

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Таїф Мохамед Алі



УДК 004.48:004.94

**МОДЕЛІ ТА МЕТОДИ РЕВЕРСНОЇ ІНЖЕНЕРІЇ ГЕННИХ
РЕГУЛЯТОРНИХ МЕРЕЖ НА ОСНОВІ ГІБРИДНИХ ІМУННИХ
СИСТЕМ**

05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Херсон – 2021

Дисертацією є рукопис.

Робота виконана в Херсонському національному технічному університеті Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор
Литвиненко Володимир Іванович,
Херсонський національний технічний університет,
завідувач кафедри інформатики і комп'ютерних наук.

Офіційні опоненти: **Гнатушенко Вікторія Володимирівна**,
Національна металургійна академія МОН України,
завідувач кафедри інформаційних технологій і систем;
доктор технічних наук, професор;

Гожий Олександр Петрович,
Чорноморський національний університет імені
Петра Могили МОН України, доктор технічних наук,
професор кафедри інтелектуальних інформаційних
систем.

Захист відбудеться «14» травня 2021 року о 13⁰⁰ годині на засіданні спеціалізованої вченої ради Д 67.052.01 у Херсонському національному технічному університеті за адресою: Україна, 73008, м. Херсон, Бериславське шосе, 24, корп. 3, ауд. 320.

З дисертацією можна ознайомитись у бібліотеці Херсонського національного технічного університету за адресою: Україна, 73008, м. Херсон, Бериславське шосе, 24, корпус 1.

Автореферат розісланий «14» квітня 2021 р.

Вчений секретар
спеціалізованої вченої ради



Н. О. Козуб

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Останні роки характеризуються значним науковим досягненням в галузі молекулярної біології — розшифрування геному людини. Постгеномна ера відзначилася, інтенсивним розвитком високоефективних експериментальних методик вивчення динамічних профілів організмів, побудови карт білок-білкових, ДНК-білкових і інших класів взаємодій. Розроблені ДНК-чипові технології дозволяють вивчати динаміку експресії тисяч генів одночасно. Розробка нових методів привела до стрімкого зростання обсягів експериментальних даних, одержуваних в молекулярній біології й генетиці, які не можна характеризувати інакше як інформаційний «вибух».

Серед закордонних учених, які зробили значний внесок у становлення й розвиток методологічних і теоретичних засад реконструкції генних регуляторних мереж (ГРМ) слід назвати: Ф. Е. Давідсон [E. Davidson], Х. Кітано [H. Kitano], Т. Чен [T. Chen], Г. Чурч [G. Church], Х. Де Лонг [H. De Jong], Т. Счлит [T. Schlitt], М.А.Саваго [M.A.Savageau], М. Счена [M. Schena], М. Счермер [M. Schermer] та інші. У галузі передобробки профілів експресій генів слід назвати: Б. Болстад [B. Bolstad], Р. Иризарри [R. Irizarry], З. Чен [Z. Chen] та інші. В Україні у розвиток теоретичних і прикладних аспектів системно-аналітичних досліджень і інформаційних технологій для обробки складних даних зробили десятки видатних дослідників, серед яких варто назвати таких учених, як: С.А. Бабічев, П.І. Бідюк, Є.В. Бодянський, В.В. Дмитренко, Ю.П. Зайченко, М.З. Згуровський, М.В. Коробчинський, Н.Д. Панкратова, В.С. Степашко, В.І. Литвиненко, О.І. Корнелюк, М.Ю. Оболенська та інші.

На разі використовують близько 10 підходів до моделювання генних мереж, серед них машинне навчання, баєсові мережі, булеві мережі, диференційні рівняння, теорія інформації, мережі Петрі, нейронні мережі, генетичні алгоритми. Кожен з цих підходів, звісно, має свої переваги та недоліки, визначення яких ускладнене недостатньою кількістю ґрунтовних оглядів у науковій літературі. Іншою проблемою є те, що згадані підходи застосовують для реконструкції невеликих мереж, що нараховують лише 10–20 генів. Зі збільшенням кількості генів обчислювальна складність зростає експоненційно. Однак задача побудови генних мереж все одно є NP-повною, тому важливою складовою оглядів повинна бути оцінка часової складності алгоритму реконструкції та аналіз алгоритмів. Таким чином, є актуальним вирішення науково - практичної задачі розробки гібридних моделей та інформаційних технологій для реконструкції генних регуляторних мереж; розробка технологій реконструкції моделей генних мереж; створення ефективних методів інтелектуального аналізу й опрацювання складних даних.

Зв'язок роботи з науковими програмами, планами, темами. Тематика розглянутих у роботі завдань лежить у площині загальнодержавних науково-технічних програм, що сформульовані в Законах України “Про наукову і науково-технічну діяльність”, “Про національну програму інформатизації”, а також відповідають планам найважливіших науково-технічних програм Міністерства освіти та науки України, зокрема: 6 – Інформатика, автоматизація

та приладобудування; 6.2.1 – Інтелектуалізація процесів прийняття рішень; 6.2.2 – Перспективні інформаційні технології і системи.

Дисертаційну роботу виконано в навчально-наукових установах України відповідно до таких науково-дослідних робіт: «Розробка гібридних нейро-фазі-імунних алгоритмів інформаційних систем та технологій для розв'язання задач в біоінформатиці та обчислювальній біології» ДР № 0116U002841 (2016 -2020 рр.); «Синтез гібридних еволюційних алгоритмів та методів для моделювання генних регуляторних мереж» номер державної реєстрації № 0116U002840 (2016 - 2020 рр.).

Мета і задачі дослідження.

Метою роботи є підвищення ефективності процесу ідентифікації динамічних генних регуляторних мереж за рахунок розробки нових моделей, методів для розв'язання задачі побудови гібридних імунних систем. Для досягнення вказаної мети розв'язувалися наступні взаємопов'язані задачі:

1. Аналіз існуючих методів і принципів реконструкції топології, моделей і динаміки генних регуляторних мереж з використанням гібридних алгоритмів.
2. Розробка індуктивної технології об'єктивної кластеризації комплексів профілів експресії генів на основі комплексного використання внутрішніх і зовнішніх критеріїв оцінки якості групування даних і комплексного критерію балансу, який враховує як характер розподілу даних в окремих задачах кластеризації, так і різницю у результатах кластеризації, отриманих із використанням рівнопотужних підмножин.
3. Розробка методів на основі гібридних імунних систем для прогнозування характеру причинно-наслідкових зв'язків генної регуляторної мережі.
4. Розробка гібридного підходу з використанням клонального алгоритму і вейвлет мереж для реконструкції генних регуляторних мереж
5. Розробка методів параметричної ідентифікації і реконструкції S-системи із застосуванням гібридного алгоритму клонального відбору.
6. Дослідження ефективності розроблених методів реверсної інженерії шляхом комп'ютерного моделювання.

Об'єктом дослідження є нелінійні нестационарні процеси в генних регуляторних мережах

Предмет дослідження – моделі та методи реверсної інженерії генних регуляторних мереж на основі гібридних імунних систем

Методи досліджень засновано на використанні положень теорії ймовірностей і математичної статистики, теорії графів, еволюційних алгоритмів, системи нелінійних диференціальних рівнянь, теорії нейронних мереж, теорії вейвлетів, методів проектування та реалізації інформаційних систем автоматизації наукових досліджень.

Наукова новизна одержаних результатів.

Вперше:

– Розроблено гібридний метод реконструкції правій частини системи звичайних диференціальних рівнянь, який використовується для опису динаміки генних регуляторних мереж, де для обчислювальної моделі обрана вейвлет-

нейронна мережа, яка навчається за допомогою алгоритму клональної селекції. Запропонований метод дозволяє отримувати оптимальні структури систем звичайних диференціальних рівнянь з мінімальною середньо квадратичною похибкою у вигляді S-систем.

– Розроблено метод трансформації, що дозволяє послідовно трансформувати простір рішень при реконструкції генних регуляторних мереж на основі даних часових рядів профілів експресії генів для пошуку взаємозв'язків між компонентами ГРМ. Запропонований метод дозволяє збільшити швидкість збіжності алгоритмів оптимізації і одночасно підвищити їх точність при вирішенні задачі параметричної ідентифікації S-системи.

Удосконалено:

– Методи гібридизації алгоритмів клонального відбору та диференційної еволюції при реконструкції генних регуляторних мереж. Запропонований підхід дозволяє збільшити збіжність і точність алгоритму оптимізації при розв'язанні задачі ідентифікації S-системи.

Отримали подальший розвиток:

– Процедура реконструкції генної регуляторної мережі завдяки використанню алгоритму клонального відбору та тригонометричної диференційної еволюції. Запропоновані методи дозволяють отримувати моделі генних регуляторних мереж з мінімальним значенням середньоквадратичної похибки, збільшити збіжність і точність алгоритму оптимізації при розв'язанні задачі ідентифікації S-системи.

– Індуктивні методи щільнісної кластеризації за рахунок об'єднання і використання двокрокових щільнісних алгоритмів, що дозволяє видаляти неінформативні гени при обробці даних мікрочіпів ДНК.

Практичне значення одержаних результатів.

Результати дисертаційного дослідження полягають в тому, що розроблені методи, алгоритми та інструментальні засоби обробки профілів експресії генів для реконструкції динамічних генних мереж, дають змогу: будувати високоефективні системи реконструкції динамічних генних регуляторних мереж за рахунок коректного використання при попередній обробці комбінованих щільнісних алгоритмів кластеризації в рамках індуктивної технології об'єктивної кластеризації з видаленням неінформативних генів, а також за рахунок обґрунтованого визначення оптимальної комбінації методів обробки даних.

Результати дисертаційної роботи впроваджені у ТДВ «Херсонський маслозавод», м. Херсон (акт від 16.01.2021 р.), для визначення якості продукції та її кластеризації у відповідності з результатами лабораторних досліджень шляхом застосування індуктивної технології об'єктивної кластеризації та попередньої обробки даних; у навчальному процесі на кафедрі програмного забезпечення систем Ужгородського національного університету, м. Ужгород (акт від 10.01.2021 р.), та на кафедрі інформатики і комп'ютерних наук Херсонського національного технічного університету, м. Херсон (акт від 15.01.2021 р.) при викладанні дисциплін «Організація баз даних і баз знань», «Інтелектуальний аналіз даних».

Особистий внесок здобувача. У роботах опублікованих у співавторстві автору дисертації особисто належать основні ідеї, теоретична та практична розробка положень, які відображені в характеристиці наукової новизни отриманих результатів. Праці [21,24] є одноосібними. У публікаціях, написаних у співавторстві, здобувачеві належать: [14, 23, 25, 26] – аналіз інформаційних технологій і систем для розв’язання задач реконструкції генних регуляторних систем; [2,3,5,6,12,13, 15, 16, 17, 19] методи та алгоритми по розробці індуктивної технології об’єктивної кластеризації профілів експресії генів; [1,4, ,22,24] – розробка і практична реалізація гібридних імунних методів реконструкції генних мереж; [7,9,20] - Розробка методів клонального відбору та тригонометричної і диференційної еволюції для розв’язання задачі ідентифікації S-системи; [8] – комплексний метод для виведення генної регуляторної мережі використовується вейвлет нейронна мережа структурно-параметричний синтез якої здійснюється алгоритмом клонального добру; [10,11]—для розв’язання задачі параметричної ідентифікації S-системи використано метод послідовна трансформація простору рішень з використанням алгоритму клонального відбору.

Апробація результатів дисертаційної роботи.

Основні наукові та практичні результати дисертаційної роботи оприлюднені й обговорені на міжнародних і національних науково-практичних конференціях та семінарах, зокрема: The 1th IEEE International Conference on Data Stream Mining & Processing. Lviv, Ukraine, 2016; IEEE 37th International Conference on Electronics and Nanotechnology(ELNANO). Kiev, 2017; Workshop Proceedings of the 8th International Conference on “Mathematics. Information Technologies. Education”, MoMLeT&DS-2019, Shatsk, Ukraine, June 2-4, 2019; Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, Ukraine, April 15-19, 2019, p. 193-210; 9-а міжнародна школа-семінар «Теорія прийняття рішень». Ужгород, УжНУ, 2016; 20-й Ювілейний Міжнародний Молодіжний Форум «Радіоелектроніка та молодь у XXI столітті» Харків 2016; XII Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (ISDMCI2016) 2016 Залізний Порт, Україна; IV Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених з автоматичного управління присвяченої дню космонавтики/ Херсон, 2016; I Всеукраїнська науково-практична конференція студентів, аспірантів і молодих учених інформаційні технології в моделюванні ІТМ-2016, Миколаїв; Міжнародної науково-практичної конференції «Інформаційні технології та взаємодії». Київ, 2016.

Публікації. За темою дисертаційної роботи опубліковано 26 наукових праць, з яких 11 публікації у фахових наукових виданнях України з технічних наук, 1 – у фахових наукових виданнях України, які індексуються у наукометричній базі Web of Science, 5 статті у закордонних фахових виданнях, які індексуються у наукометричній базі Scopus, 1 – стаття у закордонному фаховому виданні, 7 публікацій у збірниках матеріалів міжнародних і

національних конференцій, 1 свідоцтво про реєстрацію на твір, 2 публікацій є одноосібні.

Структура і обсяг дисертації. Дисертація складається з анотацій, вступу, 4 розділів, висновків, списку використаних джерел і додатків. Загальний обсяг дисертації становить 172 сторінки, у тому числі основного тексту 156 сторінок. Робота містить 43 рисунків, 28 таблиць і 4 додатки. Список використаних джерел включає 172 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дослідження, сформульовано мету та задачі дослідження, наведено відомості про зв'язок роботи з науковими програмами, планами, темами, розкрито наукову новизну та практичне значення одержаних результатів, наведено дані про апробацію результатів дисертаційної роботи.

У **першому розділі** проведено системний аналіз існуючих технологій, моделей і методів реконструкції генних регуляторних мереж. ГРМ знаходять численні застосування не тільки в біології, але і в медицині, серед яких можна виділити: виявлення генетичних захворювань, розробку нових лікарських препаратів, зменшення побічних ефектів від лікування, формування гіпотез про функціональне призначення невідомих генів і т.п. Проведено аналіз переваг та недоліків існуючих методів реверсної інженерії генних регуляторних мереж. Проведено огляд існуючих вітчизняних та закордонних публікацій.

Проведений аналіз спеціальних літературних джерел показав, що моделями, найбільш точно і повно описують структуру і поведінку ГРМ, вважаються системи диференціальних рівнянь (СДР), які обрані для подальшого аналізу в представленому дослідженні. Зазначені факти розкривають наукову проблему, окремі задачі і напрямки дисертаційного дослідження, пов'язаного з розробкою і вдосконаленням методів оцінювання структури і параметрів моделей регуляторних мереж і застосуванні даних оцінок для подальшого виявлення зв'язків і взаємодії між генами.

Другий розділ роботи присвячено розробці гібридного двох крокового алгоритму об'єктної кластеризації на основі щільнісних алгоритмів DBSCAN та OPTICS. Інтеграція щільнісних алгоритмів кластеризації з індуктивними методами дають можливість при неточних даних, використовуючи мінімум обраного критерію, знаходити вирішальне правило, точність якої менше ніж структура повної фізичної моделі. Основна ідея полягає в тому, щоб об'єднати індуктивний щільнісний DBSCAN і OPTICS алгоритм наведено на рисунку 1. Для розрахунку критерію, вибірка даних ділиться на дві рівні частини. На кожній підвибірці будується дерево кластеризації і на кожному кроці розраховується критерій балансу при однаковому числі кластерів:

$$BL = \frac{1}{MK} \sum_{j=1}^M \sum_{i=1}^K (x_{oA} - x_{oB})^2 \rightarrow \min, \quad (1)$$

де: K - кількість кластерів на даному кроці побудови дерева; M - число координат; x_{OA} - координати центрів кластерів, побудованих на частині А; x_{OB} - координати центрів кластерів, побудованих на частині В.

Індуктивна технологія
об'єктивної кластеризації

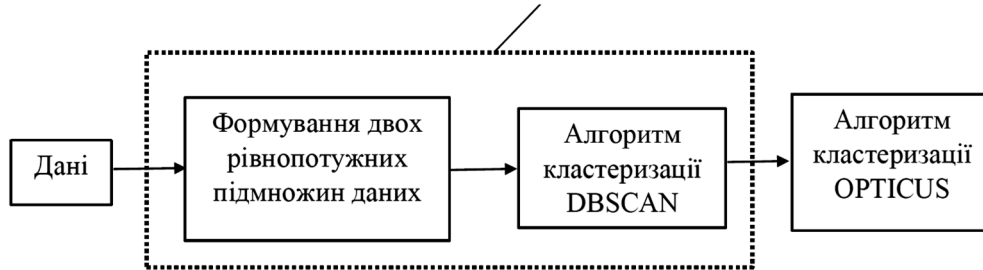


Рис. 1. Структурна схема двох крокового індуктивного методу кластеризації з використанням алгоритмів DBSCAN і OPTICUS

Запропонована процедура кластеризації передбачає наступні етапи:

Крок 1. Старт.

Крок 2. Формування вихідної множини Ω досліджуваних об'єктів. Представлення даних у вигляді матриці $n \times m$ де n – кількість строк або кількість досліджуваних об'єктів, m – кількість стовпців або кількість ознак, що характеризують об'єкти.

Крок 3. Розподіл Ω на два рівнопотужних підмножини відповідно до запропонованого алгоритмом. отримані підмножини Ω^A та Ω^B формально можна представити наступним чином:

$$\Omega^A = \{x_{ij}^A\}, \Omega^B = \{x_{ij}^B\}, j = 1, \dots, m$$

$$i = 1, \dots, n_A = n_B, n_A + n_B = n$$

Крок 4. Налаштування алгоритму кластеризації DBSCAN.

Для кожної підмножини:

Крок 5. Кластеризація даних на підмножині алгоритмом DBSCAN.

Крок 6. Фіксація результатів кластеризації при

$$eps \in [eps_{min}, eps_{max}], \Delta eps = 0,001$$

$$minPts \in [minPts_{min}, minPts_{max}], \Delta minPts = 1$$

Крок 7. Розрахунок внутрішніх критеріїв якості кластеризації для кожного результату кластеризації.

Крок 8. Розрахунок зовнішнього критерію балансу

Крок 9. Якщо модуль різниці значень зовнішніх критеріїв балансу при однаковій кількості кластерів на кожному підмножині не досягає мінімального значення, то повторюються Кроки 7-8.

інакше:

Крок 10. Фіксується кластеризація алгоритмом DBSCAN на множині Ω :

Крок 11. Налаштування алгоритму кластеризації OPTICUS.

Крок 12. Визначення кластерів даних з різною щільністю.

Крок 13. Фіксація результатів кластеризації алгоритмом OPTICUS.

Крок 14. Кінець.

Для проведення експерименту були відібрані тестові дані з бази даних обчислювальної школи східно-фінського університету [<https://cs.joensuu.fi/sipu/datasets/>], що мають складаний двовимірну просторову форму. Результати кластеризації представлені в таблиці 1:

Таблиця 1. Результати кластеризації (відсоток правильно розпізнаних даних)

Дані	Індуктивний DBSCAN+OPTICS	Індуктивний DBSCAN	DBSCAN	OPTICS
Aggregation	98,5	93,0	87,5	90,1
D31	98,1	93,3	83,4	89,2
Flame	95,7	89,8	79,6	85,4
Jain	98,8	96,2	77,0	92,4
Pathbased	98,2	96,5	81,5	83,2
R15	99,2	97,7	89,2	94,1

В **третьому розділі** розроблено гібридний метод реконструкції генної регуляторної мережі системи звичайних диференціальних рівнянь, які використовуються для опису динаміки генних мереж.

Задача реконструкції зводиться до параметричної оптимізації вибраних виразів. Недоліком фіксованого запису рівнянь є відсутність гнучкості моделі, що обмежує область її застосування. У даній роботі, для усунення зазначеного недоліку, використано систему звичайних диференціальних рівнянь (ЗДР) у вигляді S-систем. Вона нелінійна і тому здатна дуже точно відтворити поведінку генної мережі, а її структура дозволяє однозначно перетворити математичну форму запису в топологію, яку традиційно представляють у вигляді зваженого орієнтованого графа. Вершини графа відповідають генам мережі, а дуги визначають зв'язку між цими генами.

S-система є різновидом системи ЗДР такого вигляду:

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^N x_j^{g_{ij}} - \beta_i \prod_{j=1}^N x_j^{h_{ij}}, \quad (2)$$

де N – кількість генів в генній регуляторній мережі; $x_i(t)$ – змінна, що виражає зміна рівня експресії i -го гена у часі; показники ступеня g_{ij} та h_{ij} обумовлюють напрямок впливу і силу зв'язку гена x_j з геном x_i ; α_i, β_i – коефіцієнти, які беруть невід'ємні значення. З точки зору біохімічної інженерії, параметри α_i, β_i є константами швидкості, а g_{ij}, h_{ij} – кінетичними порядками.

Позитивні значення g_{ij} , свідчать про активуючий вплив гена j на ген i . Графічно це можна зобразити у вигляді дуги, спрямованої від j до i . Ідентифікація S-системи передбачає пошук оптимальних значень параметрів з множини $v = \{\alpha, \beta, g, h\}$. Для мережі, що складається з N генів, кількість параметрів, які необхідно знайти, визначається виразом $2N(N + 1)$.

Концептуальна схема запропонованого методу реконструкції генної регуляторної мережі наведений на рисунку 2.

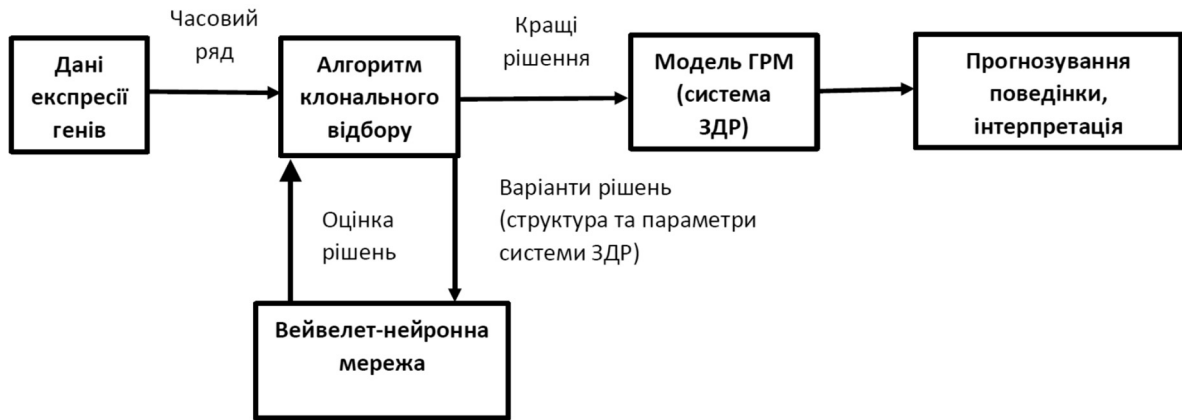


Рис. 2. Концептуальна схема реконструкції генної регуляторної мережі

В загальному вигляді, формально процес клонального відбору виглядає таким чином:

1. Алгоритм клонального відбору. У задачах оптимізації значення афінності обчислюються на підставі значень критеріїв, відображених у множину невід'ємних чисел, тобто .:

$$f : X \rightarrow \mathfrak{R}, \quad F : \mathfrak{R} \rightarrow \mathfrak{R}^+ . \quad (3)$$

Таким чином, має місце деяка функція афінності $g = F(f(x_1, x_2, \dots, x_n))$, яка визначає ступінь відповідності індивідуумів один одному. Тому управління динамікою алгоритму клонального відбору проводиться з урахуванням відносних значень афінності або рангу індивідуумів популяції.

Псевдокод алгоритму клонального відбору наведений на рис. 3.

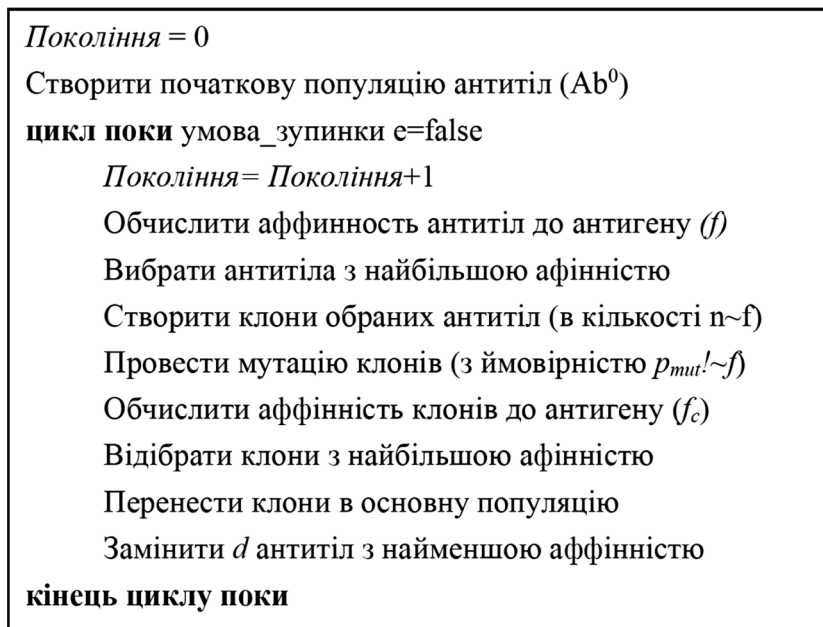


Рис. 3. Псевдокод алгоритму клонального відбору

2. Вейвлет-нейронні мережі. При вирішенні завдань синтезу і навчання вейвлет-нейронних мереж, основна увага приділяється налаштування параметрів вейвлет-нейронів, розташованих в прихованому шарі мережі, що, по суті, є

завданням оптимізації. У даній роботі пропонується використовувати алгоритм клонального відбору для оптимізації параметрів вейвлет-нейронної мережі в процесі навчання останньої. Архітектура вейвлет-нейронної мережі представлена на рис. 4.

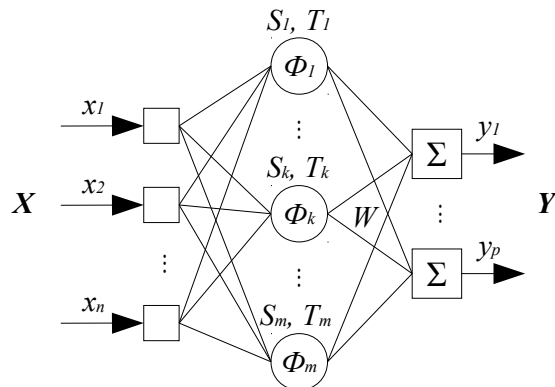


Рис. 4. Узагальнена схема вейвлет-нейронної мережі

Нейрони прихованого шару - це вейвлет-нейрони (Φ), що містять в собі налаштовуваний вейвлет. Як параметри вейвлета використовується його масштаб (s) і зсування (t) за часовою віссю. Вейвлет-нейронні мережі використовують сімейства вейвлетів, утворених від одного материнського за допомогою застосування до нього операцій масштабування і зсуву. Ґрунтуючись на представленій архітектурі, вихід нейронної мережі буде визначатися таким чином:

$$y_j = \sum_{k=1}^m w_{kj} \Phi_k(X, S_k, T_k), j = 1, 2, \dots, p, \quad (4)$$

де $X = (x_1, x_2, \dots, x_n)$ – довільний n -мірний вектор вхідних змінних; $w_{kj} \in W$ – ваги лінійного шару; $S_k = (s_{1k}, s_{2k}, \dots, s_{nk})$ та $T_k = (t_{1k}, t_{2k}, \dots, t_{nk})$ відповідно, вектори масштабу і зсуву k -го вейвлета прихованого шару мережі щодо вектора вхідних змінних; p – кількість виходів нейронної мережі.

Метод реконструкції системи ЗДР. В даному випадку права частина системи ЗДР представлена нейронною мережею. Це означає, що на вхід мережі надходять часові ряди зміни концентрації продуктів експресії генів, а вихід мережі відповідає вектору перших похідних концентрацій тих же продуктів (рис. 5).

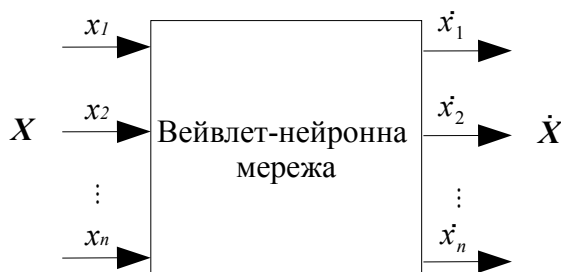


Рис. 5. Схема вейвлет-нейронної мережі для реконструкції системи ЗДР

Побудувавши цю мережу, ми зможемо знайти форму рівнянь, яка найкраще відповідає даним експресії. Для оптимізації обрано такі параметри мережі: кількість нейронів прихованого шару (m); типи вейвлетів прихованого шару (ϕ);

значення зсуву для нейронів прихованого шару (t_{ik}); значення масштабу для нейронів прихованого шару (s_{ik}); ваги вихідного шару (w_{kj}); параметри функції активації вихідного шару (a_j). Виходячи з переліку виділених параметрів, отримана наступна структура індивідуума (рис. 6). Для кодування значень параметрів ми використовували двійкову систему. Елементи рядка b_1, \dots, b_m кодують стану відповідних вейвлет-нейронів прихованого шару. Значення «0» відповідає виключеному станом (нейрон не бере участі в розрахунку вихідного значення нейронної мережі). Значення «1» - нейрон включений.

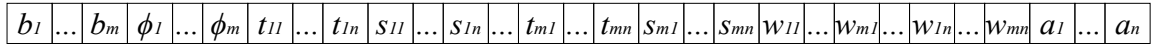


Рис. 6 Структура антитіла, що кодує вейвлет-нейронну мережу

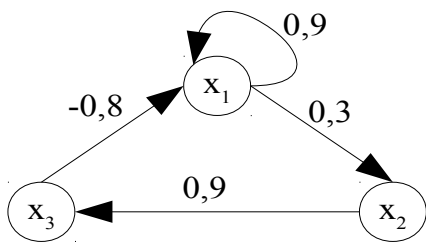
Елементи ϕ_1, \dots, ϕ_m відповідають типу вейвлета, який заданий індексом в масиві материнських вейвлетів. Навчальний алгоритм має можливість вибирати будь-який з вейвлетів як нейрона прихованого шару, внаслідок чого в рамках однієї нейронної мережі можуть використовуватися представники різних типів.

Для розрахунку афінності індивідуумів вибрано такий вираз:

$$f = \sqrt{\frac{1}{n \cdot T} \sum_{i=1}^n \sum_{j=0}^{T-1} \left(x_i^M(t_0 + j\Delta t) - x_i(t_0 + j\Delta t) \right)^2}, \quad (5)$$

де t_0 – початковий час; Δt - часовий крок; n - кількість компонентів (генів) в ГРМ; T - кількість даних часового ряду експресії; $x_i^M(t_0 + j\Delta t)$ - значення концентрації, отримані розв'язанням системи ЗДР, представленої вейвлет-нейронною мережею; $x_i(t_0 + j\Delta t)$ значення концентрації із часового ряду експресії.

Для проведення експериментів було згенеровано по три набори даних часового ряду з різними початковими умовами. У даній моделі функції SLOG2, POLYWOG1, POLYWOG2, RASP1 - материнські вейвлети. У цій мережі вузли графа є генами, а дуги показують напрямки регуляторних впливів і їх рівень (рис. 7). Позитивна вага дуги говорить про активуючий вплив, негативна вага - про інгібаторний вплив.



Матриця зв'язності

$$V = \begin{matrix} & \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix} & \begin{bmatrix} 0,9 & 0 & -0,8 \\ 0,3 & 0 & 0 \\ 0 & 0,9 & 0 \end{bmatrix} \end{matrix}$$

Рис. 7. Модель ГРМ у вигляді спрямованого зваженого графа

На основі даної мережі, побудовані часові ряди, в яких рівень експресії i -го гена $x_i(t + 1)$ обчислюється по наступних співвідношеннях:

$$x_i(t + 1) = \frac{m_i}{1 + e^{-s_i(t)}}, \quad (7)$$

$$s_i(t) = \sum_{j=0}^n w_{ij} x_j(t), \quad (8)$$

де m_i – максимальний рівень експресії i -го гена (для нашого експерименту ми взяли $m_1 = 4.0$, $m_2 = 5.0$, $m_3 = 5.0$); n – кількість генів у регуляторній мережі; w_{ij} – рівень впливу гена j на ген i . Рівень впливу задається за допомогою матриці зв'язності, яку для нашого прикладу показано на рис. 7.

Середньо квадратичні помилки навчання на кожному наборі даних склали відповідно: $f_1 = 0.059$, $f_2 = 0.05$, $f_3 = 0.063$. Графік часових рядів кращої моделі, синтезованої в другому експерименті з початковими умовами $(0.5, 0.0, 0.0)$, показано на рисунку 8.

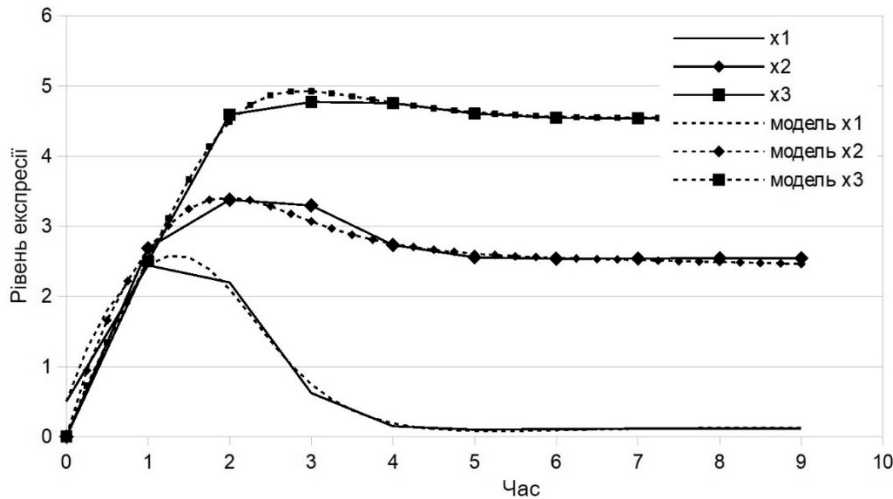


Рис. 8. Часові ряди даних з початковими умовами $(0.5, 0.0, 0.0)$

Динамічна модель, що відповідає часовим рядам, наведеним на рисунку 8, має вигляд:

$$\begin{cases} \dot{x}_1 = 11.121\Phi_1 - 9.88\Phi_2 - 6.466\Phi_3 - 6.673\Phi_4 - 13.19\Phi_5 + 6.983\Phi_6 \\ \dot{x}_2 = 1.883\Phi_1 - 6.924\Phi_2 - 7.098\Phi_3 + 3.679\Phi_4 + 0.029\Phi_5 + 4.259\Phi_6 \\ \dot{x}_3 = -5.657\Phi_1 - 12.265\Phi_2 - 6.758\Phi_3 + 10.463\Phi_4 - 9.662\Phi_5 + 9.562\Phi_6 \end{cases}, \quad (9)$$

У **четвертому розділі** розглядаються методи модифікації та гібридизації алгоритма клонального відбору для параметричної ідентифікації та реконструкції S-систем. Основна ідея методу параметричної ідентифікації S-системи з використанням модифікованого алгоритму клонального відбору полягає в процесі оптимізації, періодично, при досягненні певних умов, трансформувати простір пошуку за допомогою перерахунку проміжків можливих значень параметрів S-системи з їх одночасним центруванням щодо виявленого поточного кращого рішення. Позначимо через v будь-який з параметрів S-системи α, β, g або h . Для пошуку оптимального необхідно щоб v було поставлено в проміжку. $\{v \in: a \vee b\}$ Кінці проміжку і залежать від типу параметра і встановлюються на підставі попередніх знань про предметну область перед початком процесу оптимізації. Суть запропонованої процедури полягає в тому щоб в процесі оптимізації, періодично, при досягненні певних умов, трансформувати простір пошуку за допомогою перерахунку проміжків можливих значень v з їх одночасним центруванням щодо виявленого поточного кращого рішення

Формально процес трансформації простору рішень можна представити таким чином:

$$a^{j+1} = v_p^j \frac{s^{j+1}(b^0 - a^0)}{2}; \quad b^{j+1} = v_p^j + \frac{s^{j+1}(b^0 - a^0)}{2}, \quad (10)$$

де a^{j+1}, b^{j+1} – кінці проміжку значень параметра v для наступної трансформації; a^0, b^0 – початковий проміжок значень параметра v S-системи; v_p^j – краще значення параметра v S-системи, отримане за допомогою алгоритму оптимізації в поточній (j -тій) ітерації; s^{j+1} – коефіцієнт масштабування проміжку значень параметра v для наступної трансформації, який розраховується за формулою:

$$s^{j+1} = \begin{cases} s^j k_g, \text{ if } (v_p^j a^j) \leq \varepsilon \text{ або } (b^j v_p^j) \leq \varepsilon \\ s^j k_s, \text{ в протилежному випадку} \end{cases}, \quad (11)$$

де s^j – коефіцієнт масштабування проміжку значень параметра v в поточній (j -тій) ітерації; k_g ($k_g > 1$) – коефіцієнт розширення проміжку; k_s ($0 < k_s < 1$) – коефіцієнт стиснення проміжку; a^j, b^j – кінці проміжку значень параметра v в поточній (j -тій) ітерації; ε – порогове значення, яке фіксує факт збігу параметра v з лівим або правим кінцем проміжку.

Для отримання міри афінності антитіл до антигену обчислюється помилка моделі на часових рядах даних експресії генів. У даній роботі для розрахунку помилки застосовується такий вираз:

$$f = \text{MIN} \sum_{i=1}^N \sum_{j=1}^T \left(\frac{x_i^M(t_0 + j\Delta t) - x_i(t_0 + j\Delta t)}{x_i(t_0 + j\Delta t)} \right)^2, \quad (12)$$

де t_0 – початковий час; Δt – часовий крок; T – кількість даних часового ряду експресії; $x_i^M(t_0 + j\Delta t)$ – значення концентрації, отримані рішенням системи звичайних диференціальних рівнянь (2); $x_i(t_0 + j\Delta t)$ – спостережувані значення концентрації з часового ряду експресії.

Була проведена серія з десяти випробувань і відібраний найкращий результат ($f = 0.00112$). Графік часового ряду експресії, отриманий для кращої моделі, в порівнянні з оригінальними даними, показаний на рисунку 9.

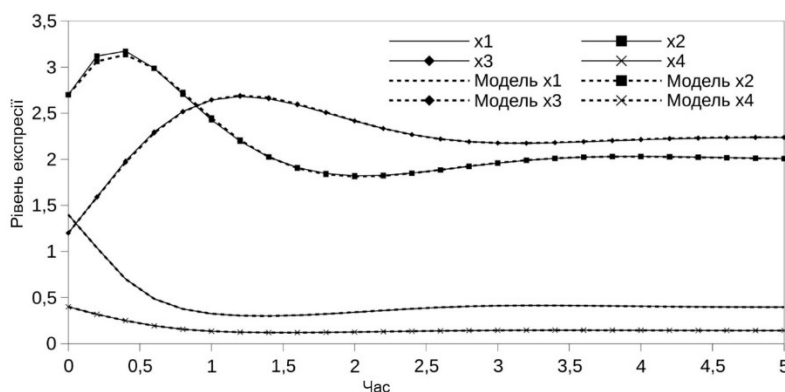


Рис. 9. Графік кращої моделі в серії з десяти випробувань

Для порівняння середніх показників обчислених параметрів S-системи розраховані їх стандартні відхилення. У таблиці 2 представлені порівняльні дані похибок моделі для серії з десяти випробувань. Для розрахунку середньої похибки використовувалася міра середнього арифметичного.

Таблиця 2. Похибки моделі при апроксимації часових рядів даних експресії генів

Похибка моделі (f)	Запропонований метод	Класичний клональний алгоритм
Мінімальна	0.001	0.03
Максимальна	0.009	0.021
Середня	0.005	0.6

З таблиці видно, що у всіх випадках похибки моделі демонструють переваги запропонованого підходу.

Запропоновано процедуру реконструкції S-системи за алгоритмом клонального відбору та диференційної еволюції. Основна відмінність алгоритму диференційної еволюції (ДЕ) від інших еволюційних алгоритмів полягає в реалізації оператора мутації з ДЕ наступним чином:

$$\mathbf{v}_i^{G+1} = \mathbf{x}_{r3}^G + F(\mathbf{x}_{r1}^G - \mathbf{x}_{r2}^G), \quad (13)$$

де $\mathbf{v}_i^{G+1}, i = 1, \dots, P$ — індивідуум, отриманий в результаті мутації; $r1, r2, r3 \in \{1, \dots, P\}$ — індекси індивідуумів, які вибираються випадково з популяції рішень в поточному поколінні такі, що $r1 \neq r2 \neq r3 \neq i$; F - масштабний коефіцієнт ($F \geq 0$).

Компоненти індивідуумів \mathbf{x}_i^G частково замінюються відповідними компонентами векторів \mathbf{v}_i^{G+1} з утворенням популяції кандидатів $\mathbf{u}_i^{G+1} = (u_{i1}^{G+1}, \dots, u_{in}^{G+1})$. Формування вектора \mathbf{u}_i^{G+1} відбувається за наступним виразом:

$$u_{ij}^{G+1} = \begin{cases} v_{ij}^{G+1}, & \text{якщо } randEvent(p_{DE}) = 1 \vee j = k \\ x_{ij}^G, & \text{в протилежному випадку} \end{cases}, j = 1, 2, \dots, n, \quad (14)$$

де $k \in \{1, \dots, n\}$ - випадковий індекс параметра, який обирається один раз для кожного індивідуума, сенс якого в тому щоб гарантувати перехід хоча б одного компонента вектору \mathbf{v}_i^{G+1} в вектор \mathbf{u}_i^{G+1} ; p_{DE} - ймовірність переходу j -того компонента вектора \mathbf{v}_i^{G+1} в вектор \mathbf{u}_i^{G+1} . Так як вираз (14) асоціюється з оператором кросинговеру в еволюційних алгоритмах, \mathbf{v}_i^{G+1} в вектор \mathbf{u}_i^{G+1} називають ймовірністю кросинговеру. В даній роботі p_{DE} названа інтенсивністю ДЕ-мутації.

Популяція нового покоління формується з популяції поточного покоління і популяції кандидатів за допомогою ДЕ-селекції:

$$\mathbf{x}_i^{G+1} = \begin{cases} \mathbf{u}_i^{G+1}, & \text{якщо } f(\mathbf{u}_i^{G+1}) \leq f(\mathbf{x}_i^G) \\ \mathbf{x}_i^G, & \text{в протилежному випадку} \end{cases}, \quad (15)$$

тобто кожен індивідуум з популяції кандидатів порівнюється з відповідним індивідуумом з поточної популяції. Якщо кандидат має меншим значенням цільової функції, він переходить в нове покоління. Інакше в нове покоління переходить поточний індивідуум.

В роботі запропонований гібридний алгоритм з використанням диференційної еволюції. В експериментах порівняли результати реконструкції ГРМ з роботою класичного і модифікованого алгоритму клонального відбору. Порівняння результатів роботи заснованих на клональній селекції, стосовно до задачі реконструкції ГРМ, що складається з чотирьох генів, представлено в таблиці 3.

Таблиця 3. Похибки моделі при дослідженні трьох методів, заснованих на клональній селекції

Похибка (f)	Досліджуваний гібридний алгоритм	Модифікований клональний алгоритм	Класичний клональний алгоритм
Мінімальна	$1.634 \cdot 10^{-10}$	0.00112	0.03
Максимальна	0.00068	0.0089	1.021
Середня	0.000272	0.00466	0.6

Модифіковані методи клонального відбору дозволяють скоротити час аналізу великих масивів даних з підвищенням точності моделі.

ВИСНОВКИ

У дисертаційній роботі наведено результати, які відповідно до поставленої мети є вирішенням науково-технічної задачі розробки гібридних імунних систем нових моделей та методів, які спрямовані на підвищення якості реконструкції та ідентифікації динамічних генних регуляторних мереж.

Основні наукові та практичні результати роботи полягають у наступному:

1. Виконано аналіз існуючих моделей та методів та моделювання генних регуляторних мереж з використанням диференціальних рівнянь. Проведено аналіз переваг та недоліків існуючих методів реверсної інженерії генних регуляторних мереж. Проведено огляд існуючих вітчизняних та закордонних публікацій. Сформульовано змістовий аналіз задачі досліджень.

2. Розроблено і практично реалізовано гібридний метод реконструкції системи правій частини системи звичайних диференціальних рівнянь, який використовується для опису динаміки генних регуляторних мереж, де для обчислювальної моделі обрана вейвлет-нейронна мережа, яка навчається за допомогою алгоритму клональної селекції. Запропонований метод дозволяє отримувати оптимальні структури систем звичайних диференціальних рівнянь з мінімальною середньо квадратичною похибкою у вигляді S-систем

3. Розроблено і досліджено метод послідовної трансформації простору рішень, що дозволяє послідовно трансформувати простору рішень при реконструкції генних регуляторних мереж на основі даних часових рядів профілів експресії генів для пошуку взаємозв'язків між компонентами генної регуляторної мережі. Запропонований метод дозволяє збільшити швидкість збіжності алгоритмів оптимізації і одночасно підвищити їх точність при вирішенні задачі параметричної ідентифікації S-системи.

4. Розроблено і досліджено методи гібридизації алгоритмів клонального відбору, та диференціальної еволюції при реконструкції генних регуляторних мереж. Запропонований підхід дозволяє збільшити збіжність і точність алгоритму оптимізації при розв'язанні задачі ідентифікації S-системи.

5. Розроблено і досліджено методи реконструкції генної регуляторної мережі алгоритмом клонального відбору та тригонометричної диференціальної еволюції. Запропоновані методи дозволяють отримувати моделі генних регуляторних мереж з мінімальним значенням середньоквадратичної похибки збільшити збіжність і точність алгоритму оптимізації при розв'язанні задачі ідентифікації S-системи.

6. Розроблені і досліджені методи індуктивної щільнісної кластеризації за рахунок об'єднання і використання двокрокових щільнісних алгоритмів, що дозволяє видаляти неінформативні гени при обробці даних мікрочіпів ДНК.

Використання результатів дисертаційної роботи дозволяє: будувати високоефективні системи реконструкції генних регуляторних систем; забезпечує високу точність і збіжність розроблених алгоритмів; зменшує час розв'язання задач реконструкції динамічних генних регуляторних мереж

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Публікації в наукових фахових виданнях України:

1. Lytvynenko V.I., Fefelov A.A., **Taif M.A.**, Lurie I.A. A Multi-Objective Immune Approach to Reconstruct Gene Regulatory Network Using Algorithm Clonal Selection. Системные технологии.- Днепропетровск. - 2015, № 6(101).- С. 67-84.
Здобувачем запропоновано використання багатоцільовий імунного підходу для реконструкції генної регуляторної мережі.
2. Babichev S., Osypenko V., **Taif M.A.**, Lytvynenko V. The using of biclustering techniques in inductive modelling systems of biological processes. Індуктивне моделювання складних систем. К.: НАН України, 2015. Випуск 7. С. 5–14.
Здобувачем запропоновано використання методів бікластеризації в системах індуктивного моделювання біологічних процесів
3. Лур'є І. А, **Таїф М.А.**, Осипенко В.В, Литвиненко В.І, Корніловська Н.В. Гібридизація алгоритму індуктивного кластер-аналізу з використанням оцінки щільності розподілу даних. Вісник Національного Університету “Львівська Політехніка” Збірник Наукових Праць No 832 Серія: Інформаційні Системи Та Мережі С. 178-190
Здобувачем запропоновано індуктивний алгоритм кластеризації з використанням оцінки щільності розподілу даних
4. Lytvynenko V.I, Fefelov A.A., **Taif M.A.**, Lurie I.A, A Multi-Objective Immune Approach to Reconstruct Gene Regulatory Network Using Algorithm Clonal Selection Системні технології. Міжвузівський збірник праць №6 (99). – Дніпропетровск, 2015. С. 61–74.
Здобувачем запропоновано використання комбінованого алгоритму клонального відбору для виведення генних регуляторних мереж з даних часових рядів

5. Babichev S., **Taif M.A.**, Lytvynenko V., Sharko A. Hybrid model of inductive clustering system of high-dimensional data based on the SOTA algorithm. Вісник Національного університету "Львівська політехніка". Серія: Комп'ютерні науки та інформаційні технології: збірник наукових праць, 2016. № 843. С. 173–179.
Здобувачем запропонована гібридна модель індуктивної кластерної системи основи алгоритму SOTA.
6. Бабічев С.А., **Таїф М.А.**, Литвиненко В.І., Фефелов А.О. Оцінка якості обробки складних даних біологічної природи на основі критеріїв ентропії. Штучний інтелект. К.: НАН України, 2016. № 2(72). С. 7–17.
Здобувачем запропоновано використання критеріїв ентропії при оцінці якості роботи алгоритмів кластеризації.
7. Fefelov A.A., **Taif M.A.**, Lytvynenko V.I., Voronenko M.A., Stepanchenko O.M., Sokur I.V., Lurie I A. Reverse engineering of the gene regulatory network with combined use of the clonal selection algorithm and differential evolution // Системні технології (2017) № 6 С. 74-84.
Здобувачем запропоновано комбіноване використання алгоритму клонального відбору і диференціальної еволюції для реверсної інженерії для реконструкції генних мереж по даним мікромасивів.
8. Фефелов А.А., Литвиненко В.И., **Таиф М.А.**, Лурье И.А. Гибридный подход при реконструкции генных регуляторных сетей // Управляющие системы и машины. — 2017. — № 3. — С. 63-72.
Здобувачем запропоновано гібридний метод реконструкції генної регуляторної мережі у часових рядах даних експресії генів, в якості моделі запропоновано використати систему звичайних диференціальних рівнянь.
9. Фефелов А.А., **Таиф М.А.**, Литвиненко В.И., Вороненко М.А. Реконструкция S-системы гибридным алгоритмом клонального отбора и дифференциальной эволюции // Управляющие системы и машины -2017. - № 6 - С. 41-51.
Здобувачем запропоновано використання моделі звичайних диференціальних рівнянь для моделювання генно-регуляторних мереж.
10. Фефелов А.А., **Таиф М.А.**, Литвиненко В.И., Лурье И.А. Параметрическая идентификация S-системы с применением модифицированного алгоритма клонального отбора // Управляющие системы и машины – 2017. - №5 - С. 43-53.
Здобувачем запропоновано метод реконструкції генних регуляторних мереж, який в поєднанні з імунним алгоритмом дозволяє підвищити швидкість і точність розв'язання задач параметричної ідентифікації S-систем.
11. Фефелов А. А., Литвиненко В. И., **Таиф М.А.**, Вороненко М. А. Объектно-ориентированная архитектура информационной системы реконструкции генных регуляторных сетей // Управляющие системы и машины. - 2017. - № 4. - С. 67–75, 82
Здобувачем запропоновано сценарій реконструкції, при якій здійснюється оптимізація просторів пошуку значених параметрів S-систем.

Публікації у наукових виданнях інших держав та у виданнях України, що включені до міжнародних наукометричних баз, зокрема SCOPUS, WoS та ін.:

12. Babichev S., **Taif M.A.**, Lytvynenko V. Inductive model of data clustering based on the agglomerative hierarchical clustering algorithm. Materials of The 1th IEEE International Conference on Data Stream Mining & Processing. Lviv, Ukraine, 2016. P.19–22.
Здобувачем запропоновано індуктивна модель кластеризації даних на основі агломеративного ієрархічного алгоритму кластеризації
13. Babichev S., **Taif M.A.**, Lytvynenko V. Estimation of the inductive model of objects clustering stability based on the k-means algorithm for different levels of data noise. Радиоелектроника, информатика, управление. Запорожье: НАН України, 2016. №4. С.54–60.
Здобувачем запропоновано методи оцінки індуктивної моделі стійкості кластеризації об'єктів на основі алгоритму k-середніх для різних рівнів шуму даних.
14. Babichev S., **Taif M.A.**, Lytvynenko V. Filtration of Nucleotide DNA Gene Expression Profiles. International frontier science letters, 2016. Vol. 8. P. 1–8.
Здобувачем представлені результати дослідження з оптимізації процесу фільтрації профілів експресії генів нуклеотидів ДНК. Оцінку якості обробки інформації здобувач проводив шляхом порівняльного аналізу результатів кластеризації оброблених та необроблених даних.
15. Babichev S., Lytvynenko V., **Taif M.A.**, Korobchynskiy M. Objective clustering inductive technology of gene expression sequences features. Communications in Computer and Information Science, 2017. P. 359–372.
Здобувачем представлена технологія об'єктивного кластеризації особливостей високомірних даних на основі методів індуктивного моделювання складних систем.
16. Babichev S., **Taif M.A.**, Lytvynenko V., Osypenko V. Criterial analysis of gene expression sequences to create the objective clustering inductive technology. Materials of 2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO). Kiev, 2017. P. 244–248.
Здобувачем представлена методика для визначення ефективності різних критеріїв оцінки якості кластеризації складних біологічних об'єктів.
17. Lytvynenko V, **Taif M.A.**, Lurie I, Krejci J, Voronenko M, Savina N. Two Step Density-Based Object-Inductive Clustering Algorithm. CEUR Workshop Proceedings of the 8th International Conference on “Mathematics. Information Technologies. Education”, MoMLeT&DS-2019, Shatsk, Ukraine, June 2-4, 2019.p.p.117-135
Здобувачем представлена архітектура цільової технології кластеризації розроблена на основі двоступеневого алгоритму кластеризації DBSCAN та OPTICS.
18. Fefelov A.O, Lytvynenko V.I, **Taif M.A.**, Savina N.B, Voronenko M.A, Lurie I.A., Boskin O.O. Hybrid immune algorithms in the gene regulatory networks reconstruction. CEUR Proceedings of the Second International Workshop on

Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, Ukraine, April 15-19, 2019, p. 193-210

Здобувачем представлені гібридні методи реконструкції генно-регуляторних мереж, які дозволяють збільшити швидкість і точність вирішення задачі параметричної ідентифікації S-системи.

Патенти:

19. Свідоцтво про реєстрацію авторського права на твір № 96203 Комп'ютерна програма «Об'єктно-індуктивний алгоритм двокрокової кластеризації на базі щільнісних алгоритмів DBSCAN I OPTICS», Лур'є І. А., Савіна Н. Б., Литвиненко В. І., Боскін О. О., **Таїф М. А.**, Корніловська Н. В.
Матеріали й тези Міжнародних і Всеукраїнських наукових та науковопрактичних конференцій
20. Литвиненко В.І, **Таїф М. А.**, Фефелов А.О, Лур'є І.А. Реконструкція генної регуляторної мережі гібридним алгоритмом клонального відбору та програмування експресії генів. Праці 9-ї міжнародної школи-семінару «Теорія прийняття рішень». Ужгород, УжНУ, 2016. С. 163–165.
Здобувачем виконано математичний опис реконструкції процесу реконструкції генної мережі алгоритмом клонального відбору.
21. **Таїф М.А.** Обратный инжиниринг генных регуляторных сетей применением иммунных алгоритмов. 20-й Ювілейний Міжнародний Молодіжний Форум «Радіоелектроніка та молодь у ХХІ столітті» Харків 2016 Том 5 С.139-140
Здобувачем виконано запропоновано методика зворотного інжинірингу генних регуляторних мереж
22. **Таїф М.А.**, Литвиненко В.І, Лур'є І.А, Фефелов А.А. Идентификация регуляторных генных сетей с помощью многокритериального алгоритма клонального отбора. Интеллектуальные Системы Принятия Решений И Проблемы Вычислительного Интеллекта (ISDMCI'2016) XII Международная научная конференция 2016 Железный Порт, Украина С. 337-341
Здобувачем виконано метод ідентифікації регуляторних генних мереж за допомогою багатокритеріального алгоритму клонального відбору
23. **Таїф М.А.**, Литвиненко В. І, Лур'є І. А, Толкачева В. В. Современные подходы классификации методов моделирования генных регуляторных сетей. Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта (ISDMCI'2016) XII Международная научная конференция 2016 Железный Порт, Украина С. 342-343.
Здобувачем проаналізовані сучасні підходи класифікації методів моделювання генних регуляторних мереж.
24. **Таїф М.А.** Моделирование регуляторной генетической сети с помощью алгоритма искусственной иммунной сети. Матеріали IV Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених з автоматичного управління присвяченої дню космонавтики/ Херсон: ХНТУ, 2016 С. 145-150.
Здобувачем виконаний аналіз засобів моделювання генних регуляторних мереж на базі імунних алгоритмів
25. Литвиненко В.І. **Таїф М.А.**, Лур'є І.А. Моделирование генной регуляторной сети с использованием S-систем и алгоритма // Збірник матеріалів I

Всеукраїнської науково-практичної конференції студентів, аспірантів і молодих учених інформаційні технології в моделюванні ІТМ-2016, Миколаїв С. 84-85

Здобувачем виконаний аналіз засобів моделювання генних регуляторних мереж з використанням S-систем

26. Бабічев С.А., **Таїф М.А.**, Литвиненко В.І. Інформаційна технологія моделювання та побудови генних регулюючих мереж на основі мікрочіпів ДНК. Праці міжнародної науково-практичної конференції «Інформаційні технології та взаємодії». Київ, 2016. С. 153–155.

Здобувачем виконаний аналіз сучасних технологій побудови генних регулюючих мереж на основі мікрочіпів.

АНОТАЦІЯ

Таїф М. А. Моделі та методи реверсної інженерії генних регуляторних мереж на основі гібридних імунних систем – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 — інформаційні технології. — Херсонський національний технічний університет, Херсон, 2021.

Дисертація присвячена вирішенню актуальної проблеми — розробці моделей і методів реконструкції генних регуляторних мереж з використанням штучних імунних систем з метою підвищення ідентифікації розроблених динамічних регуляторних мереж, створення ефективних методів інтелектуального аналізу і обробки даних.

В роботі досліджено та проаналізовано аналіз використання існуючих методів та засобів реконструкції і інформаційних технологій реконструкції генних регуляторних мереж. У роботі представлені результати досліджень практичної реалізації двоступеневих алгоритмів кластеризації DBSCAN та OPTICS в рамках об'єктивної кластеризації індуктивної технології. Архітектура цільової технології кластеризації була розроблена на основі двоступеневого алгоритму кластеризації DBSCAN та OPTICS. Реалізація технології передбачає одночасну кластеризацію даних на двох підмножинах однакової потужності за алгоритмом DBSCAN, які включають однакову кількість попарно подібних об'єктів з подальшою корекцією отриманих кластерів за алгоритмом OPTICS. Розроблений метод комбінованого використання алгоритму клонального відбору, вейвлет-нейронної мережі та звичайних диференціальних рівнянь. В запропонованому методі топологія та настройка параметрів визначається за допомогою алгоритму клонального відбору з допомогою якого налаштовуються параметри вейвлет-мережі. Вейвлет-мережа формує праву частину системи звичайних диференціальних рівнянь. Для знаходження концентрації продуктів експресії генів, які представлені способом для розв'язання задачі Коші застосовується метод Рунге-Кутти четвертого порядку. Ефективність запропонованого методу доведена експериментальними дослідженнями, які підтверджують застосовність цього підходу для виявлення взаємозв'язків між

компонентами генної регуляторної мережі. Розроблено і практично реалізовано метод параметричної ідентифікації S-системи з використанням модифікованого алгоритму клонального відбору, основна ідея якого полягає при досягненні певних умов в періодичній оптимізації, яка трансформує простір пошуку за допомогою перерахунку проміжків можливих значень параметрів S-системи з їх одночасним центруванням щодо виявленого поточного кращого рішення. Розроблено і реалізовано гібридний алгоритм з використанням диференційної еволюції. В даному гібридному клональному алгоритмі, в фаза мутації розширена операторами, запозиченими з алгоритму диференціальної еволюції. Експериментальні і порівняльні дослідження, що даний алгоритм дозволяє значно підвищити швидкість і точність рішень при розв'язанні задачі ідентифікації S-системи.

Розроблено та реалізовано комбінований алгоритм клонального відбору та тригонометричної диференційної еволюції. Експериментальні і порівняльні дослідження, підвищення швидкість і точність рішень при розв'язанні задачі реконструкції генної регуляторної системи.

Ключові слова: генна регуляторна мережа, профіля експресії генів, S-система, вейвлет-нейронна мережа, алгоритм клонального відбору, диференційна еволюція, оператор тригонометричної мутації.

ABSTRACT

Taif M.A. Models and methods of reverse engineering of genetic regulatory networks based on hybrid immune systems – Qualifying scientific work as manuscript.

Thesis for obtaining of scientific degree of the candidate of engineering science in the specialty 05.13.06 - information technologies. - Kherson National Technical University, Ministry of Education and Science of Ukraine, Kherson – 2021.

The thesis is devoted to the solution of an actual problem - the development of models and methods for the reconstruction of gene regulatory networks using artificial immune systems. The goal is to improve the identification of dynamic regulatory networks, as well as to create effective methods for intelligent analysis and data processing.

The work studies and analyzes the analysis of the use of existing methods and tools for the reconstruction of gene regulatory networks. The paper presents the results of research on the practical implementation of two step algorithms of clustering DBSCAN and OPTICS based on inductive objective clustering. The clustering architecture was developed based on the two-step clustering algorithm DBSCAN and OPTICS. Implementation of the technology involves simultaneous clustering of data on two subsets of the same power by DBSCAN algorithm, which include the same number of pairwise similar objects, followed by correction of obtained clusters by OPTICS algorithm. The developed method combines the use of clonal selection algorithm, wavelet neural network and ordinary differential equations. In the proposed method, the topology and parameter tuning are determined using the clonal selection

algorithm with which the parameters of the wavelet network are tuned. The wavelet network forms the right part of the system of ordinary differential equations. To find the concentration of gene expression products, which are represented by the method for solving the Cauchy problem, the Runge-Kutta method of the fourth order is used. The effectiveness of the proposed method has been proven by experimental studies that confirm the applicability of this approach to reveal the relationships between the components of the gene regulatory network. The method of parametric identification of S-systems has been developed and practically realized by using the modified algorithm of clonal selection which main idea is to transform the search space by recalculation of intervals of possible values of S-system parameters with their simultaneous centering on the found current best solution when reaching certain conditions of optimization. A hybrid algorithm using differential evolution is designed and implemented. In this hybrid clonal algorithm, in which the mutation phase is extended by operators borrowed from the differential evolution algorithm. Experimental and comparative studies, this algorithm can significantly improve the speed and accuracy of solutions when solving the S-system identification problem.

A combined method of clonal selection and trigonometric and differential evolution is designed and implemented. Experimental and comparative studies, increasing the speed and accuracy of solutions when solving the problem of gene regulatory system reconstruction.

Keywords: gene regulatory network, gene expression profile, S-system, wavelet neuron network, clonal selection algorithm, differential evolution, trigonometric mutation operator.

Відповідальний за випуск Козуб Н.О.

Підписано до друку 12.04.2021 р.

Формат 60x84 ¹/₁₆. Папір офсетний.

Гарнітура «Times New Roman». Друк ризографія.

Умовн. друк. арк. 1. Тираж 100 пр. Зам. 1960.

Надруковано з готового оригінал-макету.

Книжкове видавництво ФОП Вишемирський В. С.

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи: серія ХС № 48 від 14.04.2005 р.

видано Управлінням у справах преси та інформації.

Адреса: 73000, Україна, м. Херсон, вул. Соборна, 2,

тел. (050) 133–10–13, e-mail: printvvs@gmail.com, vish_sveta@rambler.ru