



## МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ

УДК 004.891.3+681.5

### СИНТЕЗ КОМБІНОВАНОГО ВИРІШАЛЬНОГО ПРАВИЛА (ВП) У КОМП'ЮТЕРНИХ СИСТЕМАХ МЕДИЧНОЇ ДІАГНОСТИКИ

**А.І. ПОВОРОЗНЮК**

Запропоновано метод синтезу комбінованого ВП у комп'ютерних системах медичної діагностики при взаємодії ієрархічних структур діагностичних ознак і станів, що діагностуються на основі аналізу апіорних умовних ймовірностей, їхніх невизначеностей та експертних оцінок структур симптомокомплексів.

#### ПОСТАНОВКА ПРОБЛЕМИ ТА АНАЛІЗ ЛІТЕРАТУРИ

Впровадження інформаційних технологій у медицину призвело до створення широкого спектру медичних інформаційних систем. На сьогоднішній день виділяють сім рівнів розвитку зазначених систем [1] — від автоматизованих медичних записів, до інтелектуальних комп'ютерних систем підтримки прийняття рішень у медицині з використанням медичних баз знань та застосуванням інтелектуальних технологій синтезу комп'ютерного діагнозу, моніторингу й оптимізації процесу лікування.

Актуальною оптимізаційною задачею, що вимагає першочергове вирішення [2, 3] є синтез ВП, відповідно до якого виконується процес діагностики (класифікації) пацієнтів для заданої множини станів (діагнозів), що діагностуються. У задачі синтезу ВП медичної діагностики використовуються два підходи.

1) У традиційній медичній діагностиці кожне захворювання (діагноз) характеризується деяким набором значень діагностичних ознак — симптомів, що утворюють симптомокомплекс цього захворювання. У неформалізованому вигляді симптоми захворювань наведено в різних медичних довідниках [4] і їх можна вважати експертною оцінкою захворювання, виробленою багатьма поколіннями лікарів. При цьому розрізняють такі типи симптомів: неспецифічні — виявляються при багатьох захворюваннях, що спричинені порушенням гомеостазу різних підсистем організму; специфічні — виявляються при даному класі захворювань; патогномонічні — характерні виключно для одного захворювання. У [1] виконано формалізацію опису симптомів під час розробки медичних баз даних (БД), а в [5] для діаг-

ностики в пацієнта одного або декількох захворювань використовується комбінаторна оптимізація при завданні симптомокомплексів бінарними векторами (але безпосереднє використання вказаного підходу можливе виключно для системи дихотомічних ознак).

2) Синтез ВП на основі аналізу навчальної вибірки. У цьому випадку використовуються такі методи: імовірнісні; на основі теорії розпізнання образів; нечіткої логіки; нейронних мереж та ін. Кожен із зазначених методів має свої переваги та недоліки, але в цих методах об'єкт діагностики представлено як «чорний ящик» й аналізується вектор діагностичних ознак. При такому представленні вихідних ознак розмірність задачі та її обчислювальний обсяг визначається розмірністю простору ознак (це обмеження в загальному переборне при підвищенні обчислювальної потужності сучасних комп'ютерів), але основним обмеженням указаних методів (особливо ймовірнісних) є підвищені вимоги до якості навчальної вибірки (підрахунок умовних імовірностей складних багаторозрядних ознак у формулі Байеса стає неможливим через відсутність необхідних даних у медичних БД). У той же час в [1] зазначається, що організм людини — складна, ієрархічна, багаторівнева, самоорганізуюча система, яка складається з взаємозалежних підсистем різного рівня підпорядкування.

Для адекватного відображення ієрархічної структури підсистем організму, автором у [6, 7] було розроблено процедуру синтезу ієрархічних структур інформативних діагностичних ознак  $S_z$  і діагностуємих станів  $S_D$ , взаємодію яких показано на рисунку. Структура  $S_D$  являє собою бінарне дерево, отримане в результаті ієрархічної кластеризації, причому коренем  $D^0$  є множина можливих діагнозів у цій предметній галузі, листами — стани  $D_i$ , що діагностуються, а гілками — кластери діагностуємих станів різного ступеня деталізації. Для синтезу  $S_D$  можуть бути використані класичні методи кластерного аналізу в просторі різнорідних діагностичних ознак. Структуру  $S_z$  було отримано в результаті перетворень  $F1: X \rightarrow S_x$ ,  $F2: S_x \rightarrow S_y$  і  $F3: S_y \rightarrow S_x$ . Перетворення  $F1$  є ієрархічною кластеризацією множини вихідних ознак  $X$  за критерієм мінімуму кореляційних зв'язків. У структурі  $S_x$  на нижньому рівні ієрархії можуть бути як окремі діагностичні ознаки, так і кластери корельованих ознак. У результаті перетворення  $F2$  визначається діагностична цінність  $I_D(x_i)$  вихідних ознак  $x_i$  щодо системи діагнозів  $\{D\}$ , виконується заміна кластерів нижнього рівня ієрархії структури  $S_x$  найбільш діагностично цінною ознакою та виконується заміна кластерів інших рівнів  $S_x$  інтегральними ознаками. Крім того, при синтезі  $S_y$  система різнорідних елементів  $S_x$  зводиться до єдиної шкали — кожна вихідна ознака  $x_i$  є непересічною множиною діагностично значимих інтервалів.

При обмеженому обсязі навчальної вибірки, заміна діапазону зміни чисельної ознаки  $\Delta = [x_{i\min}, x_{i\max}]$  послідовністю інтервалів є оптимізаційною задачею, тому що для підвищення точності апроксимації закону розподілу

$f(x)$  гістограмою  $G(x_k)$ , число інтервалів необхідно збільшувати, але при цьому в кожний із них попадає менше число точок, що знижує достовірність статистичних оцінок.

У [8] отримано вираз інтегральної помилки  $\varepsilon_k$ , що включає  $\varepsilon_a$  — помилку апроксимації та  $\sigma$  — невизначеність оцінки значень  $f(x)$  на інтервалі  $\Delta_k$ . Під час розбивки  $\Delta$  на  $m$  нерівномірних інтервалів  $\Delta_k$  і побудові відповідних гістограм

$$\varepsilon_k = \frac{\Delta_k}{\sqrt{6N_k}} \sqrt{\left(\frac{G_{k+1} - G_k}{\Delta_{k+1} + \Delta_k}\right)^2 + \left(\frac{G_k - G_{k-1}}{\Delta_k + \Delta_{k-1}}\right)^2}, \quad (1)$$

де  $N_k$  — число точок навчальної вибірки, що потрапили в інтервал  $\Delta_k$ ;  $G_k$  — ордината гістограми на інтервалі  $\Delta_k$ . У [8] запропоновано ітераційну процедуру реалізації методу градієнтного спуску для формування кількості  $m$  і значень  $\Delta_k$ , яка мінімізує  $\varepsilon_k$ .

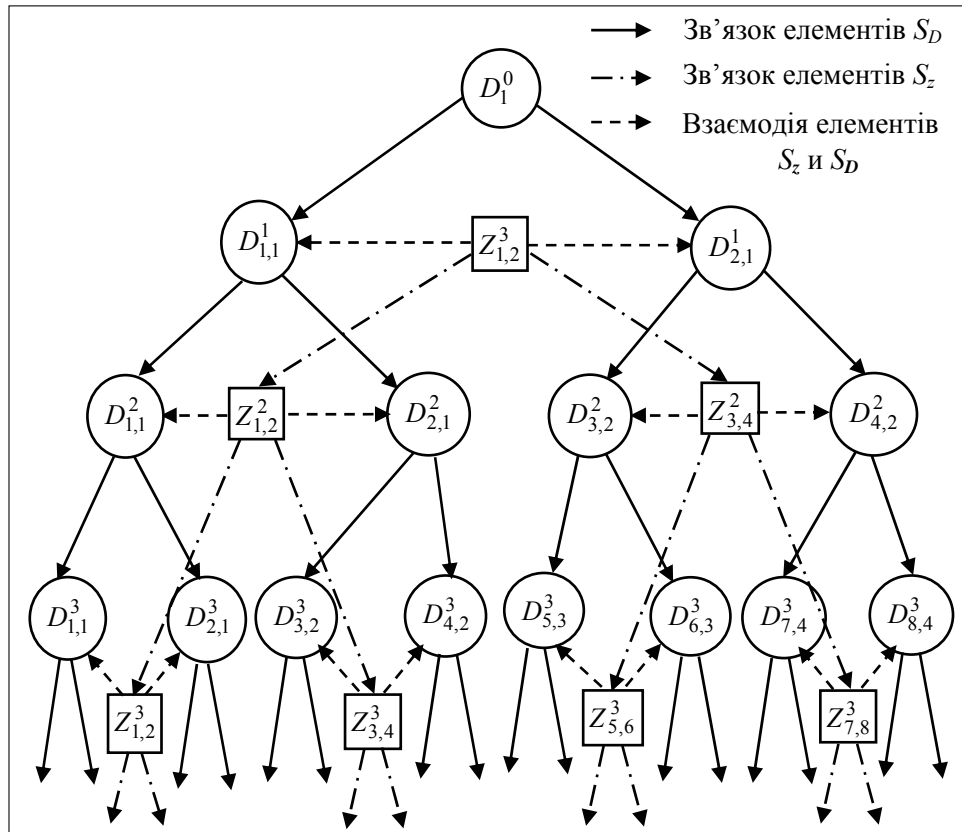


Рисунок. Взаємодія елементів ієрархічних структур  $S_z$  і  $S_D$

У результаті перетворення  $F3$  виконується узгодження топологій  $S_D$  і  $S_z$ , при цьому, кожній парі елементів  $S_D$ , що мають загального батька, ставиться у відповідність один елемент структури  $S_z$ , який є упорядкованою підмножиною елементів структури  $S_y$ , що забезпечує необхідний рівень

інформаційної повноти  $k_{\text{ін}}(X, D) = I_D(X)/H(D)$ , де  $H(D)$  (ентропія діагнозів), і однозначно задає їхній порядок включення у ВП на кожному етапі реалізації.

Реалізація комбінованого ВП при взаємодії  $S_z$  і  $S_D$ , враховуючи відзначені підходи й обмеженість навчальної вибірки, є перспективною.

**Мета роботи** — розробка комбінованого вирішального правила в кожній вершині взаємодії ієрархічних структур діагностичних ознак і станів, що діагностуються, яке враховує апіорні умовні ймовірності, їхні невизначеності, що обумовлені обмеженням об'ємом навчальної вибірки та структури симптомокомплексів у якості експертних оцінок.

### РОЗРОБКА ЙМОВІРНІСНОГО ВИРІШАЛЬНОГО ПРАВИЛА ПРИ ОБМЕЖЕНОМУ ОБ'ЄМІ НАВЧАЛЬНОЇ ВИБІРКИ

ВП диференційної діагностики двох станів  $D_q^i$  і  $D_l^i$  ( $i$  — номер етапу), будується на кожному етапі взаємодії ієрархічних структур (див. рисунок).

Для застосування ймовірнісного підходу необхідна наявність навчальної вибірки (медичної БД), яка містить інформацію про  $N$  пацієнтів, причому кожен  $i$ -пацієнт характеризується вектором діагностичних ознак  $X = (X_1^i, \dots, X_l^i, \dots, X_m^i)$  і кожному пацієнтові поставлено достовірний діагноз  $D_k^i$  із множини  $n$  можливих діагнозів  $(D_1, \dots, D_l, \dots, D_n)$ .

На підставі медичної БД формуються відповідні структури даних  $S_z$  і  $S_D$  за методикою, що описана в [6, 7]. Елементи структури  $S_D$  містять дані про апіорні ймовірності діагнозів  $P(D_i)$ , а елементи структури  $S_z$  містять гістограми умовних імовірностей належності значень ознаки  $x_i$  діапазонам  $\Delta_k$  для кожної пари діагностуємих станів свого рівня взаємодії  $G(z_{ik} / D_q)$  і  $G(z_{ik} / D_l)$ . Слід зазначити, що приведена в [6, 7] методика синтезу  $S_z$  дозволяє визначати не тільки оптимальні значення діапазонів, і ординат гістограм, але й невизначеність оцінки ординат гістограм  $\varepsilon_{ik}$  по (1), яка обумовлена апіорною невизначеністю навчальної вибірки.

Для реалізації ймовірнісного ВП, у роботі реалізовано метод послідовного аналізу (метод Вальда) на кожному етапі взаємодії ієрархічних структур  $S_z$  і  $S_D$ . Метод послідовного аналізу використовується для диференціальної діагностики й є ітераційною процедурою обстежень за допомогою системи незалежних ознак (бінарних, рангових або діагностичних інтервалів кількісних ознак). Критерієм зупинки ітераційної процедури є досягнення заданого рівня достовірності діагнозу або відмова від діагностики, якщо список діагностичних ознак вичерпаний, а заданого рівня достовірності не досягнуто.

У цьому випадку слабка залежність ознак забезпечується на етапі синтезу структури  $S_x$ , а кожен елемент структури  $S_z$  є впорядкованою за кри-

терієм діагностичної цінності множиною різнорідних діагностичних ознак, приведених до шкали діагностично-значимих інтервалів. Таким чином, при синтезі  $S_z$  виконано всі вимоги, необхідні для застосування методу послідовного аналізу.

При реалізації методу для кожного нового пацієнта вимірюється значення ознак у порядку, який заданий їхньою індексацією в елементі структури відповідного рівня  $Z_{q,l}^j$  (аналіз починається з верхнього рівня — елемента  $Z_{1,2}^1$ ), при цьому виконується диференційна діагностика станів  $D_q^j$  і  $D_l^j$ ) і на кожному  $i$ -му етапі обчислюється відношення правдоподібності

$$\Theta = \prod_i \frac{P(x_{ik} / D_q)}{P(x_{ik} / D_l)}, \quad (2)$$

яке порівнюється із порогами

$$\Theta > A, \quad (3)$$

$$\Theta < B, \quad (4)$$

де  $A$  і  $B$  — верхня і нижня границі невизначеності, необхідні для прийняття рішення.

Під час виконання (3) приймається рішення про діагноз  $D_q$ , а при виконанні (4) — про діагноз  $D_l$ . При виконанні будь-якої з умов виконується перехід на нижчий рівень ієрархії діагнозів, і виконується уточнення діагнозів. Під час невиконання (3) і (4) додається наступна  $i + 1$  ознака і процедура повторюється.

У послідовному аналізі границі прийняття рішення  $A$  і  $B$  зв'язані з помилками класифікації співвідношеннями

$$A = \frac{1 - \beta}{\alpha}, \quad B = \frac{\beta}{1 - \alpha}, \quad (5)$$

де  $\alpha$  — помилка першого роду, тобто ймовірність того, що пацієнтові з діагнозом  $D_q$  встановлено діагноз  $D_l$ ;  $\beta$  — помилка другого роду, тобто ймовірність того, що пацієнтові з діагнозом  $D_l$  встановлено діагноз  $D_q$ .

У «класичному» застосуванні послідовного аналізу не накладається жодних обмежень на рівні помилок  $\alpha$  і  $\beta$ . Їхні значення підбираються інтуїтивно за умови забезпечення антагоністичних показників якості діагностичної процедури (при збільшенні  $\alpha$  і  $\beta$  знижується достовірність діагнозу, а при зменшенні  $\alpha$  і  $\beta$  збільшується число відмов від постановки діагнозу).

Для реалізації методу послідовного аналізу для заданого набору діагнозів  $D_i$ ,  $i = \overline{1, n}$  і заданого набору діагностичних ознак  $x_j$ ,  $j = \overline{1, m}$ , необхідно розрахувати матрицю умовних ймовірностей виду

$$\mathbf{P} = \begin{pmatrix} p_{11} & \dots & p_{1j} & \dots & p_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ p_{i1} & \dots & p_{ij} & \dots & p_{im} \\ \dots & \dots & \dots & \dots & \dots \\ p_{n1} & \dots & p_{nj} & \dots & p_{nm} \end{pmatrix}. \quad (6)$$

Оскільки у ймовірнісному ВП використовуються умовні ймовірності діагностичних інтервалів різнорідних ознак, то кожен елемент матриці  $p_{ij}$  є множиною апріорних значень умовних ймовірностей  $p_{ij} = \{P(x_{ik} / D_j)\}$ , що визначаються по навчальній вибірці та використовуються у відношенні правдоподібності (2) на кожному кроці ітераційної процедури послідовного аналізу. Кількість елементів множини  $p_{ij} - m_{ij}$  залежить від типу діагностичної ознаки (для дихотомічних —  $m_{ij} = 2$ , для порядкових із числом градацій  $r$  —  $m_{ij} = r$ , для чисельних ознак значення  $m_{ij}$  визначається на етапі формування діагностично значимих інтервалів), однак, під час постановки діагнозу певному пацієнтові (конкретна реалізація ознаки  $x_{ik}$ , при якій заміряне значення ознаки  $x_i$  попадає в  $k$ -й інтервал) використовується тільки один елемент множини  $p_{ij} - p_{ijk} = (x_{ik} / D_j)$ .

Але з урахуванням обмеженого об'єму навчальної вибірки у відношенні правдоподібності (2), «істинні» умовні ймовірності замінюються їхніми оцінками — ординатами гістограм  $G(x_{ik} / D_q)$  і  $G(x_{ik} / D_l)$ , що мають відповідні невизначеності  $\varepsilon_{ik}^q$  і  $\varepsilon_{ik}^l$  обчислені по (1) (верхній індекс відповідає індексу стану, що діагностується, а нижні індекси —  $k$ -й інтервал  $i$ -ї ознаки, що виміряна у пацієнта). При заданому рівні значимості  $\xi$  (зазвичай  $\xi = 0,95$  або  $\xi = 0,99$ ) з використанням таблиць критичних значень розподілу Стьюдента, визначаються довірчі інтервали  $G(x_{ik} / D_q)$  і  $G(x_{ik} / D_l)$  ( $\delta_{ik}^q$  і  $\delta_{ik}^l$  відповідно)

$$\delta_{ik} = t_{\xi\lambda} \varepsilon_{ik}, \quad (7)$$

де  $t_{\xi\lambda}$  — критичне значення розподілу Стьюдента при рівні значимості  $\xi$  і числі ступенів свободи  $\lambda$  ( $\lambda = n - 1$ ).

Таким чином, враховуючи (7), умовні ймовірності  $P(x_{ik} / D_q)$  і  $P(x_{ik} / D_l)$ , які використовуються у відношенні правдоподібності (2), замінюються їхніми оцінками, які задано інтервальними змінами

$$\begin{aligned} P(x_{ik} / D_l) &= G(x_{ik} / D_l) \pm \delta_{ik}^l, \\ P(x_{ik} / D_q) &= G(x_{ik} / D_q) \pm \delta_{ik}^q. \end{aligned} \quad (8)$$

Підставивши (8) у (2), обчислюється відношення правдоподібності, використовуючи правила інтервальної арифметики [9], згідно із якими, для чисельних величин  $x$  і  $y$ , які задано інтервалами  $[x_H, x_B]$   $[y_H, y_B]$ , результат арифметичних операцій  $z$  теж задано інтервалом  $[z_H, z_B]$ , причому операції  $+$ ,  $-$ ,  $\times$ ,  $/$  визначено в такий спосіб

$$+: [z_H, z_B] = [x_H, x_B] + [y_H, y_B] = [x_H + y_H, x_B + y_B];$$

$$-: [z_H, z_B] = [x_H, x_B] - [y_H, y_B] = [x_H - y_H, x_B - y_B];$$

$$\times: [z_H, z_B] = [x_H, x_B] \times [y_H, y_B] = [\min(x_H \times y_H, x_H \times y_B, x_B \times y_H, x_B \times y_B), \max(x_H \times y_H, x_H \times y_B, x_B \times y_H, x_B \times y_B)];$$

$$/: [z_H, z_B] = [x_H, x_B] / [y_H, y_B] = [x_H, x_B] \times [1/y_H, 1/y_B] \quad 0 \notin [y_H, y_B]. \quad (9)$$

Зважаючи на те, що в (2) усі змінні є умовними ймовірностями і по визначенню невід'ємні, правила (9) стають простішими, тому що в операції множення  $\min$  досягається при  $x_H \times y_H$ , а  $\max$  при  $x_B \times y_B$ , і вираз (2) записується у вигляді верхніх та нижніх значень

$$\Theta_H = \prod_i \frac{G_H(x_{ik}/D_q)}{G_B(x_{ik}/D_l)} = \prod_i \frac{G(x_{ik}/D_q) - \delta_{ik}^q}{G(x_{ik}/D_l) + \delta_{ik}^l}, \quad (10)$$

$$\Theta_B = \prod_i \frac{G_B(x_{ik}/D_q)}{G_H(x_{ik}/D_l)} = \prod_i \frac{G(x_{ik}/D_q) + \delta_{ik}^q}{G(x_{ik}/D_l) - \delta_{ik}^l}, \quad (11)$$

а нерівності (3) і (4) записуються у вигляді «песимістичних» оцінок

$$\Theta_H > A, \quad (12)$$

$$\Theta_B < B. \quad (13)$$

Таким чином, використання (10) і (11) із перевіркою нерівностей (12) і (13) у методі послідовного аналізу дозволяє враховувати не тільки апріорні умовні ймовірності, але і довірчі інтервали їхніх оцінок, обумовлені істотною апріорною невизначеністю навчальної вибірки.

### ВНЕСЕННЯ ІНФОРМАЦІЇ ПРО СТРУКТУРУ СИМПТОМОКОМПЛЕКСІВ У КОМБІНОВАНЕ ВИРІШАЛЬНЕ ПРАВИЛО

Розглянемо спочатку процедуру внесення експертних оцінок про структуру симптомокомплексів під час «класичного» застосування методу послідовного аналізу.

Для формалізації інформації про структуру симптомокомплексів для заданого набору діагнозів  $D_i, i = \overline{1, n}$ , і заданого набору діагностичних ознак  $x_j, j = \overline{1, m}$ , будується матриця експертних оцінок  $E$  розмірності

$n \times m$ , у якій кожний  $i$ -й рядок описує набір симптомів (симптомокомплекс) для діагнозу  $D_i$ .

Елементи матриці  $e_{ij}$  приймають значення із множини  $\{e_3, e_2, e_1, e_0\}$  елементами якої є експертні оцінки наявності різних типів симптомів:  $e_3$  — вага патогномонічних;  $e_2$  — вага специфічних;  $e_1$  — вага неспецифічних;  $e_0$  — вага діагностичних показників, які не входять у симптомокомплекс даного захворювання. Природною умовою є виконання нерівності  $e_3 \geq e_2 \geq e_1 \geq e_0$ . Значення вагових коефіцієнтів  $e_i$  задає експерт (лікар-фахівець). У граничному випадку, коли не враховується диференціація ваг симптомів різних типів, приймаються значення  $e_3 = e_2 = e_1 = 1$ ,  $e_0 = 0$ .

Під час використання різномірних діагностичних ознак, кожен із елементів матриці  $e_{ij}$  представляється множиною  $e_{ij} = \{e_{ij}\}$ , елементи якої узгоджені з відповідними елементами множини  $p_{ijk}$ , тобто кількість елементів множин збігаються  $m_{ij}^e = m_{ij}^p$  і кожен елемент  $e_{ijk}$  відповідає елементові  $p_{ijk}$ . При цьому діагностичні ваги необхідно задавати не всій ознаці  $x_i$ , а кожному діагностичному інтервалові  $x_{ik}$ . Для дихотомічної ознаки, діагностична вага  $e_i$  задається інтервалові присутності ознаки, вага  $e_0$  — інтервалові відсутності ознаки, а для рангових і числових ознак, їхнє входження в симптомокомплекс зазвичай задається деяким термом (наприклад «висока температура»), тому за допомогою експерта необхідно уточнити границі терма  $\Delta_i = [x_{iH}, x_{iB}]$ , після чого вага  $e_i$  задається інтервалам  $x_{ik} \in \Delta_i$ , а вага  $e_0$  — інтервалам  $x_{ik} \notin \Delta_i$ .

Для реалізації комбінованого ВП необхідно використовувати обидві матриці, для чого будується матриця  $\mathbf{H}$ , елементами якої є поелементне множення матриць  $\mathbf{E}$  і  $\mathbf{P}$ . Таким чином, елементи матриці  $\mathbf{H}$  визначаються по виразу

$$h_{ijk} = e_{ijk} p_{ijk} \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad k = \overline{1, m_{ij}}. \quad (14)$$

При врахуванні розглянутого раніше методу заміни умовних ймовірностей їхніми оцінками  $G(x_{ik} / D_j)$  і довірчими інтервалами, необхідно використовувати відношення правдоподібності (10) і (11) та будувати відповідні значення матриць інтервальних оцінок  $G_B$  і  $G_H$ . Надалі, у виразах (10) і (11) замість елементів матриці  $\mathbf{P}$  будемо використовувати елементи матриці  $\mathbf{H}$ , які є зваженою мірою умовних ймовірностей системою експертних оцінок, а вирази (10) і (11) записується як

$$\Theta_H = \prod_i \frac{e_{ikq} G_H(x_{ik} / D_q)}{e_{ikl} G_B(x_{ik} / D_l)}, \quad (15)$$

$$\Theta_B = \prod_i \frac{e_{ikq} G_B(x_{ik} / D_q)}{e_{ikl} G_H(x_{ik} / D_l)}. \quad (16)$$



Це призводить до того, що умовні ймовірності тих діагностичних інтервалів  $x_{ik}$  ознак  $x_i$ , які входять у симптомокомплекс зазначеного діагнозу  $D_i$  збільшуються, а тих ознак, що не входять у симптомокомплекс, зменшуються (у граничному випадку, коли вага ознак, що не входять у симптомокомплекс  $e_0 = 0$ , умовні ймовірності зазначених ознак не враховуються). Варіюючи значеннями ваг  $e_i$ , можна задавати різний ступінь впливу окремих типів ознак, що входять у симптомокомплекс.

У розробленому в роботі методі уточнюючого діагнозу (див. рисунок), необхідно розробити діагностичні правила при всіх сполученнях взаємодії рівнів ієрархій діагностичних ознак і діагнозів. Варто нагадати, що в ієрархічній структурі ознак  $S_y$  усі рівні ієрархії (крім нижнього) є узагальненими ознаками  $\{X\}_i^k$ . Кожна узагальнена ознака  $k$ -го рівня характеризує кластер нижчого  $k-1$  рівня і є деякою функцією ознак  $k-1$  рівня, що входять у цей кластер  $x_i^k = \varphi(\{X\}_i^{k-1})$ . Для формування узагальнених ознак використовується метод головних компонент, який враховує лінійну модель зв'язків (необхідність побудови нелінійних моделей потребує додаткових досліджень в заданій предметній галузі). Таким чином, ця узагальнена ознака (перша головна компонента) є зваженою сумою вихідних ознак.

$$x_i^k = \sum_j a_j x_j^{k-1} \quad \forall j \in \{X\}_i^{k-1}. \quad (17)$$

Після визначення вагових коефіцієнтів функціональної залежності (17) можна розрахувати значення узагальнених ознак усіх рівнів ієрархії і занести їх у вихідну навчальну вибірку (значення узагальненої ознаки конкретного пацієнта є проекцією вихідних ознак на вісь головної компоненти). У такий спосіб розраховується розширена таблиця вихідних даних (навчальна вибірка), яка є основою для формування структури  $S_z$ , елементи якої включають мінімально необхідну множину інформативних вихідних ознак і узагальнених ознак усіх рівнів ієрархії синтезованої ієрархічної структури  $S_y$ . Розширена таблиця вихідних даних дозволяє для кожного виду взаємодії ієрархічних структур  $W^{km}$  (процедура постановки діагнозу пацієнтові з точністю до  $m$ -го рівня деталізації діагнозу під час використання  $k$ -го рівня діагностичних ознак) визначити матрицю умовних ймовірностей  $\mathbf{P}^{km}$  розмірністю  $N_x^k \times N_m$  ( $N_x^k$  — кількість ознак  $k$ -го рівня ієрархії ознак,  $N_m$  — кількість кластерів  $m$ -го рівня ієрархії діагнозів) безпосередньо шляхом підрахунку відповідних частот.

Для врахування експертних оцінок необхідно для кожного виду взаємодії ієрархічних структур  $W^{km}$  сформуванати відповідну матрицю експертних оцінок  $\mathbf{E}^{km}$  розмірністю  $N_x^k \times N_m$  (вихідна матриця експертних оцінок  $\mathbf{E}$  відображає взаємодію  $W^{11}$  нижніх рівнів ієрархії — вихідних ознак і діагнозів).

Розглянемо спочатку рівень взаємодії  $W^{21}$ , тобто взаємодія другого рівня ієрархії ознак і першого (нижнього) рівня ієрархії діагнозів. На цьому рівні маємо  $N_x^2 < m$  узагальнених ознак, кожна з яких отримана з відповідного кластера вихідних ознак по (17), і  $N_D^1 = n$  діагнозів. Необхідно сформуувати матрицю експертних оцінок  $E^{21}$  розмірністю  $N_x^2 \times N_D^1$  на підставі даних вихідної матриці  $E^{11}$  розмірністю  $N_x^1 \times N_D^1$  ( $N_x^1 = m$ ,  $N_D^1 = n$ ). Для цього виконуються такі дії.

1) Стовпці вихідної матриці  $E^{11}$  групуються на  $N_x^2$  груп таким чином, що кожна група відповідає кластерові вихідних ознак і узагальненій ознаці другого рівня.

2) Для кожної групи вихідної матриці обчислюється вектор-стовпець експертних оцінок узагальнених ознак по

$$e_{ij}^{21} = \sum_j a_{lj} e_{ij}^{11}, \quad \forall j \in \{X\}_l^1 \quad i = \overline{1, N_D^1}, \quad l = \overline{1, N_x^2}, \quad (18)$$

де  $a_{lj}$  — коефіцієнти функціональної залежності (17) для  $l$ -ї узагальненої ознаки.

Використання однієї і тієї ж залежності (17) як для обчислення значень узагальнених ознак, так і для обчислення їхніх експертних оцінок є логічно доцільним.

3) Зі сформованих по (18) векторів-стовпців складається матриця експертних оцінок  $E^{21}$ .

4) Визначаються елементи матриці  $H^{21}$  по виразу (14).

Шляхом аналогічних перетворень по даним матриць попереднього рівня  $E^{k-1,1}$  по (18) будуються матриці всіх наступних рівнів ієрархії ознак  $E^{k,1}$  і відповідні їм матриці  $H^{k,1}$ .

Розглянемо тепер випадок формування матриць експертних оцінок при підвищенні рівнів ієрархії діагнозів, тобто розглянемо спочатку рівень взаємодії  $W^{k,2}$  — взаємодію будь-якого  $k$ -го рівня ієрархії ознак і другого рівня ієрархії діагнозів. На цьому рівні маємо  $N_D^2 < n$  кластерів діагнозів відповідного рівня деталізації і  $N_x^k$  узагальнених ознак. Необхідно сформуувати матрицю експертних оцінок  $E^{k,2}$  розмірністю  $N_x^k \times N_D^2$  на підставі даних матриці  $E^{k,1}$  розмірністю  $N_x^k \times N_D^1$ . Для цього виконуються такі дії.

1) Рядки вихідної матриці  $E^{k,1}$  групуються на  $N_D^2$  груп таким чином, що кожна група відповідає кластерові діагнозів другого рівня ієрархічної структури діагнозів.

2) Для кожної групи вихідної матриці обчислюється вектор-рядок експертних оцінок кластерів по виразу

$$e_{ij}^{k,2} = \Psi_i e_{ij}^{k,1}, \quad \forall i \in \{D\}_l^1 \quad j = \overline{1, N_x^k}, \quad l = \overline{1, N_D^2}, \quad (19)$$

де  $\Psi e_{ij}^{k,l}$  — деяка функція об'єднання симптомокомплексів діагнозів, що входять в один кластер, у якості якої пропонуються міні складових, максимум середнє значення тощо, а конкретне значення узгоджується з експертами в цій предметній галузі.

3) Зі сформованих по (19) векторів-рядків складається матриця експертних оцінок  $\mathbf{E}^{k,2}$ .

4) Визначаються елементи матриці  $\mathbf{H}^{2,1}$  по виразу (14).

Шляхом аналогічних перетворень по даним матриці попереднього рівня  $\mathbf{E}^{k,l-1}$  по (19) будуються матриці всіх наступних рівнів ієрархії діагнозів  $\mathbf{E}^{k,l}$  і відповідні їм матриці  $\mathbf{H}^{k,l}$ .

Таким чином, розглянуто метод врахування експертних оцінок про структуру симптомокомплексів у комбінованому ВП на всіх рівнях взаємодії ієрархічних структур діагностичних ознак і станів, що діагностуються.

## ВИСНОВКИ

Розроблено метод синтезу комбінованого ВП під час взаємодії ієрархічних структур діагностичних ознак і станів, що діагностуються. Розроблене ВП дозволяє виконувати послідовний аналіз діагностичних ознак для досягнення заданого рівня достовірності діагнозу, при цьому виконується «песимістичний» прогноз, обумовлений невизначеністю апріорних умовних імовірностей під час врахування експертних оцінок структур симптомокомплексів.

## ЛІТЕРАТУРА

1. Весненко А.И., Попов А.А., Проненко М.И. Топо-типология структуры развернутого клинического диагноза в современных медицинских информационных системах и технологиях // Кибернетика и системный анализ. — 2002. — № 6. — С. 143–154.
2. Ахутин В.М., Шаповалов В.В., Иоффе М.О. Оценка качества формализованных медицинских документов // Медицинская техника. — 2002. — № 2. — С. 27–31.
3. Поворознюк А.И. Формализация этапов проектирования интеллектуальных компьютерных систем медицинской диагностики // Электронное моделирование. — Київ: ІПМЕ, 2006. — Т. 28. — № 1. — С. 85–97.
4. Справочник практического врача / Под ред. А.А. Михайлова, Л.И. Дворецкого. — М.: Новая Волна, 2001. — 528 с.
5. Тимофієва Н.К. Моделивання цільової функції в задачі клінічної діагностики на основі теорії комбінаторної оптимізації // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту. Матеріали міжнар. наук. конф. Євпаторія 18–22 травня 2009 р. Т. 1. — Херсон: ХНТУ, 2009. — С. 219–223.
6. Поворознюк А.И. Синтез иерархической структуры диагностических признаков в компьютерных системах медицинской диагностики // Вісн.

- Національного техн. ун-ту «ХПІ». Зб. наукових праць. Тематичний випуск: Системний аналіз, управління та інформаційні технології. — Харків: НТУ «ХПІ», 2003. — Т. 2, № 7. — С. 39–44.
7. *Поворознюк А.И.* Реконфигурация структуры диагностических признаков при синтезе компьютерного диагноза // Системы обработки информации. — 2007. — Вып. 2 (60). — С. 135–139.
  8. *Поворознюк А.И.* Формирование диагностических интервалов численных признаков при дифференциальной диагностике // Вісн. Хмельницького національного ун-ту. — Хмельницький: ХНУ, 2007. — 1, № 3. — С. 106–109.
  9. *Кветний Р.Н., Бойко О.Р.* Метод інтервального аналізу робастних систем управління // Матеріали ХІІІ междунар. конф. с автоматического управления (Автоматика 2006), г. Винница 25–28 сентября 2006 г. — Винница: УНИВЕРСУМ-Винница, 2007. — С. 103–108.

*Поступила 31.07.2009*