

ЗАСТОСУВАННЯ СТРАТЕГІЇ ОПТИМАЛЬНОСТІ ПРИ ВИБОРІ АЛГОРИТМІВ СИНТЕЗУ МОДЕЛЕЙ У СИСТЕМАХ БАГАТОРІВНЕВОГО СОЦІОЕКОЛОГІЧНОГО МОНІТОРИНГУ

Анотація. Стаття присвячена дослідженню ефективності використання стратегії оптимальності при визначенні алгоритму синтезу моделі у порівнянні зі стратегією придатності у процесі функціонування автоматизованої системи багаторівневого соціоекологічного моніторингу. Доведено, що поєднання в метод ідентифікації функціональних залежностей процесів індуктивного синтезу моделей об'єктів довкілля та класифікації вхідних даних з метою вибору кращого моделюючого алгоритму призводить до покращання характеристик якості результуючих моделей при зменшенні часу їх синтезу.

Ключові слова: автоматизована система багаторівневого соціоекологічного моніторингу, синтез моделі, синтезатор, локальні алгоритми перетворення інформації, ідентифікація функціональних залежностей, класифікація вхідних даних, індуктивне моделювання, стратегія придатності, стратегія оптимальності.

Аннотация. Статья посвящена исследованию эффективности использования стратегии оптимальности при определении алгоритма синтеза модели по сравнению со стратегией пригодности в процессе функционирования автоматизированной системы многоуровневого социоэкологического мониторинга. Доказано, что сочетание в метод идентификации функциональных зависимостей процессов индуктивного синтеза моделей объектов окружающей среды и классификации входных данных с целью выбора лучшего моделирующего алгоритма приводит к улучшению характеристик качества результирующих моделей при уменьшении времени их синтеза.

Ключевые слова: автоматизированная система многоуровневого социоэкологического мониторинга, синтез модели, синтезатор, локальные алгоритмы преобразования информации, идентификация функциональных зависимостей, классификация входных данных, индуктивное моделирование, стратегия пригодности, стратегия оптимальности.

Abstract. The paper is devoted to study of the efficacy of usage of an optimal strategy in determining the synthesis model algorithm compared with the strategy of applicability in the operation of an automated system of multilevel socio-ecological monitoring. It is proved that the introduction to the method of combining the identification of functional relationships of the processes of inductive synthesis of models of the environment objects and the classification of input data in order to select the best simulation algorithm leads to an improvement quality characteristics of the resulting models with decreasing the time of their synthesis.

Key words: automated system for multi-level socio-ecological monitoring, the synthesis model, synthesizer, local algorithms for data transformation, the identification of functional dependencies, the classification of input data, inductive modeling, strategy, fitness strategy for the optimal.

1. Вступ

Автоматизована система багаторівневого соціоекологічного моніторингу [1] призначена для неперервного перетворення інформації, поданої у вигляді концентрації техногенних забруднювачів у воді, повітрі, продуктах харчування та ґрунті разом із характеристиками техногенних випромінювань. У даній статті відображені результати наукових досліджень, метою яких було підвищення ефективності функціонування автоматизованих систем багаторівневого соціоекологічного моніторингу шляхом побудови алгоритмів перетворення первинної інформації на базі класифікації вхідних даних. Ці дослідження є актуальними, оскільки пов'язані із покращанням здоров'я населення шляхом підвищення ефективності управління техногенним навантаженням довкілля за рахунок удосконалення моніторингових процесів.

Розширення можливостей автоматизованих систем соціоекологічного моніторингу стримується обмеженими можливостями його науково-методичного апарату при обробці певних масивів вхідних даних низької інформативності. На сьогодні способи вирішення цієї задачі є недостатньо ефективними. Дослідження сучасного стану теорії та практики синтезу індуктивних моделей в автоматизованих системах соціоекологічного моніторингу дозволили визначити причини їх недостатньої ефективності при обробці масивів даних із значно відмінними характеристиками.

2. Аналіз останніх досліджень та публікацій

При створенні автоматизованих систем багаторівневого соціоекологічного моніторингу типовою є задача перетворення інформації про концентрацію шкідливих елементів у воді, повітрі, ґрунті та продуктах харчування, а також про характеристики техногенних випромінювань:

$$X = \{x_1, x_2, \dots, x_n\}, \quad (1)$$

де n – кількість видів забруднень, в інформацію про те, яка при таких характеристиках стану довкілля буде кількість захворювань населення на різні види хвороб:

$$Y = \{y_1, y_2, \dots, y_m\}, \quad (2)$$

де m – кількість видів захворювань.

Функціональна залежність

$$Y = f(X) \quad (3)$$

називається моделлю об'єкта моніторингу. Моделі використовуються як елементи ієрархічної структури підсистеми перетворення інформації, яка забезпечує процес обробки результатів моніторингу довкілля, і зветься локальними алгоритмами перетворення інформації (ЛАПІ) [2]. На рис. 1 подана ієрархічна структура підсистеми перетворення інформації, отримана шляхом висхідного синтезу ЛАПІ.

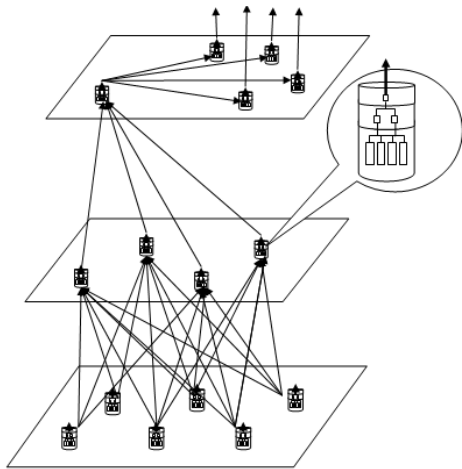


Рис. 1. Структура підсистеми перетворення інформації

Процес отримання залежності (3) називається синтезуванням моделей. Програмний модуль, який забезпечує процес синтезування, зветься синтезатором.

Оскільки функціональна залежність (3) є багатопараметричною моделлю складного об'єкта, то для її конструювання використовуються метод групового врахування аргументів (МГУА) [3] або багатшарові структури, що конструюються на його основі. При цьому індуктивні моделі використовуються як ЛАПІ за умови, що значення показника їх якості будуть не гірше наперед визначеного граничного значення. Показником якості моделей є критерій регулярності, який використовується в типовому алгоритмі МГУА. Його граничні значення визначаються на етапі проектування структури підсистеми перетворення інформації.

Таким чином, якісні характеристики окремих індуктивних моделей визначають достовірність інформації, яку отримують на виході системи моніторингу і на основі якої приймаються рішення із управління техногенним навантаженням довкілля.

Існує ряд специфічних вимог до синтезатора таких моделей. Використання індуктивних моделей як структурних елементів підсистеми перетворення інформації дозволяє вирішувати в комплексі одні з основних проблем, які виникають при проектуванні ієрархічних багаторівневих систем. Це формування ієрархічної структури підсистеми перетворення

ня інформації та координація взаємодії окремих її елементів [2]. Тому синтезатор повинен мати необхідну різноманітність – забезпечувати створення моделей, які можуть бути використані як структурні елементи на всіх рівнях перетворення інформації без значних переналаштувань за різної інформативності даних, що подаються на вхід кожного із рівнів перетворення інформації.

Основною функцією системи соціоекологічного моніторингу є забезпечення інформацією процесу прийняття рішень із управління техногенним навантаженням довкілля. У процесі формування керуючих впливів із зменшення викидів техногенних забруднювачів та оцінки наслідків запропонованих дій неодноразово виникає необхідність у використанні нових масивів даних, які до цього не були використані при синтезі структури підсистеми перетворення інформації. Зміна вектора зовнішніх завдань вимагає перебудови структури системи, оскільки певні ЛАПІ, які на даний момент присутні в системі, не забезпечують отримання нової інформації на основі додаткових даних, поданих на вхід системи. Крім того, якість частини раніше синтезованих моделей під впливом зміни властивостей вхідного масиву даних знижується і вони вже не можуть використовуватись як ЛАПІ. Таким чином, у процесі оперативного формування рішень періодично виникає необхідність підвищити якість існуючих моделей та кілька раз пересинтезувати структуру підсистеми перетворення інформації. Оскільки ця структура є ієрархічним поєднанням понад 50 моделей необхідної якості, актуальним є підвищення різноманітності синтезатора моделей. Це означає, що необхідно розширити адаптивні властивості алгоритмів синтезу моделей до зміни властивостей вхідних даних та забезпечити синтез нової структури підсистеми перетворення інформації за прийнятний час, який не обмежує оперативність прийняття рішень.

На даний час визначення того, який із множини наявних індуктивних алгоритмів буде використано у процесі синтезування моделі, відбувається за алгоритмом придатного відбору відповідно до стратегії придатності. Згідно з цією стратегією, вибраним вважається той алгоритм, значення критерію якості якого не гірше встановленого. Тобто, встановлюється певне граничне значення критерію якості моделей, і для створення ЛАПІ вибирається той індуктивний алгоритм ідентифікації функціональних залежностей, який дозволяє створити модель із заданими характеристиками. Відбувається пошук першого придатного алгоритму. Проте можлива ситуація, коли у множині наявних індуктивних алгоритмів присутні такі, які дають можливість синтезувати моделі зі значно кращими характеристиками якості. Але так як перший придатний алгоритм вже визначено, то застосовується саме він, незважаючи на гірші показники якості. Тому подальше підвищення різноманітності синтезатора можливе за рахунок застосування іншої стратегії вибору алгоритмів синтезу моделей.

3. Постановка задачі

Таким чином, ставиться задача: для вхідного масиву даних, що описує об'єкт моделювання, необхідно вибрати із наявних алгоритм, який забезпечує синтез моделі із найкращими характеристиками якості. Тобто, визначення того, який із множини наявних індуктивних алгоритмів буде використано у процесі синтезування моделі, повинно відбуватися за стратегією оптимальності.

4. Виклад основного матеріалу дослідження

Для розв'язання поставленої задачі при виборі алгоритму синтезу моделей пропонується змінити стратегію придатності на стратегію оптимальності. Передбачається вважати вибраним той алгоритм, який забезпечує показник якості синтезованих моделей близьким до кращого.

Найбільш відомим алгоритмом вибору, який реалізує стратегію оптимальності, є повний перебір. Оцінивши показники якості синтезованих моделей за всіма наявними алгоритмами, можна отримати модель із найвищим показником якості. Зважаючи на те, що множина наявних алгоритмів синтезу моделей скінченна, реалізувати в синтезаторі цей алгоритм нескладно, але необхідно враховувати, що час синтезу моделі оптимальної якості в даних умовах складає приблизно 78 секунд. Структура підсистеми перетворення інформації містить як мінімум 50 моделей. Таким чином, синтез нової структури при зміні масиву вхідних даних буде тривати близько 3900 секунд або понад годину.

Для випадків, коли час відгуку системи при зміні масиву вхідних даних не є критичною характеристикою, повний перебір залишається прийнятним алгоритмом функціонування синтезатора. Але за критерієм забезпечення оперативності рішень, які приймаються за результатами використання автоматизованих моніторингових систем, алгоритм повного перебору не є прийнятним.

Існує цілий клас задач для системи соціоекологічного моніторингу, при яких тривалість часу отримання інформації є критичною величиною. В процесі формування упереджувачих та компенсуючих заходів із зниження техногенних впливів на довкілля, необхідність використання нових даних виникає не менш, ніж 3–4 рази. Крім того, при техногенній катастрофі або під час терористичних актів, коли відбувся аварійний викид, наприклад, аміаку, хлору або інших шкідливих речовин, приходиться оперативно приймати рішення про визначення заходів щодо захисту населення. При цьому концентрація шкідливих речовин у повітрі кардинально змінюється. Відповідно якісно змінюються властивості вхідних даних. Існуючі індуктивні моделі, на яких побудована структура багаторівневої системи перетворення інформації моніторингових систем, стають неадекватними. Виникає потреба в оперативній заміні моделей у реальному режимі часу. Стоїть завдання – забезпечити оперативну роботу моніторингової системи. При інерційності системи близько години вона стає нездатною забезпечити оперативність рішень.

Для зменшення часу відгуку системи в даній роботі пропонується будувати функціонування синтезатора за новим методом адаптивного конструювання алгоритмів синтезу моделей (МАКАСМ). Задача синтезу нової моделі розв’язується як ідентифікація функціональної залежності модельованого показника від характеристик об’єкта моніторингу відповідно виразу (3). Від існуючих методів ідентифікації функціональних залежностей МАКАСМ відрізняється тим, що на першому етапі синтезу моделі відбувається класифікація вхідних даних. На рис. 2 подана функціональна схема синтезатора, створена на основі МАКАСМ.

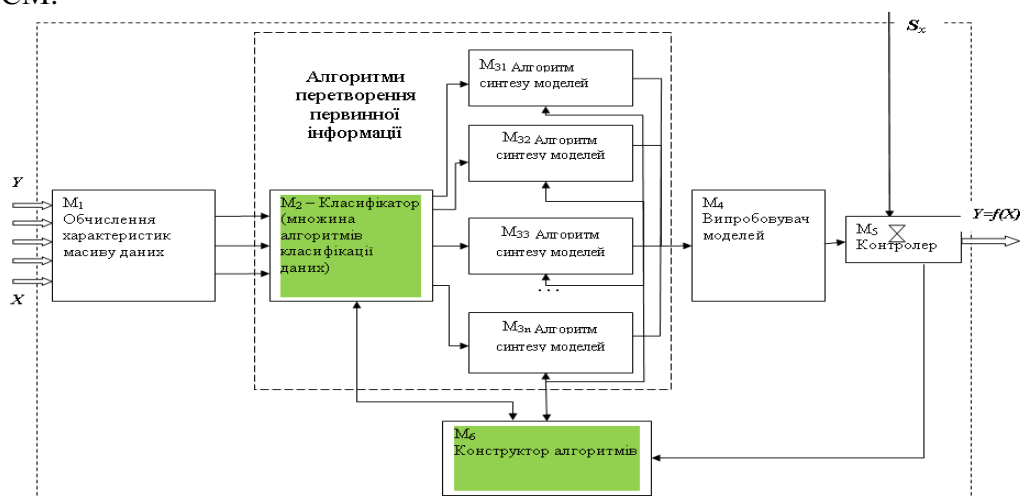


Рис. 2. Функціональна схема синтезатора моделей

На вхід синтезатора надходить масив даних первинного опису – чисельних характеристик об’єкта моніторингу. Модуль M_1 призначений для розрахунку характеристик ма-

сиву вхідних даних. Модуль M_2 реалізує алгоритм класифікації вхідних даних, який використовує їх розраховані характеристики. За результатами класифікації із виходу модуля M_2 вхідні дані направляються на вхід одного із алгоритмів синтезу моделі, які реалізуються модулями $M_{31} - M_{3n}$.

Модуль M_4 призначений для випробування синтезованої моделі та розрахунку значення її показника якості. Модуль M_5 порівнює значення показника якості синтезованої моделі та граничного значення цього показника S_x . Цей параметр модуля задається ззовні. За умови, коли якість моделі відповідає вимогам, вона направляється на вихід синтезатора для подальшого використання як ЛАПІ. У випадку, коли якість моделі незадовільна, модель на використання не подається, і запускається на виконання конструктор алгоритмів M_6 . В завдання модуля M_6 входить конструювання нових алгоритмів класифікації та синтезу моделей.

У загальному випадку задачу класифікації вхідних даних [4] можна формалізувати таким чином. Нехай задано перелік вхідних даних:

1. Множина даних M , які можуть надходити на вхід синтезатора.
2. Розбиття множини M на скінченну кількість підмножин (класів) Ω_k , ($k = \overline{1, m}$, де m – кількість класів), $M = \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \dots \cup \Omega_m$. Кожен клас визначається властивостями алгоритму, який забезпечує синтез адекватної моделі для цих даних.
3. Об'єкт класифікації ω , який є масивом вхідних даних, що надійшли в даний момент.
4. Опис об'єкта класифікації, який задається значеннями деяких характеристик x_j ($j = \overline{1, N}$, де N – кількість характеристик). Сукупність значень характеристик x_j визначає опис об'єкта $I(\omega) = \{x_1, x_2, x_3, \dots, x_N\}$.

Необхідно за значеннями з набору характеристик вхідного об'єкта $I(\omega)$ віднести його до одного з існуючих класів Ω_k , ($k = \overline{1, m}$, де m – кількість класів).

Ситуація ускладнюється тим, що діапазон зміни комбінації характеристик масивів даних, які описують окремий клас, явно визначити не вдається. Тому для розв'язання цієї слабоформалізованої задачі класифікації запропоновано використовувати нейромережі [5] в поєднанні з іншими алгоритмами класифікації, які реалізують біонічні підходи [6, 7].

Дослідження проводились при використанні 100 масивів даних, 50 з яких є результатами реальних спостережень кількості захворювань населення Черкаської області та концентрації шкідливих речовин у повітрі жилої зони. Друга половина масивів генерувалась штучно шляхом обчислення різних математичних функцій та використання генераторів значень випадкових величин.

Для проведення досліджень було сконструйовано 20 схем індуктивних ідентифікаційних алгоритмів, в основу яких лягли елементи таких алгоритмів моделювання:

1. Метод групового урахування аргументів [3].
2. Метод Степаненка [8].
3. Неперервні генетичні алгоритми [5].
4. Метод балансу пар моделей [2].
5. Методи рециркуляції та полімоделювання [2].

Дані алгоритми застосовувались при використанні чотирьох видів опорного вигляду моделі:

1. $y = a_0 + a_1x_1 + a_2x_2$.
2. $y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2$.

$$3. y = a_0 + a_1x_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_2.$$

$$4. y = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 + a_6x_1^2x_2^2,$$

де x_1, x_2 – значення вхідних параметрів опорної моделі, $a_i, i = \overline{1,6}$ – коефіцієнти моделі, що визначаються в процесі її навчання.

У табл. 1 наведені результати досліджень використання стратегії оптимальності при застосуванні повного перебору та МАКАСМ у порівнянні зі алгоритмом придатного відбору відповідно до стратегії придатності. У таблиці подано 10 об'єктів дослідження та значення критерію регулярності моделі, яка була синтезована для кожного з об'єктів при використанні тієї чи іншої стратегії. Як граничне значення показника якості, що визначає придатність моделі до використання, було використано 150 випадків захворюваності на 10000 населення (відмітка „–” у полі таблиці означає, що визначити алгоритм, який забезпечує синтез моделі із заданими характеристиками якості, не вдалося).

Таблиця 1. Порівняння стратегій придатності та оптимальності визначення алгоритму синтезу моделі за критерієм регулярності

Об'єкт дослідження	Значення критерію регулярності R при застосуванні алгоритму придатного відбору, випадків/10 000 осіб	Значення критерію регулярності R при застосуванні стратегії оптимальності, випадків/10 000 осіб	
		Алгоритм повного перебору	МАКАСМ
Захворювання органів дихання	130,8	105,8	120,8
Захворювання крові	120,7	110,9	110,9
Захворювання шлунку	–	–	–
Захворювання шкіри	–	–	–
Захворювання зору	98,7	98,7	98,7
Захворювання на бронхіт	136	124,6	124,6
Захворювання на цукровий діабет	140,2	140,2	140,2
Захворювання на туберкульоз	133,1	97,5	128,9
Захворювання нервової системи	106,5	106,5	106,5
Захворювання ендокринної системи	–	–	–
Середнє значення	123,71	112,03	118,66

Відповідно до даних табл. 1, можна зазначити, що кожен із алгоритмів вибору, розроблений на основі стратегії оптимальності, забезпечує менше значення критерію регулярності порівняно із алгоритмом, розробленим за стратегією придатності. Це свідчить про вищу якість моделей, синтезованих за алгоритмом, вибраним відповідно до стратегії оптимальності. Порівняно із алгоритмом придатного відбору, якість моделей, синтезованих із застосуванням алгоритмів, що реалізують стратегію оптимальності, покращується від 4,1% до 9,4%. Значення критерію регулярності результатів моделювання, отримані за алгоритмом повного перебору та за МАКАСМ, співрозмірні.

В табл. 2 представлені часові характеристики синтезу моделей об'єктів соціоекологічного моніторингу при використанні різних стратегій вибору алгоритму моделювання.

Таблиця 2. Тривалість синтезу моделі за вибраним алгоритмом

Об'єкт дослідження	Час синтезу моделі τ при застосуванні алгоритму придатного відбору, с	Час синтезу моделі τ при застосуванні стратегії оптимальності, с	
		Алгоритм повного перебору	МАКАСМ
Захворювання органів дихання	15	78,5	3,5
Захворювання крові	15	78,5	3,5
Захворювання шлунку	78,5	78,5	3,1
Захворювання шкіри	78,5	78,5	4,1
Захворювання зору	12,8	78,5	3,4
Захворювання на бронхіт	33	78,5	3,3
Захворювання на цукровий діабет	3,5	78,5	4
Захворювання на туберкульоз	19,8	78,5	4,2
Захворювання нервової системи	61,6	78,5	3
Захворювання ендокринної системи	78,5	78,5	4,1
Всього	396,2	785	36,2

Дані табл. 2 доводять, що в порівнянні з алгоритмом придатного відбору загальна тривалість синтезу моделей із використанням повного перебору алгоритмів збільшилась в 1,98 раз. При використанні МАКАСМ час синтезу моделі зменшився в 10,94 рази.

Таким чином, використовуючи дані табл. 1 та 2, можна стверджувати, що використання МАКАСМ при побудові синтезатора в порівнянні з алгоритмом придатного відбору дозволяє покращити якість синтезованих моделей та оперативність забезпечення інформацією процесу прийняття рішень.

Оскільки для загальної характеристики порівнюваних алгоритмів вибору важливо

врахувати обидва показники якості, пропонується використовувати узагальнюючий критерій (4). При цьому критерію регулярності R результатів моделювання надамо значення вагового коефіцієнта в 9 раз вище за значення вагового коефіцієнта часу відгуку системи τ при зміні масиву вхідних даних.

$$E = \frac{1}{R^{k_1} \cdot \tau^{k_2}}, \quad (4)$$

де R – значення критерію регулярності, τ – час синтезу моделі, k_1, k_2 – вагові коефіцієнти критеріїв R і τ відповідно.

Результати застосування показника ефективності E при порівнянні стратегій придатності та оптимальності графічно представлені на рис. 3.



Рис. 3. Значення узагальненого показника ефективності при використанні різних стратегій вибору алгоритму синтезу моделей

Рис. 3 демонструє, що застосування нового МАКАСМ на основі класифікації вхідних даних у процесі синтезу моделі об'єкта моніторингу дозволяє отримати значно вищі показники якості цих моделей у порівнянні з алгоритмами повного перебору та придатного відбору.

5. Висновки

Таким чином, використання стратегії оптимальності при виборі алгоритмів синтезу моделей у системах соціоекологічного моніторингу забезпечує підвищення якості та оперативності перетворення інформації при зміні масиву вхідних даних.

Встановлено, що новий метод адаптивного конструювання алгоритмів синтезу моделей при зміні масиву вхідних даних у порівнянні з відомим методом придатного відбору забезпечує покращання критерію регулярності на 4,1% і знижує час відгуку системи в досліджуваних умовах у 10,94 раз.

Дослідження в цій галузі доцільно продовжувати для виявлення явного вигляду узагальнюючого критерію інформативності масиву вхідних даних з метою оцінки можливостей синтезатора перед його застосуванням.

СПИСОК ЛІТЕРАТУРИ

1. Голуб С.В. Методологія створення автоматизованих систем багаторівневого соціоекологічного моделювання: автореф. дис. на здобуття наук. ступеня доктора техн. наук / С.В. Голуб. – Черкаси: Видавництво ЧНУ, 2008. – 36 с.
2. Голуб С.В. Багаторівневе моделювання в технологіях моніторингу оточуючого середовища / Голуб С.В. – Черкаси: Вид. від. ЧНУ імені Богдана Хмельницького, 2007. – 220 с.
3. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем / Ивахненко А.Г. – Киев: Наукова думка, 1981. – 296 с.
4. Теория вероятностей и математическая статистика. Математические модели: учебное пособие для студ. высш. учеб. заведений / В.Д. Мятлев, Л.А. Панченко, Г.Ю. Ризниченко [и др.]. – М.: Издательский центр «Академия», 2009. – 320 с.
5. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д., Пилинский М., Рутковский Л. – М.: Горячая линия-Телеком, 2006. – 383 с.
6. Колос П.О. Умови конструювання алгоритмів синтезу моделей в системах багаторівневого перетворення інформації / П.О. Колос, С.В. Голуб // Математичні машини і системи. – 2009. – № 4. – С. 139 – 146.
7. Колос П.А. Применение режима „свой-чужой” при распознавании алгоритмов синтеза моделей в системах многоуровневого социэкологического мониторинга / П.А. Колос // Индуктивне моделювання складних систем: зб. наук. праць. – Київ: Міжнародний науково-навчальний центр інформаційних технологій та систем НАН та МОН України, 2009. – С. 74 – 83.
8. Голуб С.В. Моделювання об'єктів моніторингу довкілля за алгоритмом Степаненка / С.В. Голуб // Електроніка та системи управління. – 2006. – № 4 (10). – С. 165 – 168.

Стаття надійшла до редакції 08.09.2010