

ANALIZA CALITATIVĂ A BAZEI DE MODELE UTILIZATĂ ÎN MANAGEMENTUL CALITĂȚII SOFTWARE

Prof. univ. dr. Ion Ivan, Asist. univ. drd. Cătălin Ploae, Conf. univ. dr. Gherasim Zenovic, Asist. univ. Marius Popa, Prep. univ. Cătălin Boja, Stud. Adrian Vișoiu

1. Diversitatea modelelor din managementul calității software

Modelele directe sunt rezultatul percepției nemijlocite a variabilelor asociate factorilor de influență. Dacă se consideră un process P căruia i se asociază variabila rezultativă y și care este determinat în evoluția sa de numeroși factori F_1, F_2, \dots, F_N cărora li se asociază, respective, variabilele X_1, X_2, \dots, X_N , un model direct are forma:

$$y = \sum_{i=1}^N a_i X_i,$$

unde a_i este un coeficient de contribuție, definit pe mulțimea $\{-1,1\}$.

Modelele econometrice reprezintă o categorie deosebit de importantă de modele utilizate preponderent pentru studierea fenomenelor și proceselor la nivel macroeconomic. Se consideră o mulțime de variabile rezultative Y_1, Y_2, \dots, Y_M , și o serie de factori F_1, F_2, \dots, F_N , cărora li se asociază respectiv variabilele X_1, X_2, \dots, X_N .

Atât pentru variabilele rezultative cât și pentru variabilele exogene se efectuează măsurători pentru k momente, rezultând un număr de MN serii de timp, unde $MN = M + N$ cu:

M - numărul de variabile rezultative;

N - numărul de variabile exogene.

Este important să se stabilească:

- lungimea seriilor de timp;
- modul în care se alege momentul T_1 ;
- uniformizarea intervalelor de culegere a datelor;
- procedurile de culegere a datelor.

Modele de optimizare liniară includ în structura lor funcții de forma

$$\{\min/\max\} f(x) = \sum_{j=1}^n a_j c_j x_j$$

unde

n - numărul de produse;

a_{ij} - coeficienții de contribuție definiți pe mulțimea $\{-1,0,1\}$;

c_j - coeficienții de multiplicare;

x_j - nivelul variabilelor asociate factorilor, care influențează procesele de maximizare sau de minimizare.

Restricțiile modelelor liniare au forma:

$$\sum_{j=1}^m a_{ij} \alpha_{ij} x_j \leq b_i, i = 1, 2, \dots, n, \text{ unde}$$

a_{ij} - coeficienții de contribuție pentru ecuația i ;

b_{ij} - nivel limitativ pentru definirea unui capăt al intervalului de variabile pentru resursele utilizate;

α_{ij} - consumul unitar de resursă de tip i pentru a realiza o unitate.???

m - numărul de restricții.

Modelele neliniare reprezintă o largă categorie utilizată în studierea interdependențelor dintre factori. Marea varietate de modele neliniare produce efecte variate în colectarea și dezvoltarea sistematizată a lor, întrucât descrierea modelelor de acest tip trebuie definite reguli care să conducă la descrieri corecte și la implementări cu grad de cuprindere deosebit de ridicat.

Sunt situații în care modelele neliniare, prin transformări convenabile sunt transformate în modele liniare.

2. Seriile de date de intrare în modelele de management a calității software

Toate modelele economice utilizează date de intrare înregistrate conform unor proceduri stricte.

În sistemul financiar-contabil există documente tipizate, există reguli de înregistrare și de calcul a indicatorilor – de exemplu, pentru completarea unei facturi există o procedură care permite ca orice persoană care dezvoltă un proces să obțină un conținut identic cu conținutul documentelor completate independent de toate celelalte persoane care reproduc, de asemenea, identic procesul respectiv. Se construiesc serii de date care include elemente de identificare și nivelurile variabilelor asociate factorilor.

Toate datele se sistematizează în tabele de forma X_Y , unde

X – numărul de linii;

Y – numărul de coloane;

Tabelele în care se înregistrează evoluția dinamicii unor fenomene economice sunt de forma K, L , unde:

K – numărul momentelor de timp T_1, T_2, \dots, T_K , pentru care se efectuează culegerea de date;

L – numărul factorilor pentru care se înregistrează niveluri ale variabilelor asociate, folosind proceduri foarte stricte.

În cazul în care datele provin din surse diferite este necesar să se specifice cum sunt reprezentate procedurile de culegere.

Apar următoarele situații:

- toate seturile de date au fost culese după aceleași proceduri, au același număr de seturi de serii, seriile au aceeași lungime, sunt echidistante, momentul de start al culegerii datelor este același, de exemplu, se înregistrează în cele 40 de județe (sunt 40 de seturi de date); datele sunt preluate ca atare întrucât sunt: corecte, complete, comparabile;
- seturile de date au fost culese după aceleași proceduri, seriile au același număr de termeni, termenii sunt echidistanți și momentele de start sunt diferite, ca în tabelul 3.

3. Complexitatea modelelor de management a calității software

Complexitatea în sens Halstead pentru un model este dată de relația:

$$C(M) = n_1 \log_2 n_1 + n_2 \log_2 n_2,$$

unde

n_1 - numărul de operanzi (variabile și coeficienți);

n_2 - numărul de operatori.

Modelul liniar, definit prin ecuația:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n,$$

conține:

- a_0, a_1, \dots, a_n - coeficienții estimați ai modelului, în număr de $n_c, n_c = n + 1$;

- x_1, x_2, \dots, x_n, y - variabilele exogene ale modelului și variabila rezultativă, în total în număr de $n_v, n_v = n + 1$;

- * - operatorii de înmulțire a coeficienților cu variabilele exogene, în număr de $n_{cv}, n_{cv} = n$;

- + - operatorii care leagă variabilele exogene formând membrul drept al ecuației modelului liniar, în număr de $n_t, n_t = n$;

- = - operatorul de atribuire. $n_a = 1$.

Efectuând înlocuirile, se obține:

$$C = 2(n+1) \log_2 [2(n+1)] + (n+1) \log_2 (n+1).$$

În cazul modelului neliniar

$$y = A \cdot B^\alpha \cdot C^\beta.$$

- operanzii sunt: $y, A, B, C, \alpha, \beta$, rezultând $n_1 = 6$;

- operatorii sunt: $=, *, ()^\alpha, *, ()^\beta$, rezultând $n_2 = 5$.

S-au notat cu $()^\alpha, ()^\beta$, ridicările la o putere, iar operatorul de înmulțire, a fost scris explicit de două ori pentru că de două ori apare în ecuație.

Complexitatea în sens Halstead a acestui model neliniar este:

$$C = 6 \log_2 6 + 5 \log_2 5 = 27.12.$$

Trebuie impus un mod unitar de calcul al complexității, pentru a nu se produce confuzii. De exemplu, în cazul modelului

$$y = (Ax + Bt)(au^2 + bz^3 + cw^4) + d,$$

înainte de a se calcula complexitatea, trebuie obținută forma prelucrată exinsă prin efectuarea calculelor algebrice:

$$y = A \cdot x \cdot u^2 + A \cdot x \cdot b \cdot z^3 + A \cdot x \cdot c \cdot w^4 + B \cdot t \cdot a \cdot u^2 + B \cdot t \cdot b \cdot z^3 + B \cdot t \cdot c \cdot w^4 + d$$

Rafinarea modelului economic conduce la scăderea complexității în sens Halstead. Dacă un model liniar are în forma inițială n variabile exogene și respectiv complexitatea

$$C_n = (n+1)[2 + 3 \log_2(n+1)],$$

prin scăderea numărului de variabile exogene la $n-k$, $k < n$, se obține o complexitate

$$C_{n-k} = (n-k+1)[2 + 3 \log_2(n-k+1)],$$

$$\Delta = C_n - C_{n-k} = 2[(n+1) - (n-k+1)] + 3(n+1) \log_2(n+1) - 3(n-k+1) \log_2(n-k+1).$$

În cazul în care prin rafinare se trece de la un model neliniar la un model liniar, de asemenea, scade complexitatea modelului.

Fie modelul neliniar:

$$M_1: y_1 = a_1x^2 + b_1z^2 + c_1xy + d_1$$

iar prin liniarizare se obține

$$M_2: y_2 = a_2x + b_2z + c_2.$$

Este important să se analizeze complexitatea relativă a modelului în raport cu o complexitate maxim posibilă care se obține prin agregarea operanzilor și operatorilor.

Modelul general al complexității în sens Halstead este:

$$C = \left(\sum_{i=1}^m f_i p_i \right) \log_2 \left(\sum_{i=1}^m f_i p_i \right) + \left(\sum_{j=1}^n h_j q_j \right) \log_2 \left(\sum_{j=1}^n h_j q_j \right),$$

unde

m - numărul de operanzi;

f_i - frecvența de apariție a operandului i ;

p_i - ponderea asociată operandului în construcția

n - numărul de operatori;

h_j - frecvența de apariție a operatorului j ;

q_j - ponderea asociată operatorului j ; în exemplul considerat, ponderea s-a construit pornind de la prioritatea operatorilor.

Complexitatea C în sens Halstead pentru modelul economic de regresie polinomială este dată de relația:

$$C = (2n+3) \log_2(2n+3) + (2n) \log_2(2n)$$

4. Generatoare de modele pentru managementul calității software

Colecțiile de modele economice evidențiază existența unor familii de modele.

Se consideră un model M care are în structură un set de variabile exogene. Trecerea la modelul M' se realizează prin includerea în setul de variabile a unui subset de variabile noi.

Trecerea de la modelul M , prin eliminarea unei variabile, conduce la obținerea unui model cu o structură mai simplă, $'M$. Atunci când modelului M se adaugă variabile, procesul este de dezvoltare, iar operatorul este notat $D(\cdot)$.

$$D(M) = M'.$$

Când se trece de la un model M la un alt model $'M$, prin eliminarea unei variabile, procesul este de simplificare, iar operatorul este $S(\cdot)$.

$$S(M) = 'M.$$

Se observă că în cazul în care variabila adăugată este X_{i+1} la modelul M , care conține deja variabilele X_1, X_2, \dots, X_i

$$D(M)_{X_{i+1}} = M'$$

$$S(M)_{X_{i+1}} = 'M$$

iar,

$$S(D(M)_{X_{i+1}})_{X_{i+1}} = M$$

În cazul în care prin simplificare informația obținută din model este în continuare relevantă, procesul se numește **rafinarea modelului**.

Generatoarele de modele liniare reprezintă mecanisme deosebit de importante pentru obținerea de modele economice reprezentative.

Se consideră variabilele independente X_1, X_2, \dots, X_n și variabila dependentă Y .

Practica economică a condus, în general, la elaborarea de modele liniare pentru că:

- fenomenele studiate urmăresc o dependență liniară;
- metodele de estimare a parametrilor sunt uzuale pentru aceste tipuri de modele;
- interpretarea rezultatelor este ușurată dacă sunt luate în calcul ipotezele de liniaritate.

Generatoarele de modele liniare cu argumente întârziat permit elaborarea de construcții care permit modelarea efectelor de antrenare multiplă care se regăsesc pe termen scurt în influențe din toate seturile.

Se consideră o mulțime formată de momente de timp echidistante $T = \{t_1, t_2, \dots, t_n\}$, unde $t_1 < t_2 < \dots < t_n$ și $t_k = t_{k-1} + \Delta, k = \overline{2, n}$. Restricția privind echidistanța nu afectează generalitatea abordării de față.

Pentru **generarea modelelor liniare cu un singur argument întârziat**, se consideră variabila independentă X și variabila dependentă Y pentru care s-au efectuat măsurători corespunzătoare momentelor de timp t_1, t_2, \dots, t_n .

Folosind metoda celor mai mici pătrate, se estimează coeficienții modelului $\hat{y}_i^{(1)} = a_0 + a_1 x_i$ și se calculează $S_1 = \sum_{i=1}^n (y_i - \hat{y}_i^{(1)})^2$. Unde \hat{y}_i este valoarea variabilei dependente estimată la momentul t_i , iar (1) este numărul de ordine al modelului.

Generatoarele de modele neliniare vizează utilizarea unor structuri de forme analitice. Se consideră un set de forme analitice dat în tabelul 1.

Forme analitice ale modelelor neliniare

Tabelul nr. 1

Nr. crt	Denumire model	Forma analitică
1.	funcția parabolică 1	$y = ax^\alpha + bx + c$
2.	funcția parabolică 1	$y = \sqrt{a + bx}$
3.	funcția parabolică 3	$y = \sqrt[3]{a + bx}$
4.	funcția parabolică 4	$y = \sqrt{a + bx^3}$
5.	parabola lui Neile	$y = \sqrt[3]{ax^\alpha + bx + c}$
6.	funcția putere	$y = ax^b$
7.	funcția exponențială 1	$y = e^{ax+b}$
8.	funcția exponențială 2	$y = ae^{bx}$
9.	funcția exponențială 3	$y = ab^x$
10.	funcția exponențială 4	$y = e^{ax} \cdot x^b$
11.	funcția logaritmică 1	$\ln y = a + b \ln x$
12.	funcția logaritmică 2	$y = a \ln x$
13.	funcția semilogaritmică	$y = a + b \ln x$
14.	funcția log-inversă	$\ln y = a + \frac{b}{x}$
15.	funcția log-log-inversă	$\ln y = a + \frac{b}{x} + c \ln x$
16.	funcția inversă	$y = a + \frac{b}{x}$
17.	funcția Prais	$y = e^{\frac{a-b}{x}}$

Nr. crt	Denumire model	Forma analitică
18.	funcția de ordinul 3	$y = \frac{x}{(ax^2 + bx + c)}$
19.	funcția hiperbolică 1	$y = \frac{1}{(a + bx)}$
20.	funcția hiperbolică 2	$y = a + \frac{b}{(c + x)}$
21.	funcția Tornqvist 1	$y = \frac{kx}{(x + a)}$
22.	funcția Tornqvist 2	$y = \frac{k(x + a)}{(x + b)}$
23.	funcția Tornqvist 3	$y = \frac{bx(x - c)}{(x + a)}$
24.	funcția Johnson	$y = e^{\frac{k - a}{(b+x)}}$
25.	funcția log-parabolică	$y = a(\ln x)^\alpha + b \ln x + c$
26.	funcția logistică	$y = \frac{k}{(1 + be^{-ax})}$
27.	funcția logistică pătratică	$y = \frac{k^2}{(1 + be^{-ax})^2}$
28.	funcția Cobb-Douglas 1	$y = Ax_1^\alpha x_2^\beta e^{\gamma x_3}$
29.	funcția Cobb-Douglas 2	$y = Ax_1^\alpha x_2^\beta$
30.	funcția CES 1	$y = A(dx_1^{-\rho} + (1 - d)x_2^{-\rho})^{-\frac{1}{\rho}} \cdot e^{\gamma x_3}$
31.	funcția CES 2	$y = A(dx_1^{-\rho} + (1 - d)x_2^{-\rho})^{-\frac{1}{\rho}}$
32.	funcția CES 3	$y = A(dx_1^{-\rho} + (1 - d)x_2^{-\rho})^{-\frac{h}{\rho}} \cdot e^{\gamma x_3}$
33.	funcția CES 3	$y = A(dx_1^{-\rho} + (1 - d)x_2^{-\rho})^{-\frac{h}{\rho}}$
34.	funcția Allen	$y = \sqrt{2hx_1x_2 - ax_1^2 - bx_2^2}$
35.	funcția Sato	$y = \frac{x_1^2 x_2^2}{(ax_1^3 + bx_2^3)}$

Pentru un set de variabile endogene format din y_1, y_2, \dots, y_M și un set de variabile exogene format din x_1, x_2, \dots, x_N se generează un model neliniar cu o necunoscută, variantele fiind

$$y_i = f(X_j),$$

obținându-se $M \cdot N$ structuri de modele.

Generatoarele de modele au menirea de a furniza liste complete de expresii analitice diferite, care să acopere întreaga diversitate a combinațiilor. Odată cu dezvoltările spre inteligența artificială a unei aplicații informatice, generatoarele de modele neliniare vor include și elemente de autoinstruire extrase din analiza colecțiilor de modele stocate în baze de modele și se va crea premisa obținerii de noi clase de modele, ceea ce corespunde realizării unui salt calitativ în structura modelelor economice.

5. Proceduri de lucru cu modele ale calității software

Până la utilizarea unui model economic, este necesară parcurgerea a numeroși pași.

Procedurile destinate estimării coeficienților au ca date de intrare:

- lungimea seriilor de date asociate factorilor de influență și variabilelor rezultative;
- seriile de date propriu-zise;

- structura modelului pentru care se estimează coeficienții, codul asociat metodei de estimare utilizate. Rezultatele oferite de aceste proceduri sunt:
- șirul de coeficienți estimați;
- nivelul sumei pătratelor diferențelor dintre nivelurile reale și cele estimate ale variabilei rezultative ;
- valori asociate rezultatului testării unor ipoteze statistice privind calitatea estimatorilor;
- informații privind derularea estimării.

Procedurile pentru operare pe seturi de date vizează realizarea:

- preluarea seturilor de date;
- omogenizarea seriilor de date prin interpolare, extrapolare și prin transformări elementare;
- concatenarea seturilor obținând seturi de date cu un număr sporit de serii de date;
- extensia seturilor de date pentru a obține serii cu un număr mai mare de termeni;
- asigurarea comprehensibilității datelor prin aplicarea de coeficienți de transformare termenilor;
- agregarea seturilor de date pentru obținerea de indicatori noi.

Datele de intrare conțin:

- numărul seriilor de date;
- lungimile seriilor de date;
- tipologiile de prelucrări.

Rezultatele obținute se concretizează în:

- noile serii de date;
- lungimile noilor serii de date;
- destinațiile de stocare a seriilor;
- informații privind calitatea procesului de operare.

Procedurile pentru generarea de date de test au menirea de a pregăti date de intrare care să fie utilizate pentru studierea modelelor. Există numeroase situații în care structurile de date existente sunt incomplete, deși sunt cunoscute tendințele de evoluție a fenomenelor și limite de variație a nivelurilor variabilelor asociate. De asemenea, lungimile seriilor de date existente, la un moment dat, sunt insuficiente pentru a produce estimări de calitate.

Procedurile pentru calculul valorilor preiau modele, coeficienți estimați și seturi de date ale variabilelor exogene și calculează termenii simulărilor, $\hat{y}_{i1}, \hat{y}_{i2}, \dots, \hat{y}_{im}$, $i = 1, 2, \dots, m$ pentru modelele M_1, M_2, \dots, M_m .

Procedurile pentru selecția de modele se folosesc în condițiile în care există:

- seria de date conținând valorile înregistrate ale unei variabile rezultative;
- seriile de date cu date estimate ale variabilei rezultative obținute prin utilizarea modelelor din baza de modele;
- procese de asigurare a compatibilității tuturor seriilor de date.

Se definește un algoritm de selectare din mulțimea modelelor a acelor care aparțin unui subinterval de omogenitate de lungime specificată.

Procedurile pentru rafinarea modelului au menirea de a reduce numărul variabilelor exogene din structura ecuației. În cazul modelelor cu mai multe ecuații, pe lângă reducerea numărului de variabile exogene, are loc și reducerea numărului de ecuații. Se obține un model economic de dimensiuni mai mici.

Rolul rafinării este de a reduce efortul de gestionare a modelului fără a pierde din calitatea estimării variabilelor rezultative.

Procedura pentru identificarea legăturilor dintre variabile pentru a stabili dacă între două variabile X și Y există relații de forma:

$$y = a \cdot x$$

$$y = \frac{x}{a}$$

$$y = (x)^a$$

$$y = x + a$$

$$y = x - a$$

$$y = \ln x$$

$$y = \sin(x)$$

Procedura constă în :

- calculul matricei de corelație
- selectarea variabilelor pentru care coeficientul de corelație este 1
- reconstruirea listei reduse a variabilelor între care nu există legături directe

Procedurile pentru regruparea datelor în funcție de ortogonalitate vizează reorganizarea seturilor de date utilizate în modele așa cum au fost furnizate de utilizatori. Pe măsură ce tabelele cu date sunt introduse, li se atribuie un număr unic de referință, iar activarea modulului conduce automat la utilizarea setului de date după numărul de referință.

Procedurile pentru stabilirea ortogonalității modelelor efectuează analiza cantitativă pentru descrierea modelelor, evidențiind:

- numărul de ecuații
- numărul de variabile distincte
- numărul de operanzi distincți
- frecvențele de apariție ale operatorilor
- pozițiile operanzilor și pozițiile operatorilor

Se compară valorile obținute și rezultă gradul de asemănare dintre modelele analizate. Dacă indicatorul de ortogonalitate $G = 0$, rezultă că modelele sunt identice. Dacă același indicator are valoarea 1, rezultă că modelele sunt total diferite.

În cazul modelelor cu grad de ortogonalitate nul se procedează la normalizarea tabelii de corespondență. Tabela de corespondență este de fapt o structură omogenă ce conține M componente, unde M indică numărul de modele diferite stocate în baza de modele. Cvartuplele incluse într-o componentă indică:

- numele ortogonal al modelului
- numărul de referință al modelului din bază
- numărul de referință a listei de variabile
- numărul de referință a setului de date

Va rezulta o nouă structură de modele, se va construi o nouă tablă de modele și se va proceda la reorganizarea listelor de referire a seturilor de date.

6. Ortogonalitatea – caracteristică a calității bazei de modele economice

Se consideră o bază de modele în care sunt deja stocate modelele M_1, M_2, \dots, M_n . Se pune problema adăugării unui nou model, M_{n+1} . Acest lucru se efectuează dacă și numai dacă modelul M_{n+1} este diferit de modelele deja existente.

Modelele cu structură identică sunt acelea care au același număr de variabile, același număr de ecuații, același număr de operatori, iar prin dezvoltarea scrierii polonoize inverse, pozițiile operanzilor și operatorilor sunt aceleași.

Modelele:

$$M_1 : y = a \cdot x + b \cdot z + c$$

$$M_2 : w = d \cdot u + h \cdot v + g$$

sunt cu structură identică, iar modelele:

$$M_3 : y = ax + be^t$$

$$M_4 : u = ce^w + dw$$

sunt identice din punct de vedere structural, adunarea fiind comutativă.

Complexitățile neponderate se calculează după următoarea formulă:

$$C(M_i) = n_i \log_2 n_i + m_i \log_2 m_i$$

unde:

- n_i numărul de operanzi ai modelului;
- m_i numărul de operatori ai modelului.

Pentru calculul complexităților ponderate, fiecare operator primește o pondere w_j , pentru exemplele prezente aici, ponderile operatorilor $=, +, *, ()^2$ fiind respectiv 1, 2, 3, 4. Formula este:

$$C(M_i) = n_i \log_2 n_i + S \log_2 S$$

unde:

$$S = \sum_{j=1}^{m_i} w_j .$$

Pentru a vedea dacă două modele sunt identice, se procedează, mai întâi, la calculul complexității neponderate, $C(M_1) = C(M_2) = C(M_3) = C(M_4) = 27,11$.

La o primă analiză, rezultă că modelele au complexități neponderate identice.

Pentru modelele M_1 și M_2 , complexitățile ponderate sunt date de relația: $C(M_1) = C(M_2) = 53,56$.

ceea ce arată că modelele M_1 și M_2 au un grad de asemănare foarte ridicat, mai mult, din punct de vedere structural sunt identice, $C(M_3) = C(M_4) = 63,61$.

Modelele M_3 și M_4 sunt identice din punct de vedere structural și diferite de M_1 și M_2 .

Modelele incluse sunt acelea care în scrierea poloneză inversă apar sub formă de subșiruri ale unor șiruri. Pentru modelele:

$$M_5 : y = ax + bz + c$$

$$M_6 : y = a \cdot x + b \cdot z + d \cdot u + c$$

se observă că M_6 se deduce din M_5 la care se adaugă termenul $d \cdot u$. Înseamnă că modelul M_5 este inclus în modelul M_6 . Complexitățile sunt: $C(M_5) = 27.11$ și $C(M_6) = 39.5$.

Raportul de asemănare, A , este dat de relația:

$$A = \frac{K}{N}$$

unde:

N - numărul maxim de componente;

K - numărul de componente identice.

Pentru modelele M_5 și M_6 , $N = \max\{11,14\} = 14$, unde valorile din paranteze reprezintă numărul de variabile și de operatori. Elementele comune se reunesc într-o mulțime $\{y, =, a, \cdot (produs), x, + (adunare), b, \cdot (produs), z, +, c\}$ al cărei cardinal este 11, rezultând $A = 0,78$.

Complexitățile ponderate sunt $C(M_5) = 53.56$ și $C(M_6) = 104.71$.

Dacă modelul liniar M_k este inclus în modelul M_j , unde:

$$M_k : y = a_0 + \sum_{k=1}^n a_i x_i$$

$$M_j : y = a_0 + \sum_{k=1}^n a_i x_i + \sum_{i=n+1}^{n+r} a_i x_i$$

complexitățile calculate sunt:

$$C(M_k) = (n+2) \cdot \log_2(n+2) + (2n) \cdot \log_2(2n)$$

$$C(M_j) = (n+r+2) \cdot \log_2(n+r+2) + 2(n+r) \cdot \log_2 2(n+r)$$

Modelele ortogonale sunt acelea în care operanzii și operatorii sunt diferiți, fără a lua în considerare operatorul de atribuire

$$M_7 : y = a \cdot b^t$$

$$M_8 : y = cx + dz + h$$

Operatorii modelului M_7 sunt constituiți în șirul: * și (), iar operatorii modelului M_8 alcătuiesc șirul: *,+,*,+. Este evident că șirurile sunt total diferite.

Operanzii din modelul M_7 se constituie în șirul: y, a, b, t, pe când cei din modelul M_8 se regăsesc în: z, c, x, d, z, h. Șirurile sunt, de asemenea, diferite.

Gradul de ortogonalitate se obține din relația:

$$O(M_k, M_j) = 1 - A(M_k, M_j).$$

Dacă indicatorul $O(M_k, M_j) = 1$, rezultă că modelele sunt ortogonale.

În cazul în care în baza de modele ce conține modelele M_1, M_2, \dots, M_n și se dorește adăugarea modelului M_{n+1} , dacă $O(M_i, M_{n+1}) = 1, \forall i = 1, 2, \dots, n$, modelul M_{n+1} este într-adevăr un model nou și prin includerea lui în baza de modele, aceasta sporește din punct de vedere calitativ.

Modelele de aceeași clasă sunt modelele care nu diferă din punct de vedere structural. Modelele:

$$M_8: y = ax + bz + cw + d$$

$$M_9: y = ax + d$$

sunt modele liniare, provin din aceeași clasă pentru că includ în alcătuirea lor: coeficienți, variabile, operatori de adunare, operatori de înmulțire și termeni de același grad egal cu 1.

Se calculează complexitatea:

$$C = n_1 \log_2 n_1 + n_2 \log_2 n_2 + n_3 \log_2 n_3$$

unde:

n_1 - numărul de termeni diferiți;

n_2 - numărul de operanzi diferiți;

n_3 - numărul de grade diferite.

În mod corespunzător, complexitățile modelelor M_8 și M_9 sunt $C(M_8) = C(M_9) = 2 \log_2 2 + 3 \log_2 3 + 2 \log_2 2 = 8.75$.

Modelele:

$$M_{10}: y = ax_1^2 + bx_2^2 + cx_3^2 + d + ex_1 + fx_2 + gx_3$$

$$M_{11}: y = ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2 + ex_5^2 + gx_6^2 + hx_7^2 + i + jx_1 + kx_2 + lx_3$$

sunt definite prin:

- coeficienții modelului M_{10} : a, b, c, d, e, f, g .
- coeficienții modelului M_{11} : $a, b, c, d, e, g, h, i, j, k, l$.
- variabilele modelului M_{10} : y, x_1, x_2, x_3 .
- variabilele modelului M_{11} : $y, x_1, x_2, x_3, x_4, x_5, x_6, x_7$.
- operatorii diferiți ai modelului M_{10} : $=, +, ()^2, *$.
- operatorii diferiți ai modelului M_{11} model are: $=, +, *, ()^2$.
- gradele termenilor modelelor M_{10} și M_{11} : $0, 1, 2$.

Se calculează complexitățile modelelor, obținându-se $C(M_{10}) = C(M_{11}) = 3 \log_2 3 + 4 \log_2 4 + 3 \log_2 3 = 17.5$.

Rezultă că M_{10} și M_{11} aparțin aceleiași clase.

Modelele asemenea sunt acelea care conțin structuri dominante. Modelele:

$$M_{12}: y = ax_1 + bx_2 x_3 + ce^{x_4} + dx_5 \cdot \log(x_6) + g$$

$$M_{13}: y = ax_1^2 + b \frac{x_2}{x_3} + ce^{x_4} + dx_5 \cdot \log(x_6) + g$$

au în structura lor termenii comuni reuniți în subexpresia $ce^{x_4} + dx_5 \cdot \log(x_6) + g$. Complexitățile neponderate ale celor două modele sunt: $C(M_{12}) = 86.03$, respectiv $C(M_{13}) = 96.21$.

Complexitatea subexpresiei comune este:

$$C(\text{common}) = 6\log_2 6 + 7\log_2 7 = 35.16.$$

Ponderea complexității subexpresiei comune a două modele M_i și M_j se calculează astfel:

$$\rho_i = \frac{C(M_i \cap M_j)}{C(M_i)}$$

$$\rho_j = \frac{C(M_i \cap M_j)}{C(M_j)}$$

Ponderea complexității subexpresiei comune în totalul complexității lui M_{12} este :

$$\rho_{12} = \frac{6\log_2 6 + 7\log_2 7}{12\log_2 12 + 12\log_2 12} = 0.4$$

și ponderea complexității subexpresiei comune în totalul complexității lui M_{13} este:

$$\rho_{13} = \frac{6\log_2 6 + 7\log_2 7}{13\log_2 13 + 13\log_2 13} = 0.36.$$

Dacă $\rho \in [0,0.82]$, modelele au în comun elemente ne semnificative.

Dacă $\rho \in [0.82,0.92]$, modelele au în comun elemente semnificative.

Dacă $\rho \in [0.92,1)$, modelele au un grad foarte ridicat de asemănare.

Pentru $\rho_i = 1$, un anumit model, M_i este inclus în alt model, M_j , cu $\rho_j > 0$. Dacă $\rho_i = \rho_j = 1$, modelele sunt identice.

În concluzie, realizarea analizei ortogonalității modelelor economice permite includerea acestora în clase de ortogonalitate a modelelor.

Modelele economice vizează laturile calitative și laturile cantitative ale evoluției unor fenomene sau ale definirii de structuri, în vederea obținerii de informații privind efectele apariției unor modificări în timp.

Orice model economic presupune înregistrarea de date pentru a vedea nivelurile caracteristicilor cu care sunt descrise dinamica unui fenomen și factorii care o determină.

Setul de date asociat unei probleme de dinamică este dat sub forma unui tabel ce include:

- coloană în care se specifică momentele de timp în care au fost efectuate măsurătorile;
- coloană de date reprezentând nivelul măsurat pentru variabila rezultativă;
- câte o coloană de date reprezentând nivelurile măsurate ale fiecărui factor de influență.

Pentru a realiza o prelucrare de calitate trebuie ca seriile incluse în tabel să aibă același număr de termeni și măsurarea lor să se realizeze exact la momentul de timp indicat în tabloul construit conform structurii date în tabelul 3.

Tabelul 3 Structură set de date specific proceselor dinamice

Moment de timp T	Variabilă rezultativă y	Factor de influență X₁	...	Factor de influență X_j	...	Factor de influență X_m
T ₁	y ₁	X ₁₁	...	X _{j1}	...	X _{m1}
T ₂	y ₂	X ₁₂	...	X _{j2}	...	X _{m2}
T ₃	y ₃	X ₁₃	...	X _{j3}	...	X _{m3}
...
T _k	y _k	X _{1k}	...	X _{jk}	...	X _{mk}
...

T_n	y_n	X_{1n}	...	X_{jn}	...	X_{mn}
-------	-------	----------	-----	----------	-----	----------

unde:

T_k – momentul de timp la care a avut loc înregistrarea valorilor;

y_k – valoarea variabilei rezultative la momentul T_k ;

X_{jk} – valoarea factorului de influență j la momentul de măsurare T_k .

Dacă diferențele

$$T_2 - T_1 = T_3 - T_2 = \dots = T_{k+1} - T_k = \dots = T_n - T_{n-1}$$

rezultă că înregistrările de date au fost efectuate la intervale egale de timp.

Pe seturile de date se definesc care afectează atât structura, cât și conținutul acestora, precum:

- adăugarea de noi linii, respectiv efectuarea de măsurători pentru momentele T_{n+1}, T_{n+2}, \dots
- adăugarea de vechi măsurători efectuate la momente ce au precedat momentul T_1 , respectiv $T_0, T-1, T-2, \dots$
- eliminarea de coloane din tabel în cazul în care variabilele nu se dovedesc a fi esențiale pentru proces.

Problema adăugării de noi variabile de influență presupune reorganizarea unui nou proces de culegere de date, pentru alte momente de timp și înregistrarea de noi niveluri pentru toate variabilele, incluzând deci și noile variabile adăugate pentru a obține un nou set de date.

Al doilea tip de seturi de date vizează o colectivitate A formată din elementele a_1, a_2, \dots, a_n , pentru care se înregistrează niveluri ale caracteristicilor C_1, C_2, \dots, C_m , al cărui structură este prezentată în tabelul 4:

Set de date pentru descrierea colectivității A

Tabelul 4

Element	Caracteristica C_1	...	Caracteristica C_i	...	Caracteristica C_m
a_1	C_{11}	...	C_{1i}	...	C_{1m}
a_2	C_{21}	...	C_{2i}	...	C_{2m}
...
a_i	C_{i1}	...	C_{ij}	...	C_{im}
...
a_n	C_{n1}	...	C_{ni}	...	C_{nm}

unde: a_i – elementul i al colectivității A și C_{ji} – valoarea caracteristicii j pentru elementul a_i al colectivității A .

Operațiile pe acest tip de set de date sunt:

- adăugarea unor noi elemente a_{n+1}, a_{n+2}, \dots
- adăugarea de noi caracteristici C_{m+1}, C_{m+2}, \dots cu reluarea măsurătorilor;
- eliminarea unor elemente din mulțime;
- eliminarea unor caracteristici, de regulă cele ale căror niveluri sunt identice pentru toate elementele colectivității.

Ca și în cazul primului tip de set de date, elementele colectivității A trebuie definite și utilizate riguros proceduri pentru măsurare.

Existența mai multor seturi de date ridică problema ortogonalității acestora atât sub aspectul structurii, cât și ceea ce privește conținutul lor.

Fie seturile de date S_i și S_j . Cele două seturi de date se caracterizează printr-o structură și un conținut.

Structura setului de date este dată de prima linie a tabelului cu ajutorul căruia setul de date se reprezintă. Conținutul setului de date constă în liniile care urmează celei de descriere a structurii.

Analiza ortogonalității seturilor de date se derulează pe mai multe niveluri. Continuarea procesului de rafinare a seturilor de date pe elemente structurale specifice acestora este condiționată de identitatea perfectă a seturilor de date într-un anumit stadiu al analizei.

Fie seturile de date S_1 și S_2 care conțin valori privind nivelurile cheltuielilor și veniturilor a două societăți comerciale SC_1 și SC_2 :

Setul de date S_1

Tabelul 5

Luna	Cheltuieli (mil. lei)	Venituri (mil. lei)
------	--------------------------	------------------------

Ianuarie	250	300
Februarie	400	350
Martie	350	350
Aprilie	550	450
Mai	200	300
Iunie	300	400

respectiv:

Setul de date S_2

Tabelul 6

Luna	Cheltuieli (mil. lei)	Venituri (mil. lei)
Martie	350	350
Iunie	300	400
Ianuarie	250	300
Mai	200	300
Februarie	400	350
Aprilie	550	450

Primul nivel al analizei ortogonalității seturilor de date S_1 și S_2 privește structura acestora. Seturile de date sunt perfect identice deoarece au același număr de linii și de coloane:

$$nl_1 = nl_2 = 7$$

$$nc_1 = nc_2 = 3$$

unde:

nl_1, nl_2 – numărul de linii ale seturilor de date S_1 , respectiv S_2 ;

nc_1, nc_2 – numărul de coloane ale seturilor de date S_1 , respectiv S_2 .

Deoarece identitatea între cele două seturi de date este identică, se trece la un nou nivel de rafinare a acestora.

În continuare se determină valorile maxime și minime a valorilor din cele două tabele.

Astfel:

$$MIN(S_1) = MIN(S_2) = 200$$

$$MAX(S_1) = MAX(S_2) = 550$$

Seturile de date fiind identice conform acestui criteriu de asemănare.

Următorul criteriu în analiză constă în determinarea valorilor minime și maxime pe fiecare linie a celor două seturi de date.

Următorul criteriu constă în aplicarea unui minim sau maxim pe fiecare coloană de valori, iar pentru șirul de valori obținut pentru un set de date se aplică o un obiectiv de sens contrar. De exemplu, pentru seturile de date considerate, se determină minimul pe fiecare coloană. Șirurile rezultate sunt:

$$S_{min1} = (200, 300)$$

$$S_{min2} = (200, 300)$$

Pentru cele două șiruri se aplică funcția obiectiv de sens opus:

$$max_{S1} = max_{S2} = max(200, 300)$$

ceea ce înseamnă că seturile de date sunt perfect identice și din acest punct de vedere.

Ultimul nivel al analizei constă în compararea valorilor în mod individual pentru cele două seturi de date prin stabilirea de relații de ordine, a proporționalității valorilor etc.

În exemplul de față, cele două seturi sunt perfect identice, orice criteriu ar fi luat în considerare. În general, un set de date este identificat în mod unic pe baza unei amprente a setului de date.

Amprenta setului de date cuprinde următoarele elemente:

- structura setului de date exprimată în număr de linii și coloane;
- valoarea minimă, respectiv maximă a valorilor conținute în tabel;

- valorile corespunzătoare criteriilor min(max), respectiv max(min);
- valoarea variabilei endogene, profit brut în cazul exemplului considerat.

Seturile de date utilizate în elaborarea de modele și în studierea proceselor economice trebuie să îndeplinească o serie de caracteristici de calitate. Acestea trebuie identificate, măsurate, influențând întreaga abordare a modelării procesului ca practică și bază a unor dezvoltări coerente.

În acest context, ortogonalitatea seturilor de date joacă un rol important în dezvoltarea unor modele cu un nivel ridicat al calității.

Bazele de modele includ:

- modele ortogonale;
- modele generative pentru o clasă, de exemplu:

$$y = a_0 + \sum_{i=1}^n a_i x_i, \quad y = a_0 + \sum_{i=1}^n a_i x_i^2, \text{ etc.}$$

- modele cu subexpresii comune cu pondere limitată;
- modele cu un nivel de includere sub o limită impusă.

Limitele sunt impuse atunci când se configurează baza de modele, pentru a se gestiona riguros varietățile de modele.

Ortogonalitatea trebuie surprinsă și în ceea ce privește seturile de date. Acestea joacă un rol important în testarea și aplicarea modelelor în realitate, contribuind la fundamentarea unor decizii corecte.

8. Concluzii

Noile instrumente trebuie mai întâi prezentate, după aceea explicate funcțiile activate pentru cele mai des întâlnite tipuri de probleme și numai după aceea se efectuează analiza comparată a rezultatelor oferite.

Baza de modele este o construcție complexă care are menirea de a reuni într-un tot unitar, asemeni unui corp omenesc, componente cu funcțiuni distincte, coezive însă, care prin interdependențele create, soluționează o mare diversitate de tipuri de probleme.

Utilizatorul își definește problemele proprii pe care și le gestionează. Este important să aibă o imagine corectă asupra problemelor pe care le reține sau le consideră noi probleme.

Reconstituirea este rezultatul reproductibilității unor succesiuni de pași, de seturi de date și cu structuri de modele cu aceleași conținut de start pe care l-a avut problema abandonată sau noua problemă obținută de la o veche problemă prin introducerea de modificări majore, atunci când problema de bază de la care s-a obținut dezvoltarea nu a fost salvată.

Caracterul deschis al acestui instrument este dat de funcțiile prin care utilizatorul definește seturi de date și modele care, dacă respectă cerințe de ortogonalitate, se includ în baza de modele sub formă de entități permanente, libere să fie referite de oricare alți utilizatori.

Problematika includerii de către utilizatori a procedurilor, a opțiunilor de prelucrare și a meniurilor în interfață întâmpină rezistența limitărilor legate de calitatea procedurilor ca software, căruia elaboratorul trebuie să-i asigure corectitudine, fiabilitate și portabilitate cel puțin la nivelul pe care îl are sistemul de gestiune a bazei de modele în ansamblul lui.

Prin includerea de proceduri ale utilizatorilor se obține un grad de eterogenitate semnificativ care se gestionează prin asigurarea transparenței în ceea ce privește proveniența procedurilor nou incluse sau prin asigurarea caracterului temporar sau prin menținerea caracterului deschis pentru texte sursă în așa fel încât alți utilizatori să testeze și să intervină pentru corectarea erorilor.

Baza de modele oferă elemente care se integrează în alte aplicații informatice. De asemenea, preia seturi de date care au rezultat în aplicațiile informatice ale utilizatorilor.

Deschiderea spre Internet evidențiază noile valențe ale modelării economice, precum și tendința de a se obține generalizări de structuri de modele în vederea extinderii utilizării lor.

Complexitatea bazei de modele are tendință de creștere pentru că:

- prin creșterea numărului de seturi de date, prin creșterea diversității acestora și prin creșterea gradului de neomogenitate, trebuie elaborate proceduri care să permită referirea unitară și proceduri pentru asigurarea asimilării din punct de vedere calitativ a acestor noi seturi de date;
- prin adăugarea de noi structuri de modele sau a noi tipologii de modele, trebuie implementați algoritmi care să asigure estimarea coeficienților, care să conducă la obținerea de rezultate finale și care să le integreze în mulțimea modelelor existente ca proces simplu de creștere a diversității;

- atunci când se propun noi tipuri de probleme utilizatorilor, trebuie să existe opțiuni vizibile în meniurile accesate și să se definească legăturile necesare între componentele deja existente în baza de modele și noile componente care se includ pentru a implementa noile tipuri de probleme.

Problematica bazei de modele are tendințe de evoluție favorabile, noile cercetări conducând spre abordări orientate către utilizarea limbajului natural și către cele mai noi rezultate din domeniul inteligenței artificiale. Dezvoltarea însă trebuie abordată gradat, pe etape pentru a asigura caracterul operațional al construcției și nu distorsiunile specifice unui produs aflat într-un perpetuu proces de elaborare.

Bibliografie

- [IVAN04] IVAN I., VIȘOIU A. – *Generator de modele liniare cu argument întârziat*, Revista de Comerț, nr. 1, 2004, pag. 47 - 50
- [IVAN04a] IVAN I., ENYEDI R., POPA M. – *Software Cloning by Translation Processes*, Lucrările Conferinței Internaționale „Cibernetica la nivel macroeconomic și microeconomic”, București, 2004
- [IVAN 03] IVAN I. – *Baze de modele pentru managementul calității software*, Contract de cercetare, ASE, 2003
- [IVAN03a] IVAN I., POPA M., CAPISIZU S., LUKACS B., FLORESCU B. – *Clonarea informatică*, Editura ASE, București, 2003
- [IVAN03b] IVAN I., POPA M., CAPISIZU S., FLORESCU B., IVAN L. – *Characteristics of the Informatics Clones*, Master of International Business Informatics Handbook, Editura ASE, București, 2003, pg. 207 – 223
- [IVAN03c] IVAN I., POCATILU P., POPA M., MIHAI T., IVAN L. – *Data Orthogonality*, Master of International Business Informatics Handbook, Editura ASE, București, 2003, pg. 235 – 249
- [NICA 02] NICA, V. T., MĂRĂCINE, V. – *Modelarea firmei*, Editura ASE, București, 2002
- [PECI94] PECICAN, E. – *Econometrie*, Editura ALL, București, 1994
- [POPA02] POPA M. – *Software pentru măsurarea gradului de asemănare a fișierelor arhivate*, ASE, București, 2002