

О.О. Летичевський, Ю.Г. Тарасіч

НЕЙРО-СИМВОЛЬНИЙ ПІДХІД У ДОСЛІДЖЕННІ БІОЛОГІЧНИХ ПРОЦЕСІВ ТА СИСТЕМ

У статті розглядається сучасний стан та перспективи застосування технології нейронних мереж та побудованих на їхній базі систем для дослідження біологічних процесів та систем, зокрема, процесів вірус-клітинної взаємодії. Зокрема, авторами розглянуто поняття нейро-символьного підходу, як такого, що поєднує використання нейронних мереж та алгебраїчного моделювання.

Застосування алгебраїчного підходу у поєднанні з нейронними мережами дозволяє виконати ефектвне формальне доведення певних властивостей об'єктів (наприклад, властивостей клітин та вірусів) або процесів, а також реалізувати пошук об'єктів чи необхідних значень їхніх параметрів, які відповідають заданим властивостям. Одним із експериментів, що розглядається, є застосування запропонованого підходу до моделювання та дослідження процесу програмованої загибелі клітини (апоптозу), викликаного певними видами вірусів, та можливості використання отриманих результатів у медичній практиці (зокрема, в лікуванні онкологічних захворювань). Основним завданням таких експериментів є аналіз та виявлення властивостей вірусів здатних запускати процес апоптозу пухлинної клітини, та, власне, визначення можливості досягнення кінцевої стадії цього процесу за заданих параметрів вірусу та клітини.

Ключові слова: нейро-символьний підхід, нейронні мережі, молекулярне моделювання, алгебраїчне моделювання, вірус-клітинна взаємодія.

О.О. Letychevskiy, Yu. H. Tarasich

NEURO-SYMBOLIC APPROACH FOR THE BIOLOGICAL SYSTEMS AND PROCESSES RESEARCH

The current state and prospects for the application of neural network technology and systems built based on neural networks for studying biological systems and processes (in particular, the processes of virus-cell interaction) were considered in the article. In particular, the concept of the neuro-symbolic approach, which combines the use of neural networks and algebraic modelling, was described.

The use of an algebraic approach in combination with neural networks allows us to obtain an effective formal proof of biological objects' properties (for example, cells' and viruses' properties) or processes, as well as to search for objects or the necessary values of their parameters that correspond to the specified properties. One of the experiments that we consider is applying the proposed approach to modelling and studying the process of programmed cell death (apoptosis) caused by certain types of viruses and considering the possibility of using the obtained results in medical practice (particularly in the treatment of oncological diseases). The main task of such experiments is to analyze and identify the properties of viruses capable of triggering tumour cell apoptosis and, in fact, to determine the possibility of reaching the final stage of this process under the given parameters of the virus and the cell.

Key words: neuro-symbolic approach, neural networks, molecular modelling, algebraic modelling, virus-cell interaction.

Вступ

Основними викликами сучасних досліджень із використанням математичних методів в біології, вірусології та медицині є удосконалення методів комп'ютерного молекулярного моделювання та застосування методів машинного навчання для визначення потенційних препаратів-кандидатів з метою боротьби з вірусами та онкологічними захворюван-

нями, вивчення можливих мутацій вірусів, про-гнозування поширення епідемій, вивчення можливих варіантів взаємодії вірусів та лі-ків з клітинами людини, обробка даних клінічних експериментів тощо.

На сьогодні розроблено досить велику кількість підходів, алгоритмів та засобів побудови моделей, що розглядають міжатомні та міжмолекулярні взаємодії,

біофізичні та біохімічні процеси з використанням математичних методів моделювання та аналізу [1]. Одним із потужних інструментів, що набуває сьогодні широкого застосування у біологічних дослідженнях, є нейронні мережі [2-4,8,9]. Основними напрямками таких досліджень є дослідження білок-білкових взаємодій вірусів та клітини-хазяїна, дослідження мікроскопічних даних з метою аналізу та вивчення перебігу процесів міжклітинної та вірус-клітинної взаємодії, пошук речовин із заданими властивостями тощо [5-14]. Для розробки інструментів прогнозування білок-білкової взаємодії вірус-хазяїн, таких як DeNovo, HOPITOR, VHPPI, InterSPPI-HVPPi та інші, розглядаються можливості застосування різних моделей машинного навчання, зокрема, машини опорних векторів, випадковий ліс, система градієнтного підсилювання та ін. [5-7]. Широкого застосування для аналізу міжклітинних та внутрішньоклітинних процесів набуває і застосування моделей глибокого навчання у флуоресцентній мікроскопії як одного з ефективних засобів вилучення знань із отриманих зображень [8,9]. Проте, не зважаючи на можливості існуючих інструментів та методів молекулярного моделювання, вчені визначають і низку недоліків, серед яких - низька точність деяких методів, обмеження рамками біологічних експериментів, наявність помилок у структурах молекул, з якими працюють програмні засоби тощо. Найбільш поширеними проблемами для реалізації та ефективної роботи нейронних мереж та систем штучного інтелекту є обмеження даних біологічних експериментів, серйозні розбіжності між розподілами навчальних та реальних тестових даних, труднощі аналізу та інтерпретації результатів, вплив на ефективність навчання, зростання кількості даних тощо. Відповідно застосування нейронних мереж/систем штучного інтелекту у більшості випадків не може гарантувати стовідсоткової достовірності результатів, хоча дійсно звужує пошук об'єктів у дослідженнях. Підвищення продуктивності та точності прогнозних моделей учені вбачають за рахунок так званого нейро-символьного підходу, який поєднує підхо-

ди, засновані на штучних нейронних мережах, та підходи, що спираються на явне представлення знань за допомогою формальних мов із можливістю застосування методів символьних обчислень. Не зважаючи на наявність вагомих здобутків у цьому напрямі [10-14], проблема розробки відповідних систем та методів залишається відкритою.

Наше бачення розв'язання проблеми продуктивності та точності інструментів моделювання біологічних процесів і систем полягає у синергії методів нейронних мереж та алгебраїчного підходу, ефективність якого доведена в інших наукових сферах [15,16]. Загальну схему запропонованого підходу, а також приклади формального представлення моделі вірус-клітинної взаємодії алгебраїчною мовою представлено у наступних розділах статті.

Нейро-символьний підхід. Загальна схема дослідження

Запропонований у статті нейро-символьний підхід передбачає комбінацію формальних методів, зокрема, алгебраїчних, та властивостей нейронної мережі. Таке поєднання дозволяє використати сильні та взаємодоповнюючі сторони обох технологій – точність для алгебраїчного підходу та швидкість класифікації або генерації для нейронних мереж. Тим самим ми усуваємо відповідні недоліки – тривалість застосування формальних методів в умовах комбінаторного вибуху та можливість помилок у класифікації нейронних мереж.

Ми розглядаємо два основні напрями та, відповідно, дві методики поєднання алгебраїчного моделювання та нейронних мереж:

1. *Застосування нейро-символьного підходу для прискорення пошуку так званих «кандидатів».* Дані експерименти спрямовані на пошук речовин із заданими властивостями або визначення наявності протягом експерименту певної події, що має виникнути під впливом визначених факторів середовища. Наприклад, пошук речовини-кандидата у вакцину проти вірусу, або визначення можливості проник-

нення вірусу у клітину та запуску її програмованої загибелі за встановлених початкових характеристик взаємодіючих агентів та середовища (температура, кислотність, наявність/відсутність електромагнітного поля тощо). У даному випадку, зважаючи на складність та масштабність моделей/експериментів, що розглядаються, використання виключно алгебраїчного моделювання може призвести до експоненційного вибуху та необмеженого часу пошуку, який може не закінчитись. Для усунення цього недоліку пропонується використовувати наступні нейронні мережі:

– Нейронні мережі, що базуються на історичних даних експерименту та передбачають результати сценарію взаємодії компонент експерименту.

– Нейронні мережі, створені на основі згенерованих сценаріїв у середовищі алгебраїчної бази знань.

Дані типи нейронних мереж розглядаються як класифікатор визначеної поведінки та прогнозують властивості середовища за сценарію, що розглядається. Наприклад, можливість активації рецепторів загибелі клітини обраною речовиною та вплив стану середовища клітини після активації рецепторів загибелі на блокування або запуск процесу апоптозу. Тут властивості речовини, наприклад, наявність електромагнітних властивостей у наночастинок після проникнення у клітину та із наведенням магнітного поля може провокувати зміну температури всередині клітини, впливаючи на зміну кількості та структури її компонент (активація та блокування певних ферментів, білків, зміна у кількості іонів кальцію, натрію, активних форм кисню (АФК), тощо).

За підказки нейронної мережі небажані напрями пошуку можуть відкидатись і швидкість моделювання та ймовірність успішного пошуку зростають. Ми додатково можемо аналізувати, наскільки близько підійшли до шуканої властивості. Загальну схему застосування даного підходу до пошуку деякого середовища із певною властивістю відображено на Рисунку 1.

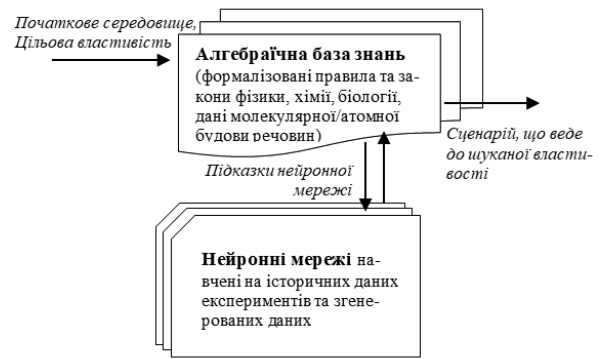


Рис. 1. Пошук властивостей

2. Застосування алгебраїчного підходу для доведення або спростування результатів класифікації або генерації нейронних мереж. Даний напрям нейросимвольного підходу стосується експерименту, коли речовина із шуканою властивістю вже згенерована нейронною мережею та треба перевірити, чи дійсно вона відповідає цій властивості (Рис.2). Для цього ми використовуємо формальне доведення або алгебраїчне моделювання.



Рис.2. Доведення або спростування властивостей речовини

Наприклад, ми маємо речовину, згенеровану нейронною мережею, визначену як потенційний кандидат на активацію рецепторів смерті клітини. У нашому підході ми використовуємо дану речовину, тобто дані про її молекулярну структуру та можливі властивості як частину вхідних даних та проводимо алгебраїчне моделювання експерименту на їхній основі. У даному випадку ми використовуємо нейронну мережу для першого етапу пошуку речовин із заданими властивостями, або вивчення алгоритму взаємодії даної речовини з клітиною (за рахунок обробки мікроскопічних зображень). А вже потім застосовуємо ал-

гебраїчне моделювання для підтвердження властивостей даної речовини.

Дана процедура дуже важлива для досліджень. Сьогодні за допомогою нейронних мереж синтезовано понад 300 000 різних речовин із шуканими властивостями [17]. Проте кожен із «кандидатів» має бути перевірений експериментально – в лабораторних умовах, що значно сповільнює дослідження. Надання ж доказового підтвердження або спростування на рівні моделювання дасть змогу скоротити час та витрати для досягнення результату.

Для обох методик основним інструментом є моделювання поведінки середовища та об'єктів, що в ньому взаємодіють. Таким середовищем, наприклад, є клітина із «зануреним» у неї вірусом, де взаємодія відбувається між вірусом або його складовими та складовими клітини (органели, ферменти, наявні у клітині іони тощо). А властивістю, що перевіряється, може бути фінальна стадія програмованої загибелі клітини. Таким чином, визначивши в алгебраїчному вигляді властивість середовища, яку ми шукаємо, за допомогою моделювання буде визначено сценарій, який приводить до появи цієї властивості. Відповідно вхідними даними для алгебраїчного моделювання є формальне представлення (представлення в алгебраїчному вигляді) середовища, що моделюється, та формалізовані властивості, яких треба досягти у процесі моделювання. Для вивчення поведінки, забезпечення багаторівневості та точності моделювання ми формуємо алгебраїчну базу знань, що представляє не лише повний формальний опис моделей, що розглядаються, а й формалізовані правила та закони фізики, хімії, біології, що дозволяє виконувати моделювання експериментів на різних рівнях абстракції – від квантових взаємодій до взаємодій між біологічними об'єктами.

Ми використовуємо як конкретне алгебраїчне моделювання так і символічне. У символічних моделях ми використовуємо не конкретні значення змінних та атрибутів середовища, що моделюється, а формули над ними. Наприклад, ми можемо вказати, що температура всередині клітини може бути від 36°C до $37,5^{\circ}\text{C}$ ($36 \leq$

$\text{cellTemperature} \leq 37.5$). У такому випадку буде виконано символічне алгебраїчне моделювання, кінцевим результатом якого буде не один сценарій, що досягає шуканої властивості, а саме всі сценарії із початкової формули, в яких ця властивість досяжна, що є основною відмінністю алгебраїчного моделювання від конкретного або імітаційного.

За інструмент моделювання використовуємо Систему створення інсерційних моделей [15,16], яка складається з інтерфейсу для формалізації предметної області та створення моделі задачі, яка буде досліджуватись у термінах формальних теорій та алгебраїчного серверу, що поєднує в собі формалізовані математичні теорії та відповідні методи, які працюють із моделями та розв'язують задачі. Основною використовуваною теорією є теорія агентів та середовищ, започаткована українським академіком О.А. Летичевським та британським науковцем Д. Гілбертом [18].

Алгебраїчна база знань молекулярних та квантових взаємодій

Для моделювання біологічних процесів та систем ми формуємо алгебраїчну базу знань, яка містить формалізовані специфікації знань з фундаментальних законів фізики, хімії та біології та алгебраїчне представлення теорій на різних рівнях абстракції – рівень квантових та атомних взаємодій, рівень молекулярних взаємодій, рівень хімічних реакцій та білкових з'єднань тощо. База знань представляє собою рівняння в алгебрі поведінок, що включає послідовності дій, які описують усі можливі взаємодії агентів та середовищ. За агентів ми визначаємо віруси, клітини, клітинні структури, білки, молекули тощо. Кожен агент має відповідні характеристики, що визначаються так званими атрибутами, які можуть приймати числові, символічні та ін. значення, тобто атрибут має відповідний тип. Сукупність значень усіх атрибутів визначають стан агента, який може змінюватись згідно з взаємодіями, можливими для заданого середовища. Агент вищого рівня

абстракції може, у свою чергу, виступати середовищем для взаємодії інших агентів.

Як приклад розглянемо формальне представлення агента типу ВІРУС. Атрибути агента типу ВІРУС представлені значеннями наступних величин: функціональний атрибут, що відображає набір білків, які входять до складу вірусу (атрибут `proteins`), атрибут `ribonucleicAcid`, що визначає (за наявності) молекулу РНК, та атрибут `deoxyribonucleicAcid`, що визначає (за наявності) молекулу ДНК, функціональний атрибут, що відображає набір жирів, які входять до складу вірусу (атрибут `lipids`), функціональний атрибут, що відображає набір вуглеводів, які входять до складу вірусу (атрибут `carbohydrates`), атрибуту дійсного типу, що визначають масу (`mass`), розмір (`size`) вірусу, атрибут `status` перелічуваного типу для визначення поточного «статусу» агента ("In Cell", "Connected to Cell" та "Virion").

Взаємодії агентів визначаються за допомогою формальних дій, які визначають зміну стану агенту або зміну значень його атрибутів. До таких дій належать, наприклад, утворення зв'язків між амінокислотами, формування білкових структур, взаємодії білків з іншими білками та речовинами. Дії визначаються в мові алгебри поведінки, що представляє собою алгебраїчні специфікації. Кожна дія має передумову її виконання та постумову, де відбувається зміна значень атрибутів агентів та середовища. Як приклад розглянемо дію `CreatePeptideBond`, яка визначає утворення пептидного зв'язку між двома амінокислотами:

```

CreatePeptideBond =
(Forall(n1:int, k1:int, n2:int, k2:int,)
((1<= n1 <= aminoasids(i).atomsNum) &&
(1<= k1 <= aminoasids(i).atomsNum) &&
(1<= n2 <= aminoasids(j).atomsNum) &&
(1<= k2 <= aminoasids(j).atomsNum) &&
n1 != k1 && n2 != k2 &&
aminoasids(i).MolOrbital(n1,k1,-1,1)==2 &&
aminoasids(i).atoms(n1).Name == O &&
aminoasids(i).atoms(k1).Name == H &&
aminoasids(j).MolOrbital(n2,k2,-1,1)==2 &&
aminoasids(j).atoms(n2).Name == N &&
aminoasids(j).atoms(k2).Name == H) ->
(Protein1.aminoAcids(1,1) = i;
Protein2.aminoAcids(1,2) =j)

```

У передумові дії виконується перевірка того, що один з амінокислотних зали-

шків має вільну аміногрупу, а інший - має вільну α -карбоксильну групу. Якщо дана передумова виконається, між амінокислотами буде утворено зв'язок, і ми «запишемо» їх у відповідний поліпептидний ланцюг. Першим записуємо амінокислотний залишок з вільною аміногрупою, другим – амінокислотний залишок з вільною α -карбоксильною групою. Аналогічним чином до ланцюжка додаються інші амінокислоти.

Усі можливі взаємодії агентів у середовищі визначаються поведінкою. Поведінковий вираз є формулою, побудованою із дій та поведінок за допомогою операцій алгебри поведінки: префіксинг «.», недетермінований вибір «+» та композиції поведінок – послідовна («;») та паралельна («||») [18]. Як приклад, рівняння поведінки, що описує взаємодію вірусу та клітини на верхньому рівні, буде записано наступним чином:

```

VIRUS_CELL_INTERACTION =
((attachmentToCellReceptors ; ((penetration-
IntoCells ; VIRUS_IN_CELL) + !penetration-
IntoCells)) + !attachmentToCellReceptors)),
де attachmentToCellReceptors – дія, що описує приєднання вірусу до клітини, penetrationIntoCells – дія, що описує проникнення вірусу в клітину, !attachmentToCellReceptors – вказує на те, що вірус до клітини не приєднався, !penetrationIntoCells – проникнення вірусу до клітини не відбувається, VIRUS_IN_CELL – поведінка, яка, складатиметься із дій, які описують процеси, що відбуваються із вірусом всередині клітини.

```

Відповідно, представивши усі структурні одиниці клітини та зовнішніх факторів у вигляді агентів та атрибутів середовища, ми отримуємо формальне представлення моделі, що розглядається. За рахунок формалізації усіх можливих взаємодій між агентами та задання різних значень атрибутів агентів та середовища, ми можемо розглядати різні сценарії поведінки агентів.

Під час моделювання буде виконано усі можливі дії агентів, тобто всі можливі комбінації дій для досягнення цілі моделювання. Такою ціллю, наприклад, може бути знаходження кількісних та якісних показників вакцин/ліків, які можуть

заблокувати вірус у клітині або заблокувати клітинні рецептори для перешкоджання проникнення вірусу до клітини; визначення можливості досягнення клітиною певного стану, наприклад, загибелі клітини внаслідок взаємодії з вірусом тощо.

Алгебраїчне моделювання експерименту

Завдання та формальний опис експерименту визначаються переліком агентів, взаємодію яких має бути розглянуто, заданими значеннями початкового стану вказаних агентів та середовища, а також властивостями, яких ми маємо досягнути. Як приклад, одним із таких експериментів є дослідження взаємодії вірусу везикулярного стоматиту з пухлинною клітиною, як такої, що здатна викликати активацію процесу її програмованої загибелі. Для моделювання даного експерименту ми використовуємо знання про біохімічні процеси всередині клітини, молекулярну будову речовин та біологічних систем, зокрема, вірусів та клітин, формалізовані правила білок-білкових, міжмолекулярних взаємодій тощо. У моделі на даному рівні враховуємо якісні та кількісні характеристики таких елементів клітини та міжклітинного середовища, як білків TRADD, FADD, RIP, TRAF2, ARAF-1, NF- κ B, cIAP1,2, білків родини Bcl-2 (ті, що беруть участь у розвитку або блокуванні апоптозу), каспаз 2,3,8,9, цитохрому, АФК, іонів Ca, Na тощо. Саме дані елементи, а також вірус, що розглядається, і виступають у ролі агентів, середовищем взаємодії яких є сама клітина. Додатково враховуються такі параметри середовища – температура, кислотність тощо. Як дії агентів розглядаємо усі можливі взаємодії між складовими вірусу та складовими клітини, формалізовані відповідно до правил визначених біологічними експериментами. Можливі взаємодії між клітинними білками та ферментами моделюються на верхньому рівні – тобто на рівні аналізу відповідності білкових доменів. Метою алгебраїчного моделювання у даному випадку є визначення можливих сценаріїв розвитку, або блокування процесу апоптозу за різних початкових значень ат-

рибутів агентів та середовища (наявність або відсутність певних типів каспаз або білків у клітині, тип та характеристики вірусу, зміна температури середовища тощо).

Як приклад, рівняння поведінки моделі, що враховує можливі взаємодії вірусу з клітиною та відповідно різні шляхи запуску процесу апоптозу клітини буде записано у такому вигляді:

```
APOPTOSIS_PROCESS =  
(CONNECTION_VIRUS_TO_CELL +  
PENETRATION_VIRUS_INTO_CELL),  
CONNECTION_VIRUS_TO_CELL =  
(BINDINGtoTNF + BINDINGtoFAS)), ...
```

У першій частині рівняння ми описуємо два можливі варіанти взаємодії вірусу з клітиною – поведінка CONNECTION_VIRUS_TO_CELL (послідовність дій та поведінок, що описують зовнішній шлях активації процесу апоптозу), або поведінка PENETRATION_VIRUS_INTO_CELL (послідовність дій та поведінок, що описують внутрішній шлях активації процесу апоптозу). Далі ми визначаємо усі набори дій та поведінок, які описують можливі взаємодії агентів та зміни стану середовища залежно від «обраного» шляху. Так, наприклад, ми вказуємо на можливі з'єднання з рецепторами смерті клітини (TNF або FAS) – поведінки BINDINGtoTNF та BINDINGtoFAS та визначаємо можливі взаємодії агентів для кожної з них.

У результаті моделювання, відповідно до заданих початкових даних експерименту, якщо, наприклад, ми маємо вірус, який активує рецептори смерті клітини TNF, буде обрано шлях, що відповідатиме поведінці CONNECTION_VIRUS_TO_CELL, і далі поведінці BINDINGtoTNF, що, у свою чергу, складається із множини дій та поведінок, які описують можливі взаємодії білків та ферментів клітини:

```
BINDINGtoTNF =  
(bindingVirusToTNFR1receptor .  
bindingTNFR1withTRADD; (bindingTRADDwithRIP ;  
((APOPTOSIS_INHIBITION) +  
(CASPASE_CYCLE_ACTIVATION))) +  
(bindingTRADDwithFADD ;  
(bindingFADDwithcFLIP; ApoptosisInhibition +  
CASPASE_CYCLE_ACTIVATION))) ,
```

Оскільки зазначена поведінка також містить недетермінізми (наприклад, дія bindingTNF-R1withTRADD визначає взає-

модію адаптерного білка TRADD із доменом загибелі TNF-R1 і далі можуть розглядатися набори дій, що призводять до блокування або запуску апоптозу), або дія binding TRADD with FADD (зв'язок адаптерного білка TRADD з білками FADD (також активують сигнальний комплекс загибелі клітини ((поведінка CASPASE_CYCLE_ACTIVATION)) або блокують процес апоптозу з'єднуючись із білком c-FLIP (дія binding FADD with FLIP))), то вибір подальшого сценарію залежить від початкових даних експерименту та від стану середовища (значень атрибутів агентів та середовища), отриманого після виконання кожної дії.

Отож, розглядаючи різні співвідношення між факторами активації та блокування процесу апоптозу та різні значення параметрів середовища, здатні впливати на перебіг відповідних реакцій – температура, концентрація АФК та іонів Ca²⁺ у клітині та, абстрагуючись від інших факторів впливу залежно від співвідношення кількості прокаспаз та білків, які блокують їхню активацію, ми бачимо різні варіанти поведінки системи. Відповідно в результаті моделювання ми отримуємо саме ту послідовність дій, яка за умови заданого початкового стану середовища приводить до підтвердження або спростування досягнення кінцевої стадії апоптозу клітини (зростання концентрації у ядрі клітини Ca²⁺ та наявність у ядрі клітини залежних протеаз (DFF, AIF, EndoGin, та ін), під дією яких відбувається фрагментація ДНК), що визначається відповідною формулою властивості:

```
(ConcentrationCa2>NormalConcentration &&
((AIFinNucleus==1) || (DFFinNucleus==1) ||
(EndoGinNucleus==1))) && DNAfragmentation==1
```

Зважаючи на складність та можливу циклічність клітинних процесів, ми зіштовхуємось із проблемою виникнення недетермінізмів, і пошук може бути нескінченим або вимагати великих обчислювальних потужностей. Саме у таких випадках постає необхідність у використанні нейронних мереж, які, аналізуючи поточний стан середовища, визначають найбільш ефективний напрямок або дію, яку треба застосовувати, або відкинуть напрямки, які ведуть до небажаного результату.

Нейронні мережі в моделюванні експерименту

Як було визначено у попередніх розділах, одним із основних завдань, що потребує розробки нейронних мереж та їх поєднання з алгебраїчним підходом, є забезпечення ефективності та швидкості пошуків сценаріїв взаємодії агентів, що розглядаються в межах експерименту. Як приклад, ми розглядаємо необхідність розробки нейромережі, що дозволить виконувати ефективний пошук сценаріїв досягнення фінальної стадії апоптозу клітини за заданих початкових станів агентів та середовища. Дана нейромережа має бути натренована для визначення так званого відношення «близькості до властивості».

Властивість речовини може визначатись формулою над її атрибутами та змістовно представляти такі поняття, як «токсичність», «кислотність», «розчинність», «теплопровідність», «електропровідність» та інші, що визначається молекулярною структурою речовини та множиною значень її характеристик, визначених відповідною алгебраїчною формулою.

Для тренування нейромережі, завданням якої є класифікація рівня відповідності властивостей речовини заданим, розглядається множина формул, що представляють дані властивості, та здійснюється моделювання, що приводить до визначеного експериментом результату шляхом виконання заданих перетворень певної початкової речовини. Для цього використовуються алгебраїчні рівняння зі створеної бази знань, що описують процеси на рівні молекулярних або квантових взаємодій. У загальному вигляді відношення близькості може представляти число, що дорівнює кількості перетворень, які необхідно виконати з використанням рівнянь взаємодій. Таким чином, маючи відповідну нейронну мережу глибокого навчання, яка є класифікатором та визначає відношення близькості між речовинами, можна на кожному кроці моделювання пошуку вибирати таке перетворення, яке дає результат найбільш близький до властивості. Наприклад, ми можемо визначати наявність магнітних властивостей наночастинок або композитів

наночастинок та розглядати модулювання фармакокінетичних процесів у пухлинній клітині. Наявність заданої властивості буде підтверджено, якщо за заданих початкових характеристик наночастинок за рахунок їхніх можливих взаємодій із ключовими металомісними білками та ферментами клітини, а також під дією магнітного поля буде підвищено температуру в середині клітини та активовано процес її програмованої загибелі.

Висновки та перспективи подальших розвідок

На даному етапі досліджень нами розроблено методологію застосування нейро-символьного підходу до моделювання експериментів, спрямованих на пошук та вивчення властивостей біологічних систем, органічних та неорганічних речовин, хімічних процесів та реакцій, що послуговуються формалізованими знаннями та законами квантової хімії, фізики, біології.

Можливі взаємодії між клітинними білками та ферментами у моделях вірус-клітинної взаємодії для поточних експериментів розглядаються на верхньому рівні абстракції, тобто на рівні визначення можливості взаємодії за рахунок наявності або відсутності певних білкових доменів відповідно до правил визначених біологічними експериментами. Окрім того, ми розглядаємо і моделі, в яких правила взаємодій між об'єктами верхніх рівнів абстракції (білок-білкові, міжмолекулярні взаємодії, тощо) є такими, що виводяться із правил міжатомних та квантових взаємодій. Хоча дослідження застосування запропонованого підходу для моделювання біологічних/біохімічних процесів та систем перебувають на стадії випробування, і наразі ми працюємо над розширенням бази формалізованих знань, перші отримані результати застосування нейро-символьного підходу до моделювання міжатомних та міжмолекулярних взаємодій, внутрішньоклітинних процесів вказують на те, що ефективність спільного використання алгебраїчного моделювання та нейронних мереж на порядок пришвидшує виконання відповідних біологічних досліджень.

Наступними кроками є дослідження необхідних початкових значень атрибутів агентів та модельованого середовища, розширення та ускладнення бази знань та моделей експериментів, навчання нейронної мережі на історичних даних результатів біологічних та біохімічних експериментів, розміщених у світових базах даних, проведення прямого конкретного алгебраїчного моделювання (для порівняння отриманих результатів з даними біологічних експериментів) та оберненого моделювання на отриманих моделях, як такого, що дозволить визначити необхідні параметри атрибутів агентів та середовища, необхідних для досягнення заданих властивостей. Зокрема, передбачається застосування нейро-символьного підходу до моделювання експериментів, пов'язаних із вивченням властивостей речовин, здатних впливати на ініціацію програмованої загибелі ракових клітин: дослідження взаємодії вірусу везикулярного стоматиту з пухлинною та здоровою клітинами; дослідження ефектів застосування наночастинок для контрольованого продукування АФК та фототермії, направлених на індукцію окисного стресу та вибірккову загибель пухлинних клітин.

Література

1. S. Pirhadi, J. Sunseri, D. Koes, Open source molecular modeling, *Journal of Molecular Graphics and Modelling*, 2016. № 69, С. 127-143.
2. B. Das, M. Kutsal, R. Das, Effective prediction of drug-target interaction on HIV using deep graph neural networks, *Chemometrics and Intelligent Laboratory Systems*, 2022. № 230.
3. J. Pineda, B. Midtvedt, H. Bachimanchi, et al, Geometric deep learning reveals the spatiotemporal features of microscopic motion. *Nat Mach Intell*, 2023. № 5, С. 71–82.
4. M. Sharma, S. Deswal, Drugs-Protein affinity-score prediction using deep convolutional neural network. *Expert Systems*, 2022. № 39(10).
5. A. Basit, W. Abbasi, A. Asif, S. Gull, F. Minhas, Training Host-Pathogen Protein-Protein Interaction Predictors, *J. Bioinform. Comput. Biol*, 2018. № 16.
6. S. Alguwaizani, B. Park, X. Zhou, D. Huang, K. Han, Predicting Interactions between Virus and Host Proteins Using Repeat Patterns and Composition of Amino Acids, *J. Healthc. Eng*, 2018, № 2018.

7. X. Yang, S. Yang, Q. Li, S. Wuchty, Z. Zhang, Prediction of Human-Virus Protein-Protein Interactions through a Sequence Embedding-Based Machine Learning Method, *Comput. Struct. Biotechnol. J.*, 2020. № 18, С. 153–161.
8. E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, Deep learning for cellular image analysis, *Nat. Methods*, 2019. № 16, С. 1233-1246.
9. M. Suomalainen, U.F. Greber, Virus Infection Variability by Single-Cell Profiling Viruses, 2021. № 13, С. 1568.
10. M. Alshahrani, M. Khan, O. Maddouri, A. Kinjo, N. Queralt-Rosinach, R. Hoehndorf, Neuro-symbolic representation learning on biological knowledge graphs, *Bioinformatics*, 2017. № 33(17), С. 2723–2730.
11. A. Chowdhury, J. Kubricht, A. Sood, P. Tu, A. Santamaria-Pang, Escell: Emergent Symbolic Cellular Language, *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, С. 1604-1607.
12. A. Agibetov, S. Matthias, Fast and scalable learning of neuro-symbolic representations of biomedical knowledge, *arXiv preprint*, 2018. 1804.11105.
13. A. Rivas, D. Collarana, M. Torrente, M. Vidal, A Neuro-symbolic System over Knowledge Graphs for Link Prediction, *Semantic Web*, 2023. С. 1 – 25.
14. M. Glauer, F. Neuhaus, T. Mossakowski, A. Memariani, J. Hastings, P. Hitzler, A. Eberhart, Neuro-symbolic semantic learning for chemistry. *Compendium of Neurosymbolic Artificial Intelligence and Applications*, 2023. С. 460-484.
15. O. Letychevskyi, V. Peschanenko, V. Volkov, Algebraic Virtual Machine and Its Applications. *Communications in Computer and Information Science*, 2022. № 1698, С. 23–41.
16. O. Letychevskyi, V. Peschanenko, Applying Algebraic Virtual Machine to Cybersecurity Tasks, *IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications*, 2022, С. 161–169.
17. A. Merchant, S. Batzner, S. Schoenholz, et al, Scaling deep learning for materials discovery, *Nature*, 2023. № 624, С. 80–85.
18. A. Letichevsky, D. Gilbert, A Model for Interaction of Agents and Environments, *Recent Trends in Algebraic Development Techniques, WADT 1999, 2000*, № 1827, С.311-328.

Одержано: 10.04.2024

Внутрішня рецензія отримана: 19.04.2024

Зовнішня рецензія отримана: 26.04.2024

Про авторів:

¹Летичевський Олександр Олександрович,
Доктор фізико-математичних наук,
старший науковий співробітник.
<https://orcid.org/0000-0003-0856-9771>.

²Тарасіч Юлія Геннадіївна,
Доктор філософії
(Інформаційні технології),
докторант.
<https://orcid.org/0000-0002-6201-4569>.

Місце роботи авторів:

^{1,2}Інститут кібернетики ім. В.М. Глушкова
НАН України,
Тел. (+38) (044) 526-20-08
E-mail: incyb@incyb.kiev.ua,
www.incyb.kiev.ua