

КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

Наведена методологія вирішення задачі прогнозування властивостей складених об'єктів на основі структурно-атрибутивної моделі знань. Запропонована методологія дозволяє обґрунтовано обрати метод виведення за аналогією та підвищити якість результатів виведення. Запропонований спосіб визначення числових характеристик множини прототипів, на основі її попереднього аналізу, може бути використаний для вибору найбільш ефективного методу виведення за аналогією.

© В.Ю. Величко, Н.М. Москалькова, 2005

УДК 681.3(16)

В.Ю. ВЕЛИЧКО, Н.М. МОСКАЛЬКОВА

МЕТОДОЛОГІЯ ВИРІШЕННЯ ЗАДАЧІ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ ОБ'ЄКТІВ МЕТОДАМИ ВИВЕДЕННЯ ЗА АНАЛОГІЄЮ НА ОСНОВІ СТРУКТУРНО-АТРИБУТИВНОЇ МОДЕЛІ ЗНАНЬ

В області автоматизації мислення та машинного навчання все частіше використовуються такі алгоритми аналізу даних, що спираються на конкретні факти поведінки об'єкта дослідження або на "прецеденти". При цьому використовується проста, але фундаментальна гіпотеза про монотонність простору рішень, яку можна виразити так: "Схожі вхідні ситуації призводять до схожих вихідних реакцій системи" [1]. Основою багатьох методів, зміст яких складає аналіз множини прецедентів (або прототипів) і знаходження серед них найбільш подібних до об'єкта дослідження (моделі), є виведення за аналогією. Різноманіття форм виведення за аналогією визначається насамперед характером інформації, яка переноситься з прототипу на модель. В одних випадках переноситься інформація, що представляє собою приписування предмету властивості [2], в інших випадках йдеться про перенос відносин [3]. Відповідно до цього Уйомов розрізняє аналогію властивостей і аналогію відносин [4]. Надалі будемо розглядати тільки аналогію властивостей, але враховуючи структуру прецедентів, а також властивості окремих об'єктів, з яких складаються прецеденти.

Сформулюємо задачу прогнозування властивостей об'єктів. Опис прецедентів здійснюється за допомогою структурно-атрибутивної моделі знань $\langle P, A, S, V \rangle$, де елементи множини A будемо називати первинними

об'єктами, елементи множини P – властивостями первинних об'єктів (первинними властивостями), елементи множини S – складеними об'єктами [5]. Складені об'єкти характеризуються деякою властивістю, яка має визначений набір значень $V = \{v_j\}$. Нехай задані множина прототипів $Q \subseteq S$ складених об'єктів, що мають відомі значення властивості, і множина M складених об'єктів, деякі властивості яких невідомі. Необхідно встановити значення властивості для кожного зі складених об'єктів $m \in M$.

Здійснимо розбиття множини прототипів Q за множиною V значень досліджуваної властивості $Q_j = \{s \in Q \mid s \text{ має значення властивості } v_j \in V\} \subseteq Q$.

Для вирішення задачі необхідно дослідити подібність моделі m до кожної підмножини прототипів Q_j . За способом формування інтегральної оцінки подібності підмножини прототипів Q_j і моделі m можна виділити два методи: загального аналізу і попарного порівняння [6].

У методі загального аналізу [7] досліджується наявність у моделі m тих первинних властивостей, що характерні для всієї множини прототипів Q_j . Внаслідок цього формується міра подібності моделі m до множини прототипів $Q_j - \mu(m, Q_j)$. При використанні методу попарного порівняння [7] для кожної моделі $m \in M$ обчислюється кортеж оцінок

$$(\rho(m, q_1), \dots, \rho(m, q_i), \dots, \rho(m, q_{Card(Q_j)})) ,$$

де $\rho(m, q_j)$ – вага моделі m щодо складеного об'єкта q_i .

Міра $\mu(m, Q_j)$ моделі m за значенням властивості v_j може бути визначена за формулою $\mu(m, Q_j) = \sum_{q_i \in Q_C^+(m) \cap Q_j} \rho(m, q_i)$, де множина $Q_C^+(m)$ – це множина прототипів, вага яких входить до C найбільших значень ваг моделі m [8].

У методі загального аналізу кількість виконуваних обчислювальних операцій менша, порівняно з використанням методу попарного порівняння. Однак метод загального аналізу доцільно використовувати для вирішення таких задач, в яких множина прототипів задовольняє певним умовам. У статті запропонований спосіб визначення числових характеристик множини прототипів на основі її попереднього аналізу. Визначені характеристики використовуються для вибору найбільш ефективного методу виведення за аналогією. Наведена методологія вирішення задачі прогнозування властивостей для структурно-атрибутивної моделі знань дозволяє обґрунтовано обрати метод виведення за аналогією та підвищити якість результатів виведення.

Визначення характеристик множини прототипів. Розглянемо деяку множину Q_j . Первинну властивість p будемо називати релевантною до множини складених об'єктів Q_j , якщо існує складений об'єкт $q \in Q_j$, до складу якого входить первинний об'єкт $a \in A$, що має первинну властивість p . Первинний об'єкт a будемо називати релевантним до значення властивості складених об'єктів v_j ,

якщо існує складений об'єкт $q \in Q_j$, який має значення властивості v_j та до складу q входить первинний об'єкт $a \in A$.

Для кожного складеного об'єкта $q \in Q_j$ за допомогою методу попарного порівняння обчислимо вагу $\rho(q, q_i) = \sum_{a \in A(q)} \sum_{p \in P(a)} k_p \cdot n_{q_i}(p)$ до кожного складеного об'єкта $q_i \in Q_j \setminus \{q\}$, де $n_{q_i}(p)$ – кратність релевантної первинної властивості p до прототипу q_i , k_p – коефіцієнт значимості первинної властивості, який визначається експертом предметної області [9]. Позначимо $\tau(q, Q_j \setminus \{q\})$ – міру подібності складеного об'єкта q до інших прототипів множини Q_j , яку обчислимо за формулою

$$\tau(q, Q_j \setminus \{q\}) = \sum_{q_i \in Q_j \setminus \{q\}} \rho(q, q_i) / \text{card}(Q_j \setminus \{q\}).$$

Таким чином, $\tau(q, Q_j \setminus \{q\})$ характеризує міру подібності прототипу $q \in Q_j$ до множини прототипів за значенням властивості, яке має q . Будемо називати $\tau(q, Q_j \setminus \{q\})$ власною мірою прототипу q .

Визначимо сукупність мір подібності прототипу q до множин прототипів за іншими значеннями властивості

$$(\tau(q, Q_1), \dots, \tau(q, Q_{j-1}), \tau(q, Q_{j+1}), \dots, \tau(q, Q_{\text{Card}(V)})),$$

де $\tau(q, Q_i) = \sum_{q_k \in Q_i} \rho(q, q_k) / \text{card}(Q_i)$, $i = \overline{1, \text{Card}(V)}$, $i \neq j$.

Будемо називати $\tau(q, Q_i)$ мірою подібності прототипу q за значенням властивості v_i .

Визначені характеристики можна використати для аналізу множини прототипів. Нехай для деякого прототипу $q \in Q_j$ існує i таке, що виконується $\tau(q, Q_j \setminus \{q\}) < \tau(q, Q_i)$. Тоді можна зробити висновок про те, що прототип q більш подібний до прототипів, що мають значення властивості v_i . Тому прогнозування значення властивості для цього об'єкта буде помилковим, якщо його вилучити з множини прототипів. Наявність таких об'єктів у множині прототипів суттєво впливає на точність прогнозування. Цей факт покладено в основу визначення наступних характеристик множини прототипів Q_j :

1. Відносна кількість прототипів, більш подібних до прототипів, що мають інші значення властивості:

$$v_j = \text{Card}(\{q \in Q_j \mid \exists i \neq j : \tau(q, Q_j \setminus \{q\}) < \tau(q, Q_i)\}) / \text{Card}(Q_j), 0 \leq v_j \leq 1. \quad (1)$$

2. $\tau(v_j)$ – міра значення властивості v_j , яку будемо визначати як середнє арифметичне значень мір прототипів q , що мають значення властивості v_j ,

$$\tau(Q_j, v_j) = \tau(v_j) = \sum_{q_k \in Q_j} \tau(q_k, Q_j \setminus \{q_k\}) / (\text{card}(Q_j) - 1) \quad (2)$$

та середнє арифметичне значення мір прототипів $q \in Q_j$ для кожного значення властивості $v_i \Big|_{i \neq j}$ за формулою

$$\tau(Q_j, v_i) = \sum_{q_k \in Q_j} \tau(q_k, Q_i) / \text{card}(Q_j), i = \overline{1, l}, i \neq j. \quad (3)$$

3. Мінімальне відносне значення різниці мір значень властивостей за формулою

$$\tilde{\tau}(v_j) = \min_{i \neq j} ((\tau(v_j) - \tau(Q_j, v_i)) / \tau(v_j)). \quad (4)$$

Якщо для деякої властивості v_i різниця мір $\tau(v_j) - \tau(Q_j, v_i)$ – мінімальна, то складені об'єкти, що мають значення властивості v_j , подібні до складених об'єктів, що мають значення властивості v_i . На основі значень v_j та $\tilde{\tau}(v_j)$ можна зробити висновок про ступінь подібності прототипів, які мають значення властивості v_j . Розглянемо використання характеристик множини прототипів для вирішення задачі прогнозування властивостей складених об'єктів.

Методологія вирішення задачі прогнозування властивостей методами виведення за аналогією. Експерт предметної області згідно зі своїми знаннями здійснює такі операції:

1. Визначення коефіцієнтів значимості k_p первинних властивостей, які впливають на обчислену вагу релевантних первинних властивостей.

2. Обрання найкращого способу підрахунку ваг первинних властивостей та первинних об'єктів. Для кожної множини прототипів Q_j за допомогою різних методів та евристичних критеріїв виведення за аналогією виконується аналіз заданих складених об'єктів та дослідження подібності первинних об'єктів. Первинні об'єкти впорядковуються за зменшенням ваги до множини прототипів Q_j . Експерт обирає, який спосіб впорядкування первинних об'єктів, що є релевантними для досліджуваного значення властивості, є найкращим. На основі обраного порядку визначається відповідний найкращий спосіб підрахунку ваг.

3. Уточнення складу та властивостей об'єктів з множини прототипів на основі обчислених характеристик v_j , $\tau(Q_j, v_i)$, $\tilde{\tau}(v_j)$ множини прототипів Q_j :

- переглянути перелік складених об'єктів $q \in Q_j$, для яких при $i \neq j$ $\tau(q, Q_j \setminus \{q\}) < \tau(q, Q_i)$ з метою перевірки правильності даних та доцільності їх

використання в множині прототипів. У першу чергу необхідно розглянути складені об'єкти, що мають значення властивості v_j , для яких v_j максимальне;

- доповнити множину прототипів складеними об'єктами, які дозволять розрізнити складені об'єкти зі значеннями властивості v_j та складені об'єкти зі значеннями властивості v_i , для яких різниця мір $\tau(v_j) - \tau(Q_j, v_i)$ – мінімальна;

- розширити простір первинних властивостей такими властивостями, що дають змогу розрізнити складені об'єкти зі значеннями властивості v_j та складені об'єкти зі значеннями властивості v_i , для яких різниця мір $\tau(v_j) - \tau(Q_j, v_i)$ – мінімальна;

- етап формування гіпотези для значень властивості v_j , для яких $\tilde{\tau}(v_j) < 0$, проводити не рекомендується, бо результат буде мати дуже низьку достовірність. Якщо у множину прототипів були внесені зміни, необхідно повторити операції етапів 1 та 2.

4. Формування вторинних властивостей складених об'єктів. Вторинною властивістю складеного об'єкта будемо називати таку ознаку, яка характеризує складений об'єкт в цілому, а не тільки окремі його компоненти. Вторинна властивість може бути визначена експертом або обчислена як функція властивостей первинних об'єктів за допомогою математичних операцій над однойменними значеннями властивостей первинних об'єктів, які є релевантними складеному об'єкта.

5. Вибір методу виведення та евристичних критеріїв виведення за аналогією. Для кожної множини прототипів Q_j розв'язуємо задачу мінімізації набору прецедентів. Найкращим вважаємо критерій та спосіб підрахунку ваг, при використанні якого отримуємо мінімальний набір прототипів. Експерт порівнює способи підрахунку ваг, що визначені при виконанні п. 2 та цього пункту, і обирає метод виведення за аналогією на основі значень v_j , $\tau(Q_j, v_i)$, $\tilde{\tau}(v_j)$.

Якщо внаслідок виконання пп. 2, 5 не вдалося обрати найкращий спосіб підрахунку ваг первинних властивостей і первинних об'єктів, то розв'язуємо задачу прогнозування властивостей для тестової множини моделей, властивості яких відомі, обраним у п. 5 методом. За результатами розв'язання задачі вибираємо найкращий спосіб підрахунку ваг.

6. Формування гіпотези про значення властивості моделей за допомогою методу виведення та евристичних критеріїв, обраних на попередніх етапах. При використанні методу загального аналізу доцільно уточнити гіпотезу про значення властивості моделі за допомогою методу попарного порівняння для подібних значень властивості складених об'єктів – Ω_j^s . На першому кроці отримуємо множину

$$\{\Omega_j^1\} = \left\{ (v_j, v_i) \mid \min_k (\tau(Q_j, v_j) - \tau(Q_j, v_k)) = \tau(Q_j, v_j) - \tau(Q_j, v_i) \right\}.$$

Для формування Ω_j^s виконуємо ітераційну процедуру: поки на деякому кроці $s \exists j \exists i \neq j : \Omega_j^s \cap \Omega_i^1 \neq \emptyset$ та $\Omega_j^s \not\subset \Omega_i^1$ виконати $\Omega_j^{s+1} = \Omega_j^s \cup \Omega_i^1$. Метод попарного порівняння виконуємо для кожної моделі, якій було прогнозовано значення властивості з множини Ω_j^s , причому до множини прототипів необхідно включити лише ті складені об'єкти з множини Q , що мають значення властивості з множини Ω_j^s . Це дозволяє водночас уточнити гіпотезу про значення властивості моделі і зменшити кількість обчислювальних операцій, порівняно з визначенням міри моделі методом попарного порівняння, за рахунок зменшення кількості прототипів, щодо яких визначається міра моделі.

7. Представлення результату виведення експерту. У випадку якщо деякі гіпотези підтвердились, за ініціативою експерта, можна додати відомості про складені об'єкти до бази фактів.

Задача прогнозування типу кристалічної структури хімічних сполук.

Розглянемо використання запропонованої методики для розв'язання задачі прогнозування типу кристалічної структури хімічних сполук структури ABCF₆. Властивість, що досліджується, мала 7 значень: Na₂SiF₆, Trirut, LiCaAlF₆, RbNiCrF₆, CsAgAlF₆, An_str, No_comp. Множина прототипів Q мала 113 прикладів, а тестова множина моделей M – 76 прикладів.

Для вибору методу виведення за аналогією було розраховано параметри $v_j, \tau(Q_j, v_i), \tau(v_j), \tilde{\tau}(v_j)$ (формули (1), (2), (3), (4)), які і наведено в табл. 1.

ТАБЛИЦЯ 1. Характеристики множини прототипів

		$\tau(v_j), \tau(Q_j, v_i)$						$\tilde{\tau}(v_j)$	v_j	
$v_j \backslash v_i$		Na ₂ SiF ₆	Trirut	LiCaAlF ₆	RbNiCrF ₆	CsAgAlF ₆	An_str			No_comp
Na ₂ SiF ₆		31,00	23,57	23,78	7,74	8,43	8,64	6,42	0,23	0,36
Trirut		25,68	37,71	29,58	12,82	6,67	10,53	9,85	0,22	0
LiCaAlF ₆		25,32	29,56	32,92	5,21	6,27	9,27	7,71	0,10	0,13
RbNiCrF ₆		8,48	13,67	6,23	29,43	28,94	10,02	14,28	0,02	0,56
CsAgAlF ₆		8,73	6,67	5,86	31,15	43,7	6,18	10,6	0,29	0
An_str		9,67	9,92	9,12	7,39	4,64	19,03	21,55	-0,13	0,93
No_comp		7,23	10,18	7,52	8,12	5,75	21,58	22,16	0,03	0,5

На основі аналізу табл. 1 можна зробити наступні висновки: 1) подібність прототипів для кожного значення властивості Trirut та CsAgAlF₆ – висока; 2) подібність прототипів для значень властивості LiCaAlF₆ та Na₂SiF₆ – середня; 3) подібність прототипів кожного значення властивості An_str, No_comp, RbNiCrF₆ – низька. Крім того, подібні прототипи, які мають різні типи кристалічних структур: Na₂SiF₆ ~ LiCaAlF₆, Trirut ~ LiCaAlF₆, RbNiCrF₆ ~ CsAgAlF₆, An_str ~ No_comp.

Порівняємо результати прогнозування для різних значень властивості, які отримано за допомогою методів попарного порівняння та загального аналізу (табл. 2).

ТАБЛИЦЯ 2. Результати прогнозування значення властивості

Значення властивості	Кількість об'єктів	Кількість об'єктів, для яких правильно спрогнозовано значення властивості	
		Метод попарного порівняння	Метод загального аналізу
Na ₂ SiF ₆	6	4 [66,67 %]	3 [50 %]
Tirut	16	15 [93,75 %]	16 [100 %]
LiCaAlF ₆	8	8 [100 %]	8 [100 %]
RbNiCrF ₆	26	25 [96,15 %]	16 [61,54 %]
CsAgAlF ₆	5	4 [80 %]	4 [80 %]
An_str	13	11 [84,62 %]	4 [30,8 %]
No_comp	2	1 [50 %]	1 [50 %]
Загалом	76	68 [89,47 %]	52 [68,42 %]

Найменші значення v_j було отримано для трьох значень властивості: LiCaAlF₆, Tirut, CsAgAlF₆. Для цих значень властивості значення $\tilde{\tau}(v_j)$ є одними з найбільших. Саме тому правильність висунутих гіпотез про об'єкти множини M склала 100% для LiCaAlF₆, 93,75% – для Tirut та 80% – для CsAgAlF₆ як за допомогою методу загального аналізу, так і методу попарного порівняння. Для інших значень властивості ступінь подібності прототипів є низькою, тому кращі результати дало використання методу попарного порівняння: Na₂SiF₆ – 66,67%, RbNiCrF₆ – 96,15%, An_str – 84,62% правильно висунутих гіпотез.

Висновки. Наведена методологія вирішення задачі прогнозування властивостей методами виведення за аналогією дозволяє обґрунтовано обрати спосіб підрахунку ваг первинних властивостей і первинних об'єктів, метод виведення за аналогією, зменшити обсяг обчислень та підвищити достовірність отриманих результатів. Точність прогнозування значення властивості склала: Na₂SiF₆ – 66,67%, Tirut – 100%, LiCaAlF₆ – 100%, RbNiCrF₆ – 96,15%, CsAgAlF₆ – 80%, An_str – 84,62%, No_comp – 50%. Загалом для всіх об'єктів тестової вибірки точність прогнозування становить 90,79%.

Експериментальні дослідження показали необхідність застосування різних критеріїв існування аналогії між моделлю і прототипом, а також різних методів виведення в залежності від характеристик множини прототипів: v_j , $\tilde{\tau}(v_j)$. Для значень властивості з низьким ступенем подібності прототипів слід використовувати метод попарного порівняння. Якщо для деякого значення властивості прототипи неподібні, необхідно внести зміни в множину прототипів або не вирішувати задачу для цього значення властивості. Подальші дослідження в цьому напрямку потрібні у використанні запропонованої методології до моделей, які потрапили до області невизначеності та невпевнених відповідей щодо різних значень досліджуваної властивості [8], що дасть змогу підвищити достовірність результатів виведення. Доцільно було б виконати статистичне обґрунтування визначення ступеня подібності прототипів (високий, середній та низький) на основі обчислених оцінок v_j та $\tilde{\tau}(v_j)$.

1. *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во Ин-та математики, 1999. – 270 с.
2. *Гладун В.П.* Партнерство с компьютером. Человеко-машинные целеустремленные системы. – Киев: Port-Royal, 2000. – 128 с.
3. *Gentner D.* Structure-Mapping: A theoretical framework for analogy // *Cognitive Science*, 7. – 1983. – P. 155 – 170.
4. *Уемов А.И.* Аналогия в практике научного исследования. – М.: Наука, 1970. – 264 с.
5. *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.
6. *Величко В.Ю., Москалькова Н.М.* Использование программного комплекса “Аналогия” для формирования гипотез о свойствах составных объектов // *Проблемы программирования*. – 2002. – № 1–2. – С. 445 – 452.
7. *Величко В.Ю., Москалькова Н.М.* Розв’язування задачі прогнозування властивостей для структурно-атрибутивних моделей за допомогою виведення за аналогією // *Штучний інтелект*. – 1999. – № 2. – С. 378 – 385.
8. *Величко В.Ю.* Розв’язання задачі прогнозування властивостей складених об’єктів на основі виведення за аналогією // *Вісн. Держ. ун-ту "Львівська політехніка" Сер. "Інформаційні системи та мережі"*. – 2001. – № 438. – С. 26 – 34.
9. *Величко В.Ю.* Прогнозирование свойств составных объектов с помощью программного комплекса “Аналогия” // *KDS-2001: Сб. науч. тр.* – Санкт-Петербург, 2001. – 1. – С. 77 – 82.

Отримано 23.05.2005