

## РОЗВ'ЯЗАННЯ ЗАДАЧІ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ СКЛАДЕНИХ ОБ'ЄКТІВ НА ОСНОВІ ВИВЕДЕННЯ ЗА АНАЛОГІЄЮ

© Величко В.Ю., 2001

**This paper is devoted to the solving task of property prediction on the basis of analogical inference. The methods and criteries of analogical inference are proposed. These methods are realized in ANALOGY software system.**

**У статті запропоновано методи і критерії виведення за аналогією. Ці методи реалізовані у програмній системі Аналогія.**

### Вступ

Виведення на основі фактів (Case-Based reasoning) є загальнопоширеною парадигмою в області автоматизації мислення та машинного навчання. При виведенні на основі фактів нова проблема вирішується за допомогою виявлення її подібності до вже вирішених проблем та адаптації відомого рішення до нової ситуації. Виведення на основі фактів складається з чотирьох основних етапів [1]. На першому етапі виконується аналіз бази відомих фактів (retrieve). На основі результатів першого етапу здійснюється формування гіпотези про властивості моделі (reuse). Далі відбувається перегляд сформованої гіпотези (revise). Підтверджена гіпотеза може бути внесена до бази фактів (retain).

Формування гіпотез може бути здійснено за допомогою індуктивного узагальнення та виведення за аналогією. При використанні індуктивного виведення на першому етапі здійснюється виведення загальної закономірності (правила), що узагальнює відомі факти. Отримане правило на другому етапі використовується для формування гіпотези про модель. При використанні виведення за аналогією посилки виведення відносяться до одного об'єкта, а гіпотеза - до іншого [2], причому кожній гіпотезі надається деяка оцінка впевненості виведення. Перший з цих об'єктів будемо називати прототипом  $q$ , а другий – моделлю  $m$ . Прикладом задачі, яка може бути розв'язана за допомогою даної технології, є задача прогнозування властивостей об'єкта дослідження. В цьому випадку на основі аналізу наявних знань про відомі властивості деякої сукупності об'єктів, опис яких представлено в базі фактів, може бути сформовано гіпотезу про властивості моделі.

У даній статті розглянемо методи формування гіпотез про властивості складених об'єктів на основі виведення за аналогією. Виведення за аналогією відіграє важливу роль в творчих процесах мислення, особливо при самонавчанні. Під складеними об'єктами будемо розуміти об'єкти, що складаються з деяких елементарних компонентів, які називатимемо первинними об'єктами. Для представлення бази знань буде використано структурно-атрибутивну модель знань.

### Структурно-атрибутивна модель знань

Для визначення структурно-атрибутивної моделі знань має бути задано кортеж  $\langle P, A, S, V \rangle$ . Елементи множини  $A$  будемо називати первинними об'єктами, елементи множини  $P$  – первинними властивостями, елементи множини  $S$  – складеними об'єктами. Складені об'єкти характеризуються деякою властивістю  $f$ , яка має визначений набір

значень  $V = \{v_j\}$ . На заданих множинах визначимо наступні відношення. Відношення  $R_1 \subseteq S \times V$  служить для опису значень властивості складених об'єктів:  $(s, v_j) \in R_1$  тоді і тільки тоді, коли складений об'єкт  $s$  має значення властивості  $v_j$ . Відношення  $R_2 \subseteq S \times A$  необхідно для представлення знань про склад відомих складених об'єктів:  $(s, a) \in R_2$  тоді і тільки тоді, коли первинний об'єкт  $a$  входить до складу складеного об'єкта  $s$ . Відношення  $R_3 \subseteq A \times P$  представляє первинні властивості первинних об'єктів:  $(a, p) \in R_3$  тоді і тільки тоді, коли первинний об'єкт  $a$  має первинну властивість  $p$ . Для представлення знань використовується семантична мережа. Вершини мережі розміщені за рівнями, а зв'язки між вершинами різних рівнів відображаються іменованими дугами, що відповідають визначеним відношенням.

Будемо називати релевантними до складеного об'єкта  $s$  ті первинні властивості, які мають первинні об'єкти, що входять до складу складеного об'єкта  $s$ . Нехай  $Z$  - деяка множина складених об'єктів. Первинну властивість  $p$  будемо називати релевантною до множини складених об'єктів  $Z$ , якщо існує складений об'єкт  $s \in Z$  та первинний об'єкт  $a \in A$  такий, що  $(s, a) \in R_2$  та  $(a, p) \in R_3$ . Первинні властивості, для яких зазначена умова не виконується, будемо називати нерелевантними.

### Теорія мультимножин

Для опису методів виведення за аналогією використаємо теорію мультимножин [3]. Множина є неупорядкований набір різних об'єктів. Неформально, мультимножина – це неупорядкований набір об'єктів, що не обов'язково різняться. Для того, щоб задати мультимножину, достатньо задати кратність елементів множини. Мультимножина  $\tilde{X}$  на множині  $X$  – це множина  $X$  разом з функцією  $r : X \rightarrow N_0$ , яка визначає кратність елементів  $X$ . Будемо позначати елементи мультимножини  $\tilde{X}$  за допомогою множини кортежів  $\{(x, r_{\tilde{X}}(x)) | x \in X\}$ . Позначка  $x \in \tilde{X}$  означає, що елемент  $x$  входить до множини  $\tilde{X}$  з ненульовою кратністю, тобто існує кортеж  $(x, r_{\tilde{X}}(x))$ , причому  $r_{\tilde{X}}(x) > 0$ .

Основні операції на мультимножинах визначаються так:

$$\text{Перетин } \tilde{X} \cap \tilde{Y} = \left\{ x^{\min(r_{\tilde{X}}(x), r_{\tilde{Y}}(x))} : x \in X \cup Y \right\}$$

$$\text{Об'єднання } \tilde{X} \cup \tilde{Y} = \left\{ x^{\max(r_{\tilde{X}}(x), r_{\tilde{Y}}(x))} : x \in X \cup Y \right\}$$

$$\text{Додавання } \tilde{X} \uplus \tilde{Y} = \left\{ x^{r_{\tilde{X}}(x) + r_{\tilde{Y}}(x)} : x \in X \cup Y \right\}$$

Множення мультимножини  $\tilde{X}$  на невід'ємне число  $\lambda$ :  
 $\lambda \bullet \tilde{X} = \left\{ x^{\lambda \bullet r_{\tilde{X}}(x)} : x \in X \right\}$ .

Причому має місце співвідношення  $\lambda \bullet \tilde{X} = \bigoplus_{i=1}^{\lambda} \tilde{X}$ . Нехай  $\tilde{B}(x)$  - довільна мультимножина. Будемо вважати, що  $\bigoplus_{x \in \tilde{X}} \tilde{B}(x) = \bigoplus_{x \in X} \left( \bigoplus_{i=1}^{r_{\tilde{X}}(x)} \tilde{B}(x) \right)$ . У зв'язку з тим, що

для операції додавання закон ідемпотентності не виконується, існує співвідношення

$$\biguplus_{x \in \tilde{X}} \tilde{B}(x) = \biguplus_{x \in X} (r_{\tilde{X}}(x) \bullet \tilde{B}(x)).$$

### Методи виведення за аналогією

Виведення за аналогією не є достовірним, а є виведенням, що має характер припущення. Тому напевно можна визначити лише необхідні умови існування подібності між моделлю та прототипом. Необхідною умовою існування подібності між двома первинними об'єктами будемо вважати наявність спільних первинних властивостей. Для двох складених об'єктів виконується необхідна умова існування подібності, якщо вони мають спільні релевантні первинні властивості. Інші, додаткові умови існування аналогії між двома складеними об'єктами визначаються залежно від евристичного критерію існування подібності та методу виведення за аналогією, які будуть описані далі. Критерії існування подібності відрізняються різними способами визначення кратності релевантних первинних властивостей. На основі значень кратності релевантних первинних властивостей буде визначено ступінь подібності двох складених об'єктів.

Введемо деякі позначення. Нехай  $Z \subseteq S$  – множина складених об'єктів,  $s \in Z$  – деякий складений об'єкт. Позначимо  $A(s) = \{a \mid (s, a) \in R_2\}$  – множина первинних об'єктів, які входять до складу складеного об'єкта  $s$ . Тоді мультимножина  $\tilde{A}(s)$  складається з наступних елементів  $\{(a, 1) \mid a \in A(s)\}$ , тобто значення  $r_{\tilde{A}(s)}(a) = 1, a \in A(s)$ . Позначимо  $P(a) = \{p \mid (a, p) \in R_3\}$  – множина первинних властивостей, які має первинний об'єкт  $a \in A$ . Тоді мультимножина  $\tilde{P}(a)$  складається з наступних елементів  $\{(p, 1) \mid p \in P(a)\}$ . Позначимо  $\tilde{A}_1(Z) = \bigcup_{s \in Z} \tilde{A}(s)$ . З того, що  $r_{\tilde{A}(s)}(a) = 1, a \in A(s)$ , випливає, що

$$r_{\tilde{A}_1(Z)}(a) = 1, a \in \tilde{A}_1(Z). \quad \text{Позначимо} \quad \tilde{A}_2(Z) = \biguplus_{s \in Z} \tilde{A}(s). \quad \text{Причому}$$

$$r_{\tilde{A}_2(Z)}(a) \geq 1, a \in \tilde{A}_2(Z).$$

У [4] були запропоновані декілька способів визначення кратності релевантних первинних властивостей на основі пошукових операцій в мережі (аналогія властивостей, структурна аналогія без обмежень, структурна аналогія з урахуванням первинного об'єкта один раз). Використаємо введені позначення та операції над мультимножинами для представлення різних способів визначення кратності релевантних первинних властивостей.

Нехай  $\tilde{P}_1(Z) = \biguplus_{a \in \tilde{A}_2(Z)} \tilde{P}(a)$ . Тоді вага  $n_Z^1(p) = r_{\tilde{P}_1(Z)}(p)$  первинної властивості

$p$  характеризує кратність релевантної первинної властивості  $p$ , яку визначено без додаткових обмежень.

Нехай  $\tilde{P}_2(Z) = \biguplus_{a \in \tilde{A}_1(Z)} \tilde{P}(a)$ . Тоді вага  $n_Z^2(p) = r_{\tilde{P}_2(Z)}(p)$  первинної

властивості  $p$  характеризує кратність релевантної первинної властивості  $p$ , яку визначено з урахуванням первинних об'єктів один раз. На відміну від першого способу кожний

первинний об'єкт, який входить до складу складених об'єктів  $Z$ , враховується один раз, бо  $r_{A_1}(a) = 1, a \in \tilde{A}_1(Z)$ .

Нехай  $\tilde{P}_3(Z) = \bigcup_{a \in \tilde{A}_1(Z)} \tilde{P}(a)$ , а  $\tilde{P}_4(Z) = \bigoplus_{s \in Z} P_3(\{s\})$ . Тоді вага

$n_Z^3(p) = r_{\tilde{P}_4}(Z)(p)$  первинної властивості  $p$  характеризує кратність релевантної первинної властивості  $p$ , яку визначено на основі аналогії властивостей. Це означає, що кратність первинної властивості  $p$ , релевантної до деякого складеного  $s$  об'єкта завжди дорівнює 1, а вага  $n_Z^3(p)$  дорівнює кількості складених об'єктів з множини  $Z$ , до яких первинна властивість  $p$  є релевантною. Таким чином, вага первинної властивості обчислюється без урахування структури складеного об'єкта.

Отже, критерії існування аналогії моделі і прототипу для структурно-атрибутивних моделей відрізняються за способами порівняння двох складених об'єктів, а саме за способами визначення кратності релевантних первинних властивостей. При цьому структура складеного об'єкта впливає на значення ваги  $n_Z^1(p)$  та  $n_Z^2(p)$ , тому відповідно будемо називати ці способи *структурна аналогія без обмежень*  $n_Z^1(p)$  та *структурна аналогія з урахуванням первинного об'єкта один раз*  $n_Z^2(p)$ . Ступінь подібності моделі і прототипу за критерієм аналогії властивостей визначається на основі загальної кількості спільних релевантних первинних властивостей, що мають модель та прототип.

При визначенні відносних ваг первинних властивостей можна враховувати потужність множини  $Z$ , тоді  $n_Z(p) = \hat{n}_Z(p) = \frac{n_Z(p)}{\text{Card}(Z)}$ . Такий спосіб визначення ваг

первинних властивостей дозволяє, не втрачаючи інформації про структуру складеного об'єкта, зменшити вплив на вагу первинної властивості потужності множини  $Z$ . При такому способі визначення вага називається нормованою по потужності множини  $Z$ .

Для визначення ваги складеного об'єкта спочатку визначаються ваги первинних об'єктів, які входять до його складу. Вага  $w_Z(a)$  первинного об'єкта  $a$  за множиною  $Z$  визначена як  $w_Z(a) = \sum_{p \in P(a)} n_Z(p)$ . Операція визначення ваги первинних об'єктів має

декілька форм, що відповідають різним способам визначення ваг первинних властивостей.

Для урахування переваги первинного об'єкта, який має більше релевантних первинних властивостей, ніж інший первинний об'єкт, для кожного первинного об'єкта  $a$  може бути обчислена кількість його релевантних первинних властивостей:  $d_Z(a) = \text{Card}(\{n_Z(p) \mid (n_Z(p) > 0) \& p \in P(a)\})$ . Тоді вага  $w_Z(a)$  первинного об'єкта  $a$  за множиною  $Z$  з урахуванням кількості його релевантних первинних властивостей визначена як  $w_Z(a) = d_Z(a) \times \sum_{p \in P(a)} n_Z(p)$ . Нормована величина

$\hat{d}_Z(a) = d_Z(a) / \text{Card}(P(a))$  відображає співвідношення кількості релевантних первинних властивостей до загальної кількості первинних властивостей первинного об'єкта  $a$ . За означенням  $\text{Card}(P(a)) \geq d_Z(a)$ , тому  $\hat{d}_Z(a) \leq 1$ . Тоді вага первинного об'єкта

визначається як  $w_Z(a) = \hat{d}_Z(a) \times \sum_{p \in P(a)} n_Z(p)$ . При такому способі визначення ваг

первинних об'єктів враховується співвідношення кількості релевантних первинних властивостей первинного об'єкта  $a$  та його нерелевантних властивостей.

Ступінь подібності складеного об'єкта  $s$  до складеного об'єкта  $q$  визначено згідно

з вагами його складових:  $\rho(s, q) = \sum_{a \in A(s)} w_q(a)$ . Ступінь подібності складеного об'єкта  $s$

до множини складених об'єктів  $Z$  визначено згідно з вагами його складових:

$\rho(s, Z) = \sum_{a \in A(s)} w_Z(a)$ . Нормована відносна вага для складеного об'єкта  $s$  щодо

складеного об'єкта  $q$  визначена як

$$\rho_q(s, Z) = \begin{cases} \rho(s, Z) / \rho(q, Z), & \text{якщо } \rho(s, Z) / \rho(q, Z) \leq 1 \\ 1, & \text{інакше} \end{cases}$$

За способом формування інтегральної оцінки подібності підмножини прототипів  $Z$  та моделі  $m$  методи дослідження подібності моделі та прототипів для структурно-атрибутивних моделей поділяються на метод загального аналізу та попарного порівняння. У методі загального аналізу досліджується наявність у моделі  $m$  тих первинних властивостей, що характерні для всієї множини прототипів  $Z$ . У результаті обчислюється міра  $\mu(m, Z) = \rho(m, Z)$  - інтегральна оцінка подібності моделі  $m$  до сукупності прототипів  $Z$ . Будемо називати  $\mu(m, Z)$  мірою моделі  $m$  за множиною  $Z$ .

Метод попарного порівняння здійснюється в два етапи. На першому етапі виконується дослідження подібності моделі  $m$  до кожного прототипу  $z_i$  з множини прототипів  $Z$ . При цьому згідно з одним з критеріїв обчислюється ступінь подібності моделі  $m$  до кожного прототипу  $z_i$ , тобто формується кортеж оцінок  $(\rho(m, z_1), \rho(m, z_2), \dots, \rho(m, z_k))$ , де  $\rho(m, z_i)$  - ступінь подібності моделі  $m$  щодо прототипу, складеного об'єкта  $z_i$ . На другому етапі методу попарного порівняння визначається інтегральна оцінка  $\mu(m, Z)$  подібності моделі  $m$  до множини прототипів  $Z$ , яка обчислюється як деяка функція від сукупності значень  $\rho(m, z_i)$  по тих прототипах  $z_i$ , що належать множині  $Z$ . Наприклад, такою функцією може бути сума

$$\mu(m, Z) = \sum_{z_i \in Z} \rho(m, z_i) \text{ чи максимум } \mu(m, Z) = \max_{z_i \in Z} \rho(m, z_i). \text{ Тобто у методі}$$

попарного порівняння здійснюється дослідження подібності моделі до кожного прототипу з множини прототипів, після чого формується інтегральна оцінка подібності моделі до множини прототипів. Метод загального аналізу є менш обчислювально складним, але дозволяє отримати лише узагальнений результат, який залежить від властивостей множини прототипів і не завжди забезпечує належну якість виведення.

Для методу попарного порівняння було розроблено додатковий критерій існування подібності складеного об'єкта  $s$  до складеного об'єкта  $q$ , який будемо називати критерієм

структурної аналогії властивостей. Нехай  $\tilde{P}_1(s) = \bigcup_{a \in A_2(s)} \tilde{P}(a)$ ,  $\tilde{P}_1(q) = \bigcup_{a \in A_2(q)} \tilde{P}(a)$  та

$\tilde{P}_5(s, q) = \tilde{P}_1(s) \cap \tilde{P}_1(q)$ . Тоді ступінь подібності складеного об'єкта  $s$  до складеного об'єкта  $q$  дорівнює  $\rho(s, q) = \sum_{p \in \tilde{P}_5} r_{\tilde{P}_5}(p)$ .

Таким чином, для структурно-атрибутивних моделей виведення за аналогією може бути здійснено двома методами (загального аналізу, попарного порівняння) з використанням запропонованих евристичних критеріїв існування аналогії двох складених об'єктів: аналогія властивостей, структурна аналогія властивостей, структурна аналогія без обмежень, структурна аналогія з урахуванням первинного об'єкта один раз. Міра подібності між моделлю та прототипом залежить від ступеня подібності моделі і прототипу, яка характеризується вагою моделі стосовно прототипу, та кількості прототипів, до яких модель є подібною.

### Розв'язання задачі прогнозування властивостей

Розроблений апарат виведення за аналогією використовується для розв'язання задач прогнозування властивостей складених об'єктів та проектування складу складених об'єктів [5]. Розглянемо розв'язання задачі прогнозування властивостей. Завдання полягає у визначенні невідомих характеристик об'єкта дослідження. Нехай задані множина прототипів  $Q \subseteq S$  складених об'єктів, що мають відомі значення властивості, і множина  $M$  складених об'єктів, властивості яких невідомі. Необхідно знайти значення властивості для кожного зі складених об'єктів  $m \in M$ .

Здійснимо розбиття множини прототипів  $Q$  за множиною  $V$  значень досліджуваної властивості. У результаті одержимо сімейство множин  $\mathcal{K} = \{Q_j\}_{j=1, \dots, N}$ , де  $Q_j = \{s \in Q \mid s \text{ має значення властивості } v_j \in V\} \subseteq Q$ . Для висунення гіпотези про значення властивості моделі  $m \in M$  необхідно досліджувати її подібність до кожної підмножини прототипів  $Q_j$ . У результаті буде отримано кортеж інтегральних оцінок подібності моделі  $m$  до кожної множини прототипів  $Q_j$  -  $(\mu(m, Q_1), \mu(m, Q_2), \dots, \mu(m, Q_n))$ . Будемо називати  $\mu(m, Q_j)$  мірою моделі  $m$  за значенням властивості  $v_j$ . Якщо міра  $\mu(m, Q_j)$  подібності моделі  $m \in M$  перевищує деякий поріг, то може бути висунута гіпотеза про те, що модель  $m$  має значення  $v_j$  властивості, що досліджується.

### Застосування

Розглянуті методи виведення було реалізовано у програмному комплексі "Аналогія". Програмний комплекс "Аналогія" – експертна система-оболонка для розв'язання задач, що пов'язані з аналізом структури об'єкта досліджень методами виведення за аналогією. Основними архітектурними компонентами програмного комплексу "Аналогія" є база знань системи, блок заповнення бази знань, діалогова система, блок аналізу бази знань, блок керування процесами розв'язування задач, словник предметної області, система шаблонів питань.

База знань системи “Аналогія” містить представлення структурно-атрибутивної моделі знань в семантичній мережі. Внесення змін до бази знань здійснюється за допомогою діалогової системи. Спількування системи з експертом здійснюється на основі обміну уточнюючих повідомлень: повідомлення користувачеві надходять у вигляді стандартних речень природною мовою, а від користувача – обмеженою природною мовою. Питання природною мовою формулюється у відповідності до системи шаблонів питань, яку можна настроювати на конкретну предметну область. Завдяки словнику термінів предметної області користувач має справу лише з функціонуванням системи на прикладному рівні, тобто всі повідомлення системи видаються в термінах області застосування системи. Блок аналізу бази знань системи “Аналогія” включає бібліотеку операцій обробки знань та модуль виведення за аналогією. Блок керування процесами розв’язування задач використовує розроблений апарат виведення за аналогією для розв’язування задач прогнозування властивостей складених об’єктів та проектування складу.

Програмний комплекс “Аналогія” було застосовано для розв’язання задач прогнозування властивостей хімічних сполук. В задачах прогнозування властивостей хімічних сполук залежності між заданими властивостями хімічних сполук та первинними властивостями хімічних елементів обумовлені значною кількістю факторів. Експериментальний аналіз методом спроб та помилок призводить до надмірних трудових, фінансових затрат та затрат часу. В зв’язку з цим надзвичайно важливо правильно спрямувати експерименти, що пов’язані з отриманням нових матеріалів [6]. Саме з цією метою поширення набули інтелектуальні системи автоматизації наукових досліджень.

Особливістю першої задачі (прогнозування типу кристалічної структури сполук ABCF6) є відносно невеликий об’єм вихідних знань про відомі сполуки. При розв’язанні цієї задачі були використані дані про 120 відомих сполук, 48 первинних об’єктів (хімічних елементів). Хімічні елементи характеризувалися 105 первинними властивостями. Досліджується властивість хімічних сполук - тип кристалічної структури, яка має 7 різних значень. Задовільні результати були отримані для 4 типів кристалічних структур: для типу Trirutile правильність висунутих гіпотез про об’єкти множини  $M$  становить 100%, для LiCaAl6 - 87,5%, для Na2Si6 – 50%, для RbNiCr6 – 65,3%. Хороші результати для об’єктів типу Trirutile були отримані як методом загального аналізу (100%), так і методом попарного порівняння (93,7%). Для інших типів кристалічних структур (LiCaAl6, Na2Si6, RbNiCr6) застосування методу загального аналізу дало незадовільні результати.

Розглянемо задачу прогнозування існування хімічних сполук складу АВ. Усі сполуки складаються з двох елементів. Складені об’єкти мають одне з двох значень властивості: утворювати сполуку або не утворювати сполуку. У задачі було представлено 89 хімічних елементів. Кожний хімічний елемент характеризувався значеннями 87 первинних властивостей. Було задано 923 прототипи, що утворюють сполуку, та 410 прототипів, що не утворюють сполуку. Для тестування було використано 692 складені об’єкти, з яких у 473 випадках утворювалась сполука, а у 219 випадках не утворювалась. Особливістю множини прототипів у цій задачі є те, що 87% хімічних елементів було представлено як у прикладах, що утворюють сполуки, так і у прикладах, що не утворюють сполуки. В зв’язку з цим ваги первинних властивостей по кожній підмножині прототипів  $n_{Q_1}(p)$  та  $n_{Q_2}(p)$ , що визначені методом загального аналізу, є приблизно однаковими. Тому метод загального аналізу не дав змоги отримати задовільних результатів.

Критерій існування аналогії двох складених об'єктів	Спосіб визначення ваги первинних об'єктів	Спосіб визначення ваги складених об'єктів	Відсоток правильних гіпотез
структурна аналогія без обмежень	$w_Z(a) = \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \sum_{a \in A(s)} w_q(a)$	60,86%
	$w_Z(a) = d_Z(a) \times \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \sum_{a \in A(s)} w_q(a)$	73,91%
	$w_Z(a) = \hat{d}_Z(a) \times \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \sum_{a \in A(s)} w_q(a)$	81,16%
	$w_Z(a) = \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \rho_q(s,Z)$	89,86%
	$w_Z(a) = d_Z(a) \times \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \rho_q(s,Z)$	86,96%
	$w_Z(a) = \hat{d}_Z(a) \times \sum_{p \in P(a)} n_Z(p)$	$\rho(s,q) = \rho_q(s,Z)$	92,75%
аналогія властивостей	незалежно від способу визначення ваги первинних об'єктів	$\rho(s,q) = \sum_{a \in A(s)} w_q(a)$	95,65%
структурна аналогія властивостей			97,10%

Для вибору найкращого способу визначення ваги первинного об'єкта та способу визначення ступеня подібності складених об'єктів випадковим чином було відібрано 69 складених об'єктів з тестової вибірки (1/10 від загальної кількості складених об'єктів тестової вибірки). Результати експериментів з використанням методу попарного порівняння наведено в таблиці.

Таким чином, найкращі результати були отримані згідно з критерієм структурної аналогії властивостей. Цей критерій разом з методом попарного порівняння використовувався для класифікації всієї тестової вибірки. Правильно було класифіковано 456 складених об'єктів, що утворюють сполуку, та 202 складені об'єкти, що не утворюють сполуку. Загалом для тестової вибірки відсоток правильних відповідей сягав 95,09%.

## Висновки

Програмний комплекс «Аналогія» є потужним засобом розв'язання широкого кола задач прогнозування невідомих властивостей складених об'єктів в різних предметних областях.

Особливістю задачі прогнозування типу кристалічної структури сполук типу ABCF6 є невеликий обсяг вихідних знань про відомі сполуки. Висока точність отриманих результатів підтверджує доцільність використання програмного комплексу «Аналогія» для розв'язання задач, коли інформація про об'єкт дослідження не є достатньою для виконання індуктивного узагальнення. Задача прогнозування існування подвійних хімічних сполук складу АВ характеризується достатнім обсягом множини прототипів (1333 складені об'єкти).



Таким чином, застосування виведення на основі аналогії є ефективним як при наявності представницької бази прецедентів, так і при невеликій кількості відомих прецедентів.

Експериментальні дослідження показали необхідність застосування різних критеріїв існування аналогії між моделлю і прототипом, а також різних методів виведення залежно від особливостей множини прототипів. В подальших дослідженнях планується зосередити увагу на тому, які властивості множини прототипів і яким чином впливають на якість результатів розв'язання задачі прогнозування властивостей.

1. A.Aamodt, E.Plaza *Case-based reasoning: foundational issues, methodological variations and System approaches*, *AI Communications* 7(1), 1994, pp.39-59 2. Уемов А.И. *Аналогия в практике научного исследования*, М.:Наука, 1970. – 264с. 3. Кнут Д.Е. *Искусство программирования для ЭВМ*, М., – 1977, Т.2 4. Величко В.Ю., Москалькова Н.М *Розв'язування задачі прогнозування властивостей для структурно-атрибутивних моделей за допомогою виведення за аналогією// Искусственный интеллект - 1999 - №2. - С.378-385* 5. Величко В.Ю., Москалькова Н.М. *Подход к решению задачи комплектации на основе вывода по аналогии// Управляющие системы и машины, - 1999 - №4. С.62-65* 6. Гладун В.П., Величко В.Ю., Киселева Н.Н., Москалькова Н.М. *Вывод гипотез о составе и свойствах объектов на основе аналогии// Искусственный интеллект, - 2000 - №1. С.44-52*

УДК 681.3

О.М. Верес

НУ «Львівська політехніка»,  
кафедра «Інформаційні системи та мережі»

## РЕПРЕЗЕНТАТИВНІСТЬ КРИТЕРІАЛЬНИХ ТЕСТОВИХ НОРМ СЛАБОСТРУКТУРОВАНОЇ ЗАДАЧІ УКЛАДАННЯ РОЗКЛАДУ

© Верес О.М., 2001

**The principles of normalization of a scale of test dancing parties are represented. The operations on analysis of allocation of test dancing parties, construction of test norms and check them representative.**

**Подані принципи нормалізації показників оптимізаційних критеріїв, множина яких формується з використанням тест-опитування.**

Після визначення множини запитань тесту виникає проблема психометрії тестів – питання про тестові норми [1]. Дві основні передумови традиційної західної тестології: питання застосування статистичних норм (квантилів розподілу балів) як діагностичної норми і питання зведення всіх емпіричних розподілів до нормальної моделі.

### Статистична природа тестових шкал

Типічний вимірювальний тест у психодіагностиці – це послідовність коротких завдань або пунктів, що дає, в результаті виконання її досліджуваними, послідовність початкових значень, яка потім приводиться до однозначної кількісної інтерпретації.

Сумарний бал по тесту підраховується за допомогою ключа: ключ встановлює числове початкове значення по кожному пункту. Вихідне по окремому завданню піддається впливу не тільки з боку фактора, що вимірюється – здібності або риси досліджуваного, але й сторонніх шумових факторів, які є іррелевантними стосовно задачі вимірювання. Основний прийом, що дозволяє ліквідувати спотворюючий вплив факторів на результат (сумарний