

УДК 61:004.651(075.8)

Андрущак І.Є.

Луцький національний технічний університет

МУЛЬТИВАРІАТИВНИЙ МЕТОД ЯКІСНОГО АНАЛІЗУ МОДЕЛІ НА ОСНОВІ КЛАСИФІКАЦІЙНИХ ПРАВИЛ.

Андрущак І.Є. Мультиваріативний метод якісного аналізу моделі на основі класифікаційних правил. У роботі запропоновано підхід якісного аналізу системи протипухлинного імунітету на основі мультиваріативного методу, що включає алгоритм послідовного покриття. При цьому на виході отримуємо множину класифікаційних правил для прогнозування форми патологічного процесу.

Ключові слова: мультиваріативний метод, якісний аналіз, протипухлинний імунітет, диференціальні рівняння із запізненням, класифікаційні правила.

Андрущак И.Е. Мультивариативный метод анализа модели на основе классификационных правил. В работе предложен подход анализа системы противоопухолевого иммунитета на основе мультивариативного метода, включая алгоритм последовательного покрытия. При этом на выходе получаем множество классификационных правил для прогнозирования формы патологического процесса.

Ключевые слова: мультивариативный метод, качественный анализ, противоопухолевый иммунитет, дифференциальные уравнения с опозданием, классификационные правила.

Andrushchak I.Ye. Multivariate method of qualitative analysis models based on rules classification. In this work proposes an approach qualitative analysis of the antitumor immunity from multivariate method comprising sequential algorithm coverage. Thus the output we get the set of classification rules for predicting forms of pathological process.

Keywords: multivariate method, qualitative analysis, antitumor immunity, differential equations with delay, classification rules.

Завданням методу є встановлення механізмів багатопараметричних впливів у моделі протипухлинного імунітету.

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:

$$\frac{dL(t)}{dt} = \alpha_L L(t) \ln \frac{\theta_L}{L(t)} - \gamma_L F(t) L(t), \quad (1)$$

$$\frac{dC(t)}{dt} = \xi(m) \alpha L(t - \tau) F(t - \tau) - \mu_C (C - C_0), \quad (2)$$

$$\frac{dF(t)}{dt} = \rho C - (\mu_f + \eta \gamma_L L(t)) F(t), \quad (3)$$

$$\frac{dm(t)}{dt} = \sigma L(t) - \mu_m m(t), \quad (4)$$

з початковими умовами при $t \in [-\tau, 0]$:

$$V(t) = V_0, F(t) = F_0, C(t) = C_0, m(t) = 0.$$

Тут α_L – коефіцієнт розмноження пухлинних клітин; θ_L – життєва ємність популяції пухлинних клітин; γ_L – коефіцієнт, що визначає ймовірність нейтралізації (руйнування) пухлинної клітини антитілом; α – коефіцієнт, що зумовлює ймовірність зустрічі антиген-

антитіло; μ_c – коефіцієнт, обернений до часу життя плазмоклітин; ρ – швидкість виробництва антитіл однією плазмоклітиною; μ_f – коефіцієнт, обернено пропорційний до часу розпаду антитіл; η – число антитіл, що вимагається на нейтралізацію одного антигена; σ – коефіцієнт, що визначає швидкість загибелі клітин за рахунок пошкоджуючої дії антигена; μ_m – коефіцієнт, що враховує швидкість відновлення пошкодженого органу; τ – фаза запізнення (час, за який здійснюється формування каскаду плазмоклітин); $\xi(m)$ – неперервна незростаюча функція ($0 \leq \xi(m) \leq 1$), що характеризує порушення нормального функціонування імунної системи через значне пошкодження органа-мішені.

Параметри:

$$\begin{aligned} p \in P = \{(\alpha_L, \theta_L, \gamma_L, \alpha, \mu_c, \rho, \mu_f, \eta, \sigma, \mu_m) : \alpha_L^{\min} \leq \alpha_L \leq \alpha_L^{\max}, \\ \theta_L^{\min} \leq \theta_L \leq \theta_L^{\max}, \gamma_L^{\min} \leq \gamma_L \leq \gamma_L^{\max}, \alpha^{\min} \leq \alpha \leq \alpha^{\max}, \mu_c^{\min} \leq \mu_c \leq \mu_c^{\max}, \\ \rho^{\min} \leq \rho \leq \rho^{\max}, \mu_f^{\min} \leq \mu_f \leq \mu_f^{\max}, \eta^{\min} \leq \eta \leq \eta^{\max}, \sigma^{\min} \leq \sigma \leq \sigma^{\max}, \\ \mu_m^{\min} \leq \mu_m \leq \mu_m^{\max}\} \subset R^{10}, \end{aligned}$$

а початкові умови

$$\begin{aligned} (L_0, F_0, C_0, m_0) \in X_0 = \{(L_0, F_0, C_0, m_0) : L_0^{\min} \leq L_0 \leq L_0^{\max}, F_0^{\min} \leq F_0 \leq F_0^{\max}, \\ C_0^{\min} \leq C_0 \leq C_0^{\max}, m_0^{\min} \leq m_0 \leq m_0^{\max}\} \subset R^4. \end{aligned}$$

Далі випадковим чином генеруватимемо початкові значення та значення швидкісних параметрів, які б належали практично обґрунтованій області. Для кожного з наборів таких параметрів здійснюється інтегрування системи (1)-(4) з отриманням відповідних траєкторій. До отриманих результатів далі застосовується алгоритм індукції дерева рішень з метою знаходження певних шаблонів для прийняття рішень.

Отже, вцілому підхід включає такі п'ять кроків.

1. Означення класів траєкторій системи. Зазначимо, що в практичних застосуваннях переважно мають справу з набагато складнішими формами поведінки, щоб охарактеризувати їх поняттями «стійка-нестійка» і відповідно вдатися до аналізу власних значень або ж експонент Ляпунова динамічної системи. Тому визначення якісних форм процесу доцільно передати до компетенції експертів. В даному випадку використовуватимемо класи, пов'язані з формами патологічного процесу: субклінічна, гостра, хронічна, летальна. Для позначення класу траєкторії вводиться атрибут класу C , який приймає одне з 4-х дискретних значень $C \in \overline{1,4}$.

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 параметру -- або початкове значення, або швидкісний параметр. Кожен рядок є набором початкових значень та швидкісних параметрів для одного запуску моделі на основі ФДР:

$$M = \left(\begin{array}{cccccccccccc} V_0^1 & F_0^1 & C_0^1 & m_0^1 & \alpha_L^1 & \theta_L^1 & \gamma_L^1 & \alpha^1 & \mu_c^1 & \rho^1 & \mu_f^1 & \eta^1 & \sigma^1 & \mu_m^1 \\ V_0^2 & F_0^2 & C_0^2 & m_0^2 & \alpha_L^2 & \theta_L^2 & \gamma_L^2 & \alpha^2 & \mu_c^2 & \rho^2 & \mu_f^2 & \eta^2 & \sigma^2 & \mu_m^2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ V_0^N & F_0^N & C_0^N & m_0^N & \alpha_L^N & \theta_L^N & \gamma_L^N & \alpha^N & \mu_c^N & \rho^N & \mu_f^N & \eta^N & \sigma^N & \mu_m^N \end{array} \right) \in R^{N \times 14}$$

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 протипухлинного імунітету. 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 категоризованому кодованому вигляді про співвідношення між початковими значеннями та між швидкісними параметрами.

$C_i \in \overline{1,4}$ – 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.

5. Застосування алгоритму послідовного покриття до співвідношень між початковими значеннями та між швидкісними параметрами. Матриця бінарних співвідношень D побудована на кроці 4, є набором навчальних даних для алгоритму послідовного покриття. Побудований набір класифікаційних правил міститиме перевірку співвідношень між початковими значеннями та швидкісними параметрами в антецедентах. В якості консеквентів правил будуть класи траєкторій моделі $C \in \overline{1,4}$.

Програмна реалізація мультиваріативного методу для моделі протиухлинного імунітету. Для реалізації методу розроблено пакет Java-класів rule.model. До складу пакету входять класи (рис.1):

beans-класи Attribute, Attribute_for_list для роботи з даними відповідних таблиць та Rule – для представлення правил. SQL-запити щодо отримання відповідних даних реалізовано в класах AttributeListPeer та TuplesPeer.

У класі Rule_set зберігається набір навчальних правил. До того ж даний клас безпосередньо реалізує алгоритм послідовного покриття. Клас містить члени: менеджер даних m_dataManager, хеш-таблиці наборів навчальних даних m_htTuples, усіх атрибутів з їх можливими значеннями m_htAtt_vals та безпосередньо множину правил m_htRule_set.

У конструкторі класу Rule_set здійснюється побудова хеш-таблиць m_htTuples та m_htAtt_vals, а також застосування алгоритму послідовного покриття – через виклик методу Sequential_covering(m_htTuples, m_htAtt_vals). Отримана множина правил виводиться в текстовий файл.

Клас Rule призначений для зберігання окремих правил. Його членами класу є дві хеш-таблиці: m_htAntecedent – для зберігання антецеденту правила та m_htConsequent – для консеквенту. За допомогою методу:

```
public void conjunctCondition(Attribute_for_list attribute, String sAttribute_value)
здійснюється кон'юнкція нової умови до правила. За допомогою методу:
public Rule copy()
```

створюється «глибока» копія правила. При цьому використовується протокол JOS (Java Object Serialization).

Підрахунок кількості позитивних та негативних навчальних наборів здійснюється у методах класу TuplesPeer.

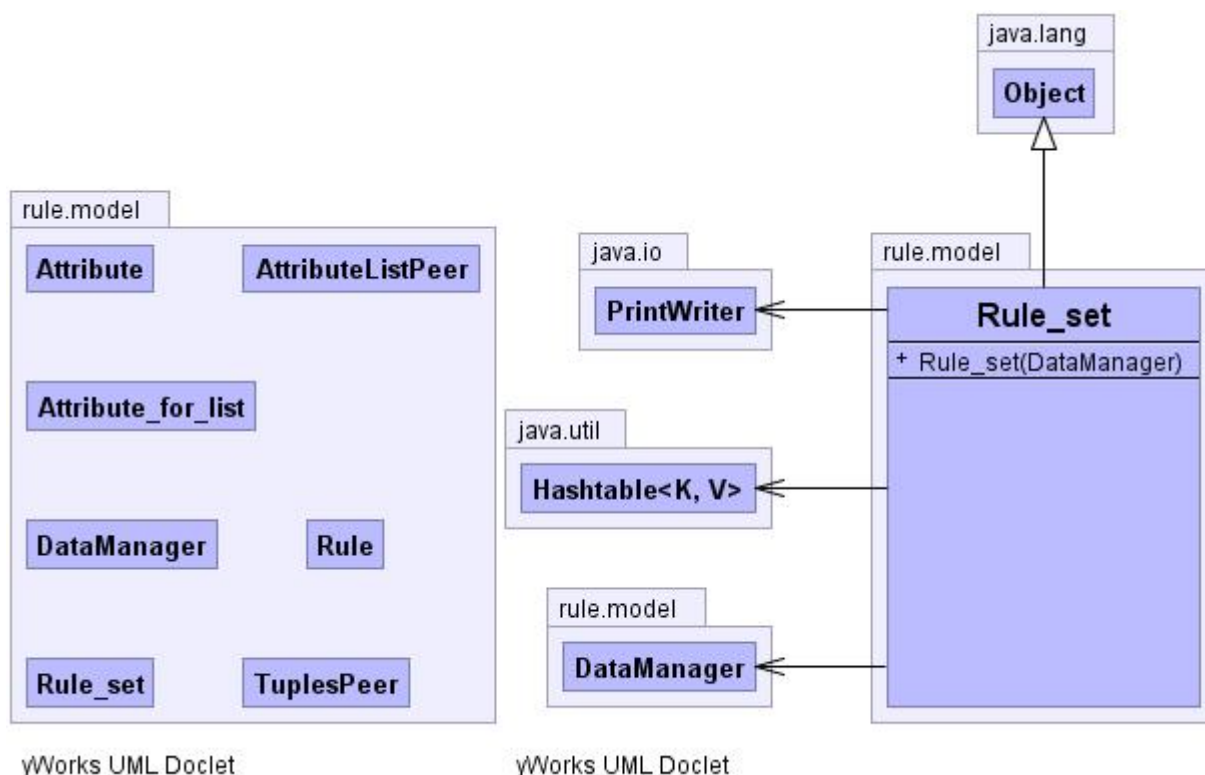


Рис. 1. Пакет rule.model

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

```
rule.model.Rule_set rule_set = new rule.model.Rule_set(rule_dataManager, sql0, rules_file_url);
```

База даних `fde`, що використовується в пакеті, реалізована в СУБД `MySQL`. Вона включає такі таблиці:

`attribute` – опис атрибутів для побудови класифікаційних правил, тобто взаємозв'язків між початковими значеннями та між швидкісними константами;

`categorized_data` – навчальні набори, що використовуються в класифікаційному алгоритмі (в даному випадку алгоритму послідовного покриття) і представляють собою матрицю D на четвертому кроці;

`init_values_values` – матриця згенерованих випадковим чином початкових значень:

$$\begin{pmatrix} V_0^1 & F_0^1 & C_0^1 & m_0^1 \\ V_0^2 & F_0^2 & C_0^2 & m_0^2 \\ \dots & \dots & \dots & \dots \\ V_0^N & F_0^N & C_0^N & m_0^N \end{pmatrix} \in R^{N \times 4}$$

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

`parameter_kind` – вид параметру;

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

rate_constants_values – матриця згенерованих випадковим чином швидкісних констант:

$$\left(\begin{array}{cccccccccc} \alpha_L^1 & \theta_L^1 & \gamma_L^1 & \alpha^1 & \mu_C^1 & \rho^1 & \mu_f^1 & \eta^1 & \sigma^1 & \mu_m^1 \\ \alpha_L^2 & \theta_L^2 & \gamma_L^2 & \alpha^2 & \mu_C^2 & \rho^2 & \mu_f^2 & \eta^2 & \sigma^2 & \mu_m^2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \alpha_L^N & \theta_L^N & \gamma_L^N & \alpha^N & \mu_C^N & \rho^N & \mu_f^N & \eta^N & \sigma^N & \mu_m^N \end{array} \right) \in R^{N \times 10}$$

Особливості програмної реалізації моделі протипухлинного імунітету. Пакет rule.model може бути використаний для широкого класу систем функціонально-диференціальних рівнянь. Для цього модель на основі функціонально-диференціальних рівнянь повинна бути реалізована у вигляді відповідного пакету Java-класів.

Даний пакет було розроблено в роботі [Гвоздецька, 2012]. Для інтеграції з пакетом decision_tree.fde в класі TumorImmunitySystemGraph, що здійснює графічну візуалізацію моделі, поряд з існуючим було додано новий конструктор, що використовує посилання за значенням на інстанцію класу MultiVariateMethod. У цьому конструкторі додатково створюється об'єкт класу JComboBox, що дозволяє вибирати форму траєкторії та запускати виконання 4-го кроку методу:

```
String[] classStrings = {"subclinical","chronic","acute","lethal"};
JComboBox m_jcbClassName = new JComboBox(classStrings);
m_jcbClassName.addActionListener(new ActionListener() {
    public void actionPerformed(ActionEvent e) {
        JComboBox jcbClass = (JComboBox) e.getSource();
        m_sClassName = (String)jcbClass.getSelectedItem();
        mvm.m_sClassName = m_sClassName;
        ((AdvancedFrame)getParent()).dispose();
        mvm.run4thStep();
    }
});
```

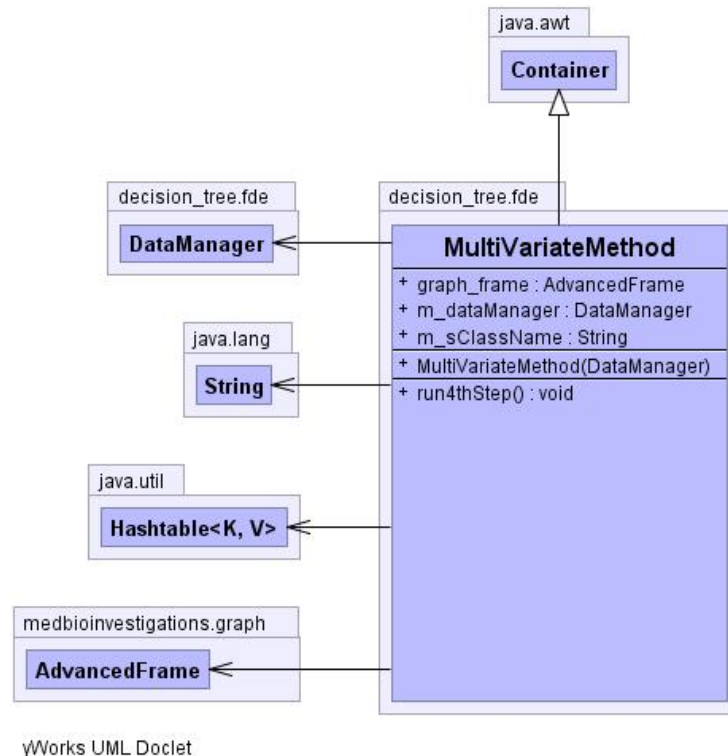


Рис.2. UML-діаграма класу MultiVariateMethod

Також слід внести відповідні зміни у таблиці бази даних fde в такому порядку:

- описати усі початкові значення та швидкісні константи в таблицях initial_values та rate_constants відповідно;

- описати залежності між початковими значеннями та між швидкісними константами, які досліджуватимемо, у таблиці attribute;

- у таблиці categorised_data створити поля у відповідності з даними таблиці attribute.

Існує чимало реалізацій алгоритму послідовного покриття побудови класифікаційних правил, зокрема у складі програм Mathcad, StatSoft Statistica і ін. Суттєвою перевагою даної реалізації є те, що алгоритм розвинено до Інтернет-проекту, де до формування навчальних наборів залучаються експерти з Інтернет-доступом, що значно розширює та балансує множину навчальних наборів.

1. Ахо А., Хопкрофт Дж., Ульман Дж., Построение и анализ вычислительных алгоритмов. – М.: Мир. – 1979. – 536 с.
2. Гвоздецька І.С. Математичні моделі пухлинного росту на основі динаміки Гомперца / Дис. на здобуття наукового ступеня кандидата техн. наук. – Тернопіль: 2012. – 130 с.
3. Марценюк В.П. Математичні моделі в системі підтримки прийняття рішень страхового забезпечення лікування онкологічних захворювань: підхід на основі динаміки Гомперца / В.П. Марценюк, І.Є. Андрушак, І.С. Гвоздецька, Н.Я. Климук // Доповіді Національної академії наук України. –2012. – №10. – С. 34 – 39.
4. Марценюк В.П. Підхід на основі актуарних математичних моделей до задач страхової медицини / В. П. Марценюк, І.Є. Андрушак, Н.Я. Климук // Медична інформатика та інженерія. Науково-практичний журнал. – 2010. – №4. – С. 85 – 87.
5. Хайрер Э., Нерсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. – М.: Мир. –1990. – 512 с.
6. Clemens Kuhn, 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 doi: 1186/1752-0509-3-83.
7. Essex B, Healy M: Evaluation of a rule base for decision making in general practice. British Journal of General Practice, 1994, 44. – P. 211 – 213.
8. Laupacis A, Secar N, Stiell IG: Clinical prediction rules: A review and suggested modifications of methodological standards. JAMA 1997; 277. – P. 88 – 494. Lea A. Segel, Mathematical Models in Molecular and Cellular Biology. CUP Archive, 1980 – 757 pp.
9. Stiell IG, Wells GA: 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, Volume 8, Issue 12, e82593.