczyn uszkodzeń (zdarzeń) pewnej klasy urządzeń (obiektów) w zależności od charakterystycznych cech (parametrów) obiektów i cech uszkodzeń.
  - analiza zachorowalności na badaną chorobę wśród ankietowanej grupy pacjentów (obiektów) w zależności od ich indywidualnych cech. itp.
- Inne aplikacje wraz z przykładami analizy zaprezentowano w [1].

## 2. Opis matematyczny.

Przedmiotem analizy jest system o następujących własnościach :

- a. Istnieje zbiór obiektów  $O$  jednorodnych, tzn. takich, które można opisać za pomocą skończonego zbioru cech  $C^O$  o dowolnych wartościach.

$O = \{o_i\}_{i=1..L_o}$	- zbiór obiektów,
$o_i$	- i-ty obiekt w zbiorze $O$ ,
$L_o$	- liczba obiektów,
$C^O = \{c_k^O\}_{k=1..L_c^O}$	- zbiór cech opisu obiektu,
$c_k^O$	- k-ta cecha opisu obiektu w zbiorze $C^O$ ,
$L_c^O$	- liczba cech opisu obiektu ze zbioru $O$ .

- b. Każdemu obiektowi  $o_i$  może zostać przyporządkowany zbiór zdarzeń losowych  $\Omega^i$ , który również można opisać za pomocą skończonego zbioru cech  $C^{\Omega}$ ; suma wszystkich zbiorów  $\Omega^i$  tworzy zbiór zdarzeń losowych  $\Omega$ .

$\Omega^i = \{\omega_j^i\}_{j=1..L_{\omega}^i}$	- zbiór zdarzeń powiązanych z obiektem $o_i$ ,
$\omega_j^i$	- j-te zdarzenie, związane z obiektem $o_i$ ,
$L_{\omega}^i$	- liczba zdarzeń związanych z obiektem $o_i$ ,
$\Omega = \{\omega_j\}_{j=1..L_{\omega}}$	- zbiór zdarzeń losowych,
$\omega_j$	- j-te zdarzenie w zbiorze zdarzeń $\Omega$ ,
$L_{\omega}$	- ogólna liczba zdarzeń,
$C^{\Omega} = \{c_k^{\Omega}\}_{k=1..L_c^{\Omega}}$	- zbiór cech opisu zdarzenia,

$c_k^\Omega$  - k-ta cecha opisu zdarzenia,

$L_c^\Omega$  - liczba cech zdarzeń.

$$c. \quad \bigcup_{i=1}^{i=L} \Omega^i = \Omega$$

tzn. nie istnieje takie zdarzenie, które nie jest powiązane z żadnym obiektem ze zbioru  $O$ .

$$d. \quad \forall i \forall k \quad \Omega^i \cap \Omega^k = \emptyset,$$

tzn. nie istnieje zdarzenie w zbiorze  $\Omega$  związane więcej niż z jednym obiektem.

e. Zbiór cech opisu zdarzenia  $C^\Omega$  składa się z co najmniej dwóch elementów, które zawsze występują w opisie zdarzenia:

- przyczyna - cecha opisu charakteryzująca przyczynę zajścia zdarzenia,
- element - cecha opisu charakteryzująca dowolną część obiektu, która jest powiązana ze zdarzeniem.

f. Istnieje skończony zbiór przyczyn występowania zdarzeń

$$P = \langle p_i \rangle_{i=1..L_p}$$

$p_i$  - i-ta przyczyna zdarzenia,

$L_p$  - liczba przyczyn.

g. Istnieje skończony zbiór elementów, które określają składowe części obiektu

$$E = \langle o_i \rangle_{i=1..L_o}$$

$o_i$  - i-ty element obiektu,

$L_o$  - liczba elementów.

h. Każdemu obiektowi  $o_i$  może zostać przyporządkowana pewna liczba dodatkowych zbiorów danych  $D^{ij}$ , opisanych zbiorem cech  $C^j$  [1]

$$D^{ij} = \langle d_k^{ij} \rangle_{k=1..L_d^{ij}}$$

$d_k^{ij}$  - k-ty element j-tego zbioru dodatkowych danych dla obiektu  $o_i$ .

$L_d^{ij}$  - liczba elementów zbioru  $D^{ij}$  dla obiektu  $o_i$ .

$$C^j = \langle c_k^j \rangle_{k=1..L_c^j}$$

$c_k^j$  - k-ta cecha dodatkowego j-tego opisu obiektu,

$L_c^j$  - liczba cech j-tego opisu,

$D^j = \langle d_k^j \rangle_{k=1..L_d^j}$  - zbiór zawierający dodatkowy j-ty opis dla wszystkich obiektów ze zbioru  $O$ ,

$d_k^j$  - k-ty element zbioru  $D^j$ .

$L_d^j$  - liczba cech dodatkowego j-tego opisu.

Podobnie jak w przypadku c i d spełnione muszą być warunki

$$\forall_j \bigcup_{i=1}^{i=L} D^{ij} = D^j \quad \text{oraz} \quad \forall_i \forall_k \forall_j D^{ij} \cap D^{kj} = \emptyset.$$

### 3. Sformułowanie problemu

#### 3.1. Struktury danych

Dla opisanego w p.2. systemu, zgodnie z uwagami zamieszczonymi we wstępie opracowania, zadanie analizy postanowiono rozdzielić na dwie zależne od siebie części:

- A. zadanie formułowania opisu systemu,
- B. zadanie ewidencjonowania ( gromadzenia ) danych i ich analizy.

#### Ad A. Formułowanie opisu systemu

Formułowanie opisu systemu, zgodnie z modelem matematycznym przedstawionym w punkcie 2, obejmuje :

- określenie  $C^o$ , czyli zbioru cech opisujących obiekt,
- określenie  $C^{\Omega}$ , czyli zbioru cech opisujących zdarzenie,
- określenie  $C^j$ , czyli J zbiorów cech opisujących dodatkowe właściwości obiektu (gdzie wymaga tego specyfika problemu [1]),
- utworzenie zbioru P zawierającego wszystkie występujące w problemie przyczyny zdarzeń,
- utworzenie zbioru E zawierającego wszystkie wymagane elementy obiektów, z którymi mogą być powiązane zdarzenia.

Ad B. Ewidencjonowanie danych

Ewidencjonowanie danych obejmuje następujące funkcje:

- tworzenie zbioru obiektów  $O$ ,

$$V_i^O = \langle v_{ki}^O \rangle_{k=1..L}^O \quad - \text{zbiór wartości cech } i\text{-tego obiektu}$$

$$v_{ki}^O \quad - \text{wartość } k\text{-tej cechy } i\text{-tego obiektu}$$

- tworzenie zbioru zdarzeń  $\Omega$ ,

$$V_i^\Omega = \langle v_{ki}^\Omega \rangle_{k=1..L}^\Omega \quad - \text{zbiór wartości cech } i\text{-tego zdarzenia}$$

$$v_{ki}^\Omega \quad - \text{wartość } k\text{-tej cechy } i\text{-tego zdarzenia}$$

- tworzenie zbiorów  $D^j$  opisujących dodatkowe właściwości obiektów, zgodnie z definicją cech dla tych zbiorów.

$$V_i^j = \langle v_{ki}^j \rangle_{k=1..L}^j \quad - \text{zbiór wartości cech } j\text{-tego opisu dla } i\text{-tego obiektu}$$

$$v_{ki}^j \quad - \text{wartość } k\text{-tej cechy } j\text{-tego opisu dla } i\text{-tego obiektu}$$

Powyższe funkcje posiadając charakter ewidencyjny muszą zapewnić :

- a. możliwość definiowania nowego elementu każdego zbioru,
- b. możliwość zmiany wartości cech dowolnego elementu wybranego zbioru,
- c. możliwość usunięcia dowolnego elementu wybranego zbioru,
- d. możliwość przeglądu całego zbioru lub jego części.

3.2. Algorytmy

W pracy jako podstawę analizy proponuje się algorytmy wyznaczania intensywności zdarzeń, które można podzielić na dwie grupy:

1. analizę intensywności zdarzeń dla wskazanego podzbioru obiektów  $O$ ,
2. analizę intensywności zdarzeń dla wskazanego podzbioru elementów  $E$  i przyczyn  $P$ .

Ad 1.

Wielkościami wejściowymi analizy intensywności zdarzeń dla wskazanego podzbioru obiektów są :

- a. podzbiór obiektów  $O_{kryt}$  taki, że  $O_{kryt} \subset O$ , określony przez sformułowane kryterium logiczne, definiujące zbiór dopuszczalnych wartości cech  $K$  obiektu

$$K = \langle k_l \rangle_{l=1..L_c^o}$$

$k_l$  - zbiór dopuszczalnych wartości  $l$ -tej cechy obiektu.

Przynależność obiektu  $o_i$  do  $O_{kryt}$  jest definiowana następującą formułą:

$$\forall_{1 \leq i \leq L^o} \quad \forall_{1 \leq l \leq L_c^o} \quad v_{li}^o \in k_l \Rightarrow o_i \in O_{kryt}$$

- b. podzbiór zdarzeń  $\Omega_{var}$  taki, że  $\Omega_{var} \subset \Omega$  określony poprzez sformułowanie kryterium logicznego, definiujące zbiory dopuszczalnych wartości cech  $W$  zdarzenia.

$$W = \langle w_l \rangle_{l=1..L_c^\Omega}$$

$w_l$  - zbiór dopuszczalnych wartości  $l$ -tej cechy zdarzenia

Podzbiór  $\Omega_{var}$  definiuje formuła

$$\forall_{1 \leq i \leq L^\Omega} \quad \forall_{1 \leq l \leq L_c^\Omega} \quad v_{li}^\Omega \in w_l \Rightarrow \omega_i \in \Omega_{var}$$

- c. podzbiór  $P_z \subset P$ ,

- d. podzbiór  $E_z \subset E$ ,

- e. funkcja  $f(o_i, \omega_j)$  realizująca standaryzację intensywności zdarzeń w zależności od indywidualnych cech obiektu  $i$  i zdarzenia; przyjęto założenie, że funkcja ta ma postać:

$$f(o_i, \omega_j) = 1.$$

Dla każdego obiektu  $o_i$  spełniającego warunek  $o_i \in O_{kryt}$  wyszukiwane są zdarzenia  $\omega_j \in \Omega_{var} \cap \Omega$ , których przyczyna należy do zbioru  $P_z$ , a określony element do  $E_z$ , i wyznaczana jest intensywność zdarzeń wg wzoru:

$$N_i = N_i + f(o_i, \omega_j) \quad (1)$$

Zgodnie z uwagą dotyczącą wartości funkcji standaryzującej  $f$  powyższy wzór upraszcza się do postaci:

$$N_i = N_i + 1 \quad (2)$$

czyli w tym przypadku intensywność zdarzeń równoważna jest liczbie zdarzeń związanych z  $i$ -tym obiektem, spełniających powyższe warunki.

Ad 2.

W funkcji analizy intensywności zdarzeń dla wskazanego podzbioru elementów  $i$  przyczyn muszą być określone te same wielkości jak w przypadku

poprzednim. Różnica wynika z przyjęcia innego algorytmu analizy intensywności, w którym nie są rozróżniane poszczególne obiekty, natomiast intensywność zdarzeń jest powiązana z elementami bądź przyczynami zdarzeń.

Dla każdego elementu zbioru  $E_x \times P_x$  lub  $(P_x \times E_x)$  wyznaczane są obiekty  $o_i \in O_{kryt}$  i powiązane z nimi zdarzenia  $\omega_j \in \Omega_{var}$  wywołane wskazanymi przyczynami i wyróżniające wskazane elementy obiektów, i wyznaczana jest intensywność zdarzeń według (2).

#### 4. Rozwiązanie problemu

Do rozwiązania formułowania opisu systemu scharakteryzowanego w p.2. wykorzystano relacyjne bazy danych, które w uproszczeniu można przedstawić w postaci tabeli, której kolumny określają cechy obiektu, natomiast wiersze określają kolejne obiekty. Wszystkie zbiory wymienione w p.2. posiadają takie właściwości. Liczba wierszy tabeli (liczba rekordów w bazie) w przypadku zbioru  $O$  określa liczbę zdefiniowanych w systemie obiektów  $L^0$ , natomiast kolejne kolumny (pola rekordu) określają cechy obiektów.

Cechy obiektu mogą być wielkościami o charakterze :

- numerycznym - pojemność skokowa silnika, liczba cylindrów, zużycie paliwa, długość odcinka przesyłu danych itp.
- opisowym - nazwa producenta, uwagi eksploatacyjne,
- logicznym - czy silnik był remontowany,
- czasowym - data produkcji, data wymiany oleju itp.

W tak sformułowanym problemie wyróżnia się następujące bazy danych :

1. baza danych zawierająca informacje o analizowanych obiektach, zwana bazą podstawową,
2. baza danych zawierająca informacje o wszystkich zdarzeniach, zwana bazą zdarzeniową,
3. bazy danych zawierające dodatkowe informacje o obiektach, zwane bazami pomocniczymi.

Powiązanie tych baz w spójny system jest realizowane poprzez wyróżnioną, wspólną dla wszystkich baz cechę, która jest wprowadzana automatycznie do struktury każdej bazy.

Opracowany zestaw programów służy do realizacji zadań określonych w p.3. Zgodnie z podziałem funkcji opisanym w tym punkcie opracowane zostały dwa programy o roboczych nazwach INSTAL i OPER, realizujące wskazane tam cele.

Oba programy zostały napisane w języku CLIPPER i skompilowane za pomocą kompilatora CLIPPER'87 firmy Nantucket. Przeznaczone są do uruchomienia na mikrokomputerach klasy IBM PC XT/AT pracujących pod kontrolą systemu operacyjnego MSXPC-DOS w wersji 3.10 lub późniejszej. Wymagają pełnego zakresu pamięci operacyjnej, tj. 640 kB. Do działania

programu niezbędne jest także wyposażenie komputera w dysk twardy o pojemności zależnej od przewidywanej wielkości baz, przeciętnie ok. 2MB. Programy pracują w trybie tekstowym, a więc mogą działać na wszystkich standardowych kartach sterowników ekranu ( m.in. CGA, EGA, Hercules). Przystosowane są do współpracy z drukarkami zgodnymi w trybie tekstowym ze standardem IBM/EPSON.

Należy podkreślić, że oprogramowanie ma charakter interaktywny i nie generuje wniosków o stanie systemu w sposób automatyczny, natomiast może dawać szybką odpowiedź na sformułowane zapytania w postaci odpowiednich zestawień oraz wyników analizy intensywności zdarzeń.

#### LITERATURA

- [1] H.Kowalowski i in.: Sprawozdanie z pracy naukowo-badawczej. "System komputerowego wspomagania analizy częstości zdarzeń losowych." Program RRI-14.
- [2] C.Delobel, M.Adiba: Relacyjne bazy danych. PWN, Warszawa 1989.
- [3] Nantucket Clipper. Reference Manual.

Recenzent: Doc.dr inż.J.Swiątek

Wpłynęło do Redakcji do 1990-04-30.

#### MICROCOMPUTER-AIDED ANALYSIS OF THE CAUSES OF STOCHASTIC EVENTS

##### S u m m a r y

A software package for computer-aided analysis of the causes of stochastic events based on the frequency analysis, is presented in the paper. A mathematical model of a system consisting of a set of objects and corresponding set of events with specific causes is given. Two algorithms of data processing are proposed. The software was written using relation databases technique.

#### МИКРОКОМПЬЮТЕРНЫЙ АНАЛИЗ ПРИЧИН СЛУЧАЙНЫХ ЯВЛЕНИЙ

##### Р е з ю м е

В работе дано описание компьютерных программ для вспомогания анализа причин явлений, путем анализа их частоты (интенсивности). Дана математическая модель системы состоящей из множества объектов и связанных с этим множеством явлении с определенными причинами. Предложено два алгоритма обработки данных. Опрограммированное реализовано опираясь на реляционных базах данных.