

В. Є. Копняк; В. Б. Мокін, д-р техн. наук, проф.

ОПТИМІЗАЦІЯ АРХІТЕКТУРИ ПРОГНОЗНОЇ ІНТЕЛЕКТУАЛЬНОЇ АГЕНТНОЇ СППР ДЛЯ ПІДТРИМКИ ЯКІСНИХ РІШЕНЬ НА ОСНОВІ ДАНИХ ПРО СТАН АТМОСФЕРНОГО ПОВІТРЯ МІСТА

Сучасні системи підтримки прийняття рішень (СППР) у сфері екологічного моніторингу мають забезпечувати високу якість рекомендацій, зокрема – гарну точність, достовірність, робастність, обґрунтованість та інтерпретованість результатів за умов динамічних, нерегулярних та неповних даних.

У статті запропоновано формалізацію задачі оптимізації архітектури прогнозної інтелектуальної агентної СППР (ІАСППР), орієнтованої на надання природномовних рекомендацій різним групам користувачів (населенню, екологам, органам влади) за принципами моделі «DPSIR» (Drivers, Pressures, State, Impacts, Responses). Введено інтегральний критерій оптимальності як середньозважену суму критеріїв якості рекомендацій, серед яких виділено такі основні: урахування достовірності даних, точності аналізу ситуації, обчислювальної ефективності та робастності, обґрунтованості й інтерпретованості рекомендацій для відповідних категорій користувачів. Доведено теорему про Парето-оптимальність архітектури, яка максимізує цей інтегральний критерій якості за виконання низки умов. Як наслідок цієї теореми сформульовано вимоги до оптимальної структури агент-бенчмарку, що з високою ймовірністю забезпечує коректний вибір оптимальної архітектури ІАСППР.

Як наслідок доведеної теореми обґрунтовано оптимальну структуру агент-бенчмарку, який названо «Air-DSS Agent Benchmark» і який генерує репрезентативну множину сценаріїв чи ситуацій, забезпечує їх збалансованість, дозволяє порівнювати різні архітектури ІАСППР та гарантує їх оцінювання за узгодженими критеріями якості.

Наведено спрощений приклад для реальних даних громадського моніторингу стану атмосферного повітря м. Вінниці за 2019 – 2025 рр., отриманих по 20 станціях з мережі моніторингу EcoCity. Було порівняно дві архітектури агентів А і В. Проведене тестування підтвердило, що архітектура А демонструє значно вищу точність класифікації сценаріїв та децю меншу похибку прогнозів порівняно з архітектурою В, що узгоджується з теоретичними передумовами. Отримані результати показують, що правильний вибір архітектури ІАСППР істотно впливає на якість рекомендацій, а розроблений агент-бенчмарк є коректним та ефективним інструментом для її оптимізації.

Ключові слова: *система підтримки прийняття рішень, оптимізація, агент, бенчмарк, інтелектуальна система, прогнозування, якість рекомендацій, стан атмосферного повітря, Парето-оптимальність.*

Вступ

Стан довкілля, зокрема атмосферного повітря, впливає на життєдіяльність людей, особливо вразливих (діти, літні люди, астматики та інші чутливі до якості повітря люди). Таким людям важливо завчасно отримати рекомендації щодо доцільності прогулянок на свіжому повітря, а за складних умов – завчасно узнати про те, що й вікна відкривати не бажано. Такі рекомендації можуть генерувати системи підтримки прийняття рішень (СППР), які мають доступ до даних моніторингу та мають прогнозні можливості щодо показників стану повітря. Сучасні СППР все більше реалізують як інтелектуальні агенти [1 – 3], які не тільки аналізують і прогнозують інформацію, а й генерують корисні рекомендації для заданої групи користувачів зручною для них природною мовою. Як правило, виділяють 3 категорії користувачів: населення (іноді, окремо виділяють ще населення, особо чутливе до стану повітря), екологи чи держекоінспектори (їм важливо знати де саме має місце перевищення для більш точної перевірки та реагування, особливо, якщо це – локальні несанкціоновані джерела), органи влади, яким цікаві рекомендації для усіх категорій користувачів для передавання інформації їх представникам і координування між собою. Іноді ще виділяють категорії власників станцій моніторингу, представників джерел

забруднень, науковців, які досліджують такі проблеми, тощо. Загалом, рекомендації формують для усіх складових популярної в ЄС так званої моделі «DPSIR» («D» – «Drivers» (рушійні сили), «P» – «Pressures» (тиск), «S» – «State» (стан), «I» – «Impacts» (вплив / наслідки), «R» – «Responses» (реакції / рішення)) [4, 5]. Структурно: D → P → S → I → R, а R впливає на кожен складову D, P, S, I окремо [4, 5]. Під час формування рекомендацій R враховуються і прогнозні дані. Для прогнозування можна використати багато різних моделей, огляд яких є у роботі [5].

Існує багато різних архітектур інтелектуальних агентних (зокрема, й мультиагентних систем) для реалізації подібних СППР, наприклад, гарний огляд майже 300 джерел про мультиагентні системи для smart-city наведено у роботі [1]. Але, як правило, такі системи розробляють під конкретну задачу. У разі, якщо треба довести оптимальність архітектури інтелектуальної агентної СППР (ІАСППР), застосовують так звані агент-бенчмарки – це еталонні набори тестів, призначені для оцінювання різних аспектів роботи інтелектуальних агентів та мультиагентних систем [6, 7]. Однак, не існує спеціалізованих бенчмарків для задачі ІАСППР, які обробляли б дані моніторингу про стан довкілля регіону за умов значної невизначеності даних та надавали б якісні рекомендації користувачам різних категорій. Більше того, в задачі, коли є багато варіантів реалізації архітектури ІАСППР, важливо вибрати оптимальний варіант. Щоб такі системи дійсно забезпечували якісну підтримку рішень, вони мають бути надійними (здатними стабільно працювати й давати повторювані результати в часі) та робастними (малочутливими до шуму, пропусків та аномалій у даних), володіти достатньо повною та адаптивною базою знань, забезпечувати високу швидкість та обчислювальну ефективність, а їхні рекомендації мають бути обґрунтованими, точними й інтерпретованими для кожного типу користувачів [9 – 11]. Ключовою передумовою досягнення такої якості є максимально можлива достовірність вхідних даних, оскільки навіть найкращі моделі та алгоритми не здатні компенсувати систематичні помилки або збої в інформаційно-вимірювальних системах (вихід з ладу датчиків, пропущені дані, збережені з помилками дані тощо), якщо ці проблеми не виявляються і не враховуються на етапі передоброблення. Особливо це актуально для міст, де основним джерелом даних є нерегулярна мережа громадських станцій з різним технічним станом, різною точністю й відсутністю формалізованих регламентів калібрування. Не менш важливим є побудова адекватної моделі для точного прогнозування даних на основі гарних вхідних даних. І потрібний модуль, який на основі вхідних і прогнозних даних правильно класифікує ситуацію (сценарій), яка має місце і для реагування на яку необхідно приймати оптимальні рішення різним категоріям користувачів. У цій ситуації постає завдання оптимально організувати архітектуру всієї агентної СППР, яка поєднує етапи оцінювання достовірності даних, класифікацію ситуацій (епізодів), прогнозування, інтерпретацію результатів для складових DPSIR-моделі та генерацію дійсно якісних природномовних рекомендацій. А для обґрунтованого вибору оптимальної архітектури такої ІАСППР потрібний спеціальний агент-бенчмарк оптимальної структури.

Мета дослідження – підвищити якість рекомендацій інтелектуальної агентної системи підтримки прийняття рішень на основі даних про стан довкілля міста у сфері екологічного управління містом шляхом вибору оптимальної архітектури прогнозної агентної СППР з використанням спеціально розробленого агент-бенчмарку оптимальної структури.

Формалізація задачі

В ідеалі можна описати багато різних моделей для усіх складових DPSIR, але в Україні є серйозні проблеми з доступом до даних і вони мали місце й у минулому (в нульових роках і раніше): реєстри джерел забруднень, координати джерел викидів, статистична звітність по джерелах утворення і джерелах викидів та, власне, викидах, по кожному підприємству, погодинні дані державного моніторингу про стан атмосферного повітря недоступні у відкритих джерелах. Навіть між державними органами та органами самоврядування є

ускладнений обмін цими даними, через недосконалість реєстрів, малу кількість станцій державного моніторингу та інші причини. Єдине, де у 20-х роках був досягнутий значний прогрес, це – розгортання української мережі громадського моніторингу стану повітря, але це не вирішує усіх проблем та й, через відсутність регламентів регулярного калібрування давачів станцій цієї мережі, їх результати не визнаються в державній системі моніторингу як об'єктивні. Обмеженість в даних не дозволяє повноцінно будувати ІАСППР для усіх складових моделі DPSIR, що суттєво звужує список можливих архітектур інтелектуальних агентів, серед яких варто шукати оптимальну.

Фактично, більшість СППР, які будуються в Україні, передбачають такий функціонал: аналізуються дані громадського оперативного моніторингу (складова «S»), оцінюються індекси забруднення атмосферного повітря (англ. «Air Quality Index» – «AQI» – це, по суті, складова «I») та формуються текстові рекомендації за певними шаблонами щодо того, що саме робити різним категоріям населення та установ. У разі, якщо буде виявлено значне відхилення від норм, тоді органи влади чи громадськість можуть переслати цю інформацію державним установам екологічного профілю, які мають акредитовані лабораторії з сертифікованими та повіреними вимірювальними приладами, які проведуть додаткове обстеження і, у разі підтвердження, передадуть іншим органам для реагування (штрафи, приписи, призупинення дії дозволу на діяльність тощо).

Нехай: X – множина точок простору (координати), T – дискретний час (у хвиликах, годинах, добах), J – індикатори стану атмосферного повітря (концентрація пилу PM2.5, PM10 тощо), H – множина громадських станцій, $S_j^*(x, t)$ – істинна (латентна) концентрація забруднювача j у точці x у момент часу t , тоді дані спостережень за показником j на станції k у моменту часу t :

$$y_{k,j}(t) = h_{k,j}(S_j^*(x_k, t)) + b_{k,j}(t) + \varepsilon_{k,j}(t), \quad (1)$$

де $h_{k,j}(\cdot)$ – характеристика давача інформаційно-вимірювальної системи (IBC); $b_{k,j}(t)$ – зміщення, через проблеми з калібруванням, старінням, пошкодженнями тощо; $\varepsilon_{k,j}(t)$ – шум.

Задача агента: оцінити не тільки S_j^* , а й – ймовірність несправності чи зміщення показів давача, довірчі інтервали для стану атмосферного повітря. Тобто будується апостеріорний розподіл $p(S^*, b, \varepsilon | Y)$ або хоча б його узагальнена оцінка $\hat{S}_j(x, t)$ з індикаторами достовірності.

ІАСППР повинна вміти класифікувати вид ситуації (сценарію) за різними класифікаціями (просторова локалізація (локальна, регіональна, глобальна), ступінь перевищення норм чи порогів індексу якості повітря, достовірність даних, ступінь ризику для громадян, конкретні види забруднення (пил Сахари, паління листя, викид підприємства) тощо) $\Phi(t_0)$ у заданий момент часу t_0 , що одразу дасть розуміння які слід надавати рекомендації, наприклад, ними можуть бути такі види:

- локальне забруднення (LOCAL)(перевищення на 1 – 2 сусідніх станціях): екологам – його шукати та локалізувати, населенню – застосовувати певні захисні засоби, обмежити перебування на свіжому повітрі тощо, органам влади – оповістити населення у певному радіусі впливу від станції, передусім, за напрямком вітру від джерела забруднення;
- глобальне забруднення (GLOBAL) (перевищення на усіх станціях міста): екологам – з'ясувати характер джерела, зібрати про нього додаткову інформацію і спробувати спрогнозувати його тривалість та наслідки впливу, населенню і органам влади – те саме, що за локального впливу, але органам влади слід оповіщати усе населення, а не тільки у певному радіусі;
- пошкодження давачів (або систем збереження їх даних) IBC (SENSOR_FAULT): власникам станцій – здійснити діагностування пристрою, можливо, замінити, органам влади – якось спробувати вплинути на власника, екологам та населенню – попередити про проблему зі спостереженнями і не турбуватись, через це, та ігнорувати ці дані;
- наявність значного перевищення на одній станції, хоча б за одним показником, хоча б 1 раз у певному інтервалі часу (ANOMAL_INTERVAL): див. вище, як для LOCAL;

– відсутність пошкоджень IBC та перевищень серед усіх поточних і прогнозних значень (NORMAL): усі категорії користувачів – інформування про поточний і прогнозний стан.

$$\Phi(t_0) \in \{\text{LOCAL, GLOBAL, SENSOR_FAULT, ANOMAL_INTERVAL, NORMAL}\}. \quad (2)$$

ІАСППР реалізує так звану політику π_θ :

$$\pi_\theta: (K_D(t_0), q) \rightarrow R_g^*(t_0, \tau), \quad (3)$$

де $K_D(t_0)$ – знання про DPSIR-систему у момент часу t_0 та дані про її складові: ряди спостережень зі станцій Y , координати станцій, оцінки $\hat{S}$ та їх довірчі інтервали, база знань про нормативи та пороги значень різних показників стану атмосферного повітря, інформація про можливі джерела забруднення, ліміти на їх викиди, метеорологічні умови тощо; q – промпт (вказівка, запит) користувача категорії $g = \{\text{public, ecologists, authorities, owner_stations, scientists, ...}\}$ на горизонт τ , наприклад такий: «Який поточний стан повітря (за показником ...) в районі...?» або «Який прогноз щодо стану повітря на завтра (3 дні, тиждень...)»?; $R_g^*(t_0, \tau)$ – набір природномовних рекомендацій (рішень) у момент часу t_0 за інтервал τ для категорії g користувачів.

Політика π_θ максимізує критерій оптимальності – функцію корисності або якості рекомендацій $U(R_g, K_D, q)$, з урахуванням відповідності фактичному та/чи прогнозному стану $\hat{S}$, дотримання екологічних нормативів та порогів для здоров'я громадян, зрозумілості й практичної цінності для категорій користувачів g . Це дозволяє вибрати оптимальну архітектуру ІАСППР A^* серед можливих m архітектур інтелектуальних агентів $A = \{a_1, \dots, a_m\}$.

Класична технологія оптимізації політики (3) – це метод машинного навчання з підкріпленням (англ. «Reinforcement Learning» – «RL»)[3, 12], де функція нагороди у момент часу t_0 за інтервал τ для категорії g враховує зменшення потенційної шкоди здоров'ю громадян та довкіллю, покращення функціонування IBC для власників станцій тощо. Але, за відсутності великих знань та обсягів детальних достовірних даних про усі зв'язки між усіма складовими системи DPSIR, що дозволяє генерувати довільні реалістичні дані для тренування RL-алгоритму, цей підхід не спрацює. Тому, більш часто для розв'язання цієї задачі використовують різні імітаційні підходи на основі наявних ретроспективних даних з певним рівнем невизначеності.

Варіанти структур агентної СППР

Опишемо декілька типових архітектур реалізації агентних СППР для поставленої задачі.

Архітектура A1. Звичайний пайплайн (послідовне підключення) π_θ^P агентів: агент, який очищає дані та виявляє проблеми, пов'язані з IBC, результатом є множина V ; агент, який за очищеними даними V робить прогноз даних за різними моделями, результатом буде множина M прогнозів з довірчими індикаторами; агент, який класифікує прогнозні дані M та визначає вид ситуації з множини (2) та за іншими класифікаціями (рівень ризику, індекс якості повітря тощо), результатом буде множина Φ_θ ; агент, який за комплексом ситуацій визначає набір рішень, на основі яких варто формувати рекомендації для заданої категорії g користувачів, результатом буде множина R_g ; агент з великою мовною моделлю (англ. «Large Language Model») – LLM_θ , який забезпечує генерування природномовного варіанта рекомендацій, з урахуванням набору рішень R_g ; тоді результат роботи моделі можна подати як композицію цих множин:

$$\pi_\theta^P = LLM_\theta \circ R_g \circ \Phi_\theta \circ M \circ V. \quad (4)$$

Архітектура A2. RAG-архітектура, що відрізняється наявністю векторної бази даних, яка містить знання з різних сфер цієї задачі та постійно оновлюється. Це може бути аналогічний пайплайн, але з більш складним інтелектуальним агентом, який формує набір оптимальних рішень RAG_g замість R_g :

$$\pi_\theta^{RAG} = LLM_\theta \circ RAG_g \circ \Phi_\theta \circ M \circ V. \quad (5)$$

Архітектура А3. Мультиагентна (МА) оркестрація, коли кожен агент може містити свою LLM, а головний агент оркеструє їх результати роботи, використовуючи різні схеми поєднання, приклади яких наведені у роботі [1].

Загалом, архітектури А1 та А2 наведені як спрощені приклади. Оскільки на етапах класифікації та формування рішень можуть враховуватись не тільки прогнози, а й – очищені дані, можуть бути й інші варіації.

В будь-якому разі, виникає задача вибору оптимальної архітектури цих ІАСПР. Обґрунтуємо яким вимогам вона повинна задовольняти, відповідно до формалізації критерію оптимальності.

Критерії якості та вимоги до оптимальної архітектури ІАСПР

Пропонуємо вибирати такі складові критерію якості, які монотонно узгоджені з практичною корисністю, тобто кожен з них приймає більші значення, то корисність зростає або не зменшується. Виходячи з наведеного вище визначення якості рішень, пропонуємо виділяти такі $K=5$ критеріїв:

- C_1 – критерій врахування достовірності даних або точність виявлення недостовірних даних, що впливають на рекомендації, який враховує наскільки архітектура враховує під час очищення даних рівень достовірності інформації окремих давачів і станцій ІВС (наприклад, частка аномальних ситуацій, правильно виявлених та «заглушених» у множині V у моделях (4) чи (5), щоб уникнути їх впливу на кінцеве рішення);

- C_2 – критерій точності аналізу ситуації, для якої надається рекомендація (наприклад, точність класифікації ситуацій чи рівня ризику в Φ_θ , точність прогнозів у M);

- C_3 – критерій обґрунтованості та інтерпретованості рекомендацій як міра узгодженості рекомендацій R_g із DPSIR-контекстом, нормативними значеннями, а також зрозумілість для відповідних категорій користувачів g ;

- C_4 – критерій робастності рекомендацій, який враховує чутливість рекомендацій до збурень у даних V , тобто наскільки мало змінюються інтерпретація епізодів та оцінка ризику у Φ_θ й загальна структура рекомендацій R_g , у разі відкидання частини вимірювань у V , або у разі введення контрольованих шумів і аномалій;

- C_5 – критерій обчислювальної ефективності, наприклад більша швидкодія з використанням менших витрат на ресурси, також характеризує здатність системи працювати в режимі, близькому до реального часу, що важливо для оперативних управлінських рішень.

Можна виділити й більше критеріїв, але вважаємо, що достатньо й цих 5. А тоді інтегральний критерій J для певної архітектури a пропонується у вигляді зваженої суми з додатними ω -вагами, які враховують пріоритетність відповідних критеріїв:

$$J(a) = \sum_{k=1}^K \omega_k C_k(a), \quad \forall \omega_k > 0, \quad \sum \omega_k = 1, \quad \forall C_k(a) \in \mathbb{R}^K. \quad (6)$$

Умови оптимальності та теорема про вибір архітектури ІАСПР

Як було зазначено вище, існує багато варіантів архітектур ІАСПР, тому важливо вміти вибирати оптимальну з них за критерієм якості рекомендацій (рішень) та з урахуванням певних обмежень та умов.

Нехай: $U(a, \varphi)$ – «справжня» практична корисність (якість) рекомендацій архітектури $a \in A$ у сценарії $\varphi \in \Phi$; $\bar{U}(a)$ – очікувана практична корисність (якість) рекомендацій архітектури a на всьому просторі сценаріїв Φ за невідомим розподілом P , яку можна наближено оцінити за експертними оцінками або інтегральними показниками ефективності системи.

Введемо декілька умов.

Умова 1 (репрезентативність тестових сценаріїв). Є скінченна вибірка з N сценаріїв (або ситуацій):

$$S = \{\varphi_1, \dots, \varphi_N\} \subset \Phi,$$

на якій обчислюються емпіричні оцінки критеріїв $\hat{C}_k(a)$, така, що:

1. S містить сценарії всіх типів, які є суттєвими для предметної області (різні типи ситуацій забруднення, різні варіанти достовірності станцій тощо).

2. Частотний розподіл типів сценаріїв у S не викривлює реальну картину (відсутнє систематичне переобтяження окремими типами), тобто є збалансованим (на практиці це перевіряють з використанням χ^2 -критерію або KL-дивергенції).

3. $\hat{C}_k(a)$ є узгодженими оцінками «істинних» критеріїв $C_k(a)$: у разі зростання N значення $\hat{C}_k(a)$ з великою ймовірністю збігаються до $C_k(a)$.

Умова 2 (монотонна узгодженість критеріїв з практичною корисністю).

Існує функція $F: \mathbb{R}^K \rightarrow \mathbb{R}$, яка строго зростає за кожною координатою – така, що:

$$\bar{U}(a) = F(C_1(a), \dots, C_K(a)).$$

Тобто, якщо для двох архітектур a_1, a_2 виконується:

$$C_k(a_1) \geq C_k(a_2) \quad \forall k \quad \exists k: C_k(a_1) > C_k(a_2),$$

тоді:

$$\bar{U}(a_1) > \bar{U}(a_2).$$

Тобто, якщо архітектура не гірша за всіма критеріями й краща хоча б за одним, тоді вона має вищу практичну корисність. Крім того, для невеликих приростів критеріїв $\Delta C_k(a) > 0$ емпірично перевіряють чи це не призводить до зменшення оцінки корисності рекомендацій: $\Delta \bar{U}(a) \geq 0$.

На практиці умова 2 перевіряється за коефіцієнтом кореляції Спірмена між значеннями критеріїв та експертними оцінками якості рекомендацій.

Умова 3 (додатні ваги). У функціоналі (6) усі ваги – строго додатні ($\forall \omega_k > 0$). Фактично, наприклад варіант $\omega_K = 0$ означав би, що цих критеріїв не K , а – $(K-1)$, тобто критерії з нульовими вагами просто ігноруються і не входять у кількість K .

Теорема.

Нехай виконуються умови 2 і 3, тоді будь-яка архітектура:

$$a^* \in \arg \max_{a \in A} J(a)$$

є Парето-оптимальною щодо вектора критеріїв $C(a)$, тобто не існує архітектури $a' \in A$, яка не гірша за a^* за всіма критеріями й краща хоча б за одним.

Якщо, крім того, виконується умова 1, а критерії оцінюються емпірично $\hat{C}_k(a)$ на репрезентативній множині сценаріїв S , тоді для достатньо великого N архітектура:

$$\hat{a}^* \in \arg \max_{a \in A} \hat{J}(a), \quad \hat{J}(a) = \sum_{k=1}^K \omega_k \hat{C}_k(a)$$

з великою ймовірністю є наближеною Парето-оптимальною щодо істинних критеріїв $C(a)$ та наближено оптимальною щодо практичної корисності $\bar{U}(a)$.

Доведення.

Теорему легко довести, якщо врахувати, що за умовою 3 зважена сума критеріїв (6) є строго монотонним відображенням за компонентами вектора критеріїв $C(a)$. Тоді одразу випливає, що жодна архітектура, яка домінує над a^* у сенсі Парето, не може мати значення інтегрального критерію, яке є меншим або дорівнює J , а отже, поява строго кращої за всіма критеріями архітектури неможлива і a^* є Парето-оптимальною. За умовою 1 та достатньо великою репрезентативною вибіркою S емпіричні оцінки $\hat{C}_k(a)$ наближено зберігають порядок між архітектурами, тож оптимальність $\hat{a}^*$ за $\hat{J}(a)$ з великою ймовірністю переноситься на істинні критерії. Аналогічно за допомогою умови 2 можна довести, що не існує архітектури, яка одночасно не гірша за всіма критеріями й має більшу практичну корисність $\bar{U}$.

Побудова оптимального агент-бенчмарку для вибору оптимальної архітектури ІАСПР

Існує багато відомих універсальних агент-бенчмарків. Наприклад [6, 7, 13 – 15]:

– Agent Bench – комплексний бенчмарк для LLM-агентів, який оцінює здатність агента виконувати складні інтерактивні задачі, що вимагають інструкційного слідування,

кодування, роботи зі знаннями й логічного міркування;

– GAIA (General AI Assistants) – складається з реалістичних завдань, які потребують інструментального використання (вебпошук, зовнішні сервіси), мультимодальності та складного міркування;

– Web Arena Benchmark – моделює реалістичні сайти (електронна комерція, форуми, контент-менеджмент), оцінює здатність агента виконувати довгі, багатокрокові веб-завдання у сфері планування, навігації, запам'ятовування та інструментального міркування.

Але не існує агент-бенчмарку для вибору оптимальної архітектури СППР для прийняття якісних рішень, пов'язаних з даними моніторингу, з урахуванням DPSIR-моделі чи хоча б її спрощених реалізацій.

Вище було виведено умови архітектури ІАСППР, яка є Парето-оптимальною на множині $\Phi \times U$ (де U – множина усіх можливих значень практичної корисності). Така архітектура є одночасно оптимальною і за інтегральним критерієм якості рекомендацій $J(a)$ на множині сценаріїв Φ , і за очікуваною практичною корисністю $\bar{U}(a)$ складових цього критерію. Використаємо ці умови для формування вимог щодо нового агент-бенчмарку (назвемо його «Air-DSS Agent Benchmark») для тестування архітектур ІАСППР, який буде з високою ймовірністю забезпечувати вибір Парето-оптимальної серед них, принаймні на множині $\Phi \times U$ та за критерієм $J(a)$. Отже, як впливає з наведеної вище теореми:

Наслідок. (Вимоги до структури агент-бенчмарку «Air-DSS Agent Benchmark»).

Якщо агент-бенчмарк оцінює архітектуру $a \in A$ за емпіричною версією інтегрального показника:

$$\hat{J}(a) = \sum_{k=1}^K \omega_k \hat{C}_k(a), \quad \omega_k > 0, \quad (7)$$

тоді бенчмарк з високою ймовірністю забезпечує, що архітектура

$$\hat{a}^* \in \arg \max_{a \in A} \hat{J}(a) \quad (8)$$

є Парето-оптимальною на $\Phi \times U$, якщо виконуються такі вимоги:

1. Репрезентативність тестових сценаріїв:
 - 1) бенчмарк повинен охоплювати всі істотні типи сценаріїв, характерні для предметної області, та таблицю їх відповідності реальним сценаріям Φ ;
 - 2) кількість даних для видів сценаріїв у бенчмарку повинна бути збалансованою і гарантувати відсутність надмірної кількості надто простих чи надто рідкісних сценаріїв;
 - 3) емпіричні критерії $\hat{C}_k(a)$ повинні бути статистично узгодженими оцінками істинних критеріїв $C(a)$, для чого бенчмарк повинен мати достатню кількість сценаріїв для стабілізації оцінок та нечутливості оцінок до крайових випадків та шумів.
2. Монотонна узгодженість критеріїв з практичною корисністю:
 - 1) збільшення кожного критерію $\hat{C}_k(a)$ не повинно зменшувати корисність, тому критерії повинні відображати справжні бажані властивості рекомендацій (точніше, за достовірними даними, швидко і бути стійкими до шумів чи випадкових аномалій тощо);
 - 2) слід підтверджувати монотонність між $\hat{C}_k(a)$ та $\bar{U}(a)$ з використанням коефіцієнта кореляції Спірмена;
 - 3) будь-яке невелике покращення критерію $\hat{C}_k(a)$ для модифікації архітектури ана на множині сценаріїв S не повинно давати зменшення корисності.
3. Лінійна інтегральна функція критерію (6) з додатними вагами.
4. Єдина методика оцінювання для усіх архітектур: застосовувати однакові сценарії S до всіх архітектур, використовувати одні й ті самі критерії та однаковий спосіб їх опрацювання.
5. Бенчмарк повинен містити сценарії різних типів складності, кожен сценарій повинен тестувати певні властивості архітектур, так, щоб критерії якості не були виродженими, тобто треба їх так підбирати, щоб:

- деякі сценарії вимагали коректності оброблення достовірності джерел (наприклад, фізично неможливі дані моніторингу на кшталт від'ємної вологості, температури повітря -100 градусів по Цельсію, концентрації пилу 3 кг/дм³ тощо),
 - деякі – високої точності розпізнавання видів сценаріїв (наприклад, слабкі піки забруднення),
 - деякі – високої робастності (великі пропуски в даних, поламани давачі, шуми),
 - деякі – обчислювальної ефективності (великі обсяги даних),
 - деякі – коректного вибору моделі прогнозу (різка зміна тренду),
 - деякі – глибокої інтерпретації ситуації (змішані сценарії, наприклад: транспорт + погода, паління листя + пил Сахари тощо).
6. Чітке визначення практичної корисності – бенчмарк має містити зовнішню або експертну оцінку:
- наскільки рекомендації архітектури відповідають моделі DPSIR;
 - наскільки вони корисні та інтерпретовані для населення, екологів, власників станцій та інших категорій з множини g;
 - наскільки правильно визначено рівень ризику (якщо його визначення передбачено);
 - наскільки рекомендації своєчасні.

Приклад побудови бенчмарку та вибору за його допомогою оптимальної архітектури ІАСПР за реальними даними моніторингу стану атмосферного повітря

Перевіримо працездатність наведених теоретичних аспектів на реальному практичному прикладі, хоча і суттєво спрощеному.

Візьмемо реальні дані з 20 станцій моніторингу стану атмосферного повітря у м. Вінниця (рис. 1). Деякі переміщались чи, після ремонту, змінювали номери, але одночасно працювало не більше 20 станцій. На практиці – менше, оскільки деякі припинили функціонування у 2020 – 2022 рр.

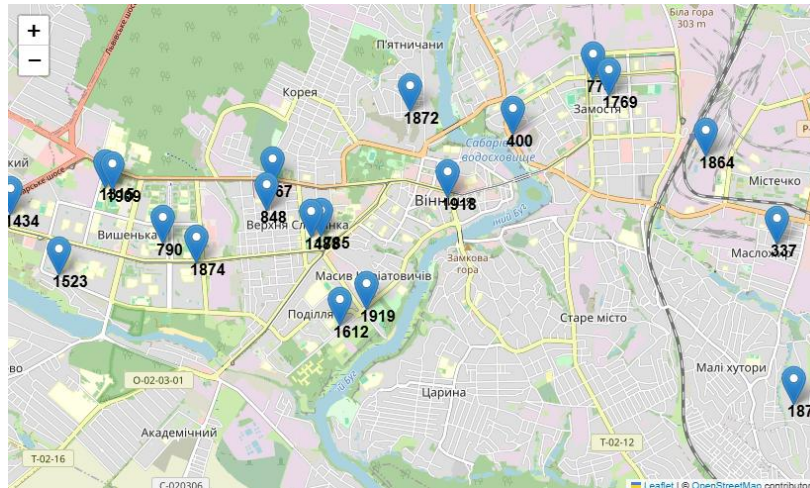


Рис. 1. Станції громадського моніторингу стану атмосферного повітря міста Вінниці, які працювали чи працюють протягом 2019 – 2025 рр., дані по яких доступні у мережі EcoCity[16]

По наведених на рис. 1 станціях отримано дані з використанням сервісу «Кабінет дослідника якості повітря України» Української громадської мережі моніторингу стану атмосферного повітря EcoCity [17], до якого автори статті мають доступ, як адміністратори цього кабінету від Вінницького національного технічного університету (ВНТУ). Використаємо дані за 2019 – 2025 рр. за показниками PM2.5, PM10, Temperature, Humidity, вже завантажені з «Кабінету дослідника якості повітря України» ВНТУ і збережені авторами в їх публічному Kaggle-датасеті [18] раніше під час підготовки їх роботи [16]. У тій же роботі [16] є частковий розвідувальний аналіз цього датасету.

Тестувались дві варіації архітектури A1:

- AgentA_QC_Prophet – агент А з контролем достовірності даних («QualityControl») та прогнозом на основі моделі Prophet:

$$\pi_{\theta 1}^P = R_g \circ \Phi_{\theta} \circ M_1 \circ V, \quad (9)$$

де M_1 – модель на базі відомої моделі часових рядів Prophet [12] без урахування сезонності та з автоматичним виявленням аномалій, більших за 90-й квантиль значень і генеруванням для них складової holiday з параметрами (prior_scale=10, lower_window = 0, upper_window=0);

- AgentB_NoQC – агент В без контролю достовірності даних:

$$\pi_{\theta 1}^P = R_g \circ \Phi_{\theta} \circ M_2, \quad (10)$$

де M_2 – «наївна» модель, яка вважає, що поточне значення показника не буде змінюватись усі наступні значення на довільний горизонт прогнозу.

Множина можливих ситуацій(2) обмежувалась 3-ма варіантами:

$$S = \{\text{SENSOR_FAULT}, \text{ANOMAL_INTERVAL}, \text{NORMAL}\} \subset \Phi.$$

Класифікація ситуацій здійснювалась за таким алгоритмом:

1. Якщо хоча б за одним із показників хоча б в одному зі значень у вхідному наборі даних було виявлено фізично неможливе значення, тоді ця ситуація класифікувалась як SENSOR_FAULT і виконувався перехід на п. 4, інакше – на п. 2.

2. Якщо хоча б за одним із показників хоча б в одному зі значень у вхідному наборі даних було виявлено перевищення верхнього чи нижнього порогових значень, що свідчить про наявність аномалії (вологість < 10, PM10 > 200 тощо), тоді ця ситуація класифікувалась як ANOMAL_INTERVAL і виконувався перехід на п. 4, інакше – на п. 3.

3. Класифікація ситуації як NORMAL і перехід на п. 4.

4. Збереження результату (множина Φ_{θ}).

За таким алгоритмом працював і бенчмарк, і обидва агенти.

В агенті AgentA_QC_Prophet для аналізу брались до уваги 7 останніх значень, разом із поточним, і прогнозувалось 1 наступне значення (горизонт прогнозу дорівнював одній добі), за яким формувались рекомендації. Прогнозування здійснювалось тільки за умови, що ситуація класифікована як NORMAL. А в агенті AgentB_NoQC для аналізу бралось тільки одне поточне значення (воно ж вважалось і прогнозним) і саме за ним аналізувалась поточна ситуація. Саме, через це, ці агенти важко назвати коректними для співставлення, але метою цього дуже спрощеного прикладу було відпрацювати всю методологію на практиці та перевірити її працездатність, а не об'єктивно розробити оптимальну архітектуру ІАСППР.

Формування набору природномовних стислих рекомендацій R_g здійснювалось за серією умов, які залежали і від класифікованої ситуації Φ_{θ} , і від категорії користувача g у промпті q .

Бенчмарк «Air-DSS AgentBenchmark» був реалізований за наведеними вище вимогами, але суттєво спрощено. У Kaggle-ноутбуку на Python було реалізовано формування варіантів ситуацій в автоматичному режимі у циклі на основі реальних даних спостережень. Для сценаріїв SENSOR_FAULT та ANOMAL_INTERVAL здійснювався пошук саме таких 7-денних інтервалів, які не містили б пропущені дані та містили б значення, які задовольняли умови для їх класифікації (див. вище). Для кожної ситуації було відомо 7-денний інтервал значень для кожного з 4-х показників та який це вид сценарію як еталонний варіант у бенчмарку.

Критерій якості (6) був спрощений максимально до критерію «точність» C_2 з двома підкритеріями. Формально, за (6) два підкритерії C_2 перетворюються на два окремих критерії а $K=2$:

$$J(a) = \omega_1 \hat{C}_1(a) + \omega_2 \hat{C}_2(a), \quad (11)$$

де $\hat{C}_1(a)$ – емпірична оцінка точності (за метрикою «accuracy» Python-бібліотеки Sklearn) збігу визначеного агентами виду ситуацій та еталонного виду ситуацій у тестах бенчмарку; $\hat{C}_2(a)$ – емпірична оцінка точності прогнозування наступного значення: якщо відносна похибка прогнозного значення за кожним показником не перевищувала еталонне значення більше, ніж на 20 %, тоді прогноз вважався успішним і це класифікувалось як 1, інакше – 0, потім аналогічно до $\hat{C}_1(a)$ точність оцінювалась за метрикою «accuracy»; оскільки

правильність класифікації ситуації значно більше впливає на точність рекомендацій, то були взято такі ваги критеріїв: $\omega_1 = 0,8$, $\omega_2 = 0,2$.

Формально, критерій, пов'язаний з достовірністю даних спостережень, непрямо впливав на результат визначення критерію точності: у разі оцінювання за даними, без урахування аналізу достовірності роботи ІВС в останній тиждень, важко було довіряти оцінкам, отриманим в поточний момент часу.

Результат оцінювання критерію (11) наведено на рис. 2.

	C1_class_acc	pm25_rmse	pm10_rmse	temp_rmse	hum_rmse	C2_pred_acc	J
agent							
AgentA_QC_Prophet	1.000000	6.660630	8.067726	2.281753	12.684616	0.50000	0.900000
AgentB_NoQC	0.421053	4.242611	5.451647	1.471058	9.458845	0.46875	0.430592

Рис. 2. Результат обчислення точності критеріїв у формулі (11) та середньоквадратична похибка прогнозування кожного показника стану атмосферного повітря кожним з агентів, усереднена по усіх сценаріях

Як видно на рис. 2, модуль класифікації ситуації AgentA_QC_Prophet є значно кращим за аналогічний модуль агента AgentB_NoQC, а прогнозні можливості агента AgentB_NoQC кращі не набагато і є такими ж поганими, як і в іншого агента, що означає, що для прогнозування реальних даних слід використовувати довші ряди спостережень та більш ефективні моделі, як наприклад, ту, що автори запропонували у статті [19] і яка забезпечила метрику $r2_score = 0,97$. Цікаво, що на рис. 2 середні похибки прогнозування AgentB_NoQC є меншими за агента AgentA_QC_Prophet, але критерій C_2 є менший, що означає, що AgentA_QC_Prophet більш часто потрапляє у потрібний інтервал $\pm 20\%$, а якщо вже помиляється, то помиляється дуже суттєво.

На рис. 3 наведено фрагмент порівняльної таблиці із Kaggle-ноутбуку на Python з параметрами ситуацій та їх еталонними видами з тестів бенчмарку і видів ситуацій, визначених агентами, архітектура яких аналізувалась.

	scenario_id	station_id	start_date	end_date	true_type	agent	pred_type	correct_cl
0	SENSOR_001	848	2021-07-13	2021-07-19	SENSOR_FAULT	AgentA_QC_Prophet	SENSOR_FAULT	1
1	SENSOR_001	848	2021-07-13	2021-07-19	SENSOR_FAULT	AgentB_NoQC	NORMAL	0
2	ANOM_002	1315	2024-01-06	2024-01-12	ANOMAL_INTERVAL	AgentA_QC_Prophet	ANOMAL_INTERVAL	1
3	ANOM_002	1315	2024-01-06	2024-01-12	ANOMAL_INTERVAL	AgentB_NoQC	NORMAL	0
4	ANOM_003	1315	2024-03-05	2024-03-11	ANOMAL_INTERVAL	AgentA_QC_Prophet	ANOMAL_INTERVAL	1
5	ANOM_003	1315	2024-03-05	2024-03-11	ANOMAL_INTERVAL	AgentB_NoQC	NORMAL	0
6	ANOM_004	1315	2024-03-05	2024-03-11	ANOMAL_INTERVAL	AgentA_QC_Prophet	ANOMAL_INTERVAL	1
7	ANOM_004	1315	2024-03-05	2024-03-11	ANOMAL_INTERVAL	AgentB_NoQC	NORMAL	0

Рис. 3. Фрагмент таблиці з тестами розробленого агент-бенчмарка за реальними даними моніторингу та класифікованими ситуаціями кожним з агентів AgentA_QC_Prophet і AgentB_NoQC

Загалом, аналіз показав, що за інтегральним критерієм J , як це видно на рис. 2, та за прикладами на рис. 3 архітектура агента AgentA_QC_Prophet є значно кращим за архітектуру агента AgentB_NoQC. Отже, оптимальною є архітектура агента AgentA_QC_Prophet (значення J є вищим більше, ніж у 2 рази).

Багато в чому цей приклад дозволив удосконалити теоретичний опис, наведений вище, та зробити його більш практично орієнтованим. Хоча приклад є суттєво спрощеним і не може претендувати на об'єктивність, але, загалом, у ньому дотримано майже усі вимоги (окрім

визначення практичної корисності рекомендацій та малої репрезентативності вибірки щодо S), які були обґрунтовані щодо агент-бенчмарка «Air-DSS Agent Benchmark» для пошуку оптимальної архітектури прогнозної ІАСППР для прийняття якісних рішень на основі даних про стан атмосферного повітря міста.

Запропонована методологія може бути ефективною і для ІАСППР інших видів довкілля, наприклад для прийняття якісних рішень на основі даних про стан масивів вод у басейні великої річки. Шляхом нескладної адаптації множин «Air-DSS Agent Benchmark» можна адаптувати і до «Water-DSS Agent Benchmark» на основі тієї ж DPSIR-моделі.

Висновки

У роботі розв'язано задачу оптимізації архітектури прогнозної інтелектуальної агентної СППР (ІАСППР), призначеної для підтримки прийняття рішень у сфері екологічного управління містом за даними моніторингу стану атмосферного повітря. На основі формалізації критеріїв якості рекомендацій, включаючи точність, достовірність, робастність та обчислювальну ефективність, сформульовано інтегральний критерій та доведено теорему про Парето-оптимальність архітектури ІАСППР, яка його максимізує. Це дозволило вивести вимоги до оптимального агент-бенчмарку та обґрунтувати його структуру як інструмента для коректного порівняння ІАСППР. Побудовано новий бенчмарк під назвою «Air-DSS Agent Benchmark», який генерує репрезентативні та збалансовані види сценаріїв дозволяє оцінювати архітектури за єдиними критеріями, які є складовими загального критерію якості рекомендацій.

За реальних даних м. Вінниці (2019 – 2025 рр.) показано, що агент із контролем достовірності даних та статистичним виявленням аномалій забезпечує більшу, ніж у 2 рази, якість класифікації та прогнозування порівняно з агентом без перевірки якості даних. Це підтверджує високу практичну значущість аналізу даних на достовірність на етапі передоброблення та використання складних інтелектуальних прогнозних моделей.

Наукова новизна роботи полягає у формалізованому обґрунтуванні критеріїв якості рекомендацій ІАСППР, доведенні теореми про Парето-оптимальність архітектури, що максимізує інтегральний критерій цієї якості, та формулюванні обґрунтованих вимог до агент-бенчмарку, який з високою ймовірністю забезпечує Парето-оптимальність вибраної архітектури ІАСППР.

Отримані результати підтверджують, що якість рішень екологічної ІАСППР суттєво залежить від архітектури прогнозних агентів, а запропонований бенчмарк є ефективним засобом оптимізації таких складних систем та може бути використаний у подальших дослідженнях і реально впроваджений, причому не тільки у сфері моніторингу та управління атмосферним повітрям міста.

СПИСОК ЛІТЕРАТУРИ

1. Nezamoddini N., Gholami A. A survey of adaptive multi-agent networks and their applications in smart cities. *Smart Cities*. 2022. Vol. 5, № 1. P. 318–347. DOI: 10.3390/smartcities5010019.
2. Multi-Agent RAG Chatbot Architecture for Decision Support in Net-Zero Emission Energy Systems / G. Gamage et al. 2024 *IEEE International Conference on Industrial Technology (ICIT)*, Bristol, United Kingdom, 2024. P. 1–6. DOI: 10.1109/ICIT58233.2024.10540920.
3. Intelligent Air Quality Control Through Continuous Policy Learning / J. D. Sinha et al. 2025 *9th International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, 2025. P. 1282–1288, DOI: 10.1109/ICISC65841.2025.11188208.
4. Environmental indicators: Typology and overview: Technical report 25/1999. Copenhagen: European Environment Agency, 1999. – 08 верес. 1999. URL: <https://www.eea.europa.eu/en/analysis/publications/tec25>.
5. Шмундяк Д. О., Мокін В. Б., Крижановський Є. М. Системний аналіз стану атмосферного повітря регіону з урахуванням впливу аномалій: монографія. Вінниця: ВНТУ, 2025. 169 с.
6. Evaluation and Benchmarking of LLM Agents: A Survey / M. Mohammadi et al. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25)*. Association for Computing Machinery, New York, NY, USA, 2025. P. 6129–6139. <https://doi.org/10.1145/3711896.3736570>.
7. Survey on evaluation of LLM-based agents / A. Yehudai et al. *arXiv preprint arXiv:2503.16416*. 2025. URL: 10.48550/arXiv.2503.16416. DOI: 10.48550/arXiv.2503.16416.

8. Mokin V., Shmundiak D., Kopniak V. Air Quality Monitoring from EcoCity. Kaggle Dataset, 2025. URL: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.
9. Turban E., Sharda R., Delen D. Decision Support and Business Intelligence Systems, 9th ed. Boston, MA, USA: Pearson Education, 2010. URL: <https://scribd.com/document/355912027/Decision-Support-and-Business-Intelligence-Systems-9th-Ed-Efrain-Turban>.
10. International Organization for Standardization, ISO/IEC 25002:2024 – Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – Quality Model Overview and Usage, Geneva, Switzerland: ISO, 2024. URL: <https://www.iso.org/standard/78175.html>.
11. International Organization for Standardization, ISO/IEC 25012:2008 – Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model, Geneva, Switzerland: ISO, 2008. URL: <https://www.iso.org/standard/35736.html>.
12. Мокін В. Б., Дратованій М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання. Вінниця : ВНТУ, 2024. 258 с. URL: <https://docs.vntu.edu.ua/card.php?id=8163>.
13. Agent Bench: Evaluating LLMs as Agents / X. Liu et al. *arXiv preprint arXiv: 2308.03688*. 2023. URL: <https://arxiv.org/abs/2308.03688>. DOI: 10.48550/arXiv.2308.03688.
14. GAIA: A Benchmark for General AI Assistants / G. Mialon et al. *arXiv preprint arXiv: 2311.12983*. 2023. DOI: 10.48550/arXiv.2311.12983. URL: <https://arxiv.org/abs/2311.12983>.
15. WebArena: A Realistic Web Environment for Building Autonomous Agents / C. Zhou et al. *arXiv preprint arXiv: 2307.13854*. 2023. DOI: 10.48550/arXiv.2307.13854. URL: <https://arxiv.org/abs/2307.13854>.
16. Копняк В. Є., Мокін В. Б., Варчук І. В. Визначення середньорічного тренду показників якості атмосферного повітря регіону за даними громадського моніторингу. *LIV Всеукраїнська науково-технічна конференція підрозділів Вінницького національного технічного університету: Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації*, Вінниця, 24-27 березня 2025 року, <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2025/paper/view/24325/20024>.
17. Інформаційний сервіс «Кабінет дослідника якості повітря України» Української мережі громадського моніторингу якості повітря «Eco City», 2025. URL: <https://archive.eco-city.org.ua/>.
18. MokinV., Shmundiak D., Kopniak V. Air Quality Monitoring from Eco City, Kaggle Dataset, Mar 2025. URL: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.
19. Метод бустингу гетероскедастичних моделей для прогнозування концентрацій пилу Сахари в атмосферному повітрі України / В. Є. Копняк та ін. *Наукові праці ВНТУ*. 2024. №2. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/733>. DOI: <https://doi.org/10.31649/2307-5376-2024-2-28-38>.

Стаття надійшла до редакції 11.12.2025.

Стаття пройшла рецензування 28.12.2025.

Копняк Володимир Євгенович – аспірант кафедри системного аналізу та інформаційних технологій, e-mail: vkorniyak@gmail.com.

Мокін Віталій Борисович – д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail: vbmokin@vntu.edu.ua.

Вінницький національний технічний університет.