

Н.К. Тимофієва, В.І. Гриценко

Аргумент цільової функції в задачі клінічної діагностики

Показано, что задача клинической диагностики разделяется на три подзадачи: структуризацию библиотеки эталонов, поиск в библиотеке эталонной информации, задачу сравнения эталонной и входной информации. Аргумент целевой функции в первой подзадаче – разбиение n -элементного множества на подмножества, во второй – размещение без повторений, в третьей – сочетание без повторений.

It is shown that the problem of clinical diagnostics is divided into three subproblems: the structuring of the library of standards, the search of the standard information in the library, the task of the comparison of the standard and entrance information. The argument of the objective function in the first subproblem is a set partitioning an n -element set into the subsets, in the other – the placing without repetitions, and in the third subproblem the combination without of iterations.

Показано, що задача клінічної діагностики розділяється на три підзадачі: структуризацію бібліотеки еталонів, пошук у бібліотеці еталонної інформації, задачу порівняння еталонної і вхідної інформації. Аргументом цільової функції в першій підзадачі є розбиття n -елементної множини на підмножини, у другій – розміщення без повторень, в третій – сполучення без повторень.

Вступ. У зв'язку з розвитком цифрової медицини діагностика на сучасному етапі набуває вагомій ролі в загальному циклі клінічних досліджень, лікуванні та профілактиці захворювань. Точність діагнозу та швидкість, з якою його можна встановити, залежить від багатьох факторів, зокрема від стану хворого, наявності даних про симптоми захворювання, лабораторних даних – з одного боку та від кваліфікації лікаря – з іншого. Складність розв'язання цієї задачі полягає в тому, що в ній виникає ситуація невизначеності, яка пов'язана: з неповною вхідною та поточною інформацією; нечітко розробленими правилами обробки та оцінки інформації, які визначаються кваліфікацією лікаря (правила прийняття оптимального рішення).

Для розв'язання задачі клінічної діагностики автоматизованим способом необхідно точно сформулювати її математичну постановку. Як правило, для її моделювання використовують стохастичні та логіко-лінгвістичні методи, лінійне цілочислове програмування, теорію розпізнавання образів і нейронні мережі [1–14]. Цю задачу іноді зводять до задачі класифікації, наприклад [12, 15]. Системи автоматизованого роз-

в'язання задач клінічної діагностики містять програми, які обслуговують бази даних, керують введенням та розпізнаванням вхідної інформації, організують пошук у бібліотеці необхідної інформації та її порівняння з вхідними даними [1, 2, 5–7, 12, 16–23]. Для генерування робочої інформації в процесі проходження захворювання використовують різні прилади [21, 22], а для її аналізу та обробки – аналітичні підходи, наприклад неоднорідні раціональні сплайни [24], оператори інтерфлєтації [25, 26] тощо.

Незважаючи на те, що проблемі клінічної діагностики присвячено багато робіт, точної математичної постановки, яка б дала змогу розробляти ефективні для її розв'язання алгоритми, ще не визначено, а одержаний глобальний розв'язок за змодельованою цільовою функцією не завжди збігається з метою дослідження. Це пов'язано з тим, що в процесі перебігу захворювання його клінічна картина змінюється, а ознаки можуть не збігатися з еталонними, тобто виникає ситуація невизначеності, пов'язана з неповною вхідною та поточною інформацією. Крім того, критерії, за якими оцінюється розв'язок, мають якісну природу. Для задання їх у кількісному значенні необхідно вводити міри подібності, які є суб'єктивними оцінками. Оскільки тут маємо перебір варіантів, ця задача належить до задач комбінаторної оптимізації. Далі покажемо, що неви-

Ключові слова: комбінаторна оптимізація, задача клінічної діагностики, комбінаторна конфігурація, цільова функція, ситуація невизначеності.

значеність в цій задачі виникає і внаслідок особливостей структури множини комбінаторних конфігурацій, які є аргументом цільової функції, та способу її моделювання.

Загальна характеристика проблеми

Діагноз захворювання можна встановити за різних умов:

- за результатами опитування і огляду пацієнта без додаткових лабораторних досліджень визначається попередній діагноз; як правило, точність попереднього діагнозу визначається реакцією пацієнта на призначене лікування, тобто експериментально; в такому разі в процесі розв'язання задачі виникає ситуація невизначеності, пов'язана з неповною вхідною та поточною інформацією, а також нечітко розробленими правилами прийняття рішень (кваліфікацією лікаря);

- діагноз визначається за результатами опитування і огляду пацієнта з урахуванням аналізів, результатів комп'ютерної діагностики тощо; якщо база даних системи містить достатньо повну інформацію, а значення змодельованих цільових функцій для знайденого розв'язку – однозначне, то ймовірність правильно поставленого діагнозу – досить висока; ситуація невизначеності може виникати внаслідок нечітко розроблених правил прийняття оптимального рішення і специфікою захворювання;

- деякі різні захворювання можуть мати подібні клінічні прояви; в цьому випадку значення змодельованої цільової функції для знайденого розв'язку – однаково; виникає ситуація невизначеності, пов'язана з певним способом моделювання цільової функції та специфікою захворювання; коректне розв'язання задачі проводиться диференціальним діагнозом з використанням додаткових критеріїв;

- клінічна картина певного захворювання може мати кілька стадій і кілька форм. В процесі розв'язання задачі виникає ситуація невизначеності, пов'язана з неоднозначністю правил встановлення діагнозу, оскільки клінічна картина певного захворювання не завжди відповідає описаній класичній.

Постановка діагнозу проводиться з використанням певних правил обробки та оптималь-

ної оцінки інформації, якими володіє лікар. Виділимо такі.

- правила, які чітко сформульовано і описано в книгах та підручниках; назвемо їх *правилами навчання*;

- правила виділення характерних ознак певного захворювання, які формуються в процесі практичної діяльності лікаря (*правила самонавчання*);

- правила виділення характерних ознак певного захворювання без лабораторних даних та комп'ютерної діагностики на рівні інтуїції завдяки особливому мисленню лікаря; цими правилами володіє обмежена категорія людей (*правила інтуїції*).

Правила прийняття оптимального рішення, що стосуються навчання і самонавчання розроблено досить ґрунтовно і формалізуються шляхом уведення цільової функції (однієї або кількох). У розроблених системах вони частково формалізовані та реалізовані. Правила, що ґрунтуються на інтуїції, досить складно формалізувати, відповідно і автоматизувати.

Незважаючи на те, що правила навчання описано досить ґрунтовно, в процесі розв'язання задачі клінічної діагностики виникає ситуація невизначеності, пов'язана із специфікою захворювання, недостатньою вхідною та поточною інформацією, неоднозначністю правил установа-лення діагнозу.

Оскільки ця задача належить до задач розпізнавання образів, то для порівняння вхідної і бібліотечної інформації необхідно ввести міри подібності. Якщо клінічна картина певного захворювання, що описується вхідною інформацією, має в бібліотеці серед усіх лише один подібний еталон, то для неї існує така міра подібності, за допомогою якої діагноз визначається однозначно, тобто задача є розв'язною. Якщо така міра подібності не дає однозначного розв'язку, то з'являється ситуація невизначеності, пов'язана із специфікою захворювання, яка може бути усунута диференціальним діагнозом.

Із сказаного випливає, що ситуація невизначеності в задачі клінічної діагностики виникає через:

- неповну вхідну та поточну інформацію;

- специфіку захворювання;
- нечітко розроблені правила прийняття оптимального рішення.

Математична постановка задачі клінічної діагностики як задачі комбінаторної оптимізації

В загальному вигляді задачу комбінаторної оптимізації сформулюємо, як у роботі [27]. Задачі цього класу, як правило, задаються однією або кількома множинами, наприклад A і B , елементи яких мають будь-яку природу. Назвемо ці множини *базовими*. Наявні два типи задач. В *першому* типі кожну з цих множин можна подати у вигляді графа, вершинами якого є її елементи, а кожному ребру поставлено у відповідність число $c_{t\xi} \in R$, $t \in \{1, \dots, n\}$, $\xi \in \{1, \dots, \tilde{n}\}$, яке називають вагою ребра (R – множина дійсних чисел). Для зручності вважатимемо, що між елементами цих множин існують зв'язки, числове значення яких назвемо *вагами*. Величини $c_{t\xi}$ назвемо *вхідними* даними і задамо їх матрицями. В *другому* типі задач між елементами заданої множини зв'язків не існує, а вагами є числа $v_t \in R$, яким у відповідність поставлено деякі властивості цих елементів, числові значення яких задаються скінченними послідовностями, що також є вхідними даними. Ці величини визначають значення цільової функції.

Для обох типів задач із елементів однієї із заданих множин, наприклад $a_t \in A$, утворюється комбінаторна множина W – сукупність комбінаторних конфігурацій певного типу (перестановки, вибірки різних типів, розбиття тощо). На елементах w комбінаторної множини W вводиться цільова функція $F(w)$. Необхідно знайти елемент w^* множини W , для якого цільова функція $F(w)$ набуває екстремального значення за виконання заданих обмежень, тобто $F(w^*) = \text{glob}_{w \in W^0} \text{extr } F(w)$, де $\text{extr} = \{\min, \max\}$, W^0 – підмножина, яка визначається обмеженнями задачі.

Уточнимо такі поняття як критерій і цільова функція.

Критерій – ознаки або властивості, які характеризують певний об'єкт або зв'язки між об'єктами і є вхідними даними.

Цільова функція – вираз, який формулюється на основі заданих критеріїв з урахуванням особливостей задачі, за яким обчислюється і оцінюється результат розв'язання задачі.

Як правило, цільову функцію ототожнюють з критеріями. Але для одних і тих же критеріїв цільову функцію можна змоделювати по-різному, тобто оцінка проводиться за різними виразами і одержується різний результат.

Задача клінічної діагностики належить до другого типу. Позначимо $A = \{a_1, \dots, a_n\}$ множину захворювань, описання яких знаходиться в бібліотеці і яку в подальшому назвемо *множиною еталонів*, де елемент a_t відповідає певному захворюванню, якому поставлено у відповідність характерні ознаки. Позначимо ознаки t -го захворювання упорядкованою множиною $V^{(t)} = (v_1^{(t)}, v_2^{(t)}, \dots, v_{q_t}^{(t)})$, де q_t – кількість ознак t -го захворювання. Вхідною інформацією в задачі клінічної діагностики є множина ознак $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$, що описує одне або кілька захворювань. Позначимо їх $B = \{b_1, \dots, b_{\tilde{n}}\}$, де b_{ξ} – захворювання, яке потрібно визначити, \tilde{n} – кількість можливих захворювань, а $q_t \neq \tilde{q}$ або $q_t = \tilde{q}$. Ознаки $\tilde{v}_r \in \tilde{V}$ вхідної інформації мають той же сенс, що і описані в еталоні ознаки $v_s^{(t)} \in V^{(t)}$, $r \in \{1, \dots, \tilde{q}\}$, $s \in \{1, \dots, q_t\}$. Назвемо однотипними такі ознаки, які характеризують один і той же параметр в еталоні і вхідних даних, наприклад значення температури, тиску, пульсу тощо.

Задача полягає у знаходженні для B з множиною ознак \tilde{V} найбільш правдоподібного одного або кількох еталонів із множини $A = \{a_1, \dots, a_n\}$, тобто за вхідними ознаками встановлюється одне або кілька захворювань b_{ξ} . Ознаки в цій задачі виконують роль критеріїв, за якими оцінюється її розв'язок.

Отже, задача клінічної діагностики розділяється на такі основні підзадачі:

- пошук інформації в бібліотеці еталонів, подібної до вхідної;
- порівняння вхідної інформації (про стан хворого, аналізи тощо) з бібліотечною для встановлення правильного діагнозу.

Перша задача для великих розмірностей повиним перебором є нерозв'язною. Для її зведення до розв'язної необхідно провести структурування бібліотеки даних про різні хвороби, розв'язавши задачу кластеризації з виділенням в окремі кластери ознак, які стосуються певного захворювання.

Аргумент цільової функції в задачі діагностики

Для побудови математичної моделі задачі клінічної діагностики з використанням теорії комбінаторної оптимізації визначимо комбінаторну конфігурацію, яка є аргументом цільової функції.

Комбінаторною конфігурацією назвемо будь-яку сукупність елементів, яка утворюється з усіх або з деяких елементів базової множини $A = \{a_1, \dots, a_n\}$ [28]. Позначимо її упорядкованою множиною $w^k = (w_1^k, \dots, w_{\eta^k}^k)$. Під символом $w_\tau^k \in A$, $\tau \in \{1, \dots, \eta^k\}$ розумітимемо як окремі елементи, так і підмножини (блоки), $\eta^k \in \{1, \dots, n\}$ – кількість елементів у w^k , $W = \{w^k\}_1^p$ – множина комбінаторних конфігурацій. Верхній індекс k ($k \in \{1, \dots, p\}$) у w^k позначає порядковий номер w^k у W , p – кількість w^k у W . Залежно від умови задачі η позначатимемо без індексу або з верхнім індексом η^k .

Оскільки в задачі клінічної діагностики аргумент цільової функції утворюється з базової множини вибиранням, розглянемо таку комбінаторну конфігурацію як вибірку. З поняттям вибірки пов'язують як саму операцію виділення підмножин заданої множини, так і її результат: вибрану підмножину [29]. Під вибіркою розуміємо вибрану підмножину (комбінаторну конфігурацію). Дві нетотожні вибірки w^k і w^i назвемо ізоморфними, якщо $\eta^k = \eta^i$. Підмножину $W_\eta \subset W$ назвемо підмножиною ізоморфних вибірок, якщо її елементи – ізоморфні вибірки [30].

Існують такі типи вибірок: *упорядковані* і *неупорядковані*. Неупорядковані це – сполучення без повторень і сполучення з повтореннями. Упорядковані це – розміщення з повтореннями і розміщення без повторень. Множина будь-якого типу вибірок складається з підмножин ізоморфних вибірок.

Виходячи з викладеного, сформулюємо таку лему.

Лема. Аргументом цільової функції в задачі порівняння вхідної інформації з бібліотечною з метою встановлення правильного діагнозу є розміщення без повторень, яке утворюється вибиранням елементів із двох базових множин $V^{(t)}$ і \tilde{V} .

Доведення. В задачі порівняння вхідної інформації з бібліотечною задано дві базові множини: $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ – множина ознак, які є вхідними даними, і $V^{(t)} = (v_1^{(t)}, v_2^{(t)}, \dots, v_{q_t}^{(t)})$ – множина ознак t -го захворювання, описаного в бібліотеці. Для елемента $\tilde{v}_r \in \tilde{V}$ з метою порівняння вибирається елемент $v_s^{(t)} \in V^{(t)}$. Утворена комбінаторна конфігурація складається з елементів двох базових множин, які в ній не повторюються і розміщуються парами. Якщо елементи різних пар поміняти місцями, то одержимо іншу вибірку. Отже, отримані комбінаторні конфігурації утворюються вибиранням і транспозицією. Так утворюється розміщення без повторень, що і потрібно було довести.

Лему доведено.

Наслідок. Якщо в задачах комбінаторної оптимізації порівнювані елементи перебувають у різних базових множинах, то комбінаторна конфігурація утворюється з елементів кількох таких множин.

Моделювання цільової функції в задачі клінічної діагностики

Для моделювання цільової функції в задачі клінічної діагностики уведемо міри подібності між елементами вхідних даних $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ і елементами множини ознак $V^{(t)} = (v_1^{(t)}, v_2^{(t)}, \dots, v_{q_t}^{(t)})$ t -го еталону $a_t \in A$.

Множина ознак $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ може описувати як одне, так і кілька захворювань. Розглянемо випадок, коли вхідна інформація стосується одного захворювання. Для нього змодельюємо цільову функцію. Позначимо виразом $u_r(v_s^{(t)}, \tilde{v}_r)$ елементарну міру подібності між елементами множин \tilde{V} і $V^{(t)}$, $r \in \{1, \dots, \tilde{q}\}$, $s \in \{1, \dots,$

..., q_l }, $l \in \{1, \dots, n^*\}$, $n^* = \min(q_b, \tilde{q})$. Оскільки ці ознаки мають різні шкали вимірів, то виберемо з них такі, щоб одержані оцінки зводилися до однієї шкали, а цільову функцію на етапі порівняння вхідних даних і еталону вважатимемо однокритеріальною.

Припустимо, що міри подібності $u_l(v_s^{(t)}, \tilde{v}_r)$ набувають значень $\{0, \dots, 1\}$. Якщо для однотипних елементів $v_s^{(t)}$ і \tilde{v}_r , $u_l(v_s^{(t)}, \tilde{v}_r) \in \{\delta, \dots, 1\}$, вважатимемо, що вони однакові, де δ – найменша величина міри подібності, для якої існує допустимий розв'язок. Однотипні елементи $v_s^{(t)}$, \tilde{v}_r , для яких $u_l(v_s^{(t)}, \tilde{v}_r) < \delta$, вважатимемо різними. Якщо множини $\tilde{V} \cap V^{(t)} = \emptyset$, то вони не містять жодних однакових елементів $v_s^{(t)}$ і \tilde{v}_r . Якщо $\tilde{V} \cap V^{(t)} \neq \emptyset$, то множини \tilde{V} і $V^{(t)}$ містять однакові елементи. Їх може бути один і більше.

Виникають такі ситуації:

- якщо $\tilde{q} = q_l$, і для будь-якого $\tilde{v}_r \in \tilde{V}$ існує в $V^{(t)}$ однаковий елемент $v_s^{(t)}$, $r = \overline{1, \tilde{q}}$, $s \in \{1, \dots, q_l\}$, а в бібліотеці не існує іншого аналогічного еталону, то задача клінічної діагностики є розв'язною.

- якщо $\tilde{q} \leq q_l$, і для будь-якого $\tilde{v}_r \in \tilde{V}$ існує в $V^{(t)}$ однаковий елемент $v_s^{(t)}$, а в бібліотеці є еталони, при порівнянні яких із вхідними ознаками значення цілової функції – однакове, то при розв'язанні задачі клінічної діагностики виникає ситуація невизначеності, яку необхідно розв'язувати диференціальним діагнозом з урахуванням додаткових критеріїв;

- якщо $\tilde{q} > q_l$, то вхідні дані описують кілька захворювань або вони містять дані, не занесені в бібліотеку. Виникає ситуація невизначеності, пов'язана з неповною бібліотечною інформацією. В цьому випадку необхідно забезпечити автоматичне внесення нової інформації, тобто на етапі розроблення програм необхідно забезпечити процес самонавчання системи.

Розглянемо задачу порівняння ознак $V^{(t)} = (v_1^{(t)}, v_2^{(t)}, \dots, v_{q_t}^{(t)})$ еталону $a_t \in A$, які визначають t -е захворювання, і вхідних ознак $\tilde{V} = (\tilde{v}_1,$

$\tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$, за якими необхідно встановити діагноз. Вважатимемо, що міри подібності між елементами $v_s^{(t)} \in V^{(t)}$ і $\tilde{v}_r \in \tilde{V}$ є вхідними даними. Їхні числові значення задамо скінченною послідовністю (комбінаторною функцією натурального аргументу), залежною від розміщення без повторень w^k . Позначимо її $\beta(f(j), w^k)|_1^{n^*} = (\beta_1(f(1), w^k), \dots, \beta_{n^*}(f(n^*), w^k))$. Якщо елементи $v_s^{(t)} \in V^{(t)}$ і $\tilde{v}_r \in \tilde{V}$ – однотипні, то значення $\beta_j(f(j), w^k) = u_l(v_s^{(t)}, \tilde{v}_r)$. В іншому разі $\beta_j(f(j), w^k) = 0$.

Оскільки вхідні дані задаються однією комбінаторною функцією, задача порівняння вхідної і бібліотечної інформації є однокритеріальною.

Для визначення подібних елементів із множин $V^{(t)}$ і \tilde{V} введемо комбінаторну функцію $\tilde{\beta}(\varphi(j), w^k)|_1^{n^*} = (\tilde{\beta}_1(\varphi(1), w^k), \dots, \tilde{\beta}_{n^*}(\varphi(n^*), w^k))$,

$\varphi(j) \in \{0, 1\}$, де $\tilde{\beta}_j(\varphi(j), w^k) = 1$, якщо $\beta_j(f(j), w^k) \geq \delta$.

В іншому разі $\tilde{\beta}_j(\varphi(j), w^k) = 0$. Кількість одиниць у функції $\tilde{\beta}(\varphi(j), w^k)|_1^{n^*}$ дорівнює $q'(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k)$.

Оцінку результату проводимо за одним критерієм, для якого запишемо такі цільові функції: середнє значення мір подібності:

$$F_1(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k) / q'(w^k), \quad (1)$$

сумарне значення мір подібності:

$$F_2(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k), \quad (2)$$

де $\sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k)$ – інтегральна міра подібності, а $\beta(f(j), w^k) = u_l(v_s^{(t)}, \tilde{v}_r)$ – елементарна міра подібності, яка визначає подібність між елементами еталону і елементами множини ознак вхідних даних. Аргументом цільових функцій (1), (2) є розміщення без повторень.

Множина W розміщень без повторень складається з підмножин W_η . Задача порівняння вхід-

них ознак з еталонними розв'язується на всій множині W . Як обумовлено в [27], закономірність зміни значень цільових функцій (1), (2) на заданому впорядкуванні підмножин W_η однакова незалежно від вхідних даних. В такому випадку в процесі розв'язання задачі виникає ситуація невизначеності, пов'язана з моделюванням цільової функції і структурою її аргумента. На підмножині W_η цільова функція змінюється так, як і на множині перестановок, і закономірність її зміни на заданому впорядкуванні вибірок залежить від вхідних даних. Якщо за функціями (1), (2) одержуємо оптимальний розв'язок для однієї і тієї ж $w^{k*} \in W_\eta$, то одержаний результат збігається з метою дослідження. В іншому разі з урахуванням функцій (1), (2) виникає ситуація невизначеності.

З метою зведення ситуації невизначеності до мінімуму введемо додаткову цільову функцію, яка визначає кількість одиниць у функції $\tilde{\beta}(\varphi(j), w^k) |_{l_1}^{n^*}$:

$$F_3(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k). \quad (3)$$

Задача порівняння ознак t -го еталону і вхідних ознак полягає в пошуку такого розміщення без повторень $w^{k*} = (w_1^{k*}, \dots, w_q^{k*})$, для якого

$$F_1(w^{k*}) = \max_{w^k \in W} F_1(w^k), \quad F_2(w^{k*}) = \max_{w^k \in W} F_2(w^k), \quad F_3(w^{k*}) = \max_{w^k \in W} F_3(w^k).$$

Розглянемо задачу перебору еталонів $A = \{a_1, \dots, a_n\}$. В цій задачі як ваги між елементами $a_t \in A$ і вхідними даними \tilde{V} виступають значення інтегральних мір подібності, одержаних за виразами (1) – (3), числові значення яких подамо матрицями $C^{(1)}$, $C^{(2)}$ і $C^{(3)}$. Номери стовпців цих матриць збігаються з номерами еталонів $a_t \in A$, розміщеними у бібліотеці. Рядок у них один і відповідає номеру один множини $V = \{\tilde{V}\}$. Оскільки при порівнянні множини ознак вхідних даних і ознак еталонів з базових множин A і V вибираються два елементи незалежно від їхнього впорядкування, то утворена комбінаторна конфігурація є сполученням без повторень. Позначимо її $w^i \in W'$, де W' – їхня всі-

ляка множина. Введемо комбінаторну (0,1)-матрицю $Q(w^i) = \|g_{st}(w^i)\|_{l_1 \times n}$. Якщо $g_{it}^k(w^k) = 1$, то з множин A і V вибраною парою $\epsilon(a_t, \tilde{V})$, в іншому разі значення $g_{it}^k(w^k) = 0$. Елементи матриці $C^{(1)}$ подамо числовою функцією $\tilde{\varphi}^{(1)}(j) |_{l_1}^{n-1}$, $C^{(2)}$ – функцією $\tilde{\varphi}^{(2)}(j) |_{l_1}^{n-1}$, $C^{(3)}$ – відповідно $\tilde{\varphi}^{(3)}(j) |_{l_1}^{n-1}$, а матриці $Q(w^i)$ – комбінаторною $\beta'(f'(j), w^i) |_{l_1}^{n-1}$.

Задача пошуку бібліотечного еталону, відповідного вхідним ознакам, полягає у знаходженні такого сполучення без повторень w^{k*} , для якого значення часткових критеріїв, за якими оцінюється результат розв'язання, були б найбільшими, тобто

$$\Phi^{(1)}(w^{k*}, w^{i*}) = \max_{w^i \in M'} \sum_{j=1}^{n-1} \tilde{\varphi}^{(1)}(j) \beta'_j(f'(j), w^i), \quad (4)$$

$$\Phi^{(2)}(w^{k*}, w^{i*}) = \max_{w^i \in M'} \sum_{j=1}^{n-1} \tilde{\varphi}^{(2)}(j) \beta'_j(f'(j), w^i), \quad (5)$$

$$\Phi^{(3)}(w^{k*}, w^{i*}) = \max_{w^i \in M'} \sum_{j=1}^{n-1} \tilde{\varphi}^{(3)}(j) \beta'_j(f'(j), w^i), \quad (6)$$

де $\tilde{\varphi}^{(1)}(j) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^{k*}) \beta_j(f(j), w^{k*}) / q'(w^{k*})$,

$\tilde{\varphi}^{(2)}(j) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^{k*}) \beta_j(f(j), w^{k*})$, а

$\tilde{\varphi}^{(3)} = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k)$. Оцінку результату про-

водимо за зваженою цільовою функцією (лінійною згорткою) $\Phi(w^{k*}, w^{i*}) = \sum_{\gamma=1}^3 \Phi^{(\gamma)}(w^{k*}, w^{i*})$.

Задача пошуку бібліотечного еталону, відповідного вхідному, оцінюється за трьома критеріями (4) – (6).

Отже, задача клінічної діагностики розділяється на підзадачі, аргументом цільової функції в яких є розміщення без повторень і сполучення без повторень. Ці комбінаторні конфігурації утворюються вибиранням елементів із двох базових множин. Як видно з постановки задачі перебору еталонів, пошук еталону, подібного до вхідного \tilde{V} , потребує повного перебору. Цю

задачу можна звести до розв'язної шляхом структуризації бібліотеки еталонів за певними ознаками, наприклад за типом захворювання. Кожне захворювання має описуватися множиною ознак, які задають клінічні форми і стадії. Окремо варто виділити захворювання, які характеризуються однаковими ознаками, а також визначити однакові для групи захворювань ознаки, що дозволить звужувати область оптимального пошуку. Тобто, на етапі структуризації бібліотеки розв'язується задача кластеризації, аргументом цільової функції в якій є розбиття n -елементної множини на підмножини.

Описана математична постановка моделює різні ситуації і за виразами (4)–(6) дозволяє знаходити подібні еталони, які описують одне або кілька захворювань. Якщо для вхідних даних $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_q)$ знайдено кілька еталонів із A , для яких значення $q'(w^k) < \tilde{q}$, причому $V^{(t)} \cap V^{(d)} = \emptyset$, $t, d \in \{1, \dots, n\}$, $t \neq d$, то вони описують різні захворювання. Якщо $V^{(t)} \cap V^{(d)} \neq \emptyset$, то необхідно проводити диференціальний діагноз.

Висновок. Для розв'язання кожної з описаних підзадач розробляються незалежні алгоритми, які працюють як вбудовані процедури в ітераційному режимі. Тобто задача клінічної діагностики потребує для розв'язання розроблення гібридних алгоритмів.

При формалізації задачі клінічної діагностики з використанням теорії комбінаторної оптимізації описано ситуацію невизначеності, пов'язану із структурою аргументу цільової функції і вибраною мірою подібності. Для виходу з цієї ситуації необхідно вводити додаткові цільові функції.

1. *Бейли Н.* Математика в биологии и медицине. – М.: Мир, 1970. – 326 с.
2. *Мельников В.Г.* Медицинская кибернетика. – Киев: Вища шк., 1978. – 239 с.
3. *Долгополов И.Н.* Система управления процессами клинической диагностики в компьютерных сетях в условиях неопределенности // Проблемы управления и информатики. – 2008. – № 5. – С. 133–148.
4. *Грищенко А.В.* Диагностирование заболеваний на базе нейронных сетей // Искусств. интеллект. – 2006. – № 4. – С. 281–289.

5. *Петрухін В.О.* Математичні моделі, алгоритми і системи збору, обробки та інтерпретації медичної інформації: Автореф. дис... д-ра техн. наук / Ін-т кібернетики ім. В.М. Глушкова НАН України. – К., 2005. – 36 с.
6. *Хачатурідзе Т.М.* Інформаційна технологія підтримки прийняття рішень в системі моніторингу при діагностиці захворювань: Автореф. дис... канд. техн. наук / Нац. авіаційний ун-т. – К., 2006. – 21 с.
7. *Савінов І.М.* Розробка способів та засобів підвищення точності медичних інформаційно-діагностичних систем для неврологічних і нейрохірургічних клінік: Автореф. дис... канд. техн. наук / Вінницький держ. техн. ун-т. – Вінниця, 1999. – 17 с.
8. *Марценюк В.П., Кравець Н.О., Ковальчук О.Я.* Про збіжність еволюційних алгоритмів у задачах медичної діагностики // Искусств. интеллект. – 2002. – № 4. – С. 37–41.
9. *Пилипенко М.В.* Моделі та алгоритми діагностики в медичних інформаційних системах: Автореф. дис... канд. техн. наук: 05.13.06 / Херсон. нац. техн. ун-т. – Херсон, 2005. – 23 с.
10. *Мисик А.В.* Аналіз одновимірних та двовимірних діагностичних даних методами штучних нейронних мереж: Автореф. дис... канд. фіз.-мат. наук / Харків. нац. ун-т ім. В.Н. Каразіна. – Харків, 2004. – 19 с.
11. *Володарський Є.Т., Кошева Л.О.* Методологія формування комп'ютерної моделі для медичного діагностування // Електроніка та системи упр. – 2007. – № 2. – С. 121–126.
12. *Поворознюк А.І.* Система підтримки прийняття рішень в медицині на основі структурної ідентифікації об'єктів діагностики // Матеріали XIV міжнар. конф. з автоматичного управління «Автоматика-2007». – Севастополь, 10–14 вер. 2007 р. – Севастополь: Б.и., 2007. – С. 58–61.
13. *Прокопчук Ю.А.* Формалізм для описання медичних діагностических задач // Матеріали XIII міжнар. конф. з автоматичного управління «Автоматика-2006». – Вінниця, 25–28 вер. 2006 р. – Вінниця: Б.и., 2006. – С. 325.
14. *Разработка модели Байесовской сети для решения задач диагностики вирусных гепатитов / М.А. Вороненко, В.И. Литвиненко, Ф.Б. Рогальский и др.* // Матеріали міжнар. наук. конф. «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту. Ч. II». – Євпаторія, 18–22 трав. 2009 р. – Україна, Євпаторія: Б.и., 2009. – С. 268–274.
15. *Кардаш Я.А.* Математична постановка завдання класифікації для інтелектуальних систем // Матеріали XVI Всеукр. наук. конф. «Сучасні проблеми прикладної математики та інформатики». – Львів, 8–9 жовт. 2009 р. – Львів: Б.и., 2009. – С. 94–95.
16. *Мацуга О.М.* Інформаційна технологія обробки неоднорідних медичних даних для підтримки прийняття рішень під час діагностики: Автореф. дис...

- канд. техн. наук / Дніпропетр. нац. ун-т. – Дніпропетровськ: Б.и., 2007. – 18 с.
17. Волкова С.О., Трунов О.М. Аналіз методів і засобів підвищення якості та надійності систем медичної діагностики // Мат. машини і системи. – 2008. – № 2. – С. 158–164.
 18. Власюк А.І., Месюра В.І., Власюк Б.А. Автоматизована віртуальна система діагностики стану організму людини // Вісн. Вінниц. політехн. ін-ту. – 2004. – № 3. – С. 75–79.
 19. Компоненти інформаційних технологій для медицини / А.Є. Батюк, С.Є. Батюк, С.І. Пилипчик та ін. // Матеріали міжнар. наук. конф. «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» (ISDMCI'2009). – Євпаторія, 18–22 трав. 2009 р. – Україна, Євпаторія: Б.и., 2009. – С. 135–137.
 20. Весненко А.И., Попов А.А., Проненко М.И. Топология структуры развернутого клинического диагноза в современных медицинских информационных системах и технологиях // Кибернетика и систем. анализ. – 2002. – № 6. – С. 143–154.
 21. Пат. України № 31160А. – Спосіб електропунктурної діагностики / В.З. Тиднюк, Т.О. Радкевич, В.І. Гриценко, Б.О. Береговський та ін. – Опубл. 15.12.2000, Бюл. № 7–11.
 22. Тиднюк В.З., Алев Л.С., Ходаковский Н.И. Полевые взаимодействия и особенности разработки устройств для диагностики и коррекции состояния биологически активных точек // Кибернетика и систем. анализ. – 2006. – № 2. – С. 157–165.
 23. Файнзильберг Л.С., Беклер Т.Ю. Применение математического моделирования в исследовании нового метода медицинской диагностики // Вісн. Нац. техн. ун-ту «Харківський політехнічний інститут»: Зб. наук. пр. Тем. випуск: Інформатика і моделювання. – Харків: НТУ «ХПІ». – 2011. – № 36. – С. 183–188.
 24. Patient-specific vascular NURBC modeling for isogeometric analysis of blood flow / Zhang Yongjie, Bazilevs Yuri, Goswami Samrat et al. // Comput. Meth. Appl. and Eng. – 2007. – 196, N 29–30. – P. 2943–2959.
 25. Литвин О.М. Методи обчислень. Додаткові розділи. – К.: Наук. думка, 2005. – 332 с.
 26. Першина Ю.І. Математична модель 3D тіла в комп'ютерній томографії // Питання оптимізації обчислень (ПОО-XXXV): Пр. міжнар. симп. Україна, Крим, Велика Ялта, Кацивелі, 24–29 вер. 2009 р. – Київ, 2009. – Т. 2. – С. 188–193.
 27. Тимофієва Н.К. Теоретико-числові методи розв'язання задач комбінаторної оптимізації: Автореф. дис... докт. техн. наук / Ін-т кібернетики ім. В.М. Глушкова НАН України, Київ, 2007. – 32 с.
 28. Тимофеева Н.К. О способах образования аргумента целевой функции в задачах комбинаторной оптимизации // Кибернетика и систем. анализ. – 2002. – № 6. – С. 96–103.
 29. Рыбников К.А. Введение в комбинаторный анализ. – М.: Изд-во МГУ, 1985. – 308 с.
 30. Тимофеева Н.К. Об особенностях формирования и упорядочения выборки // Кибернетика и систем. анализ. – 2004. – № 3. – С. 174–182.

Поступила 10.02.2012
Тел. для справок: (044) 502-6365 (Київ)
E-mail: tymnad@gmail.com
© Н.К. Тимофієва, В.І. Гриценко, 2012

Н.К. Тимофеева, В.И. Гриценко

Аргумент целевой функции в задаче клинической диагностики

Введение. В связи с развитием цифровой медицины диагностика на современном этапе приобретает важную роль в общем цикле клинических исследований, лечении и профилактике заболеваний. Точность диагноза и скорость, с которой его можно поставить, зависит от многих факторов, в частности от состояния больного, от наличия данных о симптомах заболевания, лабораторных данных с одной стороны и от квалификации врача – с другой. Сложность решения этой задачи заключается в том, что в ней возникает ситуация неопределенности, связанная: с неполной входной и текущей информацией; нечетко разработанными правилами ее обработки и оценки, которые определяются квалификацией врача (правила принятия оптимального решения).

Для решения задачи клинической диагностики автоматизированным способом необходимо точно сформулировать ее математическую постановку. Как правило, для ее

моделирования используют стохастические и логико-лингвистические методы, линейное целочисленное программирование, теорию распознавания образов, а для ее решения – генетические алгоритмы и нейронные сети [1–14]. Эту задачу иногда сводят к задаче классификации, например [12, 15]. Системы автоматизированного решения задач клинической диагностики включают в себя программы, обслуживающие базы данных, контролирующие введение и распознавание входных данных [1, 2, 5–7, 12, 16–23]. Для генерирования дополнительной информации в процессе протекания заболевания используют различные приборы [21, 22], а для ее анализа и обработки – аналитические подходы, например неоднородные рациональные сплайны [24], операторы интерфлетации [25, 26] и др.

Несмотря на то, что проблеме клинической диагностики посвящено много работ, точная математическая постановка, которая бы позволяла разрабатывать эффектив-

ные для ее решения алгоритмы, еще не определена, а полученное по смоделированной целевой функции глобальное решение не всегда совпадает с целью исследования. Это объясняется тем, что в процессе заболевания его клиническая картина изменяется, и его признаки могут не совпадать с эталонными, т.е. возникает ситуация неопределенности, связанная с неполной входной и текущей информацией. К тому же критерии, по которым оценивается решение, имеют качественную природу. Для задания их в количественном значении необходимо вводить меры сходства, которые являются субъективными оценками. Поскольку здесь наблюдается перебор вариантов, эта задача относится к задачам комбинаторной оптимизации. Далее будет показано, что неопределенность в этой задаче возникает и вследствие особенной структуры множества комбинаторных конфигураций, которые есть аргументом целевой функции, и способа ее моделирования.

Общая характеристика проблемы

Диагноз заболевания может устанавливаться при разных условиях:

- по результатам опроса и осмотра пациента без дополнительных лабораторных исследований определяется предварительный диагноз; как правило, точность предварительного диагноза определяется реакцией пациента на назначенное лечение, т.е. экспериментально; в процессе решения задачи возникает ситуация неопределенности, связанная с неполной входной и текущей информацией, а также нечетко разработанными правилами принятия решения (квалификацией врача);

- диагноз определяется по результатам опроса и осмотра пациента с учетом анализов, результатов компьютерной диагностики и др.; если база данных системы содержит достаточно полную информацию, а значение смоделированных целевых функций для найденного решения – однозначно, то вероятность правильно поставленного диагноза достаточно высока; ситуация неопределенности может возникать вследствие нечетко разработанных правил принятия оптимального решения и спецификой заболевания;

- некоторые заболевания могут иметь подобные клинические проявления; в этом случае значения смоделированной целевой функции для найденных решений разных заболеваний – одинаковы; возникает ситуация неопределенности, связанная со способом моделирования целевой функции и спецификой заболевания; корректное решение задачи проводится дифференциальным диагнозом с использованием дополнительных критериев;

- клиническая картина определенного заболевания может иметь несколько стадий и форм; в процессе решения задачи возникает ситуация неопределенности, связанная с неоднозначностью правил установления диагноза, поскольку клиническая картина заболевания не всегда соответствует описанной классической.

Постановка диагноза проводится с использованием разработанных правил обработки и оптимальной оценки информации, которыми владеет врач. Выделим такие:

- правила, четко сформулированные и описанные в книгах и учебниках; назовем их *правилами обучения*;

- правила выделения характерных признаков заболевания и установление корректного диагноза, которые вырабатываются в процессе практической деятельности врача (*правила самообучения*);

- правила выделения характерных признаков заболевания без лабораторных данных и компьютерной диагностики на уровне интуиции благодаря особенному мышлению врача; этими правилами владеет ограниченная категория людей (*правила интуиции*).

Правила принятия оптимального решения, которые относятся к обучению и самообучению, разработаны достаточно обстоятельно и формализуются путем введения целевой функции (одной или нескольких). В разработанных системах они частично формализованы и реализованы. Правила, которые базируются на интуиции, достаточно сложно формализовать, соответственно и автоматизировать.

Несмотря на то, что правила обучения описаны достаточно обстоятельно, в процессе решения задачи клинической диагностики возникает ситуация неопределенности, связанная со спецификой заболевания, недостаточной входной и текущей информацией, неоднозначностью правил установления диагноза.

Поскольку задача клинической диагностики относится к задачам распознавания образов, то для сравнения входной и библиотечной информации необходимо ввести меры сходства. Если признаки клинической картины определенного заболевания имеют в библиотеке лишь один подобный эталон, то для них существует такая мера сходства, с помощью которой диагноз определяется однозначно, т.е. задача разрешима. Если такая мера сходства не дает однозначного решения, то появляется ситуация неопределенности, связанная со спецификой заболевания, которая может быть устранена дифференциальным диагнозом.

Из изложенного следует, что ситуация неопределенности в задаче клинической диагностики возникает при неполной входной и текущей информации; в связи со спецификой заболевания и нечетко разработанными правилами принятия оптимального решения.

Математическая постановка задачи клинической диагностики как задачи комбинаторной оптимизации

Сформулируем математическую постановку задачи клинической диагностики и смоделируем для нее целевую функцию в рамках теории комбинаторной оптимизации.

В общем виде задачу комбинаторной оптимизации сформулируем, как в работе [27]. Задачи этого класса, как правило, задаются одним или несколькими множествами, например A и B , элементы которых имеют любую природу. Назовем эти множества *базовыми*. Имеется два типа задач. В *первом* типе каждое из этих множеств можно представить в виде графа, вершины которого – его элементы, а каждому ребру соответствует число $c_{t\xi} \in R$, $t \in \{1, \dots, n\}$, $\xi \in \{1, \dots, \tilde{n}\}$, называемое весом ребра (R – множество вещественных чисел). Для удобства по-

ложим, что между элементами этих множеств существуют связи, числовое значение которых назовем *весами*. Величины $c_{i\xi}$ назовем *входными* данными и зададим их матрицами. Во *втором* типе задач между элементами заданного множества связей не существует, а весами выступают числа $v_i \in R$, которым в соответствие поставлены некоторые свойства этих элементов. Их числовые значения задаются конечными последовательностями, которые есть входными данными. Эти величины определяют значение целевой функции.

Для обоих типов задач из элементов одного из заданных множеств, например $a_i \in A$, образуется комбинаторное множество W – совокупность комбинаторных конфигураций определенного типа (перестановки, выборки разных типов, разбиения и пр.). На элементах w комбинаторного множества W вводится целевая функция $F(w)$. Необходимо найти элемент w^* множества W , для которого целевая функция $F(w)$ принимает оптимальное значение при выполнении заданных ограничений, т.е. $F(w^*) = \text{glob extr}_{w \in W^0 \subset W} F(w)$, где $\text{extr} = \{\min, \max\}$, W^0 – подмножество, определяемое ограничениями задачи.

Уточним такие понятия как критерий и целевая функция.

Критерий – признаки или свойства, характеризующие определенный объект или связи между объектами и являются входными данными.

Целевая функция – выражение, которое формулируется на основе заданных критериев с учетом специфики задачи, по которому вычисляется и оценивается результат решения задачи.

Как правило, целевую функцию отождествляют с критериями. Но для одних и тех же критериев целевую функцию можно смоделировать по-разному, т.е. оценка проводится по разным выражениям и получается разный результат.

Задача клинической диагностики относится ко второму типу. Обозначим $A = \{a_1, \dots, a_n\}$ множество заболеваний, описание которых имеется в библиотеке и которое в дальнейшем назовем *множеством эталонов*, где элементу a_i соответствует определенное заболевание, характеризующее некоторыми признаками. Обозначим признаки t -го заболевания упорядоченным множеством $V^{(t)} = (v_1^{(t)}, v_2^{(t)}, \dots, v_{q_t}^{(t)})$, где q_t – количество признаков t -го заболевания. Входной информацией в задаче клинической диагностики будет множество признаков $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$, описывающее одно или несколько заболеваний. Обозначим их $B = \{b_1, \dots, b_{\tilde{n}}\}$, где b_{ξ} – заболевание, которое следует определить, \tilde{n} – количество возможных заболеваний, а $q_t \neq \tilde{q}$ или $q_t = \tilde{q}$. Признаки $\tilde{v}_r \in \tilde{V}$ входной информации имеют тот же смысл, что и заданные в эталоне $v_s^{(t)} \in V^{(t)}$, $r \in \{1, \dots, \tilde{q}\}$, $s \in \{1, \dots, q_t\}$. Назовем однотипными такие признаки, которые характеризуют один и тот же параметр в эталоне и входных данных, например значение температуры, артериального давления, пульса и др.

Задача заключается в нахождении для B с множеством признаков \tilde{V} наиболее правдоподобного одного или нескольких эталонов из множества $A = \{a_1, \dots, a_n\}$, т.е. по входным признакам устанавливается одно или несколько заболеваний b_{ξ} . Признаки в этой задаче выполняют роль критериев, по которым оценивается ее решение.

Итак, задача клинической диагностики разделяется на такие основные подзадачи:

- поиск информации в библиотеке эталонов, подобной входной;
- сравнение входной информации (данные о состоянии больного, лабораторные анализы и пр.) с библиотечными с целью установления правильного диагноза.

Первая задача для больших размерностей полным перебором – неразрешима. Для ее сведения к разрешимой задаче необходимо провести структуризацию библиотеки данных о разных болезнях, решив задачу кластеризации с выделением в отдельные кластеры признаков определенной болезни.

Аргумент целевой функции в задаче диагностики

Для построения математической модели задачи клинической диагностики с использованием теории комбинаторной оптимизации определим комбинаторную конфигурацию – аргумент целевой функции.

Комбинаторной конфигурацией назовем любую совокупность элементов, которая образуется из всех или некоторых элементов базового множества $A = \{a_1, \dots, a_n\}$ [28]. Обозначим ее упорядоченным множеством $w^k = (w_1^k, \dots, w_{\eta^k}^k)$.

Под символом $w_i^k \in A$, $\tau \in \{1, \dots, \eta^k\}$ подразумеваем как отдельные элементы, так и подмножества (блоки), $\eta^k \in \{1, \dots, n\}$ – количество элементов в w^k , $W = \{w^k\}_1^p$ – множество комбинаторных конфигураций. Верхний индекс k ($k \in \{1, \dots, p\}$) у w^k обозначает порядковый номер w^k в W , p – количество w^k в W . В зависимости от условия задачи η обозначаем без индекса или с верхним индексом η^k .

Поскольку в задаче клинической диагностики аргумент целевой функции образуется из базового множества выбором, рассмотрим такую комбинаторную конфигурацию как выборки. С понятием выборки связывают как саму операцию выделения подмножеств заданного множества, так и ее результат: выбранное подмножество [29]. Под выборкой подразумеваем выбранное подмножество (комбинаторную конфигурацию). Две нетождественные выборки w^k и w^j назовем изоморфными, если $\eta^k = \eta^j$. Подмножество $W_{\eta} \subset W$ назовем подмножеством изоморфных выборок, если ее элементы – изоморфные выборки [30].

Существуют такие типы выборок: *упорядоченные* и *неупорядоченные*. Неупорядоченные – это сочетания без повторений и сочетания с повторениями. Упорядоченные – это размещения с повторениями и размещения без повторений. Множество любого типа выборок состоит из подмножеств изоморфных выборок.

Учитывая изложенное, сформулируем такую лемму.

Лемма. Аргументом целевой функции в задаче сравнения входной информации с библиотечной с целью установления правильного диагноза есть размещение без повторений, которое образуется выбором элементов из двух базовых множеств $V^{(i)}$ и \tilde{V} .

Доказательство. В задаче сравнения входной информации с библиотечной задано два базовых множества: $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ – множество признаков, которое задает входные данные, и $V^{(i)} = (v_1^{(i)}, v_2^{(i)}, \dots, v_{q_i}^{(i)})$ – множество признаков t -го заболевания, описанного в библиотеке. Для элемента $\tilde{v}_r \in \tilde{V}$ с целью сравнения выбирается элемент $v_s^{(i)} \in V^{(i)}$. Образованная выборка состоит из элементов двух базовых множеств, которые в ней не повторяются и размещаются парами. Если элементы разных пар поменять местами, то получим другую выборку. Следовательно, полученные комбинаторные конфигурации формируются выбором и транспозицией. Так образуется размещение без повторений, что следовало доказать.

Лемма доказана.

Следствие. Если в задачах комбинаторной оптимизации сравниваемые элементы находятся в разных базовых множествах, то комбинаторная конфигурация образуется из элементов нескольких таких множеств.

Моделирование целевой функции в задаче клинической диагностики

Для моделирования целевой функции в задаче диагностики введем меры сходства между элементами входных данных $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ и элементами множества признаков $V^{(i)} = (v_1^{(i)}, v_2^{(i)}, \dots, v_{q_i}^{(i)})$ t -го эталона $a_i \in A$.

Множество признаков $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$ может описывать как одно, так и несколько заболеваний. Рассмотрим случай, когда входная информация касается одного заболевания. Для него смоделируем целевую функцию. Обозначим выражением $u_i(v_s^{(i)}, \tilde{v}_r)$ элементарную меру сходства между элементами множества \tilde{V} и $V^{(i)}$, $r \in \{1, \dots, \tilde{q}\}$, $s \in \{1, \dots, q_i\}$, $l \in \{1, \dots, n^*\}$, $n^* = \min(q_i, \tilde{q})$. Поскольку эти признаки имеют разные шкалы измерений, то выберем такие меры сходства, чтобы полученные оценки сводились к одной шкале, а целевую функцию на этапе сравнения входных данных и эталона смоделируем как однокритериальную.

Положим, что меры сходства $u_i(v_s^{(i)}, \tilde{v}_r)$ принимают значения $\{0, \dots, 1\}$. Если для одноподобных элементов $v_s^{(i)}$ и \tilde{v}_r , $u_i(v_s^{(i)}, \tilde{v}_r) \in \{\delta, \dots, 1\}$, примем, что они одинаковы, где δ – наименьшая величина меры сходства, для которого существует допустимое решение. Одноподобные элементы $v_s^{(i)}$, \tilde{v}_r , для которых $u_i(v_s^{(i)}, \tilde{v}_r) < \delta$, считаем разными. Если множества $\tilde{V} \cap V^{(i)} = \emptyset$, то они не содержат одинаковых элементов $v_s^{(i)}$ и \tilde{v}_r . Если $\tilde{V} \cap V^{(i)} \neq \emptyset$, то множества \tilde{V} и $V^{(i)}$ содержат одинаковые элементы.

Возникают такие ситуации:

- если $\tilde{q} = q_i$ и для любого $\tilde{v}_r \in \tilde{V}$ существует в $V^{(i)}$ одинаковый элемент $v_s^{(i)}$, $r = \overline{1, \tilde{q}}$, $s \in \{1, \dots, q_i\}$, а в библиоте-

ке не существует другого аналогичного эталона, то задача клинической диагностики – разрешима;

- если $\tilde{q} \leq q_i$ и для любого $\tilde{v}_r \in \tilde{V}$ существует в $V^{(i)}$ одинаковый элемент $v_s^{(i)}$, а в библиотеке находятся эталоны, при сравнении которых со входными признаками значение целевой функции – одинаково, то при решении задачи клинической диагностики возникает ситуация неопределенности, которую необходимо решать дифференциальным диагнозом с учетом дополнительных критериев;

- если $\tilde{q} > q_i$, то входные данные описывают несколько заболеваний или они содержат данные, которые не занесены в библиотеку. Возникает ситуация неопределенности, связанная с неполной библиотечной информацией. В этом случае необходимо обеспечить автоматическое внесение новой информации, т.е. на этапе разработки программ необходимо обеспечить процесс самообучения системы.

Рассмотрим задачу сравнения признаков $V^{(i)} = (v_1^{(i)}, v_1^{(i)}, \dots, v_{q_i}^{(i)})$ эталона $a_i \in A$, которые определяют t -е заболевание, и входных признаков $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{q}})$, по которым необходимо установить диагноз. Положим, что меры сходства между элементами $v_s^{(i)} \in V^{(i)}$ и $\tilde{v}_r \in \tilde{V}$ есть входными данными. Их числовые значения зададим конечной последовательностью (комбинаторной функцией натурального аргумента), которая зависит от размещения без повторений w^k . Обозначим ее $\beta(f(j), w^k) \Big|_1^{n^*} = (\beta_1(f(1), w^k), \dots, \beta_{n^*}(f(n^*), w^k))$. Если элементы $v_s^{(i)} \in V^{(i)}$ и $\tilde{v}_r \in \tilde{V}$ – одноподобные, то значение $\beta_j(f(j), w^k) = u_i(v_s^{(i)}, \tilde{v}_r)$. Иначе $\beta_j(f(j), w^k) = 0$.

Поскольку входные данные задаются одной комбинаторной функцией, то задача сравнения входной и библиотечной информации – однокритериальна.

Для определения подобных элементов из множеств $V^{(i)}$ и \tilde{V} введем комбинаторную функцию $\tilde{\beta}(\varphi(j), w^k) \Big|_1^{n^*} = (\tilde{\beta}_1(\varphi(1), w^k), \dots, \tilde{\beta}_{n^*}(\varphi(n^*), w^k))$, $\varphi(j) \in \{0, 1\}$, где $\tilde{\beta}_j(\varphi(j), w^k) = 1$, если $\beta_j(f(j), w^k) \geq \delta$. Иначе $\tilde{\beta}_j(\varphi(j), w^k) = 0$. Количество единиц в функции $\tilde{\beta}(\varphi(j), w^k) \Big|_1^{n^*}$ равняется $q'(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k)$.

Оценку результата проводим по одному критерию, для которого запишем такие целевые функции: среднее значение мер сходства

$$F_1(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k) / q'(w^k), \quad (1)$$

суммарное значение мер сходства

$$F_2(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k), \quad (2)$$

где $\sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k) \beta_j(f(j), w^k)$ – интегральная мера сходства, а $\beta(f(j), w^k) = u_i(v_s^{(i)}, \tilde{v}_r)$ – элементарная мера сходства, определяющая подобие между элементами эталона и элементами множества признаков входных данных. Аргументом целевых функций (1), (2) есть размещение без повторений.

Множество размещений без повторения W состоит из подмножеств W_η . Задача сравнения входных признаков с эталонными решается на всем множестве W . Как оговорено в [27], закономерность изменения значений целевых функций (1), (2) на заданном упорядочении подмножеств W_η одинакова, независимо от входных данных. В этом случае в процессе решения задачи возникает ситуация неопределенности, связанная с моделированием целевой функции и структурой ее аргумента. На подмножестве W_η целевая функция изменяется так же, как и на множестве перестановок, и закономерность ее изменения на заданном упорядочении выборок зависит от входных данных. Если по функциям (1), (2) получаем оптимальное решение для одного и того же $w^{k*} \in W_\eta$, то можно допустить, что полученный результат совпадает с целью исследования. В противном случае с учетом функций (1), (2) возникает ситуация неопределенности.

С целью сведения ситуации неопределенности к минимуму введем дополнительную целевую функцию, определяющую количество единиц в функции $\tilde{\beta}(\varphi(j), w^k) \Big|_1^{n^*}$:

$$F_3(w^k) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^k). \quad (3)$$

Задача сравнения признаков t -го эталона и входных признаков заключается в нахождении такого размещения без повторений $w^{k*} = (w_1^{k*}, \dots, w_q^{k*})$, для которого $F_1(w^{k*}) = \max_{w^k \in W} F_1(w^k)$, $F_2(w^{k*}) = \max_{w^k \in W} F_2(w^k)$, $F_3(w^{k*}) = \max_{w^k \in W} F_3(w^k)$.

Рассмотрим задачу перебора эталонов $A = \{a_1, \dots, a_n\}$. В этой задаче весом между элементами $a_t \in A$ и входными данными \tilde{V} служит значение интегральных мер сходства, полученных по выражениям (1)–(3), числовые значения которых представим матрицами $C^{(1)}$, $C^{(2)}$ и $C^{(3)}$. Номера столбцов этих матриц совпадают с номерами эталонов $a_t \in A$, размещенных в библиотеке. Строка у них одна и соответствует номеру эталона $V = \{\tilde{V}\}$.

Поскольку при сравнении множества признаков входных данных и признаков эталонов из базовых множеств A и V выбираются два элемента независимо от их упорядочения, то образованная комбинаторная конфигурация есть сочетание без повторений. Обозначим ее $w^i \in W^i$, где W^i – их всевозможное множество. Введем комбинаторную (0,1)-матрицу $Q(w^i) = \|g_{st}(w^i)\|_{1 \times n}$. Если $g_{st}^k(w^k) = 1$, то из множеств A и V выбрана пара (a_t, \tilde{V}) , иначе значение $g_{st}^k(w^k) = 0$. Элементы матрицы $C^{(1)}$ представим числовой функцией $\tilde{\varphi}^{(1)}(j) \Big|_1^{n-1}$, $C^{(2)}$ – функцией $\tilde{\varphi}^{(2)}(j) \Big|_1^{n-1}$, $C^{(3)}$ – соответственно $\tilde{\varphi}^{(3)}(j) \Big|_1^{n-1}$, а матрицы $Q(w^i)$ – комбинаторной $\beta'(f^i(j), w^i) \Big|_1^{n-1}$.

Задача поиска библиотечного эталона, соответствующего входным признакам, заключается в нахождении такого сочетания без повторений w^{k*} , для которого значения частичных критериев, по которым оценивается результат решения, были бы наибольшими, т.е.

$$\Phi^{(1)}(w^{k*}, w^{i*}) = \max_{w^i \in M} \sum_{j=1}^{n-1} \tilde{\varphi}^{(1)}(j) \beta'_j(f^i(j), w^i), \quad (4)$$

$$\Phi^{(2)}(w^{k*}, w^{i*}) = \max_{w^i \in M} \sum_{j=1}^{n-1} \tilde{\varphi}^{(2)}(j) \beta'_j(f^i(j), w^i), \quad (5)$$

$$\Phi^{(3)}(w^{k*}, w^{i*}) = \max_{w^i \in M} \sum_{j=1}^{n-1} \tilde{\varphi}^{(3)}(j) \beta'_j(f^i(j), w^i), \quad (6)$$

где $\tilde{\varphi}^{(1)}(j) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^{k*}) \beta'_j(f(j), w^{k*}) / q'(w^{k*})$, $\tilde{\varphi}^{(2)}(j) = \sum_{j=1}^{n^*} \tilde{\beta}_j(\varphi(j), w^{k*}) \beta'_j(f(j), w^{k*})$, а $\tilde{\varphi}^{(3)} = \sum_{j=1}^{n^*} \beta'_j(\varphi(j), w^k)$. Оценку результата проведем по взвешенной целевой функции $\Phi(w^{k*}, w^{i*}) = \sum_{\gamma=1}^3 \Phi^{(\gamma)}(w^{k*}, w^{i*})$. Задача поиска библиотечного эталона, соответствующего входному, оценивается по трем критериям (4) – (6).

Таким образом, задача диагностики разделяется на подзадачи, аргументом целевой функции в которых есть размещение без повторений и сочетание без повторений. Эти комбинаторные конфигурации образуются выбором элементов из двух базовых множеств. Как видно из постановки задачи перебора эталонов, поиск эталона, подобного входным признакам, требует полного перебора. Эту задачу можно свести к разрешимой путем структуризации библиотеки эталонов по определенным признакам, например по типу заболевания. Каждое заболевание должно быть описано множеством признаков, которые задают клинические формы и стадии. Отдельно необходимо выделить заболевания, характеризующиеся одинаковыми признаками, а также определить одинаковые для группы заболеваний признаки, что позволит сузить область оптимального поиска, т.е. на этапе структуризации библиотеки решается задача кластеризации, аргументом целевой функции в которой есть разбиение n -элементного множества на подмножества.

Описанная математическая постановка моделирует различные ситуации и по выражениям (1)–(6) позволяет находить подобные эталоны, описывающие одно или несколько заболеваний. Если для входных данных $\tilde{V} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_q)$ найдено несколько эталонов из A , для которых значение $q'(w^k) < \tilde{q}$, причем $V^{(t)} \cap V^{(d)} = \emptyset$, $t, d \in \{1, \dots, n\}$, $t \neq d$, то они описывают разные заболевания. Если $V^{(t)} \cap V^{(d)} \neq \emptyset$, то необходимо проводить дифференциальный диагноз.

Заключение. Для решения каждой из описанных подзадач необходимо разрабатывать независимые алгоритмы, которые работают как встроенные процедуры в итерационном режиме, т.е. задача клинической диагностики требует для своего решения разработки гибридных алгоритмов.

При формализации задачи клинической диагностики с использованием теории комбинаторной оптимизации определена ситуация неопределенности, связанная со структурой аргумента целевой функции и выбранной мерой сходства. Для выхода из этой ситуации необходимо вводить дополнительные целевые функции.