

KAZIMIERZ DYRKA

CENTRUM OB LICZENIOWE
POLITECHNIKI WROCLAWSKIEJ

WROCLAW

ANALIZA STATYSTYCZNA DANYCH EKSPLOATACYJNYCH

Model systemu

Podano formalny opis obiektów i schematy podstawowych procedur systemu analizy statystycznej danych w oparciu o model danych systemu informacyjno-informatycznego przedsiębiorstwa przemysłowego. Model wykorzystano przy budowie podsystemu informatycznego.

1. Rola i miejsce systemu ASDE w systemie informatycznym

Identyfikacja i poznanie obiektów i zjawisk eksploatacji jest jednym z pierwszych i koniecznych przedsięwzięć w programie optymalizacji eksploatacji maszyn i pojazdów. Pozwala bowiem określić obiekt /zjawisko/ poprzez jego własności, co jest niezbędne nie tylko do rozwiązania, ale i postawienia zadań optymalizacji.

Wprowadzenie do systemów zarządzania eksploatacją maszyn i pojazdów nowoczesnych środków technicznych - maszyn cyfrowych i systemów informatycznych stwarza materialne przesłanki zastosowania do praktyki zarządzania metod statystyki, teorii podejmowania decyzji itp.

System "Analiza Statystyczna Danych Eksploatacyjnych", nazywany dalej systemem ASDE, ma stanowić praktyczne narzędzie syntetycznego opisu danych i modelowania danych, a w dalszym ciągu - oceny obiektów i procesów eksploatacji na bazie statystycznych modeli uzyskanych na drodze analizy danych zbieranych w systemie informacyjnym przedsiębiorstwa.

Aby określić rolę systemu informatycznego w zakresie badań obiektów i zjawisk, założymy, że:

1. Istnieje system informacyjny wraz z systemem informatycznym ewidencji obiektów i zdarzeń.
2. Wyniki dowolnego planowanego eksperymentu zawarte są w bazie danych systemu informatycznego lub dadzą się uzyskać z informacji tam zawartych.

Planowanie eksperymentu w systemie ASDE rozumiane jest w sensie określenia populacji i jej cech podlegających badaniu, wielkości i sposobu wyboru próbki, metody pomiaru cech. Zakres możliwych do przeprowadzenia eksperymentów określa system informatyczny, którego istnienie postulowane jest w punkcie 1. i zależy, między innymi, od zakresu i stopnia szczegółowości zbieranych danych.

Wykonanie eksperymentu ma w systemie znaczenie czysto pojęciowe. W założeniu systemu eksperyment jest wykonywany w trakcie eksploatacji. Z chwilą zaplanowania eksperymentu jest on wykonany, a wyniki jego są zawarte w bazie systemu.

Analiza wyników eksperymentu polega na zastosowaniu wybranych metod opisu danych oraz metod i algorytmów wnioskowania statystycznego /patrz [2] /.

Udostępnienie danych empirycznych, metod i algorytmów w omówionym zakresie - to podstawowe cele systemu.

Cele postawione przed systemem implikują zakres i treść prac informatycznych oraz zadania i elementy systemu informatycznego.

Ze względu na spełniane zadanie można wyróżnić dwa podstawowe elementy:

1. System definiowania obiektów
2. System opisu i analizy danych.

2. System definiowania obiektów

Niech $E = \{E_1, \dots, E_n\}$ będzie zbiorem obiektów systemu ewidencji,
 $X = \{X_1, \dots, X_m\}$ - zbiorem obiektów systemu ASDE, wreszcie
 $F = \{F_1, \dots, F_s\}$ - zbiorem przekształceń, takich że:

$$\bigvee_{i=1, \dots, m} \bigwedge_{k_1, \dots, k_1} F_{k_1} \cdot F_{k_2} \cdot \dots \cdot F_{k_1}(E) = X_i$$

$k_1 \in \{1, \dots, s\}$

Definicja. System $\langle E, F, X \rangle$ nazywamy systemem definiowania obiektów.

2.1. Obiekty systemu ewidencji

Dla dalszych rozważań przyjmijmy, że elementami zbioru E są^{1/}:

1/ Określenie zbioru obiektów systemu ewidencji przyjęto zgodnie z koncepcją ogólnego systemu informatycznego realizowanego w C.O. Politechniki Wrocławskiej dla branży górniczej.

1. Obiekty systemu bazowego eksploatacji.
2. Obiekty systemu zarządzania eksploatacją.
3. Warunki eksploatacji systemu bazowego.
4. Typy zachowań wyróżnionych obiektów systemu bazowego eksploatacji.
5. Zmienne elementarne.
6. Relacje.

Aby pojęcia definiowane w dalszym ciągu opracowania miały sens fizyczny, dopuszcza się następującą interpretację obiektów systemu ewidencji:

a/ Obiektami systemu bazowego eksploatacji maszyn i pojazdów mogą być:

- grupy maszyn,
- zespoły maszyn,
- części zamienne i materiały eksploatacyjne,
- obsługa maszyn,
- stanowiska remontowe itp.

b/ Obiektami systemu zarządzania eksploatacją mogą być:

- zjednoczenia,
- przedsiębiorstwa,
- komórki organizacyjne itp.

c/ Typy zachowań obiektów scharakteryzowane są przez elementy abstrakcyjnego zbioru S , zwanego dalej przestrzenią stanów, przy czym zakłada się, że:

/i/ obiekt może znajdować się tylko w jednym ze stanów zbioru

$$S = \{s_1, \dots, s_n\},$$

/ii/ wyróżniona jest klasa stanów określająca zachowanie obiektu sprawnego,

/iii/ przebywanie obiektu w stanie nieaktywnym^{2/} nie zmienia sprawności obiektu.

d/ Zmienne systemu ewidencji /zmienne elementarne/ opisują wyróżnione aspekty działania obiektów systemu bazowego lub obiektów systemu zarządzania. Zakłada się, że wszystkie występujące zmienne opisują globalne efekty działania w rozłącznych przedziałach czasu o długości Δ . Zmienne elementarnymi mogą być:

- ilość pracy wykonanej przez obiekt w przedziale czasu

$$\Delta_i = \langle i\Delta, (i+1)\Delta \rangle,$$

- ilość zużytych materiałów określonego typu przez obiekt w przedziale

$$\Delta^*i = \langle i\Delta^*, (i+1)\Delta^* \rangle,$$

- pracochłonność przy odnowie obiektu w przedziale

$$\Delta^{**}i = \langle i\Delta^{**}, (i+1)\Delta^{**} \rangle,$$

- czas pracy obiektu w przedziale czasu Δ_i itp.

2/ Stan $s_i \in S$ nazywamy aktywnym, jeśli istnieje stan $s_j \in S$ taki, że przejście obiektu ze stanu s_i do stanu s_j ma charakter losowy. Pozostałe stany z S nazywamy nieaktywnymi.

e/ Obiekty systemu ewidencji typu "relacje" określają zależności pomiędzy innymi obiektami tego systemu. W szczególności relacje określają związki obiektów systemu bazowego z obiektami zarządzania i zmiennych elementarnych z innymi obiektami systemu.

Relacje mogą być opisane przez funkcje zdaniowe np.

"obiekt 1 przebywał w przedziale czasu $\langle t_1, t_2 \rangle$ w stanie s",

"obiekt 1 jest częścią składową obiektu k", zdania np.

"zmienna $X_{11}(u)$ opisuje ilość pracy wykonanej przez obiekt 1 w przedziale czasu Δ " i itp.

Relacje reprezentowane są w systemie ewidencji poprzez odpowiednie struktury danych i system kodowy informacji.

2.2. Obiekty systemu ASDE

Zbiór obiektów systemu ASDE zdefiniowano indukcyjnie wg schematu:

1. Zmienne systemu ewidencji są obiektami systemu ASDE.
2. Jeśli każda ze zmiennych Z_1, \dots, Z_n jest obiektem systemu, to i zmienne

$$U_k = f_k(Z_1, \dots, Z_n)$$

są obiektami systemu. Funkcje f_k dobiera się tak, aby zmienne U_k miały określony sens fizyczny i były zmiennymi pewnego modelu, np. modelu teorii niezawodności.

Wprowadźmy oznaczenia:

- C_r - podzbiór zbioru liczb naturalnych o licznosci r ,
 L - zbiór obiektów systemu bazowego eksploatacji i systemu zarządzania,
 $K/1/$ - zbiór cech obiektu $1 \in L$,
 $X_{11}/k/$ - zmienne systemu ewidencji opisujące k -tą cechę obiektu 1 w przedziale czasu Δ ,

$$W = \bigcup_{1 \in L} \bigcup_{k \in K/1/} \bigcup_i X_{11}/k/ ,$$

- $M_i/\alpha/$ - podzbiór elementów zbioru L posiadających w chwili i własność " α ",
 $T_1/s/$ - zmienna systemu ewidencji określająca czas przebywania obiektu 1 w stanie s

$$T = \bigcup_{1 \in L} \bigcup_{s \in S} T_1/s/$$

- Q - relacja równoważności określona w przestrzeni stanów S ,
 S/Q - zbiór klas abstrakcji,
 s_k'' - elementy zbioru S/Q

- $N_{11}/s/$ - ilość wystąpień obiektu 1 w stanie s w przedziale czasu Δ i

$$N = \bigcup_{l \in L} \bigcup_{s \in S} \bigcup_1 N_{11}/s/$$

Niech X_V będzie zbiorem zmiennych takich, że:

1. $V \in X_V$.

2. $\bigvee_{l \in C_r} X_{11}/k/ \in X_V \Rightarrow \left(\sum_{l \in C_r} X_{11}/k/ \right) \in X_V$.

3. $\bigvee_{l \in M_1/\alpha/} X_{11}/k/ \in X_V \Rightarrow \left(\sum_{l \in M_1/\alpha/} X_{11}/k/ \right) \in X_V$.

Niech X_T będzie zbiorem zmiennych takich, że:

5. $T \subset X_T$

6. $\bigvee_{s \in S'_k} T_1/s/ \in X_T \Rightarrow \left(\sum_{s \in S'_k} T_1/s/ \right) \in X_T$.

7. $\bigvee_{l \in M_1/\alpha/} T_1/s/ \in X_T \Rightarrow \left(\sum_{l \in M_1/\alpha/} T_1/s/ \right) \in X_T$.

Niech X_N będzie zbiorem zmiennych takich, że:

8. $N \subset X_N$.

9. $\bigvee_{l \in C_r} N_{11}/s/ \in X_N \Rightarrow \left(\sum_{l \in C_r} N_{11}/s/ \right) \in X_N$.

10. $\bigvee_{l \in M_1/\alpha/} N_{11}/s/ \in X_N \Rightarrow \left(\sum_{l \in M_1/\alpha/} N_{11}/s/ \right) \in X_N$.

11. $\bigvee_{s \in S'_k} N_{11}/s/ \in X_N \Rightarrow N_{11}(s'_k) \in X_N$.

Niech \bar{X} będzie najmniejszym zbiorem zmiennych takich, że:

12. $X_V \subset \bar{X}, X_T \subset \bar{X}, X_N \subset \bar{X}$.

13. $Z_{11}(k'), Z_{11}(k'') \in \bar{X} \Rightarrow \frac{Z_{11}(k')}{Z_{11}(k'')} \in \bar{X}$

Definicja.^{3/} Dowolny podzbiór X zbioru \bar{X} nazywamy zbiorem obiektów systemu ASDE. Klasa obiektów zbioru X zdefiniowana powyżej może być, w miarę potrzeb, rozszerzona poprzez dołączenie do definicji innych reguł tworzenia zmiennych.

W prezentowanym modelu zakres reguł obrano tak, aby umożliwić:

- dla ustalonego obiektu agregację zmiennych w czasie /warunki 2,6,9/,
- dla ustalonego przedziału czasu, agregację zmiennych opisujących określony aspekt eksploatacji grupy obiektów /warunki 3,7,10/,
- dla ustalonego przedziału czasu i ustalonego obiektu agregację cech obiektu /warunki 11,13/.

2.3. Procedury systemu definiowania obiektów

Obiekty systemu ewidencji i systemu ASDE są identyfikowane na podstawie wystąpień odpowiadających im atrybutów i wystąpień określonych struktur danych.

Zadaniem procedur systemu definiowania obiektów jest nadanie wartości obiektom /zmiennym/ systemu ASDE i organizacja określonych struktur danych. Można wyróżnić następujące grupy procedur:

- procedury wyszukiwania danych,
- procedury obliczania wartości funkcji.

Przydatność i efektywność metod wyszukiwania informacji jest ściśle związana z organizacją i wielkością zbiorów danych systemu ewidencji. Wykorzystano również programy systemu zarządzania bazą danych DMS-2.

Procedury obliczania wartości funkcji w systemie ASDE definiowane są wg schematów:

Schemat 1.

Niech X będzie zbiorem ciągów zaobserwowanych wartości zmiennych elementarnych postaci $\{x_1, \dots, x_n\}$, A funkcją zdaniową, której zmiennymi są atrybuty obiektów. Na zbiorze X określimy klasę transformacji F'_A w sposób następujący:

$$1. F'_A \left(\{x_1\} \right) = x_1,$$

$$2. F'_A \left\{ x_1, \dots, x_n, x_{n+1} \right\} = \begin{cases} F'_A \left(\{x_1, \dots, x_{n-1}, x_n + x_{n+1}\} \right), & \text{jeśli } A, \\ F'_A \left(\left(\{x_1, \dots, x_n\} \right), x_{n+1} \right), & \text{jeśli } \sim A. \end{cases}$$

Każda transformacja otrzymana w wyniku ustalenia A jest procedurą systemu.

Schemat 2.

Niech S' będzie zbiorem ciągów wystąpień stanów ustalonego obiektu

^{3/} Definicja określa zbiór obiektów w sposób formalny. W pewnych przypadkach obiekty mogą nie posiadać sensownej interpretacji.

postaci $\{s_1, \dots, s_n\}$, B wyróżnionym podzbiorem zbioru stanów S. Na zbiorze S' określimy klasę transformacji F_B^n w sposób następujący:

$$1^\circ F_B^n(s_1) = \begin{cases} 1, & \text{jeśli } s_1 \in B, \\ 0, & \text{jeśli } s_1 \notin B, \end{cases}$$

$$2^\circ F_B^n(s_1, \dots, s_n, s_{n+1}) = \begin{cases} F_B^n(s_1, \dots, s_n) + 1, & \text{jeśli } s_{n+1} \in B, \\ F_B^n(s_1, \dots, s_n), & \text{jeśli } s_{n+1} \notin B. \end{cases}$$

Każda transformacja otrzymana w wyniku ustalenia B jest procedurą systemu.

3. System opisu i analizy danych

Zastosowanie statystyki w systemach informatycznych może odnosić się do jednej z klas zdań:

1. opisu danych,
2. analizy danych,
3. prognozowania^{4/}.

Zadanie typu "opis danych" ma miejsce wówczas, gdy duży zbiór informacji oryginalnych /zbiór informacji systemu ewidencji/ chcemy skondensować w postaci niewielkiej liczby charakterystyk opisowych. Jest rzeczą ważną, aby redukcję danych tak przeprowadzać, żeby za pomocą odpowiednio dobranych opisowych charakterystyk w próbie^{5/} uzyskać możliwie najwięcej informacji zawartych w zbiorze danych oryginalnych.

W systemie ASDE jako charakterystyki opisowe w próbie przyjęto zwykle charakterystyki rozkładu wartości w próbie /np. [2], [3]/:

- pierwszy i drugi moment,
- współczynnik asymetrii i eksces,
- wartości skrajne i rozstęp,
- współczynnik zmienności,
- empiryczne rozkłady gęstości i empiryczne dystrybuanty.

Zadanie typu "analiza danych" ma miejsce wówczas, gdy wnioskujemy o właściwościach populacji na podstawie próby i polega na wyznaczeniu statystycznego modelu danych.

Podstawowe zadania rozwiązywane przy modelowaniu danych to:

- estymacja parametrów,
- weryfikacja hipotez statystycznych,
- analiza korelacji i regresji.

4/ Zadania typu "prognozowanie" nie wchodzi w zakres istniejącej wersji systemu ASDE.

5/ W prezentowanym modelu jako próbę interpretuje się ustalony podzbiór zbioru danych systemu ewidencji.

W omawianym modelu systemu ASDE jako populację można przyjąć zbiór obiektów na przykład: zbiór części określonego typu, lub zbiór wartości cechy np. ilość wykonywanej przez obiekt pracy na zmianie. Zakres analizy danych obejmuje badania zgodności rozkładów empirycznych z teoretycznymi rozkładami obserwowanych zmiennych losowych. Na podstawie znanej próbki losowej, drogą estymacji parametrów wybranych rozkładów, a następnie weryfikacji hipotez o rozkładzie, wybiera się rozkład "najlepiej" opisujący daną próbkę. Współczynnikiem jakości pozwalającym na porównanie kilku nieodrzuconych rozkładów może być maksymalny poziom istotności, na którym hipoteza o zgodności danego rozkładu z empirycznym nie będzie jeszcze odrzucona. W systemie informatycznym ASDE w zakresie analizy danych wykorzystano program PAS1 opracowany przez zespół Ośrodka Informatyki Politechniki Świętokrzyskiej /[3]/.

Lista rozkładów, których zgodność rozkładów empirycznych może być badana przy użyciu programu PAS1 obejmuje:

- rozkład normalny,
- rozkład logarytmo-normalny,
- rozkład wykładniczy,
- rozkład Weibulla,
- rozkład gamma,
- rozkład Erlanga,
- rozkład Rayleygha,
- rozkład Maxwella,
- rozkład beta,
- rozkład Poissona.

Lista testów do badania zgodności rozkładów w programie PAS1 obejmuje:

- test Kołmogorowa dla prób dużych,
- test chi-kwadrat dla prób dużych o licznosci > 30 ,
- test Cochraha dla prób o licznosci > 2 .

4. Uwagi końcowe

Prezentowany model został wykorzystany przy realizacji podsystemu informatycznego, który omówiono w [4].

Sformalizowanie podstawowych definicji i operacji pozwoliło na rozwiązania informatyczne niezależne od interpretacji danych i przyszłych zastosowań wyników.

Model analizy statystycznej danych wzbogacony o schematy wnioskowania oparte na statystycznej teorii decyzji wraz z podsystemem informatycznym realizującym jego algorytmy może stanowić podstawę wdrożenia do praktyki kierowania eksploatacją metod statystyki opartych na dużych ilościach różnorodnych danych zbieranych w trakcie bieżącej eksploatacji maszyn. Dane z bieżącej eksploatacji mogą stanowić naturalną bazę informacji do weryfikacji teoretycznych modeli eksploatacji i aktualizacji parametrów.

Wstępna eksploatacja podsystemu informatycznego opartego na przedstawionym modelu wskazuje na przydatność danych eksploatacyjnych dla celów analizy statystycznej poza zakresem małych wartości zmiennych.

LITERATURA

- [1] Aleksandrowicz-Kamionka Z., i inni, Przetwarzanie informacji o niezawodności urządzeń technicznych, Część I, Zeszyt 1, Kielce 1977.
- [2] Benjamin J.R., Cornell C.A., Rachunek prawdopodobieństwa, statystyka matematyczna i teoria decyzji dla inżynierów, WNT, Warszawa 1977.
- [3] Cramér H., Metody matematyczne w statystyce, Warszawa 1958.
- [4] Zak D., Podsystem informatyczny analizy statystycznej danych eksploatacyjnych, Komunikat Centrum Obliczeniowego Politechniki Wrocławskiej, Wrocław 1978.

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЛУАТАЦИОННЫХ ДАННЫХ

РЕЗЮМЕ

В статье представлена формальная модель описания объектов для цепей статистического анализа, схемы главных процедур, базирующих на данных информатически-информационной системы предприятия. Модель использована при программировании информатической подсистемы.

STATISTICAL ANALYSIS OF EXPLOITATION DATA

Summary

The description of the objects and the fundamental procedures for statistical analysis of data is presented. The construction of the model is based on the informational system of the industrial works. The described model was used for building a computer system.