

ФОРМАЛЬНО-МЕТРИЧНІ АСПЕКТИ КЕЙС-БАЗОВАНИХ РЕАЛІЗАЦІЙ ПРИ ВИРІШЕННІ ТЕХНОЛОГІЧНИХ ПРОБЛЕМ БУРІННЯ

*Івано-Франківський національний технічний університет нафти і газу, м. Івано-Франківськ, Україна

Анотація. *Означено сутність застосування кейс-базованих міркувань при вирішенні технологічних проблем процесу буріння, що в кінцевому підсумку дозволяє всі описані операції фактично звести до оперування із значеннями, представленими в формі сутностей з накладеними обмеженнями, які можуть бути стратифіковані як за кількістю, так і якістю виконаних представлень для введених технологічних параметрів. Таким чином, кейс-базовані міркування (розмірковування на основі прецедентів) застосовано як одну з ефективних методологій побудови знання-орієнтованих систем, де центральним елементом є минулий експертний досвід у формі кейсів (прецедентів). Із системної точки зору важливо, що чим більше буде таких кейсів, тим вища якість машинних міркувань, що очікувано повинні привести до рішення проблеми, де вже існуючі кейси в базі служать для уточнення проєктованого рішення та дозволяють ефективну адаптацію вже існуючих результатів у формі кортежів значень технологічних параметрів до певних заданих початкових умов у формі накладених систем обмежень, де рішення вважається правильним, якщо воно задовольняє накладену систему обмежень. Описано процес формування простору рішень для виділених технологічних проблем із накладеними обмеженнями. Процес видобування кейсів базується на добре обґрунтованих відношеннях між індексованими сутностями, що описуються відповідними формальними семантиками обчислюваних відповідей. З даною метою виконано аналіз впливаючих моделей на проєктований простір рішення технологічної проблеми в контексті представленого ймовірного підходу до кейс-базованого виведення, що є суттєвим розширенням випадку статистичних розмірковувань, оскільки в такому розширенні експертні оцінки в формі коефіцієнтів впевненостей оперуються відповідно до відношення схожості на множині гомогенізованих представлень предметної області.*

Ключові слова: кейси, кейс-базовані міркування, обмеження, технологічні проблеми, релевантність, рішення кейсу, буріння свердловин, інтелектуальна система.

Аннотація. *Раскрыта сущность применения кейс-основанных размышлений при решении технологических проблем процесса бурения, что в конечном итоге позволяет все описанные операции фактически свести к оперированию со значениями, представленными в форме сущностей с наложенными ограничениями, которые могут быть стратифицированы как по количеству, так и по качеству выполненных представлений для введенных технологических параметров. Таким образом, кейс-основанные рассуждения (рассуждения на основе прецедентов) применены как одна из эффективных методологий построения знание-ориентированных систем, где центральным элементом является прошлый экспертный опыт в форме кейсов (прецедентов). С системной точки зрения важно, что чем больше таких кейсов, тем выше качество машинных размышлений, которые ожидаемо должны привести к решению проблемы, где уже существующие кейсы в базе служат для уточнения проектируемого решения и позволяют эффективную адаптацию уже существующих результатов в форме кортежей значений технологических параметров в определенных начальных условиях в форме наложенных систем ограничений, где решение считается правильным, если оно удовлетворяет наложенную систему ограничений. Описан процесс формирования пространства решений для выделенных технологических проблем с наложенными ограничениями. Процесс добычи кейсов базируется на хорошо обоснованных отношениях между индексированными сущностями, которые описываются соответствующими формальными семантиками вычисляемых результатов. С данной целью выполнен анализ влияющих моделей на проектируемое пространство решений технологической проблемы в контексте представленного вероятностного подхода кейс-базирующегося вывода, что является существенным расширением случая статистических рассуждений, поскольку в таком расширении экспертные оценки в форме коэффициентов уверенности оперируются в соответствии с отношением сходства на множестве гомогенизированных представлений предметной области.*

Ключевые слова: кейсы, кейс-основанные рассуждения, ограничения, технологические проблемы, релевантность, решение кейса, бурение скважин, интеллектуальная система.

Abstract. *The essence of application of case-based considerations solving technological problems of the drilling process is defined, which ultimately allows all described operations to be actually reduced to the operation with values presented in the form of entities with imposed restrictions, which can be stratified both in terms of quantity and quality of performed domain entities representations for the introduced technological parameters. Thus, case-based considerations (reasoning based on precedents) are used as one of the effective methodologies for constructing knowledge-oriented systems, where the main element is the past expert experience in the form of cases (precedents). From the system point of view it is important that the more the number of such cases will be higher, the higher will be the quality of machine considerations that should be expected to solve the problem, where already existing cases in the cases base serve to clarify the projected solution and to allow the effective adaptation of existing results in the form of tuples of values for technological parameters to certain specified initial conditions in the form of imposed constraints systems, where the current solution is considered to be correct, if it satisfies the imposed system of constraints insofar. It is described the process of forming a solution space for selected technological problems with imposed constraints. The process of extracting of the cases is based on well-grounded relations between indexed entities, which are described by the corresponding formal semantics of computed results. For this purpose, the analysis of impacting models on the projected space of the solutions of the technological problem in the context of the presented probabilistic approach to case-based inference is made, which is a significant extension of the case of statistical reasoning, since in such extension, expert estimates in the form of confidence coefficients are operated accordingly to the similarity relation on the homogeneous set of representations for the subject domain area.*

Keywords: cases, case-based reasoning, constraints, technological problems, relevance, case solving, well drilling, intelligent system.

1. Вступ

Техніка кейс-базованих міркувань (міркування на основі прецедентів відповідно) є однією із найбільш просунутих у штучному інтелекті, оскільки вона дозволяє ефективно поєднати основні типи базових концепцій (методи міркувань і методи машинного навчання для вирішення проблем), базуючись на попередньому досвіді в формі кейсів-прецедентів-випадків [1–6]. Важливим показником успішності застосування даної методології є саме складні, комплексні проблеми, рішення яких обов'язково потребує минулого досвіду, і без такого чіткого вираження експертно-орієнтованого досвіду побудова рішення є взагалі неможливою, якщо прагнути розумного прийняття рішення з максимальним обґрунтуванням та верифікацією даного процесу в цілому й зокрема. В області інтелектуальних систем прийняття рішень методології міркування на основі прецедентів є способом ефективного включення експертного досвіду у процес прийняття рішень. Саме з цієї точки зору дана методологія є складовою методологією для процесів побудови знання-орієнтованих систем. Це шлях ефективної адаптації минулого експертного досвіду до рішення нових проблем. У свою чергу рішення нових проблем дає змогу генерувати певний досвід, що відповідно підсилюватиме інтелектуальність системи. Таким чином, у початковому наближенні суть методології може бути зведена до машинного навчання у процесі рішення нових проблем. Для того, щоб вирішити певну проблему, необхідний певний, скажімо, мінімальний набір знань (у базі знань), у той же час, вирішивши певну проблему, ми виконуємо оновлення бази знань шляхом внесення нових консистентних входжень у формі сутностей знань, що розширяє всю область застосування та здатність системи вирішувати відповідні нові проблеми або навіть цілі класи проблем. Формуючи в такій послідовності процес побудови знання-орієнтованих систем (ЗОС), слід мати на увазі, що структура знань повинна не тільки описувати сигнатуру відповідних кейсів, але й описувати процеси відображення елементів даних сигнатур, які не відповідають концепції знань, що формують відповідний минулий досвід експерта, виконувати обчислення відповідних рівнів схожості, визначати

відповідні способи адаптації правильних, коректних та задовільних рішень для виділених проблем шляхом їх можливої модифікації на рівні параметрів входження та їх граничних значень [7–9]. Тому, за аналогією з експертами-людьми, можна стверджувати, що система повинна відпрацювати певну чи деяку базову кількість кейсів-прецедентів, які можна розглядати відповідно як базові, щоб отримати деякий мінімальний рівень «кваліфікації» в термінах експертів-людей. Таким чином, для кожної предметної області відповідно можна виділити деякі базові набори кейсів-прецедентів, що можуть інтерпретуватися також як типові, найбільш загальні. На рівні бази знань для бази кейсів ми отримуватимемо деяке ядро бази знань, що може утворювати його множину інерції у процесі очікуваних модифікацій. Важливим фактом, зрозумілим із загальної теорії баз знань, є те, що такі базові входження ми можемо внести відразу без потреби навчання системи по кейсах-прецедентах. Причому кейс-входження будуть відразу гарантовано істинними. З точки зору штучного інтелекту, важливим фактом буде також те, що така початкова ініціалізація системи в такій формі задаватиме основний градієнт та відповідні сценарії її розумної поведінки при рішенні проблем у предметній області, де недослідженою частиною проблеми є поєднання методології кейс-прецедентного розмірковування із специфікаціями теорії представлення та задоволення (дотримання) обмежень [10–11].

Тому *метою* даної статті є отримання додаткових метричних топологій описаних вище процесів у формі можливості специфікації (опису) окремих кейсів шляхом накладання відповідних (релевантних) систем обмежень. Відповідно, база кейсів верхнього рівня буде описуватися певною сукупністю обмежень у формі релевантних множин та систем, що в цілому утворюватиме певну ієрархію обмежень як максимально жорсткого випадку представлення і формування конгломерації обмежень, як наслідок допущення нестрогості стосовно обмежень на виділених або всіх рівнях відповідно.

2. Основна частина

Виходячи з заданої проблеми (технологічної проблеми в області буріння нафтових і газових свердловин) [7–9] методологію міркування (розмірковування) слід застосовувати для того, щоб видобути минулі схожі кейси з метою повторного або модифікованого застосування для вирішення технологічних проблем (наприклад, вибору режиму буріння: форсованого, оптимального, раціонального і т.д.) $Sol(TP)$, що в них закладено: $Sol(TP)|Case_i$ або $Sol(TP)|Case_j^{mod}$, де $i, j \in N$. Як тільки рішення буде знайдено, то можна застосувати відповідні методи машинного навчання для покращання загалом знань системи, базованих на прецедентно-кейсовому експертному досвіді. З програмно-прагматичної точки зору, слід приєднати до бази кейсів заданий новий кейс, а до бази знань відповідно – новий експертний досвід:

$$Sol(TP_j) \Big| = \left[\underbrace{CaseBase^{init.}}_{CB^{init.}} \Big| + Case^i \Big| = \left[KB_{init.}^{CB} \Big| + \underbrace{KB_i^{Case^j}}_{KnowledgeBase} \Big| \right]_{i, j \in N} .$$

Оскільки в досліджуваній предметній області буріння нафтових і газових свердловин мають справу з операціями з технологічними регламентами найвищої строгості, то виходитимемо в подальшому з виключної преференційності для випадку саме ієрархії обмежень:

$$CH = Comb^{CS} [\dots CS_i \dots CS_j \dots], \text{ де } CS_{ij} = Comb^{CSet} [\dots CSet_{i1} \dots CSet_{i2} \dots].$$

Спосіб функціонування операторів $Comb()$ визначатиметься, безумовно, вибраним способом та параметрами компарації як по окремих обмеженнях, так і по їх сукупностях у

цілому. Зрозуміло, що найпростіше представлення для процедури компарації отримується, коли обмеження мають виключно числове вираження і найскладніше вираження – у випадку текстових представлень.

Розглянемо питання побудови сигнатури типового кейсу у предметній області буріння НГС (нафтових і газових свердловин). Для цього будемо розглядати послідовності параметрів (tp -параметри в загальному), що описують технологічний процес: керовані параметри (tcp -параметри), некеровані параметри (ucp -параметри), збудуючі параметри (dcp -параметри) та вихідні результуючі параметри (ocp -параметри). Початкова сигнатура матиме такий вигляд:

$$Sig^{init} = \begin{pmatrix} tcp_1, & tcp_2, & \dots, & tcp_{n_1} \\ ucp_1, & ucp_2, & \dots, & ucp_{n_2} \\ dcp_1, & dcp_2, & \dots, & dcp_{n_3} \\ \hline ocp_1, & ocp_2, & \dots, & ocp_{n_4} \end{pmatrix}.$$

Для кейсів із обмеженнями зрозуміло, що природною формалізацією є стилістика інформаційно-пошукових задач на основі обмежень [10–11]: (X, D, C) , де X – множина змінних, D – доменне значення, C – множина обмежень. Зрозумілим також є контроль перебігу технологічного процесу, а конкретніше, зокрема, певних його станів $TP.State^i$, $i \in N$. Кожен технологічний параметр із початкової сигнатурної матриці $TP_i = \{tcp_{i_1}, ucp_{i_2}, dcp_{i_3}, ocp_{i_4}\}$, де, відповідно, $i_1 \in [1; n_1]$, $i_2 \in [1; n_2]$, $i_3 \in [1; n_3]$, $i_4 \in [1; n_4]$ для $n_1, n_2, n_3, n_4 \in N$ прийматиме відповідні значення з наборів, що належатимуть деякому домену, і матиме відповідний вплив на вихідні параметри при заданих обмеженнях у формі множин, систем та ієрархій.

Розглянемо основні стадії функціонування запропонованої методології при рішенні нових технологічних проблем шляхом просування кейс-запиту (запитового кейсу).

1. Видобування минулих кейсів, схожих до кейс-запиту на основі індуктивних (або дедуктивних) методів та методик. Причому процес навчання з кейсів виконується безпосередньо при їх застосуванні.

2. Повторне використання – рішення, формовані у множині кращих відібраних кейсів, будуть використані для побудови нового рішення для кейс-запиту із додаванням потенційно нових узагальнень та специфікацій, що будуть необхідним наслідком модифікацій існуючих рішень до нових умов, відповідно, до нових систем обмежень. Наведемо інтерпретацію для нашого випадку кейсів на основі обмежень. Рішення полягатиме в послідовній множині компарації виду

$$Comp(CS.Case \rightarrow Query, CS_i.Case_i)_{i \in N} \cdot \mapsto^{i_{max}}$$

Очевидно, що мінімальний результат матиме представлення виду

$$\min_{\mapsto^{i_{max}}} \left[\min_{i \in N} \left[Comp(CS.Case \rightarrow Query, CS_i.Case_i)_{i \in N} \right] \right].$$

Для інших випадків виходитимемо з результату, отриманого у вигляді

$$[CS \rightarrow Case^i] \rightarrow Sol\{V_{i_1}, V_{i_2}, V_{i_3}, V_{i_4}\} \Big|^{sat} [CS.Case \rightarrow Query].$$

3. Ревізія рішення – рішення, отримане від кейс-запиту, ще раз оцінюється на пред-

мет задоволення вихідних параметрів $\left[\begin{array}{ccc} \dots & tcp_{i_1}^{j_1} & \dots \\ \dots & ucp_{i_2}^{j_2} & \dots \\ \dots & dcp_{i_3}^{j_3} & \dots \end{array} \right] = [\dots ocp_{i_4}^{j_4} \dots]$, де $[1 \leq i_k \leq n_k]$,

$[1 \leq j_k \leq n_k]$, $n_{1..4} \in N$.

4. Утримання рішення. Залежно від результатів п.3 буде виконано додавання певного кейсу до бази кейсів і нових знань до бази знань:

$$\left[[Sol \rightarrow Case]^+ \rightarrow CaseBase \right] \mid - \left[[Sol \rightarrow KB]^+ \rightarrow KB_{case}^{all} \right].$$

Таким чином, специфікації описів проблем відображаються у відповідні індекси, що дозволяє реалізацію видобування кейсів на основі розмірковувань у термінах апроксимації та подібності (схожості). Механізм функціонування відображення для індексів такий: новий кейс для рішення представляється відповідною специфікацією проблеми:

$$Case.Query = \langle TP, TP.Spc \rangle + Sol(Case.Query).$$

Причому пошук рішення кейсу виконується у просторі описів проблеми шляхом пошуку схожих специфікацій проблеми. Таким чином, виконання $Sol(Case_i) \mid = [Case_i]$ пошуку кейсу у просторі описів проблем відповідатиме опису певного кейсу і буде аналогічним твердженням про те, що рішення попередніх кейсів є основою рішення заданої нової проблеми: $Case_i \mid = [Case.Query]$. Відповідним важливим наслідком буде також те, що рішення $Sol(Case_i)$, збережене в $Case_i$, буде значимим і для $Case.Query$. Описана процедура не включає адаптації. Зрозуміло, що такий механізм не матиме відповідної гнучкості у предметній області, оскільки рішення не адаптується, а просто копіюється. Для внесення функції адаптації необхідно будувати процес видобування кейсів на індексах замість прямих специфікацій кейсів.

Введемо деяку кодууючу шему $Schema()$, яка обчислюватиме індекси для кожного з кейсів, реалізуючи представлення відповідного абстрактного процесу:

$$Schema(Case_i, .Query) = \langle Schema(TP, .Query), Schema(Sol) \rangle,$$

де $Sol = Solve(Case.Query)$.

Зрозуміло, що кодууюча схема типу вищевказаної значною мірою залежатиме від особливостей імплементації самої кейс-базованої системи міркувань (КБМ-система), а саме від способу представлення кейсів, представлення індексів, опису домену та ін. Отже, окреслимо формальні властивості, що повинні бути дотримані в кожному з випадків:

$$\text{якщо } \langle Case_i \rangle \rightarrow \langle Case.Query \rangle,$$

$$\text{modi } Schema(Case_i) \rightarrow Schema(Case.Query).$$

Відповідно, як наслідок, моделі для кейсів $Md(Case_i)$ і $Md(Case.Query)$ зберігатимуться у відповідних моделях для індексів $Md(Schema())$, а саме в $Md(Schema(Case_i))$ і $Md(Schema(Case.Query))$, що відповідатиме властивості монотонності:

$$\text{якщо } Md(Case_i) \rightarrow Md(Case.Query),$$

$$\text{modi } Md(Schema(Case_i)) \rightarrow Md(Schema(Case.Query)).$$

Таким чином, відповідно до даної властивості, рішення $Sol()$ може бути знайдено шляхом пошуку по базі кейсів CB вздовж виміру « $\mid=$ » між індексами (на множині індексів). Видобутий кейс із заданим індексом задовольняє шуканий новий індекс і буде рішенням проблеми $Sol(TP)$ без гарантованої 100-відсоткової впевненості, як наслідок, вибраного способу апроксимації, оскільки алгоритм видобування виконує апроксимація на основі відношення « $\mid=$ » між кейсами, коли виконується порівняння індексів кейсів. У термінах розміркування на основі кейсів та подібностей (схожостей) це полягатиме у представленні відношення $Schema(Case_i) \mid= Schema(Case.Query)$, що може бути також переадресовано, якщо взяти до уваги, що кожний індекс містить у собі фактично два компоненти: $\langle Schema(TP) \rangle$, $\langle Schema(TP.Spc) \rangle$. Таким чином, видобування кейсів може бути базованим на добре обґрунтованих відношеннях між індексами, що включає відповідні формальні семантики.

$$\begin{cases} Schema(TP.Query) \mid= Schema(TP_i), \\ Schema(TP.Spc_i) \mid= Schema(TP.Spc_i.Query), \end{cases}$$

$$\begin{cases} Schema(TP.Query) \mid= Schema(TP_i), \\ Schema(TP.Spc_i) \mid= Schema(TP.Spc_i.Query). \end{cases}$$

Таким чином, ми отримаємо два відповідних алгоритми видобування кейсів.

Подальший розвиток такого представлення шляхом об'єднання ймовірностей приведе до необхідного розгляду відповідної множини кейсів, і шукана апроксимація досягатиметься за рахунок кластеризації атрибутів даних (параметрів) шляхом групування відповідних кейсів, що мають схожі релевантні властивості.

Таким чином, імовірнісні моделі для кейс-базованих міркувань (КБМ) визначатимуться заданою множиною змінних (параметрів) виду $\langle tp_1 \cdots tp_n \rangle$. Відповідно, кожен кейс $\overline{case_i}$ буде деякою інстанцією в формі підстановки релевантних значень. Для таких

змінних (параметрів) матимемо представлення виду $\overline{case_i} = \left(\underbrace{tp_1 = v_1, \dots, tp_n = v_n} \right)$.

Тоді можна бачити, що, згідно з введеними формалізмами, база кейсів $Cbase$ буде множиною з m незалежних та співрозмірних розподілених наборів даних у формі підстановок. На наступному кроці множина кейсів може бути кластеризована в L груп, що дозволить отримати відповідний розподіл імовірностей, які представляються відповідними кластерами. Тоді матимемо, що для всіх L $prb(\overline{Case_i}) \mid X = X_L$ виражатиме ймовірність того факту, що кейс належатиме до кластерів X_L , де X – змінна ймовірнісної природи, співвіднесена із введеними кластерами. Таким чином, на наступному кроці можна стверджувати, що кожен кейс може бути апроксимований ваговою сумою відповідних розподілів виду $prb(\overline{Case_i}) = \sum_{l=1}^L prb \mid X = X_L \mid prb(\overline{Case_i} \mid X = X_L)$. Припускаючи, що параметри tp_i у середині кожного кластера є незалежними, матимемо, що

$$prb(\overline{Case_i}) = prb(tp_1 = v_1, \dots, tp_n = v_n) = \sum_{l=1}^L prb(X = X_L) \prod_{i=1}^n prb(tp_i = V_i \mid X = X_L).$$

На основі такої моделі стає можливим рішенням різного типу ймовірнісно-

базованих розміркувань власне при побудові рішення типу $Sol(TP)^{prb}$. Таким чином, після виконання перевірки бази кейсів $Cbase$ і встановлення $\overline{Case_{Qm}}$ як запитового кейсу, на основі вищесказаного матимемо можливість означити прогнозуючий розподіл як $prb(\overline{Case_{Qm}}|Cbase) = prb(\overline{Case_{Qm}}|Cbase, \Delta)$, де Δ позначає відповідні параметри моделі $\Delta = (cd, cp) = (\delta^{cd}, \delta^{cp})$, де δ^{cd} – параметри, що описують встановлений кластерний розподіл $\delta^{cd} = (cd_1, \dots, cd_l | cd_i = prb_i(X = x_i))$ а δ^{cp} – параметри, що належать до умовних імовірностей кластерів стосовно значень змінних $\delta^{cp} = \{cp_{ij}\}$, причому кожна конструкція $\{cp_{ij}\}$ є деякою множиною параметрів загального виду. Для дискретних змінних із визначеною кардинальністю $Card(D_i, tp_i)$ в CSP – нотації відповідно матимемо, що

$$cp_{ij} = (cp_{L_1}, \dots, cp_n, \dots, cp_{L_1}, \dots, cp_{L_n}) \square Card(D_i, tp_i) \square prb(tp_i = v_i | X = x_j).$$

Враховуючи початкове припущення про те, що бази кейсів $Cbase$ є явно незалежні при заданому Δ , матимемо, що $prb(\overline{Case_{Qm}}|Cbase) = prb(Case_{Qm}|\Delta)$.

З точки зору реального технологічного процесу, який протікає в умовах невизначеності, необхідно визначати значення неінстанційованих змінних tp_i , які є рішенням пошукової проблеми для кейс-запиту при деяких заданих інстанційованих значеннях, що утворюють опис проблеми,

$$Sol(Q_{TP} = TP) = uninst(\dots tp_i \dots) | inst(\dots tp_{i-2} \dots).$$

Такий вид представлення дозволяє серед усіх можливих значень для tp вибрати найбільш імовірні значення.

У свою чергу, нечітка модель базуватиметься на степені належності попереднього кейсу до множини достатньо схожих кейсів по відношенню до поточної проблеми. Відповідно до такого принципу, весь процес розмірковувань базуватиметься на відповідному рівні схожості на противагу рівню інстанціцій і матиме безпосередній стосунок до означення рівня схожості та способу його вимірювання для простору опису стану проблеми $TP.State$ та простору рішення $Sol.Space$, відповідно. Важливо, що $TP.State, Space$ і $Sol.Space$ обидва є нечіткими відношеннями, означеними в $[0, 1]$ і є застосовними до пар «атрибут (параметр)–значення» як виду відповідного представлення на множині кейсів:

$$\forall TP_i, Sol_i, TP_j, Sol_j \in Cbase, Sol.Space(TP_i, TP_j) \leq TP.State.Space(sd_i, Sol_i, Sol_j).$$

Таке представлення означає, що відношення схожості для простору технологічної проблеми обмежує відношення схожості у просторі рішення. Таким чином, наприклад, якщо дві проблеми є схожими, то їх рішення повинні бути такими ж схожими, як і відповідні описи проблем. Таким чином, вирішуючи деяку нову проблему

$$Case_{Query} = \langle TP_{Query}, Sol_{Query} \rangle,$$

де рішення є початково невідомим, саме накладені обмеження визначатимуть множину можливих значень для Sol_{Query} виду

$$Sol_{Query} = \bigcap \overline{Case} (TP, Sol) \in Cbase \{ Sol_{Query} \in TP | Sol.Space(TP_i, TP_{Qm}) \leq TP(Sol_i, Sol_{Qm}) \}.$$

Для того, щоб охарактеризувати КБМ на рівні оцінки схожості кейсів, необхідні ві-

дповідні описи КБМ на рівні інстанціацій. Таким чином, у контексті сказаного, КБМ – системи слід описувати на рівні інстанціацій кортежем виду (St, Sol, Ψ) , де St, Sol – множини відповідних (різних) пар «атрибут(параметр)–значення», що представляють описи проблеми, та релевантного рішення відповідно. Відображення Ψ виконує присвоєння потенційних рішень до наявних проблем $\Psi : St \rightarrow Sol$ відповідно так, що для деякого кейсу $Case^i$ матимемо, що $Case^i = \langle tp_i, \Psi(tp_i) \rangle \in St \times Sol$. Таким чином, кейс-базований інформаційний фреймворк матиме вигляд

$$InffCB = \langle (St, prb_{St}), Sol, \Psi, Q_{St}, Q_{Sol}, Cbase \rangle,$$

де опису стану проблеми St присвоєна міра ймовірності prb_{St} , яка дозволяє розглядати відповідну послідовність пар описів «проблема–рішення» $\langle TP - Sol \rangle$ в St як реалізацію релевантної послідовності вхідних параметрів, що характеризуються відповідними мірами ймовірності:

$$(prb_{St})^{n+1} = \dots prb_{St} TP.Space(St^{n-1}) \dots$$

Така розмірність визначатиме ймовірнісний простір, в якому даний вивід (інференція кейсів) має місце. Зміст введених позначень відповідає вищесказаному, а саме $Cbase$ – база кейсів, $\Theta_{St}, \Theta_{Sol}$ є функціями схожості (подібності) у просторі опису проблеми $St.Space$ і просторі рішення $Sol.Space$ відповідно. Введені функції використовуються для відображення системи на рівні подібності (схожості). Виходячи з того, що $St.Space$ є скінченним, стає можливим визначення множини ступенів подібності для вхідних та вихідних параметрів (Sol_{St} та Sol_{Sol}), що може бути виражено як

$$\begin{cases} Sol_{St} := \{\Theta_{St}(tp_i, tp_j) \mid tp_i, tp_j \in St.Space\}, \\ Sol_{Sol} := \{\Theta_{Sol}(\Psi(tp_i), \Psi(tp_j)) \mid tp_i, tp_j \in St.Space\}. \end{cases}$$

Таким чином, сигнатурний відбиток системи по відношенню до Ψ – структури на рівні подібності (схожості) представлятиметься заданим профілем подібності $Sp : sd_{st} \rightarrow [0, 1]$, де sd – ступінь подібності відповідно, що може бути проінтерпретовано таким чином:

$$Sp(v) := \left| in :: TP.State \right|_{tp_i, tp_j \in St} \left| St(tp_i, tp_j) = v^{\Theta_{Sol}(\Psi(tp_i), \Psi(tp_j))} \overbrace{\Theta_{Sol}(\Psi(tp_i), \Psi(tp_j))}^{out, Sol} \right|, \Theta_{St} \inf out.Sol .$$

Отже, можна бачити, що профіль подібності (sp) має відносно просту структуру, яка дозволяє формулювання гіпотези подібності $S.hyp$ як функції $S.hyp 0,1 \rightarrow 0,1$, що цілком у стилі абдуктивного підходу до побудови рішення:

$$\underbrace{(\Theta_{St}(tp_i, tp_j) = v)}_{in :: TP.State} \rightarrow \underbrace{(\Theta_{Sol}(\Psi(tp_i), \Psi(tp_j)) \geq S.hyp(v))}_{out :: Sol.Space} . S.hyp .$$

Таке представлення, відповідно, інтерпретується як факт того, «що чим більше схожим є два стани проблеми, тим більше схожими будуть також і результати».

Для заданого механізму кейс-базованого виведення (розмірковування на основі прецедентів) inf_{CB} для бази кейсів $Cbase$ і запитової проблеми TP_{Qm} структура подібності (схожості) технологічної проблеми (inf_{CB}, TP_{Qm}) буде визначатися відповідним профілем:

$$(TP, \Theta_{st}, \Theta_{Sol}) \sim TP_{SS}(Cbase, TP_{Qm}) \mid \{V_{ij}^3 = (V_{ij}^1, V_{ij}^2) \mid 1 \leq i \leq n, 1 \leq j \leq n\},$$

де $V_{ij}^1 = \Theta_{st}(TP_i, TP_j)$, $V_{ij}^2 = \Theta_{Sol}(Sol_i, Sol_j)$. Схожим чином можна визначити структуру результатів (SS – структура подібності, OS – структура результату) як множини відповідних значень:

$$TP_{OS}(Cbase, TP_{Qm}) = TP_{SS}(Cbase, TP_{Qm}) \cup \{Sol_j \mid 1 \leq j \leq n\}.$$

У той же час слід акцентувати увагу додатково на виключній структурі саме ймовірного підходу, при якому основна задача полягає не у прогнозуванні самих значень, а у прогнозуванні чистих імовірностей, що дозволяє отримати нове представлення виду $TP(Sol_{Qm} = Sol \mid TP_{OS}(Cbase, TP_{Qm}))$, яке інтерпретується таким чином, що ймовірність того, що результат $\Psi(TP_{Qm})$ задається через деяку підстановку Sol , визначається інформацією, заданою через $TP_{OS}(Cbase, TP_{Qm})$. Логічним завершенням описаної процедури кейс-базованого виведення в термінах імовірнісної моделі є виконання відображення отриманого розподілу ймовірностей для простору результатів на відповідний рівень інстанціації.

Нехай $Case_{Qm}$ є запитовим кейсом. Тоді, застосовуючи методологію КБМ, отримуємо множину видобутих ранжованих кейсів $Cset$. Ранжування здійснюється відповідно до подібностей і доцільностей (корисності) кейсів. З даної точки зору, кожен кейс, відповідно, повинен отримати ряд міток (анотацій) $Case_i = (TP_i, Sol_i, out.tp_i^{set})$, де застосування рішення Sol_i дає відповідну їй множину результатів $out.tp_i^{set}$. Така множина результатів може інтерпретуватися також у термінах оцінок для множини видобутих кейсів, що задовольняють задані умови вибору. Таким чином, у даному контексті можна застосовувати функцію доцільності для оцінки результатів згідно з шемогою $uf(out.tp_i^{set}) > 0$ для бажаних результатів, і $uf(out.tp_i^{set}) < 0$ – в усіх інших випадках. Для побудови формалізації виходитимемо із того, що КБМ – система підтримки прийняття рішень (КБМПР) у своєму розпорядженні має обмеження – обмежену кількість альтернатив (рішень) $alt = \bigcup_{Case_i \in Cbase\{Sol_i\}}$. Беручи до уваги необхідність використання характеристик доцільності в описі структури систем КБМПР, слід обчислювати доцільність кожної дії по відношенню до поточної проблеми TP_{Qm} у формі суми подібностей та доцільностей для множини видобутих кейсів:

$$uf(alt_i) = uf_{Qm} \mid_{Cset} (alt_i) = \sum SmbR(TP_j, TP_{Qm}), uf_{(tp_j, out.tp_j)}^{out}(tp_j, alt_j, out.tp_j) \in Cset.$$

Таким чином, усі потенційні дії можна розташувати відповідно до їх доцільності, і система вибиратиме відповідно дію із максимальною доцільністю із усіх наявних.

Крім того, опис кейсу може включати також результат інстанціації рішення, тобто того факту, чи така інстанціація дає бажаний результат, чи ні. Тобто, в найбільш загальному випадку кейс представлятиметься деяким кортежем виду $\langle TP, Sol, Out.tp \rangle$, де TP – власне технологічна проблема, Sol – її очікування рішення, $Out.tp$ – очікувані результати. Проте представлення аж ніяк не може претендувати на повну вичерпність щодо складових компонент кейсів. Наприклад, за аналогією з експертними системами (рис. 1) може також включатися компонента пояснень, що дасть можливість підсилити складові зворотного зв'язку з користувачем. Звичайно, зрозуміло, що як і у випадку інших видів систем штучного інтелекту, найбільшу складність щодо реалізації представлятимуть саме випадки включення текстових і графічних даних (у формі зображень), оскільки їх обробка сама по собі вже потребує складних алгоритмів та рішень, а в багатьох контекстах таких алгоритмів може і не існувати в досконалому каталогізованому вигляді.



Рисунок 1 – Контейнери предметних знань

Важливою задачею є також представлення компонентів (складових) кейсу на семантичному рівні. Для цього слід виконувати чітко розділення між особливостями об'єктів, самими об'єктами та реляційними об'єктами як такими. В той же час особливості можуть розглядатися як присвоєння «атрибут (параметр) – значення» – (TP_i, V) – тобто вони фактично є описами властивостей, виділених як ключові в даному контексті. У той же час атрибути визначаються в мережевому семантичному словнику. Змінні співвідносять до атрибутів і тому вони (змінні) є або числовими, або переозначаються і уточнюються у словнику. Таким чином, об'єкти є атрибутами, згрупованими в форму більш комплексних даних для забезпечення функціональності розмірковувань на більш високому рівні. Отже, коли виконати збагачення представлень знань засобом реляцій між об'єктами, тоді, очевидно, буде отримано певний вид реляційних представлень. У той же час всі компоненти кейсу із кортежу $\langle TP, Sol, Out.tp \rangle$ можуть бути, відповідно, представлені або властивостями, або об'єктами, або реляційними об'єктами з метою тримання моделі знань, що вимагається в домені застосування [9].

Використовуються дві основні моделі знань для представлення опису проблеми: це пари «атрибут-значення» та реляційні об'єкти. Самі пари «атрибут-значення» є найпростішим і найбільш часто вживаним видом представлень. Відповідно, опис проблем задається у вигляді послідовності особливостей – (ft^1, \dots, ft^n) . Нехай маємо послідовність таких пар: $\dots \langle tp_i, v_i \rangle \dots$, а саме виділимо окремі ланки послідовності $\{\langle tp_1, v_1 \rangle, \dots \langle tp_i, v_i \rangle, \dots \langle tp_n, v_n \rangle\}$. Для виконання такого комплексного присвоєння буде потрібна деяка послідовність значень змінних $\langle v_1 \dots v_i \dots v_n \rangle$. Тоді початкова структура типового кейсу знань матиме вигляд

$$\left\{ \underbrace{\langle \{otp_1\} \dots \{otp_m\} \rangle}_{Out.tp} \underbrace{\langle \{ctp_1 = v_1, \dots, ctp_n = v_n\} \rangle}_{Dcase} \right\}_{Kcase}$$

де $Kcase$ – кейс знань, утворений із кейсу даних $Dcase$ шляхом оцінювання його впливу на очікувану множину параметрів на виході, де описаний початковий кейс результатів, може бути уточнено: $Out.tp = \{otp_1 = v_1^1, \dots, otp_m = v_m^m\}$, де v_i^i – це очікуване на виході значення, іншими словами, очікуване значення вихідного параметра.

Таким чином, основна задача для моделювання полягає у представленні впливу ініціалізації керованих параметрів на очікуваний стан інстанціації вихідних параметрів. Проте, дана задача суттєво ускладнюється двома факторами:

- 1) вплив кожного конкретного вхідного параметра на послідовність вихідних пара-

метрів чи на один певний конкретний параметр, який залежить від установлених значень інших вхідних параметрів;

2) множина вхідних параметрів поділятиметься на підмножини та сингелтони (виділені параметри) залежних, незалежних та вільних параметрів.

Рішення даної проблеми у пропонованому дослідженні виконується шляхом накладання та контролю обмежень у рамках методології CSP (інформаційно-пошукових задач на основі обмежень). Представлення інформації про буріння нафтогазових свердловин (НГС) виконується у вигляді кейсів.

Таблиця 1 – Пари «атрибут-значення» для випадку буріння НГС

Атрибут (параметр)	Значення
Родовище	Атазівське
Тип буріння	Роторне
Наявність конденсату	Так
.....

Таблиця 2 – Представлення об'єктів

Опис родовища	Кількість пробурених свердловин 12
	Газових свердловин 5. Нафтових свердловин 15
.....
Характеристики типової свердловини	Середня глибина буріння 100. Кількість використаних доліт 10
.....

Основна мета полягає в побудові узагальнених кейсів та кейсів-прототипів. Зокрема, використання прототипів дозволить значно підсилити процедуру видобування кейсів та розширить її функціональні можливості. Всі компоненти – особливості (feature components) можуть бути згруповані у групи, і, таким чином, кейси представлятимуться засобами об'єктів. Представлення в формі реляційних об'єктів, як правило, потребуватиме візуалізації в формі дерева або графа.

Інформація щодо рішень кейсу *Sol* залежатиме від того, які задачі встановлено при побудові рішення. В більшості випадків основна мета – це прогнозування певної мітки або класу, тобто задачі класифікації в загальному формулюванні. Тому для заданої множини міток *Lset* матимемо, очевидно, що $Sol \in Lset$. У більшості випадків використовується саме вид бінарної класифікації з двома мітками типу $\langle true, false \rangle$, $\langle positive, negative \rangle$, яким, власне, і позначається результат. У випадку буріння НГС заслуговує на увагу не бінарна, а саме мультикласова класифікація, коли міток більше, ніж дві, а саме, наприклад:

$$\langle yes, no, unknow \rangle, \langle low, medium, high \rangle, \langle veryslow, slow, medium, high, forced \rangle.$$

Різна кількість міток у даному випадку цілком природна, враховуючи рівень складності та невизначеності, з яким ми маємо справу у предметній області буріння НГС, де процеси характеризуються саме як невідтворювані та такі, що протікають в умовах невизначеності, нечіткості та неповноти. Таким чином, мультикласові мітки включатимуть присвоєння підмножини існуючих міток, що фактично і буде рішенням виділеної проблеми $Sol \in Lset$.

У загальному випадку можуть розглядатися також й інші знання, наприклад, такі, які необхідні для аргументації кейсів на етапі їх ревізії (генерації зворотного зв'язку) або знання забезпечення підтримки, необхідні на етапах перегляду і збереження. У процесі розвитку і підтримання контенту самі контейнери знань можуть бути також аналізовані з

точки зору їх наповнення, що відкриває можливість оголошення типів знань відповідно до того, як система проводить процес розмірковування. Важливо також відмітити, що представлена стратегія КБМ за своєю суттю є концептуальною методологією, яка для імплементації потребуватиме використання потенційно кількох або навіть багатьох методологій штучного інтелекту рівня практичної реалізації.

3. Висновки

У даному дослідженні розкрито сутність кейс-базованих процедур підтримки прийняття рішень оператором технологічного процесу (в нашому випадку буріння свердловин на нафту і газ). Представлена концепція кейс-базованості дозволяє концентрувати експертний досвід роботи у предметній області буріння свердловин на нафту і газ і задовольняє критеріям основних джерел знань, потрібних для функціонування кейс-базованих міркувань як таких у цілому та зокрема. Серед таких джерел слід виділити семантичні словники предметної області, бази кейсів у формі баз знань, метрики схожості (подібності), контейнери адаптації та налаштування. В загальному випадку слід виходити з того, що технологія кейс-базованої підтримки прийняття рішень виконує ранжування необхідних технологічних дій, де виділені цілі можуть інтерпретуватися як рішення відповідних регламентованих технологічних проблем. Проведено моделювання очікуваної сутності кейсів даних та знань, описано типологію очікуваних рішень, представлено оцінку потенційних результатів. Тому можна стверджувати, що коли немає однорідного представлення для всіх наявних кейсів, то технологічні параметри не можуть бути локалізовані згідно з їх позиціями в регламенті, тому необхідним є використання імен атрибутів технологічних параметрів разом із указанням відповідного метричного шляху, що допомагатиме у виконанні ідентифікації таких атрибутів, починаючи від початкового кореневого(вузлового) представлення. Виділено та проаналізовано властивості кейсів та баз кейсів. Показано, що коли рівень компетентності експертів предметної області низький, то це означатиме, що не спостерігається перекривання описів поточних кейсів з описами інших кейсів у базі, які можуть пропонувати відмінні альтернативні рішення. Також можна буде подібним чином виконувати побудову шем подання та інтерпретації предметних знань, що буде темою подальших досліджень.

СПИСОК ДЖЕРЕЛ

1. Abdrabou E.A.M.L., Salem A.-B.M. Case-based reasoning tools from shells to object-oriented frameworks. *InICCOMP*. 2008. Vol. 67. P. 781–786.
2. Bergmann R., Schaaf M. Structural case-based reasoning and ontology-based knowledge management: A perfect match. *Journal of Universal Computer Science*. 2003. Vol. 9, N 7. P. 608–626.
3. Bierer A., Hofmann M. Dimensions of case-based reasoner quality management / ed. L. McGinty, D. Wilson. *Case-Based Reasoning Research and Development: Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 2009. Vol. 5650. P. 90–104. DOI: 10.1007/978-3-642-02998-1, 53.
4. Gronau N., Laskowski F. Using case-based reasoning to improve information retrieval in knowledge management systems / ed. E. Menasalvas, J. Segovia, P. Szczepaniak. *Advances in Web Intelligence: Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 2003. Vol. 2663. P. 94–102. DOI: 10.1007/3-540-44831-4, 9.
5. Hüllermeier E. *Case-Based Approximate Reasoning*. Berlin: Springer, 2007. 63 p.
6. Hüllermeier E. Schlegel P. Preference-based CBR: first steps toward a methodological framework. *Proc. of the 19th International Conference on Case-Based Reasoning Research and Development: ICCBR'11*. London, UK: Springer-Verlag, 2011. P. 77–91. DOI: 10.1007/978-3-642-23291-6_8 38, 63.
7. Sheketa V.I., Melnyk V.D., Romanyshyn Y.L., Chesanovskyy M.S. The Construction of Technological Problems Cases for the Purpose of Intelligible Control. *Perspective technologies and methods in MEMS design (MemsTech'2016): XIIth International Conference (Lviv-Polyana, 20–24 th April 2016)*. Lviv, 2016. P. 96–99.

8. Sheketa V.I., Pikh V.Y., Styslo T.R., Chesanovsky M.S. Modeling methodology for knowledge-based systems of wells drilling control. *Perspective technologies and methods in MEMS design (Mem-sTech'2017)*. XIIIth International Conference (Lviv–Polyana, 20–23 th April 2017). Lviv, 2017. P. 56–58.
9. Чесановський М.С. Формально-алгоритмічна та програмна імплементація моделей кейсів даних про процес буріння. *Математичні машини і системи*. 2018. № 4. С. 44–55.
10. Barták R., Dechter R. *Constraint Processing*. New-York: Morgan Kaufmann Publisher, 2003. 210 p.
11. Tsang E. *Foundations of Constraint Satisfaction*. London and San Diego: Academic Press, 1993. 421 p.

Стаття надійшла до редакції 05.12.2018