УДК 61:004.651(075.8)

І.Є. Андрущак

ПІДХІД ЯКІСНОГО АНАЛІЗУ ДИНАМІЧНОЇ СИСТЕМИ НА ОСНОВІ ДЕРЕВА РІШЕНЬ

У роботі запропоновано підхід до якісного аналізу динамічної системи на основі функціонально-диференціальних рівнянь. При цьому на виході отримуємо дерево рішень для прогнозування форми траєкторії системи. Підхід доведено до програмної реалізації у пакеті Java-класів.

Реальні динамічні математичні моделі залежать від чималої кількості параметрів, які мають різний вплив на поведінку моделі. Аналіз локальної чутливості дозволяє встановити параметри, що найбільше впливають на «силу сигналу». Біфуркаційний аналіз показує на те, від яких параметрів залежить якісна поведінка моделі. На жаль більшість методів аналізу моделей не можуть розглядати комбінації параметрів, оскільки простір для аналізу параметрів істотно зростає. Такі обмеження є дуже суттєвими оскільки переважно в моделях змінюються одночасно кілька параметрів (факторів).

Патологічні процеси можуть бути змодельованими системами звичайних диференціальних рівнянь. Аналіз моделей має на меті встановити головні характеристики системи, такі як чутливість до певних параметрів, або ж зміни у поведінці траєкторій. Чисельний аналіз моделі прагне кількісно або якісно описати відгуки системи на стимули і встановити умови, що визначають форму траєкторії.

Аналіз локальної чутливості встановлює зміни у виходах моделі при варіації параметрів моделі і у такий спосіб визначаються найвпливовіші фактори. Якісні методи аналізу моделі вивчають певні стійкі стани системи як виходи моделі, розглядаючи їх як різні форми патологічного процесу, як наприклад, субклінічну або ж летальну. Для невеликих моделей можна знайти в аналітичному вигляді умови притягання до певних стійких станів. Складніші моделі вимагають чисельних методів якісного аналізу відгуків системи. При цьому досліджується, чи варіація певного параметру веде до змін стійкості.

Багато якісних та кількісних методів аналізу систем є «уніваріативними», тобто зміна в поведінці системи або її виходах аналізується одночасно відповідно до варіації щодо одного параметра. Тому такі методи можуть виявити як важливі параметри для траєкторій системи лише окремі параметри. Це не відображає багатопараметричні стратегії керування системами. Певні форми патологічних процесів, такі як гостра або ж хронічна, летальна, не можуть бути представлені ланцюжком однопараметричних умов. Рішення про форму патологічного процесу приймається на основі загальної реакції системи, а не впливу на «певну групу клітин». Отже, мультиваріативний метод аналізу є адекватнішим у порівнянні з уніваріативним підходом. Він має враховувати одночасні зміни кількох параметрів, що призводять до відгуку системи певної форми, а отже і встановлювати певні стратегії лікування, що включають кооперацію деяких популяцій.

В даній роботі представимо автоматизований мультиваріативний метод, що визначає умови на комбінації параметрів. Системи звичайних диференціальних рівнянь використовують як параметри швидкісні константи та початкові концентрації певних популяцій клітин.

Швидкісні константи зазвичай вважаються сталими при визначеній температурі. Водночас початкові концентрації популяцій клітин можуть суттєво відрізнятися залежно від стадії, типу клітин і ін. У роботі [Yvonne Koch, 2013] запропоновано мультиваріативний підхід на основі дерева рішень, що досліджує лише вплив початкових концентрацій у системах ЗДР при фіксованих швидкісних константах. У реальних моделях рідко коли можна використати значення швидкісних сталих з літератури, або безпосередньо їх виміряти. Як правило знаходять їх оцінки виходячи з часових рядів концентацій популяцій клітин.

Тому метою даної роботи є розробити мультиваріативний метод якісного аналізу моделі на основі звичайних диференціальних рівнянь, що зводиться до дерева рішень і враховує як швидкісні константи, так і початкові умови.

Мультиваріативний метод якісного аналізу моделі на основі функціонально-диференціальних рівнянь (ФДР)

Завданням методу є встановлення механізмів багатопараметричних впливів у моделі ФДР.

Загальні ідеї методу були розроблені в роботі [Koch, 2013] для випадку ЗДР. У даній роботі його буде розвинено для ФДР. При цьому використовуємо підхід Монте-Карло, який полягає у випадковій генерації параметрів та побудові на їх основі моделі ФДР. Далі застосовують алгоритм індукції дерева рішень. Зауважимо, що метод роботи [Koch, 2013] застосовувався лише для дослідження впливу початкових умов на траєкторію системи ФДР. В даній роботі метод буде розвинено до швидкісних параметрів системи функціонально-диференціальних рівнянь також.

Отже, припускається існування моделі на основі ФДР при початкових значеннях та швидкісних параметрах із заданих інтервалів:

$$\frac{dx(t)}{dt} = f(x(t), x_t, p),\ t > 0, \qquad (1)$$

$x(s) = x_0 \equiv const$, $s \in [-\tau, 0]$, $\tau > 0$ – сталі початкові умови.

Параметри:

$$p \in P = \{(p_1, p_2, ..., p_m):$$
$$p_j^{\min} \le p_j \le p_j^{\max},\ j = \overline{1,m}\} \subset R^m,$$

а початкові умови:

$$x_0 \in X_0 = \{(x_{0,1}, x_{0,2}, ..., x_{0,n}):$$
$$x_{0,i}^{\min} \le x_{0,i} \le x_{0,i}^{\max},\ i = \overline{1,n}\} \subset R^n.$$

Тут $x(t) \in R^n$, $x_t \in C^1[-\tau, 0)$, $f: R^n \times C^1[-\tau, 0) \times R^m \to R^n$. Використане позначення для вектор-відрізка траєкторії:

$$x_t = \{x(t+s): s \in [-\tau, 0)\}.$$

Далі випадковим чином генеруватимемо початкові значення та значення швидкісних параметрів, які б належали практично обгрунтованій області. Для кожного з наборів таких параметрів здійснюється інтегрування системи ФДР (1) з отриманням відповідних траєкторій. До отриманих результатів далі застосовується алгоритм індукції дерева рішень з метою знаходження певних шаблонів для прийняття рішень.

Отже, в цілому підхід включає такі п’ять кроків.

1. Означення класів траєкторій системи. Зазначимо, що можуть бути означені будь-які класифікаційні критерії залежно від моделі, що аналізується, та певної наукової проблематики. Наприклад, при аналізі моделей патологічних процесів використовуватимемо класи, пов’язані з формами патологічного процесу: субклінічна, гостра, хронічна, летальна. Для позначення класу траєкторії вводиться атрибут класу C, який приймає одне з K дискретних значень $C \in \{1, ..., K\}$.

2. Генерація матриці випадкових початкових значень та швидкісних параметрів. Для того, щоб дослідити весь простір початкових значень та швидкісних параметрів щодо генерації класів траєкторій, визначених на першому кроці, генерується матриця випадкових початкових значень та швидкісних параметрів на основі ймовірнісних розподілів у визначених інтервалах. У даній роботі ми припускаємо, що початкові значення та швидкісні параметри розподілені рівномірно на інтервалах. Кожен стовпчик відповідає множині значень одного параметра – або початкове значення, або швидкісний параметр. Кожен рядок є набором початкових значень та швидкісних параметрів для одного запуску моделі на основі ФДР:

$$M=\begin{pmatrix} x_{0,1}^{1} & x_{0,2}^{1} & \vdots & x_{0,n}^{1} & p_1^1 & p_2^1 & \vdots & p_m^1 \\ x_{0,1}^{2} & x_{0,2}^{2} & \vdots & x_{0,n}^{2} & p_1^2 & p_2^2 & \vdots & p_m^2 \\ \cdots & \cdots & & \cdots & \cdots & \cdots & & \cdots \\ x_{0,1}^{N} & x_{0,2}^{N} & \vdots & x_{0,n}^{N} & p_1^N & p_2^N & \vdots & p_m^N \end{pmatrix}\in$$

$$\in R^{N\times(n+m)}.$$

3. Запуск моделі і класифікація набору вхідних даних. Кожен набір початкових значень та швидкісних параметрів, згенерованих на другому кроці, використовуються в якості входу для моделі на основі ФДР. Чисельне інтегрування рівнянь здійснюється за допомогою методу Адамса [Хайрер, Нерсет]. Вихідні траєкторії класифікуються на основі критеріїв, запропонованих на першому кроці. Виходячи з результатів класифікації наборам початкових значень і швидкісних параметрів приписуються відповідні атрибути класів.

4. Побудова матриці залежностей між початковими значеннями та між швидкісними параметрами. Метод припускає, що для форми траєкторій системи співвідношення між початковими значеннями та між швидкісними значеннями є набагато важливішими порівняно з їх абсолютними значеннями. Тому будується матриця, що включає інформацію у категорованому кодованому вигляді про співвідношення між початковими значеннями та між швидкісними параметрами, згенерованими на кроці 2.

$C_l\in\{1,...,K\}$, $l=\overline{1,N}$ – значення атрибуту класу, пов’язані з відповідними формами траєкторій.

Отже, на даному кроці чисельні значення початкових значень та швидкісних параметрів трансформуються у категоріальні значення атрибутів наборів навчальних даних. Оскільки ймовірність рівності випадкових чисел дорівнює нулю, то матриця D виглядає свого роду «бінаризацією» співвідношень між початковими значеннями та між швидкісними параметрами. Тобто матриця D включатиме лише значення 0 та 2.

$$D=\begin{pmatrix} x_{0,1}\otimes x_{0,2} & \vdots & x_{0,1}\otimes x_{0,n} & \vdots & x_{0,n-1}\otimes x_{0,n} & p_1\otimes p_2 & \vdots & p_1\otimes p_m & \vdots & p_{m-1}\otimes p_m & C \\ x_{1,2}^{1} & \vdots & x_{1,n}^{1} & \vdots & x_{n-1,n}^{1} & p_{1,2}^{1} & \vdots & p_{1,m}^{1} & \vdots & p_{m-1,m}^{1} & C_1 \\ x_{1,2}^{2} & \vdots & x_{1,n}^{2} & \vdots & x_{n-1,n}^{2} & p_{1,2}^{2} & \vdots & p_{1,m}^{2} & \vdots & p_{m-1,m}^{2} & C_2 \\ \cdots & & \cdots & & \cdots & \cdots & & \cdots & & \cdots & \cdots \\ x_{1,2}^{k} & \vdots & x_{1,n}^{k} & \vdots & x_{n-1,n}^{k} & p_{1,2}^{k} & \vdots & p_{1,m}^{k} & \vdots & p_{m-1,m}^{k} & C_k \end{pmatrix}.$$

Тут

$$x_{i,j}^{l}=\begin{cases}0, & if \quad x_{0,i}^{l}<x_{0,j}^{l}\\ 1, & if \quad x_{0,i}^{l}=x_{0,j}^{l}\\ 2, & if \quad x_{0,i}^{l}>x_{0,j}^{l}\end{cases},\quad i,j=\overline{1,n},\ i\neq j,\ l=\overline{1,N};$$

$$p_{i,j}^{l}=\begin{cases}0, & if \quad p_i^l<p_j^l\\ 1, & if \quad p_i^l=p_j^l\\ 2, & if \quad p_i^l>p_j^l\end{cases},\quad i,j=\overline{1,m},\ i\neq j,\ l=\overline{1,N}.$$

5. Застосування алгоритму індукції дерева рішень до співвідношень між початковими значеннями та між швидкісними параметрами. Матриця бінарних співвідношень D, побудована на кроці 4, є набором навчальних даних для алгоритму індукції дерева рішень. Побудоване дерево рішень міститиме перевірку співвідношень між початковими значеннями та швидкісними параметрами у своїх вузлах. В якості листків дерева будуть класи траєкторій моделі $C \in \{1,...,K\}$.

Програмна реалізація. Для реалізації методу розроблено пакет Java-класів decision_tree.fde. До складу пакету входять класи (рис. 1).

Attribute – представляє атрибути, що відповідають взаємозв'язкам між початковими значеннями та швидкісними параметрами;

AttributeListPeer – клас-слуга по обслуговуванню запитів менеджера даних щодо атрибутів взаємозв'язків;

Attribute_for_list – клас, що містить усю інформацію про атрибути взаємозв'язків, причому у вигляді хеш-таблиць, що використовується при визначенні умов поділу;

DataManager – клас – менеджер даних для отримання інформації з бази даних через посередництво відповідних класів-слуг;

DecisionTree – клас, що реалізує метод індукції дерева рішень на основі таблиці взаємозв'язків між початковими значеннями та між швидкісними константами (Додаток 1);

InitialValuesPeer – клас-слуга для генерації випадкових початкових значень;

MultiVariateMethod – клас для реалізації мультиваріативного методу, представленого в роботі – головний клас пакету;

RateConstantsPeer – клас-слуга для генерації випадкових швидкісних констант у відповідності з інформацією в таблиці rate_constants;

TuplesPeer – клас-слуга для формування і обробки навчальних наборів, що використовуватиметься у класифікаційному алгоритмі.

У класі MultiVariateMethod (рис. 2) здійснюється генерація випадкових значень параметрів (крок 2):

```
M_x0 = dm.getRandomInitialValues();
M_rateConstants =
dm.getRandomRateConstants();
```

Далі запускається аплет інтегрування системи ФДР. При цьому експерт здійснює вибір форми отриманої траєкторії (крок 3). Після цього запускається крок генерації матриці взаємозв'язків параметрів

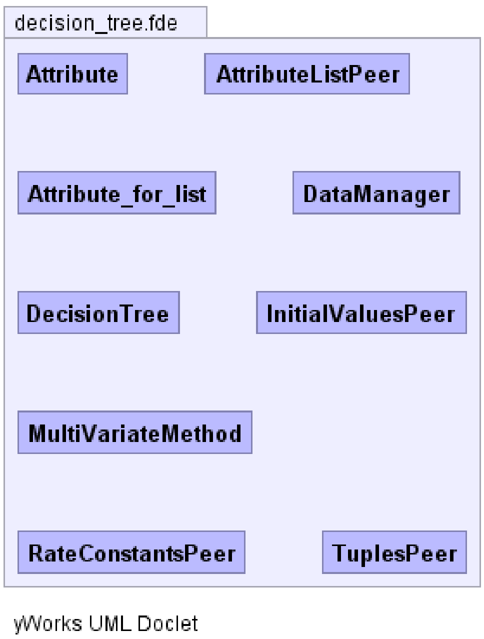

Рис. 1. Пакет decision_tree.fde

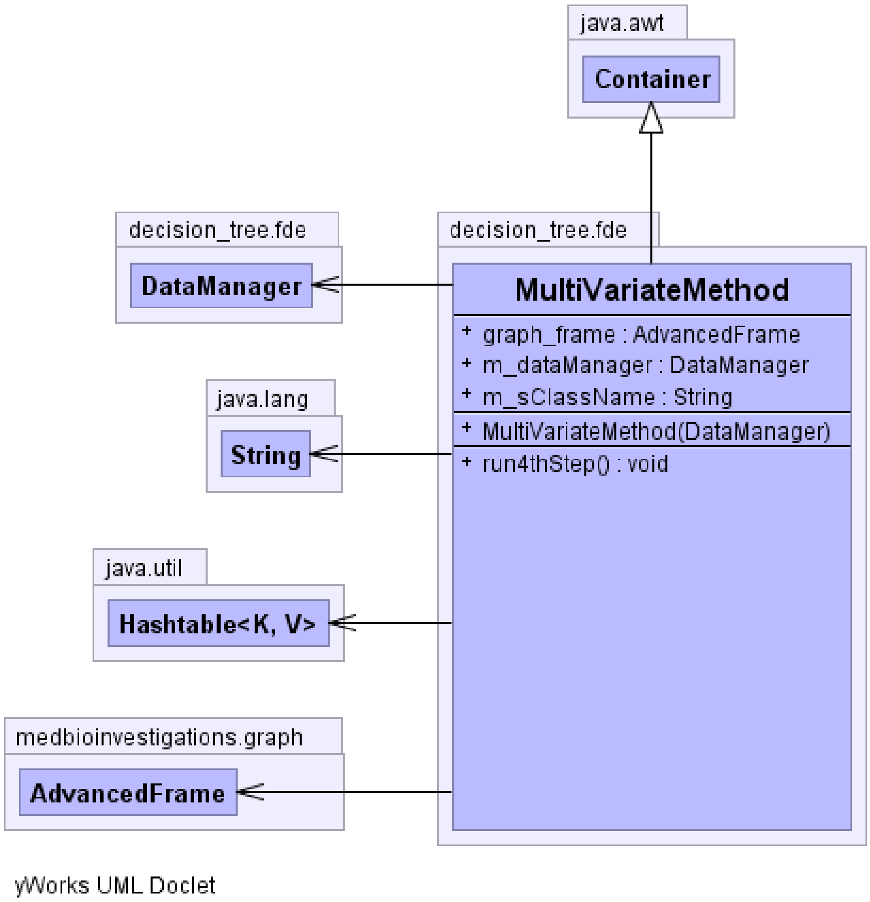

Рис. 2. UML-діаграма класу MultiVariateMethod

(крок 4). Зауважимо, що послідовність кроків 2–4 може виконуватися як завгодно багато разів. У будь-який момент користувач можу запустити алгоритм індукції дерева рішень (крок 5):

dtDecision_tree_FDE = new decision_tree.fde.DecisionTree(dmtnRoot, dataManager_FDE, htAttribute_list, "Information gain").

База даних fde, що використовується в пакеті, реалізована в СУБД MySQL. Вона включає такі таблиці (рис. 3):

attribute – опис атрибутів для побудови дерева рішень, тобто взаємозв'язків між початковими значеннями та між швидкісними константами;

categorized_data – навчальні набори, що використовуються в класифікаційному алгоритмі (в даному випадку індукції дерева рішень) і представляють собою матрицю D на четвертому кроці;

init_values_values – матриця згенерованих випадковим чином початкових значень:

$$\begin{pmatrix} x_{0,1}^{1} & x_{0,2}^{1} & \vdots & x_{0,n}^{1} \\ x_{0,1}^{2} & x_{0,2}^{2} & \vdots & x_{0,n}^{2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{0,1}^{N} & x_{0,2}^{N} & \vdots & x_{0,n}^{N} \end{pmatrix} \in R^{N\times n}$$

initial_values – опис початкових значень (включаючи мінімальні та максимальні значення);

parameter_kind – вид параметра;

rate_constants – опис швидкісних констант (включаючи мінімальні та максимальні значення);

rate_constants_values – матриця згенерованих випадковим чином швидкісних констант:

$$\begin{pmatrix} p_{1}^{1} & p_{2}^{1} & \vdots & p_{m}^{1} \\ p_{1}^{2} & p_{2}^{2} & \vdots & p_{m}^{2} \\ \cdots & \cdots & \cdots & \cdots \\ p_{1}^{N} & p_{2}^{N} & \vdots & p_{m}^{N} \end{pmatrix} \in R^{N\times m}.$$

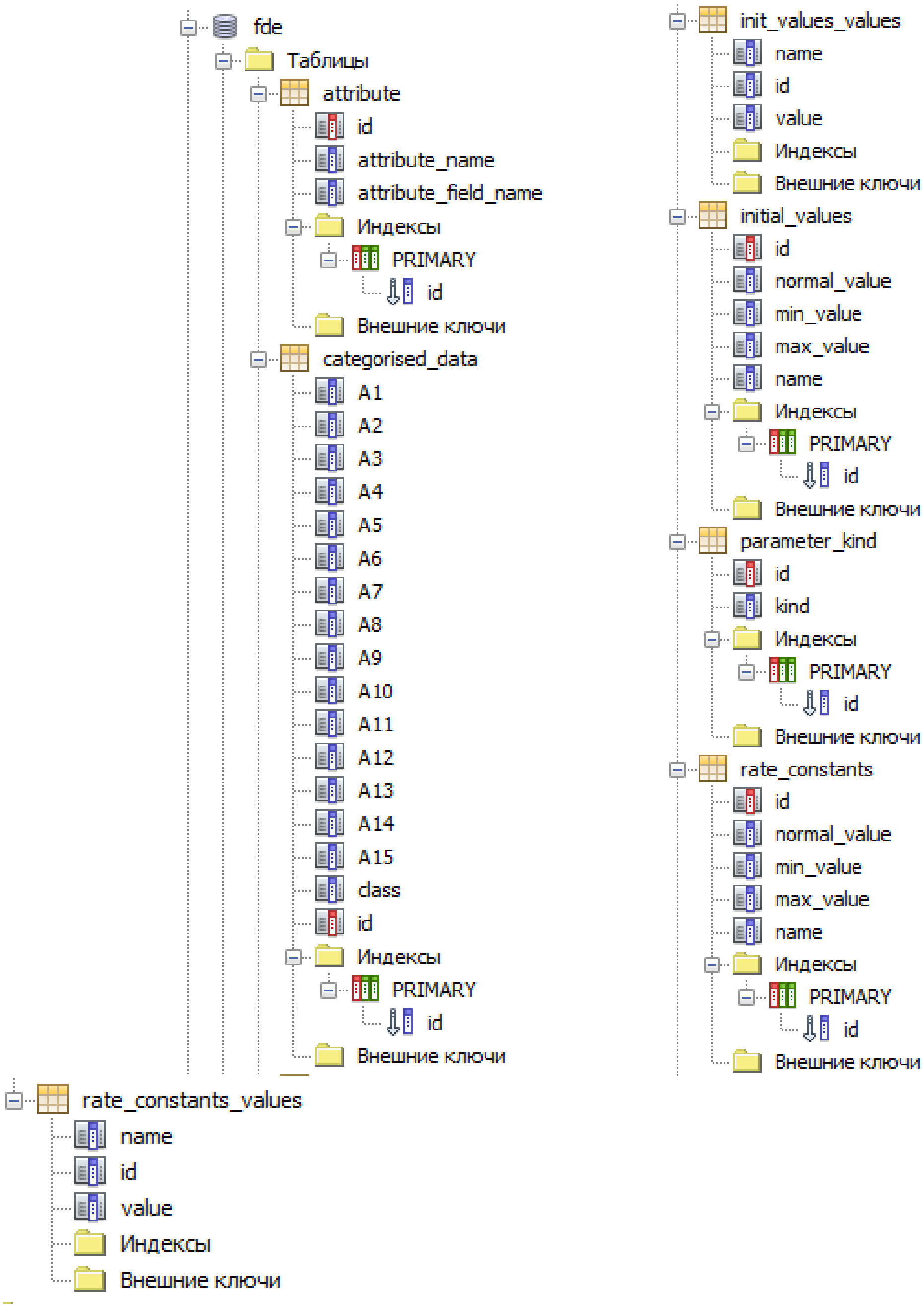

Рис. 3. Таблиці бази даних fde

Висновки

На п’ятому кроці алгоритму до набору навчальних даних можуть бути застосовані інші алгоритми data mining – наприклад, з метою отримання класифікаційних правил – метод послідовного покриття. Набори класифікаційних правил також можуть бути отримані виходячи з дерева рішень.

1. *Хайрер Э., Нерсетт С., Ваннер Г.* Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. – М.: Мир, 1990. – 512 с.
2. *Aho A., Hopcroft J., Ullman J.* Construction and analysis of computational algorithms. – M.: World. – 1979. – 536 p.
3. *Gvozdetska I.S.* Mathematical models of tumor growth based on the dynamics Gomperz // Dis. for the degree of Candidate of Sc. sciences. – Ternopol. – 2012. – 130 p.
4. *Clemens Kühn, Christoph Wierling, Alexander Kuhn, Edda Klipp, Georgia Panopoulou, Hans Lehrach and Albert J Poustka*. Monte Carlo analysis of ODE Model of the Sea Urchin Endomesoderm Network // BMC Systems Biology. – 2009. – 3:83.
5. *Essex B., Healy M.* Evaluation of a rule base for decision making in general practice // Brit-

ish Journal of General Practice. – 1994. – 44. – P. 211–213.
6. *Hayrer E., Nersett S., Wanner G.* Solutions of ordinary differential equations. Non-rigid task. – M .: Peace, 1990. – 512 p.
7. *Laupacis A., Secar N., Stiell I.G.* Clinical prediction rules: A review and suggested modifications of methodological standards. JAMA 1997. – 277. P. 488–494.
8. *Lea A. Segel.* Mathematical Models in Molecular and Cellular Biology. CUP Archive, 1980. – 757 p.
9. *Stiell I.G., Wells G.A.* Methodologic Standards for the Development of Clinical Decision Rules in Emergency Medicine. Annals of Emergency Medicine. – 1999. – 33:4. – P. 437–447.
10. *Yvonne Koch, Thomas Wolf, Peter K. Sorger, Roland Eils, Benedikt Brars.* Decision-Tree Based Model Analysis for Efficient Identification of Parameter Relations Leading to Different Signaling States // PLOS ONE | www.plosone.org, December 2013, Vol. 8, Issue 12, e82593.

Одержано 05.09.2014

Про автора:

Андрущак Ігор Євгенович,
кандидат технічних наук,
доцент кафедри
«Комп’ютерних технологій професійного навчання».

Місце роботи автора:

Луцький національний технічний університет
м. Луцьк,
вул. Кічкарівська, 31.
Тел.: (068) 329 00 00,
(093) 452 00 00,
(066) 409 00 00.
E-mail: westa@land.ru