

## ***Класифікація ситуацій в умовах невизначеності за допомогою наближених множин***

У статті розглянуто питання створення теоретичного підґрунтя для класифікації вхідних ситуацій в прецедентних системах в умовах невизначеності. Використання засобів теорії наближених множин дало змогу вирішити питання належності ситуації до того чи іншого класу на основі доступної інформації. Перевага наведеного підходу полягає в тому, що для оцінки невизначеності ситуації не потрібно ніякої попередньої або додаткової інформації відносно даних.

The question of theoretical bases creation for uncertainty classification of entrance situations in case-based systems is considered in this paper. Using the rough set theory has enabled to solve the problem on a belonging of a situation to any class on the basis of accessible information. The advantage of this approach is that any previous or additional information concerning the data is not necessary for uncertainty estimation of situation.

Останнім часом при розробці систем підтримки прийняття рішень все частіше використовують моделі на основі прецедентів (Case-based reasoning, CBR) [1]. Механізми CBR успішно застосовуються для вирішення широкого кола задач класифікації, діагностики, прогнозування, планування і проектування в різних предметних галузях.

Як показано в [2], застосування прецедентної моделі дозволяє підвищити ефективність прийняття рішень у проблемних ситуаціях шляхом пошуку й адаптації рішень, прийнятих раніше в подібних ситуаціях і збережених в формі прецедентів у спеціально побудованому інформаційному сховищі – сховищі прецедентів.

Використання прецедентних систем особливо доцільно в умовах невизначеності ситуацій, що виникають, та неповноти інформації.

Особливістю прецедентних систем є використання подібності поточної ситуації до деякої аналогічної ситуації, що відбулася раніше (*прецеденту*), збереженої в сховищі прецедентів.

Процес прийняття рішення на основі прецедентів, як правило, складається з чотирьох фаз:

- пошуку та витягу подібних

прецедентів;

- вироблення рішення;

- перегляду і модифікації (адаптації) прийнятого рішення;

- збереження знов сформованого прецеденту.

На першій фазі використовуються моделі подібності ситуацій, найбільш часто застосовуються відношення нечіткої структурної подібності [3].

Відібрані на цій фазі прецеденти разом з поточною ситуацією далі розглядаються спільно.

На другій фазі для ухвалення рішення про вибір того або іншого рішення використовується модель процесу або модель предметної області, що дозволяє розглядати прецеденти не як самостійні одиниці знань, а всередині деякої математичної або імітаційної моделі, що є для прецедентів контекстом, або оболонкою. Контекст може являти собою поверхневу або глибинну модель знань, що дозволяє побудувати симбіоз системи, заснованої на прецедентах, із системою, заснованою на моделі [4].

На третій фазі виконується контроль правильності прийнятого рішення, причому якщо прийняте рішення про подальші дії спричинило результат, який і очікувався, формується новий прецедент, що складається з поточної ситуації і планів її

подолання, та передається на четверту фазу безпосередньо.

Якщо ж є розбіжність між результатом, якого очікували, і результатом, який одержали, виробляється необхідна модифікація (адаптація) прийнятого рішення або вибір іншого з можливих, причому на четверту фазу передається прецедент неуспішного (або неточного) вирішення поточної ситуації.

Четверта фаза призначена для збереження знов утвореного прецеденту в сховищі прецедентів.

Типовий прецедент може бути визначений як структура, яка складається з опису задачі, що характеризує ситуацію на момент активізації прецеденту, і рішення, що містить список дій, пов'язаних з вирішенням задачі, а також (можливо) опис ситуації, що буде мати місце після вибору прецеденту.

Якщо множина ситуацій, що розглядаються як гіпотези при визначенні подальших дій, в реальних умовах виявляється значною, то для її обмеження треба використовувати механізм класифікації [5]. Але використання методів теорії нечітких множин, як це представлено в [3], не дозволяє проводити ефективну класифікацію ситуацій, тому що для отримання правил класифікації потрібна апріорна інформація, яка може бути отримана тільки за допомогою експертів.

Таким чином, питання розробки методів та засобів класифікації ситуацій в прецедентних системах, які працюють в умовах невизначеності та неповноти даних, є актуальним.

Як один з підходів до дослідження цього питання можуть бути використані засоби теорії наближених множин.

Теорія нечітких множин і теорія наближених множин є незалежними і пропонують альтернативні підходи до оцінки невизначеності, як показано в [6]. Головна перевага теорії наближених множин полягає в тому, що для оцінки невизначеності ситуації не потрібно ніякої попередньої або додаткової інформації відносно даних (подібно ймовірності в статистиці, ступеня

приналежності або значення можливості в теорії нечітких множин).

Іншою перевагою є порівняна легкість обробки інформації і відносно прості алгоритми [7]. В основу концепції теорії наближених множин покладено ідею неточних знань. У цьому підході знання неточне, якщо воно містить неточні концепти.

Показано [6], що неточні концепти можуть бути визначені приблизно в наявному, доступному знанні з використанням двох точних концептів, що мають назву *нижнього* і *верхнього наближення*. Нижнє наближення концепту складається з усіх об'єктів, які необхідно належать концепту, в той час як верхнє наближення концепту складається з усіх об'єктів, що можливо належать розглянутому концепту. Простір між нижнім і верхнім наближенням є граничною областю концепту, він складається з усіх об'єктів, що не можуть класифікуватися з упевненістю до концепту або його доповнення, з використанням доступних знань.

Метою статті є дослідження способів оцінки невизначеності вхідної інформації та класифікації ситуацій в прецедентних системах, що діють в умовах невизначеності та неповноти інформації.

Нехай існує множина (можливо, нескінченна) ситуацій  $U = \{s_1, s_2, \dots, s_n\}$ . На вході прецедентної системи ситуації треба класифікувати як належні до того чи іншого класу прецедентів. Наша задача полягає в тому, щоб забезпечити за допомогою тієї інформації, яка є доступною, достовірною або можливо достовірною, віднесення вхідної ситуації до того чи іншого класу ситуацій, не використовуючи або мінімально використовуючи знання експертів.

Задамо систему  $RS$ , що класифікує, як четвірку виду  $RS = \{U, AR, VL, f\}$ , де  $U$  – множина  $U = \{s_1, s_2, \dots, s_n\}$  ситуацій,  $AR$  – множина параметрів (параметри в  $AR$  представлені двома множинами: множиною характеристик об'єктів  $\Theta_F$  і множиною відношень між об'єктами  $R_O$ ,  $AR = \Theta_F \cup R_O$ ),  $LV$  – множина значень параметрів, що

визначена як

;

де  $LV_m$  – область визначення параметра  $m$ ,  $f: U \times AR \rightarrow VL$  – кінцева функція, така, що  $f(s_i, q) \in VL_q$  для кожного  $q \in AR, s_i \in U$ .

Нехай існує множина  $IND$ , така, що  $IND \subset AR, s_i, s_j \in U$ , тоді бінарне відношення нерозрізненості (еквівалентності) ситуацій  $IND$  може бути визначене в такий спосіб:

$$f(s_i, m) = f(s_j, m)$$

Дві ситуації  $s_i$  і  $s_j$  є нерозрізненими по множині параметрів  $IND$  у  $RS$  тоді і тільки тоді, коли  $f(s_i, m) = (s_j, m)$  для кожного  $m \in IND$ . Для кожної ситуації  $s_j \in U$  відношення  $\tilde{IND}$  визначає множину нерозрізнених (або подібних) з нею ситуацій, що позначається як  $[s_i]_{IND}$ . Множина  $U/\tilde{IND}$  (або  $\{B_i, \text{випадку}\}$ ) може бути розділена на класи еквівалентності

що є  $IND$ -елементарними множинами  $RS$ . Таким чином,  $\tilde{IND}$  – це відношення нерозрізненості, що задане на  $U$  для кожного  $IND \subset AR$ .

Система  $RS$  має вибірковість тоді і тільки тоді, коли всі атоми в  $RS$  є множинами, що складаються з одного елемента, тобто  $AR$  є відношенням тотожності. При використанні такого підходу класи ситуацій можуть бути визначені через відносно невелике число параметрів (або навіть один). Це відповідає особливостям когнітивних процесів обробки інформації ОПР – складністю одночасного сприйняття великого числа параметрів, наявності тенденції зосереджуватися тільки на декількох, найбільш важливих характеристиках.

Нехай  $S$  – деяка підмножина множини ситуацій  $U, S \subseteq U$ , а простір наближення  $SP$  задано парною  $SP = \{U, \tilde{IND}\} \subseteq \{U, \tilde{IND}\}$ . Области нижнього  $\underline{IND}(S)$  і верхнього  $\overline{IND}(S)$  наближень визначимо як:

$$\underline{IND}(S) = \{s_i \in U \mid [s_i]_{IND} \cap S \neq \emptyset\}$$

Нижнє наближення  $\underline{IND}(S)$  є об'єднанням всіх тих  $IND$ -елементарних множин, кожна з яких міститься в  $S$ , верхнє наближення  $\overline{IND}(S)$  є об'єднанням тих елементарних множин, кожна з яких має непорожнє перетинання з  $S$ .

Множину  $\tilde{IND}(S) = \underline{IND}(S)$  назвемо  $IND$ -сумнівною (недовизначеною) областю в  $SP = \{U, \tilde{IND}\}$ .

Якщо задано підмножини  $S_1 \subseteq U$  і  $S_2 \subseteq U, \underline{IND}(S_1) \subseteq S_1 \subseteq \overline{IND}(S_1)$ , нижнє і верхнє наближення в  $SP$  мають наступні властивості [2]:

$$\underline{IND}(\emptyset) = \overline{IND}(\emptyset) = \emptyset$$

$$\underline{IND}(U) = \overline{IND}(U) = U$$

$$\underline{IND}(-S_1) = -\overline{IND}(S_1)$$

$$\overline{IND}(-S_1) = -\underline{IND}(S_1)$$

$$\overline{IND}(S_1 \cup S_2) = \overline{IND}(S_1) \cup \overline{IND}(S_2)$$

$$\overline{IND}(S_1 \cap S_2) = \overline{IND}(S_1) \cap \overline{IND}(S_2)$$

$$\underline{IND}(S_1 \cup S_2) \supseteq \underline{IND}(S_1) \cup \underline{IND}(S_2)$$

$$\underline{IND}(S_1 \cap S_2) \subseteq \underline{IND}(S_1) \cap \underline{IND}(S_2)$$

$$\underline{IND}(\underline{IND}(S_1)) = \overline{IND}(\underline{IND}(S_1)) = \tilde{IND}(S_1)$$

$$\overline{IND}(\overline{IND}(S_1)) = \underline{IND}(\overline{IND}(S_1)) = \tilde{IND}(S_1)$$

якщо  $S_1 \subseteq S_2$ , то  $\underline{IND}(S_1) \subseteq \underline{IND}(S_2)$ ;

якщо  $S_1 \subseteq S_2$ , то  $\overline{IND}(S_1) \subseteq \overline{IND}(S_2)$ .

Використовуючи  $\tilde{IND}(S_i)$  відношення нерозрізненості  $\tilde{IND}(S_i)$  можна визначити функцію приналежності наближеної множини:

$$\mu_S^{IND}(s_i) \in [0,1]$$

$$\begin{aligned} \underline{IND}(S) &= \{s_i \in U : \mu_s^{IND}(s_i) = 1\}; \\ \overline{IND}(S) - \underline{IND}(S) &= \{s_i \in U : 0 < \mu_s^{IND}(s_i) < 1\}; \\ \overline{IND}(S) &= \{s_i \in U : \mu_s^{IND}(s_i) > 0\}. \end{aligned}$$

Функція приналежності  $\mu_s^{IND}$  наближеної множини має наступні властивості:

$$\begin{aligned} \mu_s^{IND}(s_i) &= 1 \Leftrightarrow s_i \in \underline{IND}(S); \\ \forall s_i \in U, \mu_{U-S}^{IND}(s_i) &= 1 - \mu_s^{IND}(s_i); \\ \mu_s^{IND}(s_i) &= 0 \Leftrightarrow s_i \in (U - \overline{IND}(S)); \\ \forall s_i \in U, \mu_{s_1 \cup s_2}^{IND}(s_i) &= \max(\mu_{s_1}^{IND}(s_i), \mu_{s_2}^{IND}(s_i)); \\ 0 < \mu_s^{IND}(s_i) < 1 &\Leftrightarrow s_i \in (\overline{IND}(S) - \underline{IND}(S)); \\ \forall s_i \in U, \mu_{s_1 \cap s_2}^{IND}(s_i) &= \min(\mu_{s_1}^{IND}(s_i), \mu_{s_2}^{IND}(s_i)). \end{aligned}$$

Математичний апарат теорії наближених множин дозволяє здійснити початкову, “грубу” класифікацію, в ньому відсутні формальні засоби для оцінки близькості ситуацій, що знаходяться в проміжку між верхнім і нижнім наближеннями. Оскільки класична теорія наближених множин досить чутлива до можливих помилок класифікації [3], очевидна необхідність запровадження такої оцінки.

Оцінку невизначеності можливо дати за допомогою використання узагальненої моделі наближених множин [4].

Дамо визначення системи, що класифікує з невизначеністю, та представимо спосіб оцінки невизначеності в такій системі.

Система  $URS$ , що класифікує з невизначеністю, може бути представлена у вигляді  $URS = \langle U, AR, VL, h, u, d \rangle$ , де  $U, AV, VL, h$  відповідають представленому вище визначенню системи  $RS$ , що класифікує;  $u$  – функція, що визначає достовірність реєстрації параметра  $m$  в ситуації  $s$ :  $u(m, s_j) \in [0.0, 1.0]$ ;  $d$  – функція, що визначає ступінь значимості параметра  $m$  для ситуації  $s$ :  $d(m, s_j) \in [0.0, 1.0]$ .

У випадку, якщо параметр  $m$  для  $s_i$  зареєстровано достовірно,  $u(m, s_i) = 1$ , якщо про значення параметра  $m$  нічого не відомо

–  $u(m, s_i) = 0$ , проміжні оцінки характеризують сумнівне значення  $m$ .

Значення функції  $d$  визначимо наступним чином:  $d(m, s_i) = \frac{|E_s \cap E_s^m|}{|E_s|}$ ;

$$E_s^m \subseteq E_s, \quad \forall s_i \in E_s^m, u(s_i, m) = 1$$

де  $E_s$  – клас еквівалентності для  $s_i$ .

$$\text{Значення } \sum_{E_s} d_i \times u_i$$

можливо використати для визначення позитивного класу  $POS$ , (множини таких  $s_i$ , які характеризуються наявністю істотних для  $E_s$  параметрів  $m$ ) (а значення

– для визначення негативного класу  $NEG$  (множини таких  $s_i$ , які характеризуються відсутністю істотних для  $E_s$  параметрів  $m$ ).

Для оцінки невизначеності ситуації необхідно використати відносну помилку класифікації [3]. У цьому випадку визначається гранична область  $BND$  між позитивною і негативною областю, відповідно до заданих факторів класифікації.

Нехай  $S$  – непорожня підмножина множини ситуацій  $U$ . Ситуація  $s_i$  певно належить  $S$  у випадку, якщо  $s_i \in POS(S)$ , певно не належить  $S$ , якщо  $s_i \in NEG(S)$ , можливо належить  $S$ , якщо  $s_i \in BND(S)$ . Границі класів  $POS$ ,  $NEG$ ,  $BND$  визначаються з урахуванням оцінки відносної помилковості класифікації.

Оцінку відносної помилковості класифікації множини  $S$  щодо позитивного класу  $POS$  і негативного класу  $NEG$  можна визначити за допомогою коефіцієнтів  $C_P(S)$  і  $C_N(S)$ :

$$C_N(S) = \left( \sum_{i=1}^{n_s} (d_i \times u_i) \right) / \left( \sum_{i=1}^{n_s} d_i \right) ;$$

Нехай кортеж

$$SP = \{U, \tilde{IND}_{P,N}\}$$

є простором наближення,

$$\tilde{IND}_{P,N} = \{E_1, E_2, \dots, E_i\}$$

є множиною класів еквівалентності відношення  $\tilde{IND}_{P,N}$ . Функція приналежності класу еквівалентності  $E$  підмножині  $S$  має вигляд:

$$\mu_S^E = \frac{|(S \cap E)|}{|S|}.$$

Тоді позитивний клас  $POS(S)$  може бути заданий як об'єднання тих елементарних множин, чії критерії класифікації гарантують, що відносна помилка  $C_P(E)$  буде менше, ніж  $\mu_S^E$ :

$$POS(S) = \cup \{E \in \tilde{IND}_{P,N} : C_P(E) < \mu_S^E\}.$$

Негативний клас  $NEG(S)$  визначимо як об'єднання тих елементарних множин, чії критерії класифікації гарантують, що відносна помилка  $C_N(E)$  буде менше, ніж  $\mu_S^E$ :

$$NEG(S) = \cup \{E \in \tilde{IND}_{P,N} : C_N(E) < \mu_S^E\}.$$

Гранична область  $BND_{P,N}(S)$  представлена об'єднанням тих елементарних множин, чия класифікація не належить позитивній чи негативній областям множини  $S$ :

$$BND_{P,N}(S) = \cup \{E \in \tilde{IND}_{P,N} : E \notin POS, NEG\}$$

При формуванні гіпотез логічного висновку в якості цільових розглядаються тільки ситуації, що приналежні  $POS$ -області або  $BND$ -області, за умови  $POS(S) = \emptyset$ .

### Висновки

Результатом дослідження є створення теоретичного підґрунтя для класифікації вхідних ситуацій в прецедентних системах. Використання засобів теорії наближених множин дало змогу вирішувати питання належності ситуації до того чи іншого класу на основі тієї інформації, що є доступною системі, та зменшити потребу втручання експертів у процес прийняття рішень. Подальший розвиток можливий в сторону комбінування наведеного способу класифікації з попередньою концептною кластеризацією сховища прецедентів, що дозволить збільшити відносну точність

### Література

1. Aamodt A., Plaza E. Case-Based Reasoning: foundational issues, methodological variations and system approaches // AI Communications. – 1994. – Vol. 7. – № 1. – P. 39-59.
2. Нечипоренко О.А. Использование технологии Case-Based Reasoning в проектировании программных систем // Перспективные информационные технологии и информационные среды. – Таганрог. – 2002. – № 3. – С. 27-32.
3. Ходаков В.Е., Шерстюк В.Г., Козуб Н.А., Дидык А.А. Приложение теории нечеткого подобия к представлению знаний в интеллектуальных системах // Вестник ХГТУ. – Херсон. – 1998. – № 2(2). – С. 138-141.
4. Gomez A., Pavon R., Laza R., Corchado J.M. Integrating Rule Based Systems with Change along Time in a CBR Environment // Simposio de Informatica y Telecomunicaciones SIT'02. – Lissabon. – 2002. – P. 193-202.
5. Pawlak Z. Rough Classification // International Journal of Man-Machine Studies. – 1984. – Vol. 20(5). – P. 469-483.
6. Pawlak Z. Rough Sets, Rough Relations and Rough Functions // Fundamenta Informaticae. – 1996. – Vol. 27. – № 2(3). – P. 103-108.
7. Pawlak Z. Vagueness - a Rough Set View // Structures in Logic and Computer Science. – 1997. – P. 106-117.
8. Pawlak Z., Jerzy W., Slowinski R., Ziarko W. Rough Sets // Communications of ACM. – 1995. – № 38(11). – P. 88-95.