

**Яременко В.С.**

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»

## ОГЛЯД НАЯВНИХ МУЛЬТИАГЕНТНИХ СИСТЕМ ДЛЯ ЗАДАЧ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

*Проведено аналіз мультиагентних систем глибинного аналізу великих об'ємів даних, що були розглянуті як сукупність інтелектуальних агентів і зв'язків між ними. Показано, що процес аналізу включає сукупність етапів з інтеграції даних, попереднього опрацювання даних, класифікації даних, кластеризації даних, визначення асоціативних правил і представлення даних у формі бази знань. У роботі проведено дослідження з побудови методологічних основ визначення функціональності моделі мультиагентної системи. У якості параметрів оцінки виступали такі: оцінка роботи з даними та формування бази знань, оцінка організації архітектури та виконання процесів системою, оцінка ефективності навчання агентів, оцінка взаємодії функціональних елементів системи, оцінка взаємодії «людина-машина» й оцінка ефективності використання апаратних і програмних ресурсів інформаційних систем.*

**Ключові слова:** мультиагентна система, інтелектуальний агент, глибинний аналіз даних, нейромережвий алгоритм, кластеризація, інформаційна система, мережвий ресурс.

**Постановка проблеми.** Відповідно до стандартного визначення агент є програмним алгоритмом, що запускається з метою виконання певного завдання відповідно до запиту користувача чи іншого програмного блоку. Сучасні системи інтелектуального (глибинного) аналізу даних (data mining) зазвичай розглядаються як сукупність агентів і зв'язків між ними. Агент такої системи визначається як функціональний елемент, що здатний до аналізу та впливу на середовище мережевого ресурсу інформаційної системи (далі – ІС), частиною якої він є.

Інтелектуальний аналіз даних є комплексним процесом із визначення закономірностей і моделювання правил організації великих об'ємів даних. Аналіз на базі агента зумовлює застосування системою автономних інтелектуальних агентів із метою підтримки процесу пошуку та прийняття рішень, що має призводити до збільшення рівня продуктивності та масштабованості програмного комплексу. Завдяки параметрам автономності, мобільності й адаптивності використання інтелектуальних агентів вважається ефективним рішенням для паралельного та розподіленого аналізу даних [1].

Функціональні можливості роботи агента як програмного комплексу можна розділити на три групи (рис. 1):

– агент реагування (reactive agent);

– проактивний агент (proactive agent);

– інтерактивний агент (social or interactive).

Режим роботи «агент реагування» обмежується впливом алгоритму на взаємодію типу «машина-машина» (M2M: machine-to-machine) і реагуванням на конкретні запити чи запуском відповідно до заданого режиму. У проактивному режимі агент самостійно аналізує потоки даних, на цьому рівні додається аналіз взаємодії типу «людина – машина» (P2M: person-to-machine). В інтерактивному режимі агент, крім того, здійснює вплив на інших агентів і здатен впливати на взаємодію типу «людина – людина» (P2P: person-to-person). Слід зауважити, що мультиагентна система зазвичай реалізує всі три наведені режими роботи. Упровадження таких систем дозволяє збільшити функціональність з'єднання M2M-типу та зменшити обсяг P2M-контактів, що призводить до вивільнення людського ресурсу, зокрема спеціалістів, що задіяні в рутинних операціях із підтримки стабільної роботи IT-інфраструктури [1].

Із метою визначення ключових аспектів побудови мультиагентних систем було проведено аналіз останніх досліджень і публікацій, присвячених методам глибинного аналізу даних. Так, у роботах [1–3] загальні принципи побудови інтелектуального агента як функціонального вузла системи глибинного аналізу даних. Дослідження [4–9] присвячені особливостям побудови мультиагентних систем і перевагам підходів, моделей і архі-

тектур, що можуть бути використані в процесі їх розроблення. Також у роботах [5; 9–11] були розглянуті та проаналізовані підходи, що останнім часом застосовуються у побудові мережевих сервісів на основі мультиагентних систем розподіленого глибинного аналізу великих об'ємів даних.

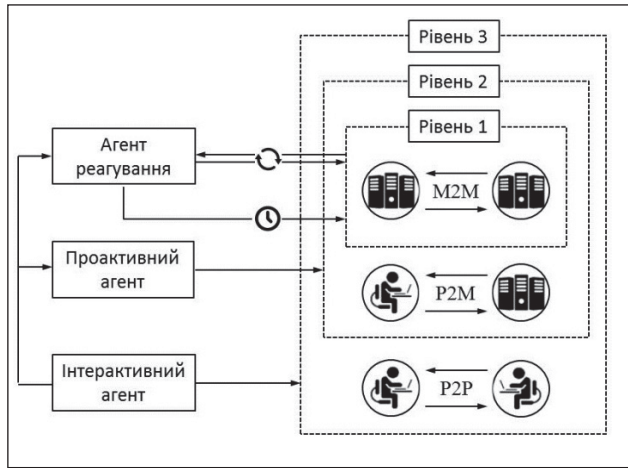


Рис. 1. Базові режими роботи мультиагентної системи

**Постановка завдання.** Проведений аналіз досліджень мультиагентних систем дає можливість побудувати цілісну методологію, що може слугувати універсальним інструментом розроблення архітектури алгоритмів на основі інтелектуальних агентів відповідно до поставлених задач. Тому метою роботи є аналіз підходів, що використовуються під час створення моделей інтелектуального аналізу великих об'ємів даних у рамках розподілених систем.

### 1. Використання інтелектуального агента під час аналізу даних

На рис. 2 показано, що глибинний аналіз даних із боку агента є двостороннім процесом. З одного

боку, він включає аналіз даних агентом, що пере-дує етапу прийняття рішень, а з іншого – оброблені дані використовуються для навчання й оптимізації алгоритму самого агента. Він складається з оброблення даних за допомогою агента, а також пошуку даних для агентів [1–3]. Процес аналізу включає таке:

- інтеграцію даних;
- попереднє оброблення даних;
- класифікацію даних;
- кластеризацію даних;
- визначення асоціативних правил;
- представлення даних у формі бази знань.

При цьому представлення даних у формі бази знань використовується для вдосконалення логіки агента.

Оцінка ефективності застосування інтелектуального агента під час глибинного аналізу даних залежить від аспектів роботи системи, що підлягають перевірці. Зазвичай виділяють такі рівні оцінки:

- рівень роботи з даними та формування бази знань;
- рівень організації архітектури та виконання процесів системою;
- рівень ефективності навчання агентів;
- рівень взаємодії функціональних елементів системи;
- рівень P2M-взаємодії;
- рівень ефективності використання апаратних і програмних ресурсів ІС.

Інтеграція інтелектуальних агентів за умови адекватної організації системи суттєво збільшує продуктивність роботи комплексу та спрощує його структуру. Під час вивільнення людського ресурсу робота з такими системами надає широкі можливості для використання її профільними

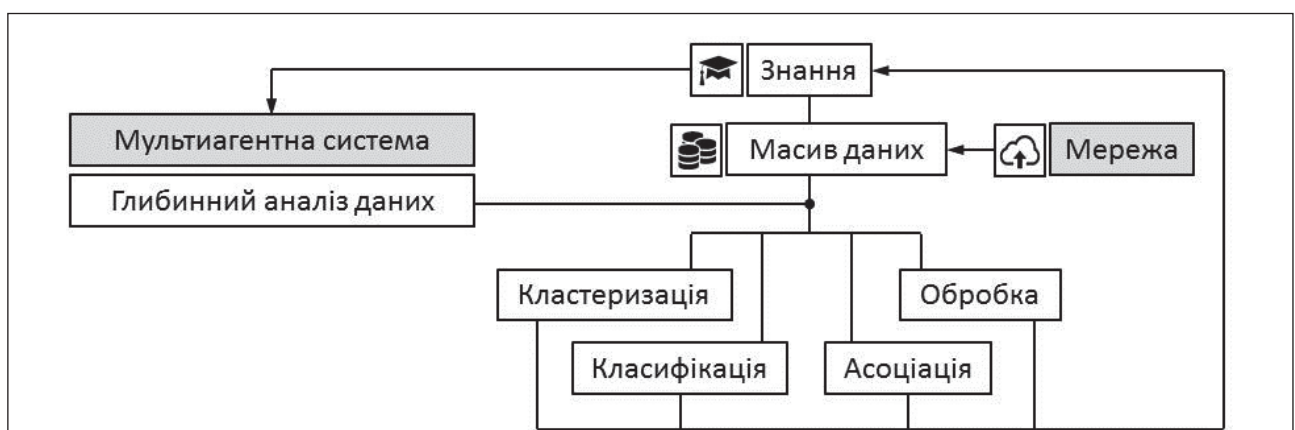


Рис. 2. Схема глибинного аналізу даних мультиагентною системою зі зворотним зв'язком

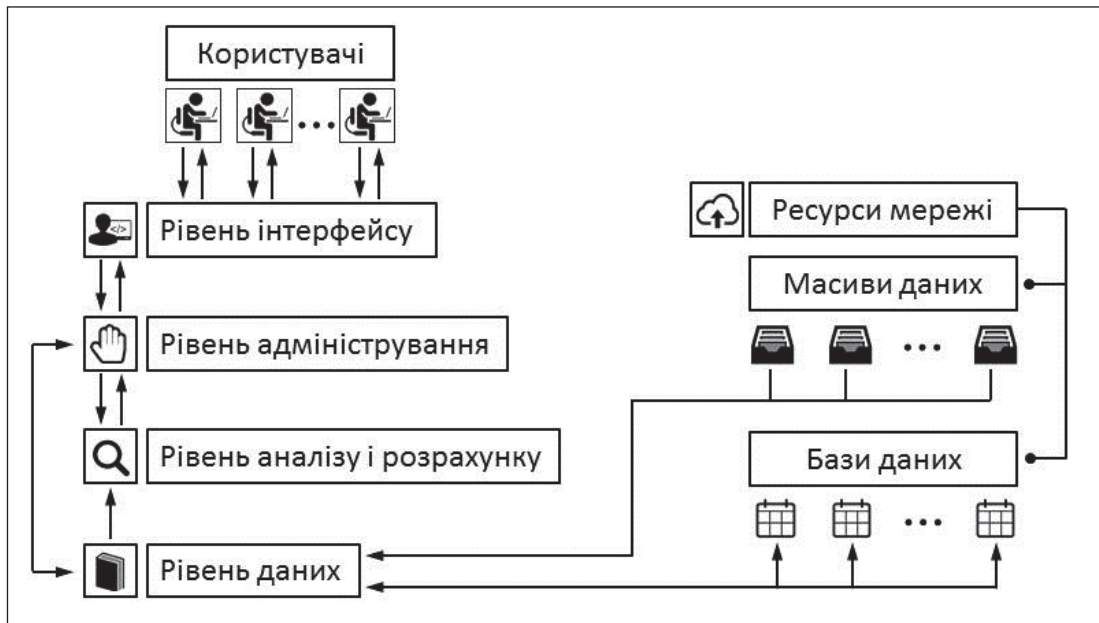


Рис. 3. Чотирирівнева архітектура мультиагентної системи

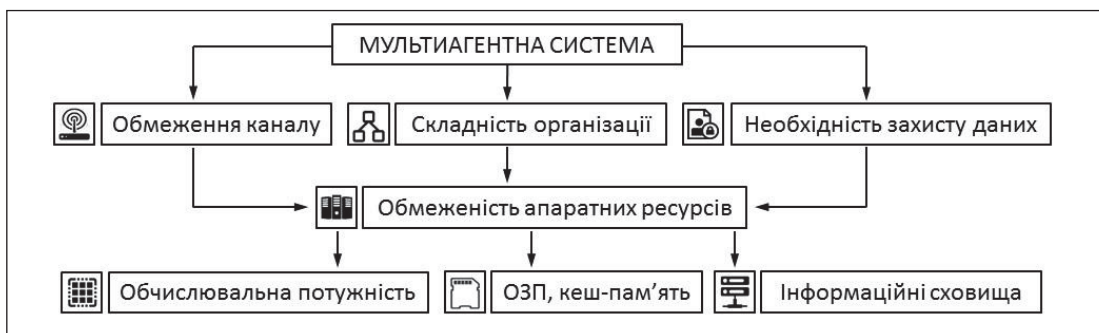


Рис. 4. Особливості організації мультиагентної системи розподіленого аналізу даних

фахівцями та мультидисциплінарними експертами. Така система підвищує якість отриманих знань, спрощує процес ідентифікації моделей, а також сприяє процесу прийняття рішень у режимі реального часу [1].

## 2. Побудова мультиагентної системи аналізу даних

Як було вказано вище, сучасні системи глибокого аналізу даних містять у собі певну кількість багатофункціональних агентів, а отже, можуть розглядатися як мультиагентні. Базова структура, представлена на рис. 3, складається з чотирьох рівнів:

- 1) рівень інтерфейсу;
- 2) рівень адміністрування;
- 3) рівень аналізу й розрахунку;
- 4) рівень даних.

Завдання, що необхідно виконати, визначається користувачем на рівні інтерфейсу. На цьому

рівні активуються агенти інтерфейсу, що перетворюють запит користувача на інструкції для машинних алгоритмів системи [4–7]. На рівні аналізу й розрахунку за допомогою відповідних агентів відбувається глибокий аналіз даних. При цьому дані отримуються з масивів і баз даних на рівні, що оперується агентами даних. Агенти рівня адміністрування при цьому керують процесом вирішення задач; таким чином, цей рівень слугує зв'язком між іншими рівнями, а також використовується для вибору оптимального методу й отримання статистичних даних роботи системи.

Організація мультиагентного розподіленого глибокого аналізу даних є нетривіальною задачею, що вимагає побудови ефективної інфраструктури мережевого ресурсу ІС і вирішення значної кількості проблем, що пов'язані між собою (рис. 4), серед яких можна виділити такі:

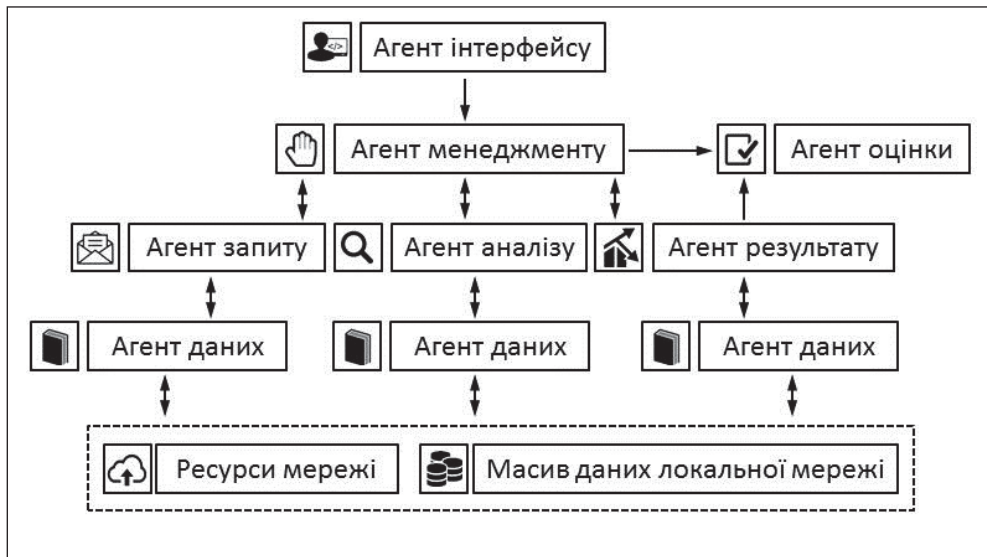


Рис. 5. Архітектура мультиагентної системи розподіленого аналізу даних

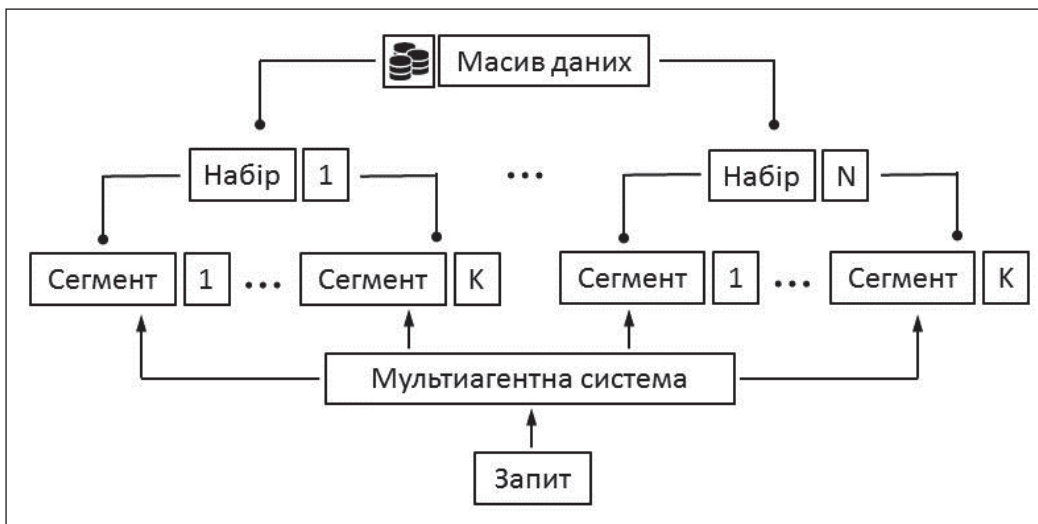


Рис. 6. Алгоритм роботи мультиагентної системи розподіленого аналізу даних

- обмеження на пропускну спроможність каналів передачі даних;
- складність організації багатокомпонентної системи, що складається з автономних агентів;
- загроза нелегального доступу до конфіденційних даних;
- обмеження на потужність обчислювальних ресурсів.

Така система має забезпечувати відповідний рівень масштабованості, інформаційного захисту, надійності функціонування, адаптивності й автономності процесу отримання даних агентами.

Під час розподіленого мультиагентного глибокого аналізу даних (MADM: Multi-Agents Systems In Distributed Data Mining) зазвичай використовуються такі агенти [5–9]:

- агент інтерфейсу, що, як було вказано, відповідає за зв'язок із користувачем;
- агент менеджменту, що отримує запити від агента інтерфейсу, формує план із їх виконання й відповідає за їх завершення, а також синхронізує діяльність інших агентів;
- агент даних, що передає дані з наявних джерел до агентів аналізу, результату та запиту;
- агент аналізу, що реалізує алгоритм глибокого аналізу на основі інформації, наданої модулем знань;
- агент результату, що отримує результат пошуку даних від агента аналізу та представляє ці результати через шаблони звітів;
- агент оцінки, що містить імена та параметри всіх зареєстрованих агентів;

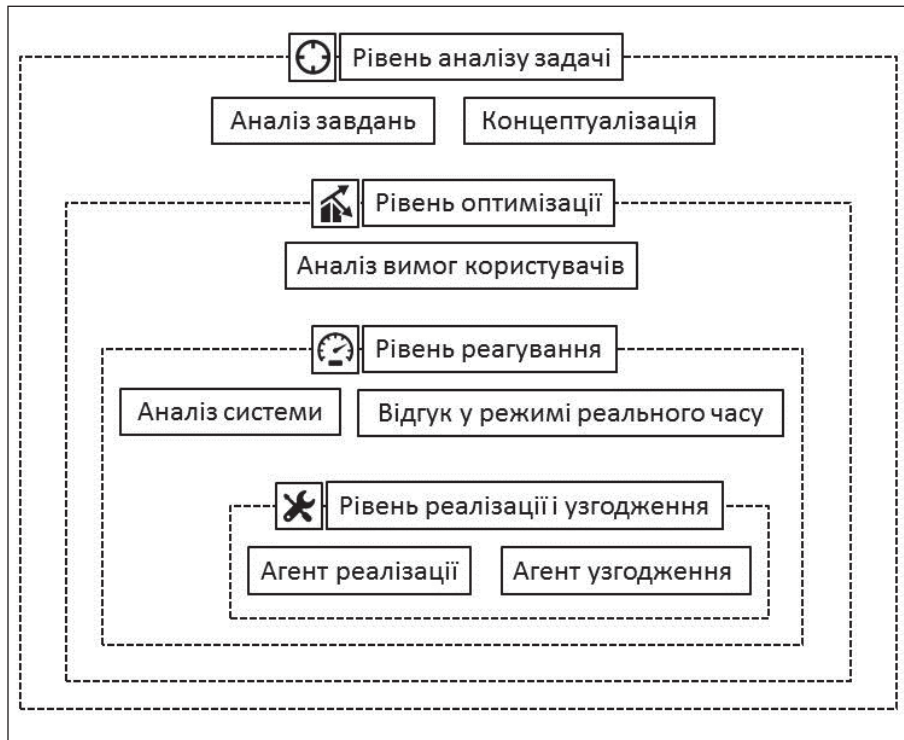


Рис. 7. Багаторівнева схема оптимізації системи глибокого аналізу даних

– агент запиту, що створюється для кожного запиту користувача та використовує правила бази знань для створення запитів належної форми.

Типова архітектура організації мультиагентної системи розподіленого аналізу даних показана на рис. 5.

### 3. Підтримка й оптимізація мультиагентної системи аналізу даних

Для забезпечення ефективної роботи мультиагентної системи глибокого аналізу даних необхідно провести попередній етап поділу та кластеризації великих об’ємів даних, що підлягають аналізу [5; 8–9]. Зазвичай масив даних розділяють на  $N$  наборів, кожен із яких поділяється на  $K$  сегментів (рис. 6), при цьому

$$K = \frac{N}{q}, \text{де } q \in [1; N]. \quad (1)$$

Із кожним сегментом працює окремий агент, що на основі його аналізу генерує відповідні моделі та правила.

Після цього етап кластеризації може бути реалізований шляхом визначення відстані між параметрами об’єктів відповідно до обраної метрики. Так, відстань між параметрами  $x_i$  та  $x_j$  може бути визначена відповідно до метрики Мінковського

$$D_{ij} = \left[ \sum_{i=1}^d |x_i - x_j|^{\frac{1}{n}} \right]^n \text{де } n, d - \text{натуральні числа,} \\ \{\mathbb{N}\} \quad (2)$$

або Манхетенської метрики:

$$D_{ij} = \sum_{i=1}^d |x_i - x_j|. \quad (3)$$

Для оптимізації мультиагентної системи використовуються протоколи взаємодії з користувачами на рівні інтерфейсу. При цьому схема оптимізації системи глибокого аналізу даних складається з таких шарів (рис. 7):

- рівень аналізу задачі;
- рівень оптимізації;
- рівень реагування;
- рівень реалізації й узгодження.

На першому рівні розглядається сама постановка завдання, створюється шаблон рішення і здійснюється концептуалізація конкретних задач. На рівні оптимізації проводиться аналіз запитів і оцінки роботи системи користувачем. На рівні реагування система дає відповідь на виклики відповідно до правил із бази знань. Агенти в цьому шарі не проходять етап навчання, тому що від них вимагається реакція в режимі реального часу. Останній, найбільш глибокий рівень, надає можливість прискорити процес видобування даних за допомогою паралельного чи розподіленого оброблення.

Як показано на рис. 8, оптимізація мультиагентної системи використовує як основу базу переконань. Далі агент оптимізації передає дані до агента прогнозування. На початку організації

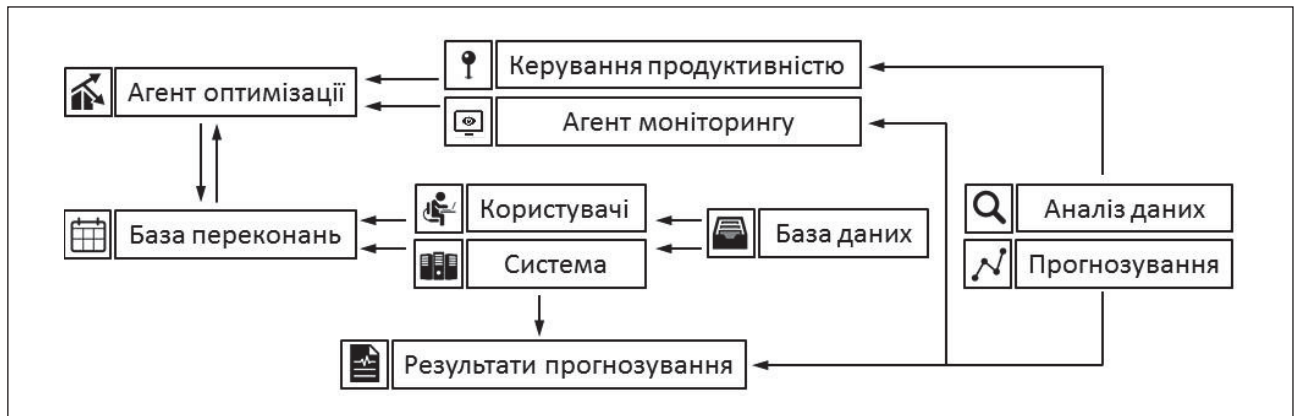


Рис. 8. Оптимізація прогнозування в глибинному аналізі даних

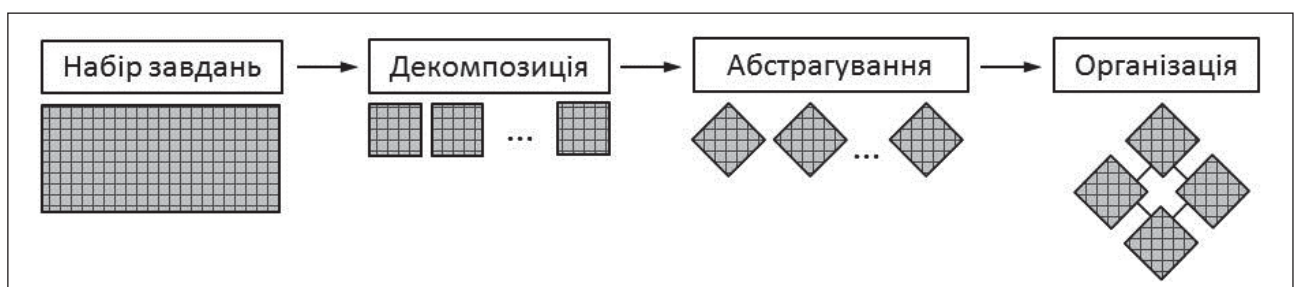


Рис. 9. Алгоритм паралельного глибинного аналізу нечітких наборів даних

роботи системи користувач має забезпечувати зворотний зв'язок, за якого відгуки повідомляються агенту моніторингу. Таким чином, навчання відбувається на основі фактичного зворотного зв'язку, а отже, у ролі агента моніторингу може виступати користувач, що спеціалізується на мультиагентних системах. У будь-якому разі агент моніторингу та керування продуктивності повинен виконувати певні завдання для досягнення мети, зокрема, кількісно оцінювати відгуки інших користувачів. На основі зворотного зв'язку можна побудувати модель, що на базі статистичних даних екстраполює функціональний зв'язок між системними прогнозами та реакцією користувачів. Основним обмеженням на цьому етапі є якість зворотного зв'язку (як на рівні P2P, так і на рівні P2M). Інше обмеження міститься в просторі пошуку, що може не відповідати обмеженим обчислювальним ресурсам. Агент продуктивності може бути функціональним вузлом прийняття рішень, користувачем, який викликається координатором, або координатором (відповідно до виконання задачі оптимізації системи). Для прогнозування зазвичай використовуються нейромережеві моделі (далі – НММ) [12], що навчаються на основі інформації, накопиченої в базі даних системи. Нейромережевий аналіз охоплює як про-

гнозування роботи машинних алгоритмів, так і відгуки користувачів відповідно до вивчення типових патернів людської поведінки.

Через те, що користувач є частиною системи, а оптимізація базується на його відгуках, у мультиагентних системах активно використовуються технології м'яких обчислень (soft computing), зокрема нечітка логіка [7].

#### 4. Організація мультиагентної системи паралельного аналізу нечітких даних

Для аналізу неточних або неповних даних і пошуку взаємозв'язків між ними в мультиагентних системах використовуються грубі набори (rough sets). Для наборів даних цього типу не використовується додаткова інформація, як-то статистика, імовірнісні залежності тощо [1–11]. Мультиагентні системи ефективно працюють із грубими наборами в рамках підходу паралельного аналізу даних. Базовими етапами, що здійснюються під час реалізації цього підходу (рис. 9), є такі:

- декомпозиція;
- абстрагування;
- організація.

Аналогічно до розділення масиву даних, що аналізується, на набори, систему можна розділити на багато підсистем, кожна з яких може бути представлена власним субагентом. При цьому кожен

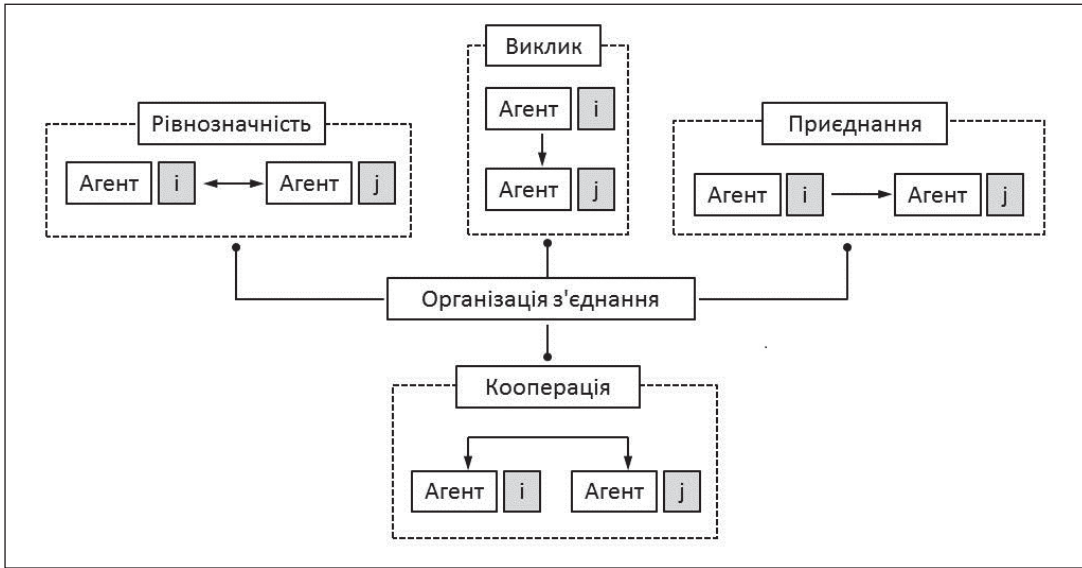


Рис. 10. Основні види взаємозв'язку між агентами

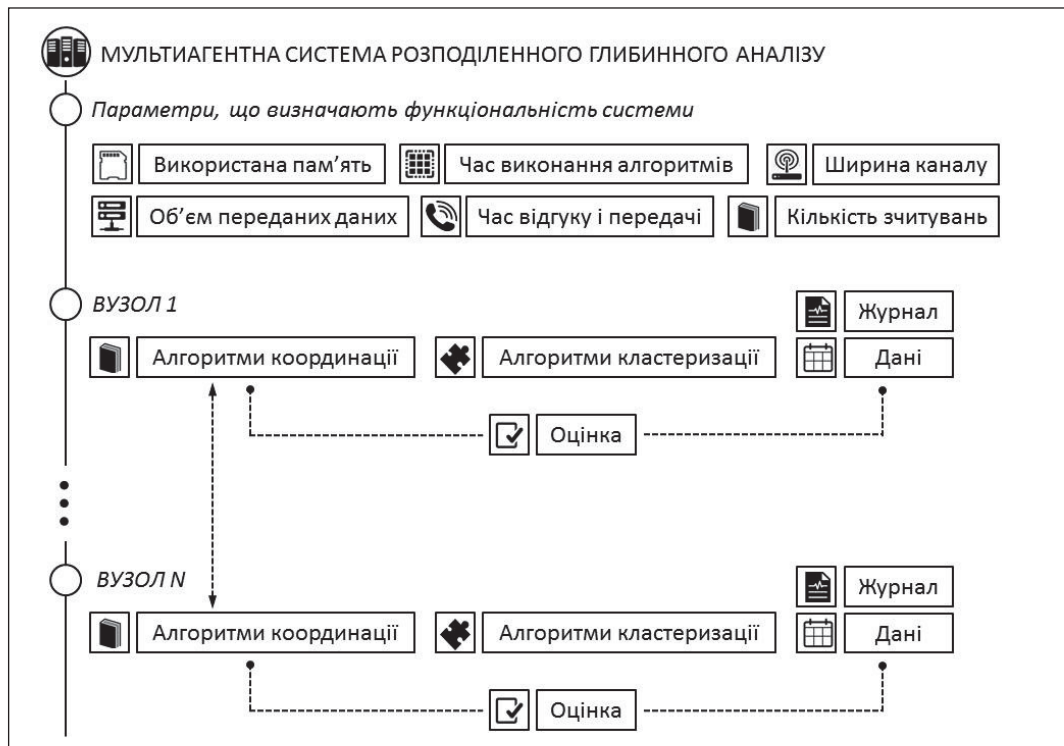


Рис. 11. Загальна схема роботи мультиагентної системи розподіленого глибокого аналізу

субагент включає набір завдань, алгоритм аналізу й інтерфейс зв'язку з іншими субагентами. Згідно із цим підходом, кожна субінформаційна система є інтелектуальним агентом, що використовує власний метод глибокого аналізу даних. Використовуючи різні типи взаємозв'язків між агентами, можна отримати остаточні правила прийняття рішень.

Основними типами взаємозв'язку між агентами є рівнозначність (equivalence), приєднання (inclusion),

кооперація (cooperation), виклик (calling). На основі цих правил можна побудувати мультиагентну систему будь якої складності (рис. 10).

Таким чином, загальна схема роботи мультиагентної системи розподіленого глибокого аналізу базується на взаємодії між функціональними вузлами, що складаються із субагентів, які взаємодіють між собою (рис. 11). Оптимізація системи в режимі реального часу відбувається за

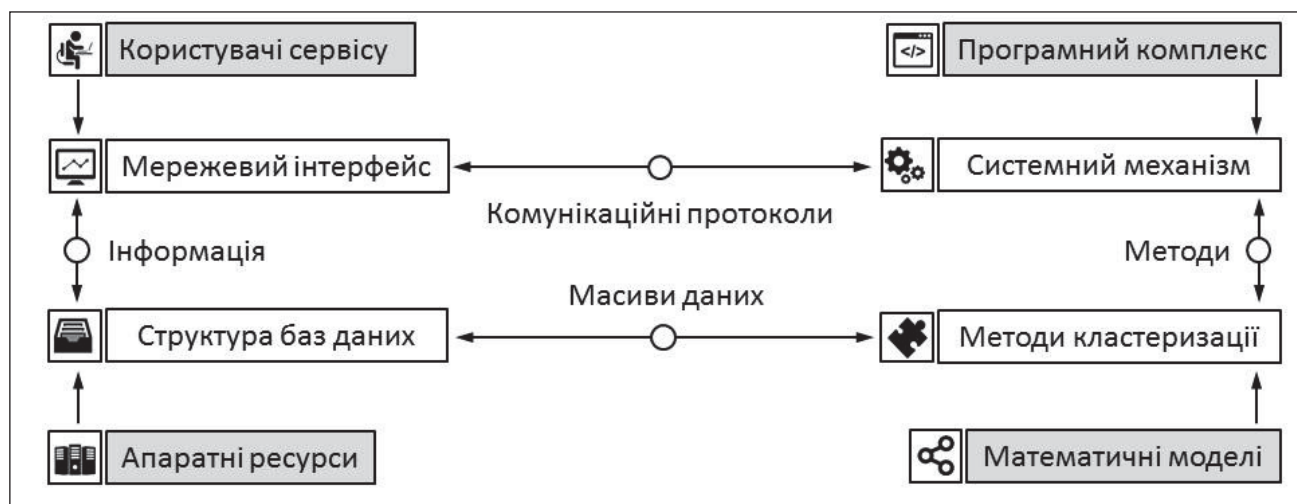


Рис. 12. Веб-сервіс на основі мультиагентної системи

рахунок роботи агента ефективності, який відповідає за взаємодію між іншими агентами й аналіз статистики.

Мультиагентну систему розподіленого аналізу даних можна розглядати як веб-сервіс із доступом і розмежуванням доступу користувачів глобальної мережі (рис. 12).

**Висновки.** У роботі проведено аналіз мультиагентних систем глибокого аналізу великих об'ємів даних. Показано, що процес аналізу включає сукупність етапів з інтеграції, попереднього оброблення, класифікації, кластеризації даних, а також визначення асоціативних правил і представлення даних у формі бази знань.

Проведене дослідження побудови методологічних основ визначення функціональності моделі мультиагентної системи. Указані основні проблеми організації веб-сервісу з використан-

ням мультиагентної системи глибокого аналізу даних. Основними є обмеження на пропускну спроможність каналів передачі даних, складність організації багатокomпонентної системи, що складається з автономних агентів, загроза нелегального доступу до конфіденційних даних і обмеження на потужність обчислювальних ресурсів.

Зауважено, що на початку організації роботи системи користувач має забезпечувати зворотний зв'язок, виступаючи в ролі агента моніторингу. Агент моніторингу та керування продуктивністю повинен кількісно оцінювати відгуки інших користувачів. Зазначається, що для прогнозування зазвичай використовуються нейромережеві моделі, що навчаються на основі інформації, накопиченої в базі даних системи. Побудовані схеми організації мультиагентної системи паралельного аналізу нечітких даних.

#### Список літератури:

1. Data Mining. (2017). Larsen & Keller Educ, 3rd ed., Saltham: Neruda.
2. Fariz A., Abouchabaka J., & Rafalia N. Using multi-agents systems in distributed data mining: a survey. *Journal of Theoretical & Applied Information Technology*, 2015. 73(3), Pp. 427–440.
3. Cao L., Weiss G., & Philip S. A brief introduction to agent mining. *Autonomous Agents and Multi-Agent Systems*, 2012. 25(3), Pp. 419–424.
4. Russell S., & Norvig P. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited. 2016.
5. Corchado J. Trends in practical applications of agents and multiagent systems: 9th International Conference on Practical Applications of Agents and Multiagent Systems. Berlin: Springer. 2011.
6. Al-Barky M., Ali J. Intelligent mining agent, in 8th International Conference on Computing Technology and Information Management (ICCM), 2012. 1. P. 23.
7. Serrano E., Rovatsos M., & BotíA J. A. Data mining agent conversations: A qualitative approach to multiagent systems analysis. *Information Sciences*, 230, 2013. Pp. 132–146.
8. Kazík O., Pešková K., Pilát M., & Neruda R. Meta learning in multi-agent systems for data mining. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 02*, 2011, August. Pp. 433–434. IEEE Computer Society.



9. Sharma D., & Shadabi F. Multi-agents based data mining for intelligent decision support systems. In Systems and Informatics (ICSAI), 2014, 2nd International Conference on IEEE. (2014, November). Pp. 241–245.
10. Corchado J. Trends in practical applications of agents and multiagent systems. 9th International Conference on Practical Applications of Agents and Multiagent Systems. Berlin: Springer. 2011.
11. Chaimontree S., Atkinson K., & Coenen F. (2011, May). A multi-agent based approach to clustering: harnessing the power of agents. In International Workshop on Agents and Data Mining Interaction. Springer, Berlin, Heidelberg. Pp. 16–29
12. Pandey T., Panda N., & Sahu P. Improving performance of distributed data mining (DDM) with multi-agent system. IJCSI Int. J. Comput. Sci, 2(9), 2012. Pp. 74–82.

## ОБЗОР СУЩЕСТВУЮЩИХ МУЛЬТИАГЕНТНЫХ СИСТЕМ ДЛЯ ЗАДАЧ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

*Проведен анализ мультиагентных систем глубинного анализа больших объемов данных, которые были рассмотрены как совокупность интеллектуальных агентов и связей между ними. Показано, что процесс анализа включает совокупность этапов по интеграции данных, предварительной обработке данных, классификации данных, кластеризации данных, определению ассоциативных правил и представлению данных в форме базы знаний. В работе проведено исследование по построению методологических основ определения функциональности модели мультиагентной системы. В качестве параметров оценки выступали оценка работы с данными и формирование базы знаний, оценка организации архитектуры и выполнения процессов системой, оценка эффективности обучения агентов, оценка взаимодействия функциональных элементов, оценка взаимодействия «человек – машина» и оценка эффективности использования аппаратных и программных ресурсов информационных систем.*

**Ключевые слова:** мультиагентная система, интеллектуальный агент, глубинный анализ данных, нейросетевой алгоритм, кластеризация, информационная система, сетевой ресурс.

## AN OVERVIEW OF EXISTING MULTI-AGENT SYSTEMS FOR DATA MINING PROBLEM SOLVING

*The paper analyzes the multi-agent systems for data mining of large amounts of data, which are considered as a set of intellectual agents and relationships between them. It is shown that the process of data analysis includes the stage of data integration, pre-processing, classification, clustering, the definition of associative rules and the presentation of data in the form of a knowledge base. As evaluation parameters are used such parameters as: evaluation of work with data and the knowledge base creation, evaluation of system architecture organization and system processes execution, agents learning efficiency, evaluation of functional elements interaction, evaluation of “man – machine” interaction, efficiency of hardware and software resources usage. The example of a web-service indicates the problems that arise during a multi-agent system for data mining development process.*

**Key words:** multiagent system, intelligent agent, data mining, neural network algorithm, clusterization, information system, network resource.