КОМПЬЮТЕРНАЯ МАТЕМАТИКА

Математические модели в биологии и медицине

УДК 577.1

О.А. ВАГІС

ДОСЛІДЖЕННЯ ПОКАЗНИКІВ ЗДОРОВ'Я НА ОСНОВІ ЛАНЦЮГІВ МАРКОВА

Розглянуто одну з методик моделювання показників здоров'я населення на основі математичного апарату ланцюгів Маркова. Ця методика дозволяє за початковими даними обчислювати набори безумовних імовірностей та здійснювати імовірнісний прогноз стану здоров'я окремої людини або груп населення на майбутні періоди життя.

Вступ. Застосування математичного моделювання в медицині і, зокрема, в охороні здоров'я пов'язано із значним розширенням можливостей обчислювальної техніки. Оскільки за своєю природою біологічні послідовності, до яких відносяться показники здоров'я, є дискретними об'єктами, то для їх опису необхідно використовувати апарат теорії імовірностей і математичної статистики.

Дослідження методів розпізнавання складних дискретних об'єктів зумовлена широким спектром нових, що виникли за останні часи на межі математичної та інших наук, зокрема, біології та медицини. Розроблено методи розпізнавання дискретних об'єктів на таких структурах, як байєсівські мережі, моделі ланцюгів Маркова, незалежні ознаки [1, 2]. За допомогою моделей ланцюгів Маркова здійснено розпізнавання прогресії гліом головного мозку на основі модифікованих показників швидкості осідання еритроцитів (МШОЕ), які є біофізичним маркером прогнозу збільшення пухлин [3]. Оскільки показники МШОЕ змінюються за часом, то їх можна вважати залежними випадковими величинами, тому доцільно було перевірити роботу байєсівської процедури розпізнавання на моделі ланцюгів Маркова. Аналіз результатів розпізнавання показав перевагу цієї моделі у порівнянні з методом найближчих сусідів [2].

Моделі Маркова описують процес зміни стану системи за часом, тому мають широке застосування для дослідження показників здоров'я окремих пацієнтів або груп населення [4, 5].

Статистичні дані щодо тривалості життя, причин смертності, захворюваності за класами хвороб та інші показники збираються всіма медичними закладами. Для підвищення ефективності роботи медичних установ розробляються моделі, що дозволяють порівнювати показники здоров'я населення різних регіонів, перевіряти гіпотези щодо діагнозу та розвитку хвороби пацієнта, прогнозувати протікання хвороби залежно від результатів обстеження хворого і методики лікування. Все це можливо реалізувати завдяки використанню імовірнісно-статистичних методів, а саме математичного апарата ланцюгів Маркова та інформації з медичних баз даних. Наведемо основні поняття цієї теорії.

Ланцюгом Маркова називають послідовність випробувань, у кожному з яких об'єкт (система) приймає тільки одне з k станів $E_1, E_2, ..., E_k$, причому умовна імовірність $p_{ij}(s)$ того, що на s-му кроці об'єкт знаходиться в стані j, за умови, що після попереднього $(s-1)$-го кроку він знаходився в стані i, не залежить від результатів попередніх кроків або іспитів. Тобто умовна імовірність $p_{ij}(s)$ – перехід із стану i в стан j не залежить від номеру s, і її можна записати скорочено – p_{ij}. Оскільки перше випробування не має попереднього, то початковому випробуванню відповідає безумовна імовірність $a_j = P(E_j)$ – імовірність здійснення стану об'єкта E_j при початковому випробуванні. Це і є ланцюг, де кожна наступна ланка з'єднана лише з попередньою і залежить лише від неї.

Нехай число станів об'єкта (системи) кінцеве і дорівнює k: $E_1, E_2, ..., E_k$. Усі можливі стани системи несумісні, тобто в кожен момент часу система може знаходитись тільки в одному з розглянутих станів. Випробування проходять через рівні проміжки часу, номер випробування називають номером кроку і він є параметром часу. Умовні імовірності p_{ij} називають перехідними імовірностями із стану E_i в стан E_j. Множину перехідних імовірностей $\{p_{ij}\}$ упорядковують у вигляді стохастичної матриці P:

$$\mathrm{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix}. \quad (1)$$

Оскільки ці стани складають повну систему подій, то сума імовірностей кожної строки цієї матриці дорівнює 1, тобто

$$\sum_{j=1}^{k} p_{ij} = 1, \ (i = 1,...,k). \quad (2)$$

Ланцюг Маркова, в якому перехідні імовірності не залежать від номера кроку, $p_{jm}(r) = p_{jm}(s) = p_{jm}$, називають однорідним. Якщо $p_{jm}(r) \neq p_{jm}(s)$, ланцюг Маркова називають неоднорідним.

Розглянемо простий приклад коли однорідний ланцюг Маркова включає лише два можливих стани здоров'я хворого: E_1 – покращення стану та E_2 – погіршення стану в післяопераційний період. Стохастичну матрицю переходів P можна записати у вигляді:

$$\begin{array}{cc} & \begin{array}{cc} E_1 & E_2 \end{array} \\ \mathrm{P} = \begin{array}{c} E_1 \\ E_2 \end{array} & \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}. \end{array}$$

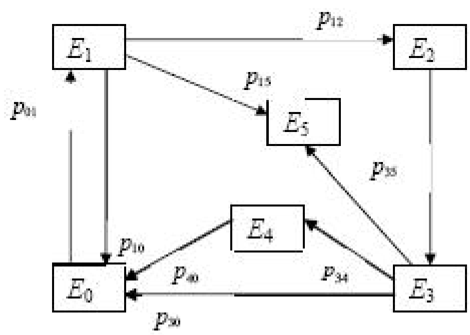
Розглянемо пари «день-стан» і відповідні імовірності переходу зі стану у стан за певний проміжок часу. Якщо в попередній день у хворого спостерігалося покращення стану $-E_1$, то сьогодні покращення стану слід очікувати з імовірністю $p_{11} = 1-\alpha$, а погіршення – $p_{12} = \alpha$. Аналогічно, якщо попереднього дня спостерігався стан E_2, то покращення наступить з імовірністю $p_{21} = \beta$, а погіршення стану буде залишатися з імовірністю $p_{22} = 1-\beta$.

У ході лікувально-діагностичного процесу стан хворого може бути описано великою кількістю різноманітних симптомів і показників: температура, тиск, дані рентгенологічного характеру, лабораторні дослідження і т. і. Різноманітні характеристики станів змінюються в різні моменти часу з різними темпами, і тому лікар, на основі накопиченого досвіду, на кожному етапі процесу може обрати деяку підмножину елементів показників для діагностики та перевірки математичними методами, зокрема на основі моделей Маркова.

Для аналізу ланцюгів Маркова використовують графи станів. Розглянемо елементарний граф шести станів здоров'я людини (рисунок), де E_0 – здоровий або відчуває себе таким; E_1– відчуває хворобу, але лікується самостійно; E_2 – звернувся до лікаря, проводиться обстеження; E_3 – діагноз поставлено, лікування; E_4 – реабілітація; E_5 – летальний вихід. Стрілками на графі вказані ненульові перехідні імовірності. Перехідна імовірність p_{11} на графі не вказана, означає що пацієнт знаходиться в стані E_1 і залишиться в цьому стані на наступному кроці. Якщо

$$p_{10} + p_{11} + p_{12} + p_{13} + p_{14} + p_{15} = 1,$$

то згідно рисунка $p_{10} + p_{11} + p_{12} + 0 + 0 + p_{15} = 1$. Тоді перехідна імовірність $p_{11} = 1-(p_{10} + p_{12} + p_{15})$ і аналогічно: $p_{22} = 1 - p_{23}$, $p_{33} = 1-(p_{30} + p_{34} + p_{35})$, $p_{44} = 1 - p_{40}$, $p_{55} = 1$. З рисунка видно, що із стану E_5 не виходить жодної стрілки, попавши в нього один раз система залишається в ньому назавжди, цей стан називається поглинаючим. На основі значень імовірностей, вказаних на графі, формується стохастична матриця станів. Представлена схема є спрощеною моделлю реального розвитку станів хворого.



РИСУНОК

Розглянемо приклад для n станів системи. Нехай $E_1, E_2, ..., E_n$ складають однорідний ланцюг Маркова. На кожному кроці матриця перехідних імовірностей p_{ij} є стохастична матриця P, елементи якої відомі (1). Оскільки початковий іспит не має попереднього, то йому відповідає безумовна імовірність $\alpha_j = P(E_j)$ – імовірність E_j на початку іспитів. Вектор $\{\alpha_j\}$ задано. Розглянемо алгоритм находження імовірностей станів на першому, другому та k-му кроці. Позначимо безумовну імовірність знаходження системи в стані E_j на k-му кроці, як $p_j(k)$. Тоді для всіх можливих станів $E_1, E_2, ..., E_n$ на кожному кроці k отримуємо вектор відповідних імовірностей:

$$p_1(k), p_2(k), ..., p_n(k), \qquad (3)$$

де $k = 1, ..., n$. Для вектора початкових імовірностей $\{\alpha_j\}$ використовуємо ті ж позначення:

$$p_1(0), p_2(0), ..., p_n(0), \qquad (4)$$

тобто початкові імовірності є безумовними імовірностями на нульовому кроці: $\alpha_j = p_j(0), \quad j = 1, ..., n$. Оскільки початковий іспит обов’язково закінчується одним з можливих станів і всі стани попарно несумісні, то виконується умова нормування (2):

$$\sum_j p_j(0) = 1.$$

Якщо на початку система знаходиться у відомому стані E_m, то початкові імовірності мають вигляд:

$$p_1(0) = 0\,;\; p_2(0) = 0\,;...;\; p_m(0) = 1\,;...;\; p_n(0) = 0.$$

Знайдемо вектор безумовних імовірностей (3) після першого кроку при $k=1$. Ці імовірності є перехідними імовірностями з стану E_m в будь-який інший стан E_j, $j=1,...,n$: $p_1(1)=p_{m1}$; $p_2(1)=p_{m2}$;...; $p_n(1)=p_{mn}$. Ці імовірності складають m-ю строку стохастичної матриці P.

У випадку, коли вектор початкових імовірностей (4) є випадковим, тобто не визначений стан, в якому система могла б бути на нульовому кроці, імовірності (4) складають повну систему гіпотез. Отже, імовірності станів системи на першому кроці можна знайти за формулою повної імовірності:

$$\{p_j(1)\}=\{p_j(0)\}\cdot P.$$

Далі аналогічно, вважаємо імовірності $p_1(1), p_2(1),..., p_n(1)$ як повну систему гіпотез, знаходимо імовірності на другому кроці $p_1(2), p_2(2),..., p_n(2)$ за формулою:

$$p_j(2)=\sum_{i=1}^{n} p_i(1)p_{ij}.$$

Сукупність усіх безумовних імовірностей на другому кроці у матричному вигляді:

$$\{p_j(2)\}=\{p_j(1)\}=\{p_j(0)\}\cdot P^2,$$

де P^2 – квадрат стохастичної матриці.

У загальному випадку на *k*-му кроці отримуємо:

$$\{p_j(k)\}=\{p_j(0)\}\cdot P^k. \qquad (5)$$

Для неоднорідного ланцюга Маркова, коли стохастична матриця змінюється на кожному кроці, співвідношення (5) для вектора безумовних імовірностей на *k*-му кроці має вигляд:

$$\{p_j(k)\}=\{p_j(0)\}\cdot P(1)\cdot P(2)\cdot ...\cdot P(k), \qquad (6)$$

де $\mathrm{P}(m)$ – стохастична матриця системи для *m*-го кроку.

Використання математичного апарата ланцюгів Маркова дозволяє за початковими даними обчислювати безумовні імовірності та, відповідно, здійснювати імовірнісний прогноз станів здоров’я для окремих пацієнтів, а також і різних груп. Для обчислення вектора безумовних імовірностей на *k*-му кроці використовується формула (6), де *k* – кількість станів системи (здоров’я).

В теперішній час статистичні дані стосовно смертності, причин різноманітних захворювань населення збираються в багатьох розвинених країнах в електронному вигляді. Ці дані використовуються для дослідження показників здоров’я на майбутнє, а також для розуміння поточних показників та впливу на них екології та якості життя.

Для дослідження станів здоров'я груп населення за класами захворювань все населення можна класифікувати за віком t_i та станом здоров'я E_j. Якщо E_0 позначити як «здоровий», а E_n – «летальний вихід», то стани системи $E_1, E_2, ..., E_{n-1}$ класифікуються за групами захворювань у відповідності з міжнародною класифікацією хвороб (МКБ), яка обов'язкова для всіх країн-членів Всесвітньої організації охорони здоров'я. Будь-який з векторів безумовних імовірностей є імовірнісним прогнозом для відповідного віку. Сукупність векторів безумовних імовірностей, знайдених на кожному кроці обчислень, можна записати у вигляді матриці, у якій кожен рядок представляє віковий інтервал, а кожен стовпчик – стан здоров'я. Елементами, що знаходяться на перетині *i*-го рядка і *j*-го стовпчика, є імовірності на *i*-му віковому інтервалі опинитися в стані E_j. Важливою умовою застосування методики є несумісність станів пацієнта, тобто один пацієнт у кожний момент часу може знаходитись лише в одному стані. Для забезпечення цього обмеження обирається алгоритм формування вимог для кожного стану згідно з МКБ та з залученням експертів.

Здобуті вектори безумовних імовірностей можна інтерпретувати не тільки як імовірнісний прогноз на майбутнє для одного індивіда або групи, але і як багатовимірний показник поточного стану здоров'я. Якщо прогноз покаже, що при збереженні поточних умов половина чоловічого населення, що нині належить до покоління новонароджених, досягне смертності, наприклад, у 55 років, то ця інформація дає можливість оцінити потенціал цього покоління та вплинути на зміни умов життя для підвищення показника тривалості життя в майбутньому. Розглянута методика дозволяє не тільки констатувати фактичні показники, але і вносити корективи та моделювати отриманий ефект від них.

Висновки. Розроблена методика з використання математичного апарата ланцюгів Маркова дозволяє за початковими даними обчислювати безумовні імовірності та, відповідно, здійснювати імовірнісний прогноз станів здоров'я, як окремої людини, так і різних груп населення на подальші періоди життя. Ця методика може мати практичну реалізацію на основі персоніфікованих баз даних про стани здоров'я та смертність населення.

А.А. Вагис

ИССЛЕДОВАНИЕ ПОКАЗАТЕЛЕЙ ЗДОРОВЬЯ НА ОСНОВЕ ЦЕПЕЙ МАРКОВА

Показана одна из методик моделирования показателей здоровья населения на основе математического аппарата марковских цепей. Эта методика позволяет по начальным данным вычислять наборы безусловных вероятностей и осуществлять вероятностный прогноз состояния здоровья отдельного человека или групп населения на последующие периоды жизни.

A.A. Vagis

RESEARGH OF HEALTH INDICATORS BASED ON MARKOV CHAINS

A technique of health indicators of population modeling based on mathematical apparatus of Markov chains is presented. It allows computing the sets of unconditional probabilities by the initial data and carrying out a probabilistic forecast of the health status of an individual or groups of people for succeeding periods of life.

1. Гупал А.М., Сергиенко И.В. Оптимальне процедуры распознавания. К.: Наук. думка, 2008. С. 231.
2. Гупал А.М., Сергиенко И.В. Симметрия в ДНК. *Методы распознавания дискретных последовательностей*. К.: Наук. думка, 2016. С. 227.
3. Гридина Н.Я., Вагис А.А., Тарасов А.Л. Процедуры распознавания на байесовских сетях для показателей скорости оседания эритроцитов. *Компьютерная математика*. 2013. № 2. С. 97 – 104.
4. Кант В.И. Математические методы и моделирование в здравоохранении. М.: Медицина, 1987. С. 224.
5. Медик В.А., Токмачев В.С. Математическая статистика в медицине. М.: *Финансы и статистика*. 2007. С. 800.

Одержано 25.10.2017

Про автора:

Вагіс Олександра Анатоліївна,
доктор фізико-математичних наук, старший науковий співробітник,
провідний науковий співробітник Інституту кібернетики імені В.М. Глушкова НАН України.