

ГЕНЕТИЧНИЙ АЛГОРИТМ РОЗВ'ЯЗАННЯ ЗАДАЧІ МАРШРУТИЗАЦІЇ В МЕРЕЖАХ

С.Д. Погорілий, Р.В. Білоус

Київський національний університет імені Тараса Шевченка,
03022, Київ, проспект Академіка Глушкова, 2, корпус 5.

Тел.: (044) 526 0522.

e-mail: romanrpd@gmail.com, sdp@rpd.univ.kiev.ua

Розглянуто можливість формалізації багатокритеріальної задачі пошуку оптимальних шляхів у комп'ютерній мережі, яка представлена у вигляді графа. Запропоновано генетичний алгоритм маршрутизації як задачі багатопараметричної оптимізації. Запропоновано методику та викладено особливості застосування генетичних операцій описаного алгоритму.

Investigated formalization possibility of multi-criteria optimal path problem in a computer network which is represented as a graph. A genetic algorithm for routing problem as multi-parametric optimization is offered. Proposed a method and expounded the features of application of genetic operations for the described algorithm.

Вступ

Однією з найпоширеніших функціональних задач мереж є задача про найкоротші шляхи: пошук шляху між двома визначеними вершинами графа, який відповідає найменшому значенню певного функціонала за визначеним критерієм. Це фундаментальна задача багатьох застосувань у сферах транспорту, маршрутизації, комунікації. Існує низка класичних алгоритмів розв'язання цієї задачі (Белмана-Форда, Дейкстри) [1, 2]. Для зменшення часу виконання цих алгоритмів сформовано низку їх паралельних версій [3], що дозволяють підвищити продуктивність роботи поширених алгоритмів маршрутизації.

Класичні алгоритми розв'язання задач маршрутизації оперують лише одним параметром оптимізації – вагою (ціною) шляху, що виражає сукупність його адитивних характеристик. Проте, зазвичай, існує декілька параметрів, які характеризують кожну гілку мережі, наприклад, пропускна здатність, затримка, швидкість передачі, навантаження, надійність які можна поділити на адитивні та неадитивні. Таким чином, у сучасних мережах з'явилась необхідність розв'язання задачі про найкоротші шляхи з кількома критеріями оптимізації [4]. Класичні алгоритми не застосовні до задач такого типу, що належать до класу NP-складних. Обчислювальні затрати на розв'язання таких задач експоненційно зростають із ростом розмірності оброблюваних графів. Тому виникає необхідність формування нових підходів та алгоритмів розв'язання задач пошуку оптимальних шляхів з багатьма критеріями, одним із яких є генетичний підхід.

Постановка задачі

В задачі про найкоротші шляхи задається зважений орієнтований граф $G=(V,E)$, де V – множина вершин, $E \in V \times V$ – множина ребер графа. В загальному випадку існує декілька вагових функцій $\omega_1, \dots, \omega_k : E \rightarrow R$, кожна з яких відповідає певному критерію оптимізації. Довільний шлях $p = v_i \rightarrow v_j$ складається з послідовності ребер $\langle v_i, v_{i_1} \rangle, \dots, \langle v_{i_k}, v_j \rangle \in E$ і може бути представлений у вигляді послідовності вершин графа, що належать шляху $p = \langle v_i, v_{i_1}, \dots, v_j \rangle$.

Вершини $v_i, v_{i_1}, \dots, v_j \in V$, причому кожна вершина входить до шляху лише один раз. Наприклад, на рис. 1 послідовність (1, 4), (4, 5), (5, 6), (6, 9) є шляхом з вершини 1 до вершини 9 і відповідає представленню (1, 4, 5, 6, 9).

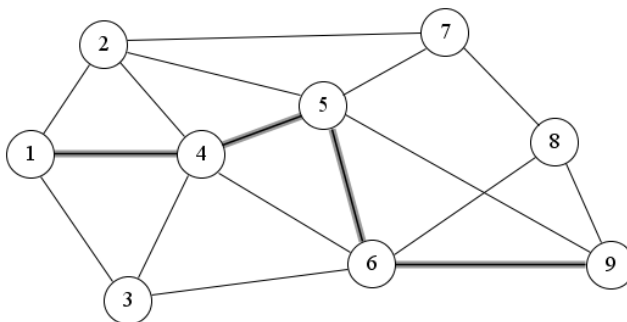


Рис. 1. Шлях (1, 4, 5, 6, 9)

Нехай індекс s відповідає початковій а d – кінцевій вершинам шуканого шляху $p = v_i \rightarrow v_j$. Визначимо $x_{i,j}$ як

$$x_{i,j} = \begin{cases} 1, & \text{якщо ребро } (i, j) \text{ входить до шляху;} \\ 0 & \text{в іншому випадку.} \end{cases}$$

Нехай загальна кількість критеріїв оптимізації задачі k . За кожним критерієм можна обчислити певний функціонал (цільову функцію), який відповідає якості шляху з точки зору алгоритму маршрутизації і визначається як

$$C_m(p) = F_m(\omega_m(i, j), x_{i,j}), m=1..k, (i, j) \in E. \quad (1)$$

Для адитивних характеристик шляху (затримка, довжина), що використовуються як метрики сучасних алгоритмів маршрутизації F_m є сумою значень вагової функції ребер, які входять до шляху p . Для неадитивних характеристик шляху (пропускна спроможність, надійність, навантаження) функціонал F_m є складною функцією від багатьох параметрів і може враховувати не тільки стан з'єднань, але й стан маршрутизаторів мережі, зміну середовища передачі даних та ін.

Позначимо множину всіх можливих шляхів між вершинами v_s та v_d як P . В загальному випадку задача про найкоротший шлях між двома визначеними вершинами в графі з багатьма критеріями може бути сформульована наступним чином:

$$\min_p C_m(p) = F_m(\omega_m(i, j), x_{i,j}), m=1..k, (i, j) \in E, \quad (2)$$

$$\sum_{\substack{j=s \\ j \neq i}}^d x_{i,j} - \sum_{\substack{j=s \\ j \neq i}}^d x_{j,i} = \begin{cases} 1, & \text{якщо } i = s, \\ -1, & \text{якщо } i = d, \\ 0 & \text{в іншому випадку,} \end{cases} \quad (3)$$

$$\sum_{\substack{j=s \\ j \neq i}}^d x_{i,j} = \begin{cases} \leq 1, & \text{якщо } i \neq d, \\ 0, & \text{якщо } i = d. \end{cases} \quad (4)$$

Умови (3) та (4) вимагають, щоб шуканий шлях не містив циклів. Умова (2) вимагає, щоб цільова функція за кожним критерієм оптимізації по всіх можливих шляхах $p = v_s \rightarrow v_d \in P$ досягала найменшого значення на шуканому шляху.

Метод кодування

Розглянемо певний шлях $p = \langle v_s, v_i, v_j, \dots, v_d \rangle$, де $v_s, v_i, v_j, \dots, v_d \in V$, що є проміжним розв'язком задачі. Хромосома генетичного алгоритму, яка є представленням розв'язку p , складається з послідовності цілих додатніх чисел, що відповідають ідентифікаторам вершин графа, через який проходить шлях p . Кожний локус (положення) хромосоми визначає порядковий номер вершини (ідентифікатор якої збігається зі значенням відповідного гену) у шляху. Гени першого та останнього локусів хромосоми зарезервовані за початковою та кінцевою вершинами шляху відповідно. Загальний вигляд хромосоми зображено на рис. 2.

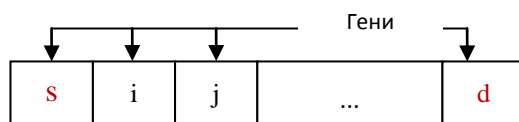


Рис. 2. Загальний вигляд хромосоми

Повернемося до прикладу (див. рис. 1). Нехай $s=1, d=9$. На рис. 3 зображено хромосому, яка відповідає шляху (1, 4, 5, 6, 9).

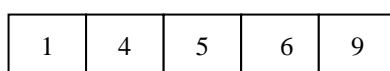


Рис. 3. Приклад хромосоми

Таким чином, довжина хромосоми може змінюватись від 2 до $|V|$. Таке представлення розв'язку задачі дещо ускладнює формування генетичних операцій кросоверу та мутації, проте дозволяє однозначно визначити шлях за його представленням і навпаки.

Початкова популяція

Першим етапом роботи генетичного алгоритму є вибір початкової популяції та її розміру. Початкова популяція, що складається з великої кількості хромосом, дозволяє набагато швидше знайти розв'язок достатньої точності. Проте такий підхід вимагає значних обчислювальних затрат для реалізації алгоритму: що більший розмір популяції, то більше пам'яті обчислювальної системи буде витрачено, і час виконання усіх генетичних операцій (кросовер, мутація, відбір) значно зростатиме. Розмір популяції повинен експоненційно зростати залежно від складності поставленої задачі (наприклад довжини хромосоми) для отримання достатньо хорошого результату. Однак останні дослідження у сфері генетичних алгоритмів показали, що хороший результат за прийнятний час можна отримати використовуючи популяції значно меншого розміру. Експеримент доводить, що для задачі про найкоротші шляхи у графі достатній розмір популяції $N = 2|V|$. Важливим питанням також є вибір способу формування початкової популяції. Існує два основних підходи вирішення цієї проблеми:

- генерація випадкових хромосом;
- формування початкової популяції користуючись класичним алгоритмом.

Перший варіант дозволяє значно спростити процес створення початкової популяції алгоритму, зменшуючи таким чином час роботи алгоритму вцілому. Такий підхід також доцільний, якщо не існує класичного алгоритму розв'язку вихідної задачі.

Другий підхід вимагає значних обчислювальних затрат, проте дозволяє досягнути бажаного результату роботи генетичного алгоритму за значно меншою кількістю ітерацій.

Для описаної вихідної задачі при кількості параметрів оптимізації $k > 2$ в загальному випадку не тільки не існує класичного алгоритму розв'язку, але й точного розв'язку взагалі. Тому для задачі пошуку найкоротших шляхів з багатьма параметрами оптимізації доцільним є формування початкової популяції випадковим чином. Однак застосування класичного алгоритму пошуку найкоротших шляхів за одним із критеріїв дозволить значно скоротити час роботи алгоритму та покращити його параметри збіжності.

Повернемось до формулювання задачі (1). Для позначення важливості кожного критерію оптимізації введемо вагові коефіцієнти $\delta_i, i=1 \dots k$. Для формування початкового покоління хромосом застосовується алгоритм Дейкстри за одним із критеріїв оптимізації задачі, ваговий коефіцієнт якого δ_i є найбільшим.

Генетичні операції

Вибір батьківської пари хромосом. Існує декілька підходів до вибору батьківської пари [5]. Найбільш простий з них – панміксія. Цей підхід являє собою випадковий вибір батьківської пари. В цьому випадку особина може стати членом кількох пар. Незважаючи на простоту, такий підхід є універсальним для розв'язання різних класів задач. Однак він достатньо критичний до чисельності популяції, оскільки ефективність алгоритму, що реалізує такий підхід, знижується з ростом чисельності популяції.

Селективний спосіб вибору особин в батьківську пару полягає в тому, що "батьками" можуть стати тільки ті особини, значення пристосованості яких не менше середнього значення пристосованості по популяції. Такий підхід забезпечує більш швидку збіжність алгоритму. Однак через швидку збіжність селективний відбір батьківської пари не підходить, коли ставиться задача визначення кількох екстремумів, оскільки для таких задач алгоритм, як правило, швидко збігається до одного з розв'язків. Крім того, для деякого класу задач із складним ландшафтом пристосованості швидка збіжність може перетворитись у передчасну збіжність до квазіоптимального розв'язку. Цей недолік в запропонованій моделі скомпенсований завдяки використанню відповідного механізму відбору – турнірного.

Кросовер. Генетична операція кросоверу формує нове покоління розв'язків задачі користуючись генами попереднього покоління. В задачі про найкоротші шляхи кросовер являє собою обмін частковими шляхами обраних хромосом для формування нового шляху. Застосування класичної операції кросоверу не можливе, оскільки довжина хромосоми змінюється і кожен ген не відповідає певній визначеній характеристиці розв'язку задачі.

Алгоритм кросоверу для описаного вище представлення вимагає присутності хоча б одного спільного гену (вершини) у хромосомах, що схрещуються. Спільні гени є точками кросоверу, а часткові шляхи, що сполучають їх між собою, підлягають обміну між хромосомами-батьками.

Застосування операції кросоверу не вимагає, щоб спільні гени хромосом знаходились в однакових локусах хромосоми. Таким чином, операція кросоверу не залежить від положення спільних вершин у шляху.

Повернемось до графа, зображеного на рис. 1. Початковою та кінцевою вершинами оберемо вершини під номером 1 та 9 відповідно. Нехай випадковим чином було сформовано початкове покоління розв'язків задачі. Для операції схрещування (кросоверу) оберемо два довільні шляхи з множини розв'язків. Розглянемо хромосоми, які підлягають операції кросоверу. Хромосома А (рис. 4, а) відповідає шляху (1, 2, 5, 6, 9) (рис. 4, б).

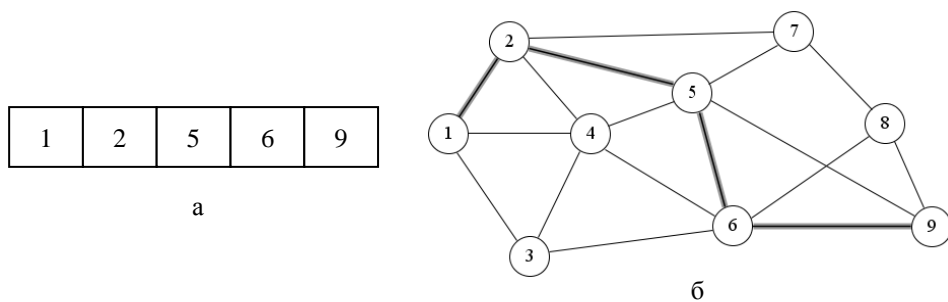


Рис. 4. Хромосома А та відповідний шлях на графі

Хромосома Б (рис. 5, а) відповідає шляху (1, 4, 5, 7, 8, 9) (рис. 4, б).

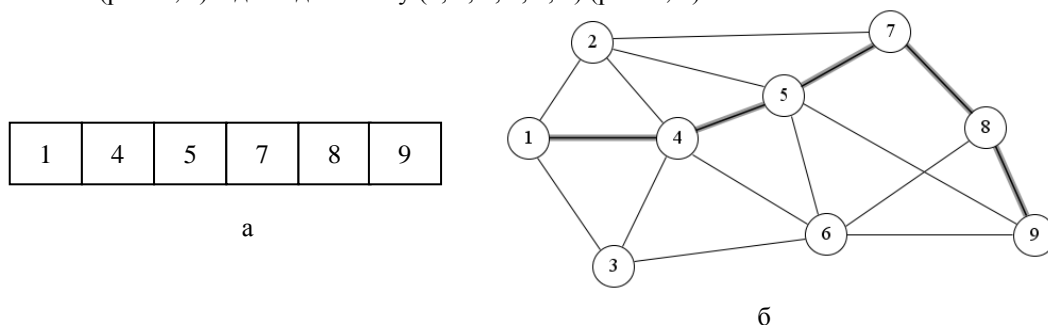


Рис. 5. Хромосома Б та відповідний шлях на графі

Як видно з рис. 4, 5, у хромосом А та Б є спільний ген, що відповідає вершині 5. Ген, що відповідає спільній вершині, є точкою кросоверу. Процедура обміну частковими шляхами та формування наступного покоління розв'язків зображена на рис. 6.

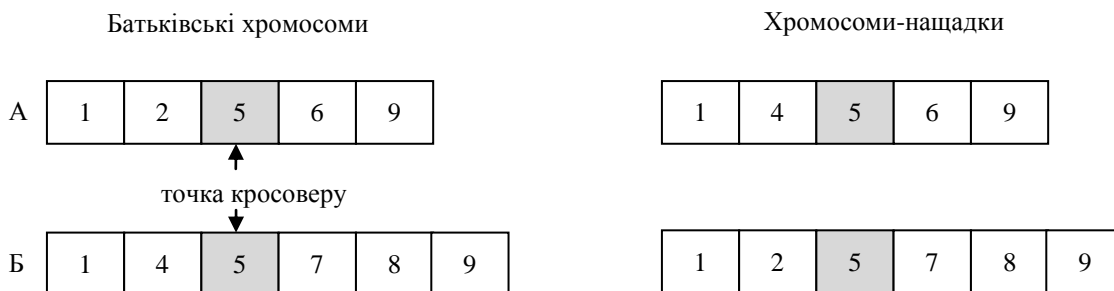


Рис. 6. Одноточковий кросовер хромосом А та Б

Залежно від кількості спільних вершин у батьківських хромосомах можна виділити одноточковий та багатоточковий оператори кросоверу. Приклад багатоточкового кросоверу зображено на рис. 7. Шляхи (1, 2, 4, 5, 7, 8, 9) та (1, 3, 4, 6, 5, 9) мають дві спільні вершини (окрім початкової та кінцевої вершин шляху) – 4 та 5. Ці вершини є точками кросоверу відповідних хромосом.

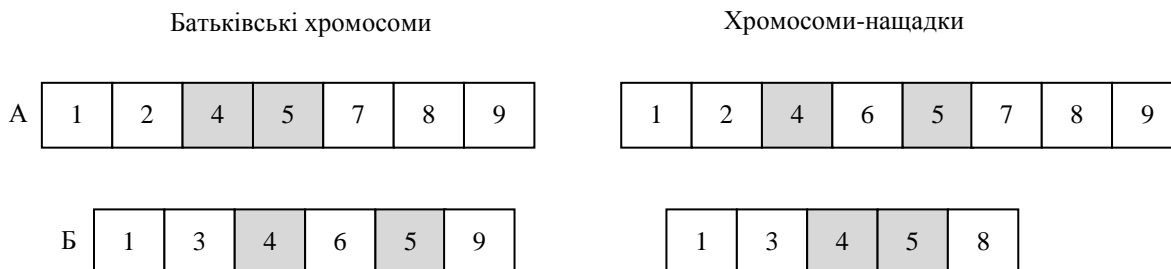


Рис. 7. Двоточковий кросовер хромосом А та Б

Внаслідок операції кросоверу можливе формування шляхів, що містять цикли. Оскільки такі шляхи не задовольняють умову (3) задачі, їх необхідно відкинути і виключити з множини розв'язків. Тому після операції кросоверу всі хромосоми-нащадки підлягають перевірці на присутність циклів у відповідних їм шляхах на графі. Хромосоми, що не проходять перевірку, відкидаються і не приймають участі в операції відбору.

Мутація. Основне завдання операції мутації в генетичному алгоритмі – випадкова зміна розв'язків задачі з метою розширення множини пошуку цільової функції. В класичному виконанні операція мутації полягає у випадковій зміні одного чи декількох генів. Зважаючи на особливості кодування, в задачі про найкоротші шляхи класичний механізм операції мутації не застосовний [6].

Опишемо алгоритм «одноточкової» мутації на прикладі графа, розглянутого вище. Оберемо довільний шлях $p = (1, 2, 4, 6, 8, 9)$ на графі, що є розв'язком задачі на певному етапі роботи генетичного алгоритму (рис. 8). Цей шлях відповідає хромосомі В певного покоління генетичного алгоритму.

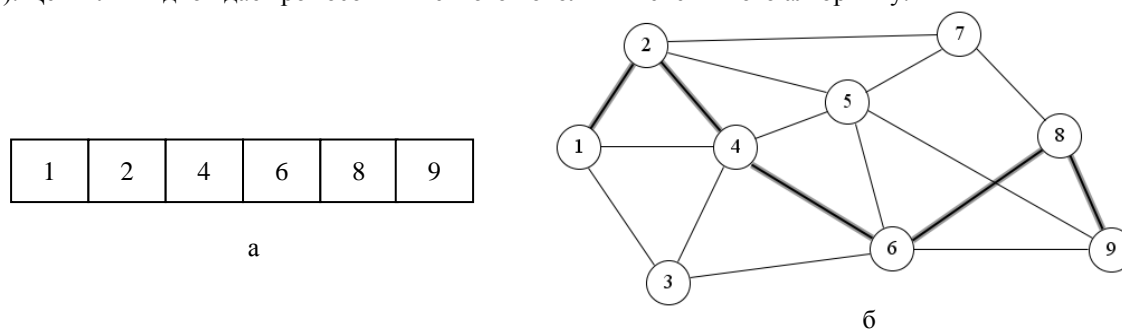


Рис. 8. Хромосома В та відповідний шлях на графі

Для формування операції мутації довільним чином обирається дві вершини v_i та v_j шляху p . Нехай $P_{(i,j)}$ – множина усіх можливих шляхів між вершинами v_i та v_j . Оберемо довільний шлях $p_{i,j} \in P_{(i,j)}$ і замінімо ним проміжний шлях між вершинами v_i та v_j в p . Сформований таким чином альтернативний розв'язок задовольняє умовам задачі і може бути включений до наступного покоління генетичного алгоритму. З точки зору представлення, операція мутації відповідає заміні послідовності генів у хромосомі між положеннями генів v_i та v_j альтернативною послідовністю, що відповідає шляху $p_{i,j} \in P_{(i,j)}$.

Оберемо в хромосомі В довільним чином точки мутації – вершини 4 та 8. Одним із альтернативних шляхів між цими вершинами є шлях (4, 5, 7, 8, 9). Використавши його для операції мутації, одержимо шлях (1, 2, 4, 5, 7, 8, 9). Загальний вигляд операції мутації для хромосоми В зображено на рис. 9.

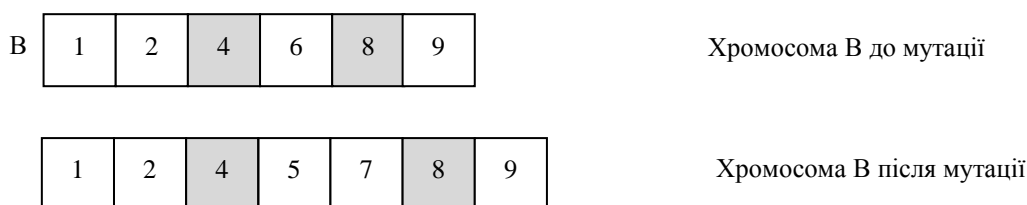


Рис. 9. Операція мутації хромосоми В

Як і при операції кросоверу результат операції мутації також може не задовольняти умову (3) задачі. В цьому випадку, аналогічно попередньому, вводиться операція перевірки результату, а розв'язки, що містять цикли, відкидаються.

Відбір. Операція відбору (відтворення) спрямована на покращення середньої якості популяції генетичного алгоритму. Така можливість досягається за рахунок того, що «кращі» особини мають більші шанси переходу до наступного покоління. Запропонований генетичний алгоритм використовує метод турнірного відбору [7].

Нехай N – кількість особин популяції. Характеристика кожної i -ї особини популяції (фенотип) виражається значенням функції пристосованості $F(i)$. Для відбору певного індивіда створюється група з M ($M \geq 2$) випадково обраних особин. Індивід з найбільшою пристосованістю в групі відбирається, решта – відкидається. Переваги такого підходу:

- відсутність збіжності до локального екстремуму;
- відсутність стагнації;
- відбір не вимагає глобального впорядкування усіх особин покоління;
- немає необхідності явного обрахунку функції пристосованості.

Визначимо відхилення певного розв'язку задачі від ідеального розв'язку як зважену L_h -норму:

$$r(\bar{C}; h, \bar{\delta}) = \|\bar{C} - \bar{C}^*\|_{h, \bar{\delta}} = \left[\sum_{j=1}^k \delta_j^h |C_j - C_j^*|^h \right]^{1/h}, \quad (5)$$

де $\bar{\delta} = (\delta_1, \dots, \delta_k)$, $\bar{C} = (C_1, \dots, C_k)$. $\bar{C}^* = (C_1^*, \dots, C_k^*)$ – точний розв'язок задачі.

Параметр h визначає характер впливу відхилень за кожним окремим критерієм. При $h = 1$ на загальне відхилення головним чином впливає сума відхилень розв'язку за всіма критеріями, при збільшенні значення параметра h до $h = \infty$ підвищується роль відхилення за кожним критерієм окремо. Такий підхід дозволяє гнучко змінювати не тільки вплив кожного окремого критерію, але й їх кореляцію.

Для невеликої кількості параметрів оптимізації задачі точний розв'язок можна знайти користуючись будь-яким класичним алгоритмом. Однак уже при $k > 2$ загального алгоритму розв'язку задачі не існує. В такому випадку доцільно використати найкращий проміжний розв'язок $\bar{C}' = (C'_1, \dots, C'_k)$. Найкращий проміжний розв'язок – це точний розв'язок, що відповідає поточній популяції генетичного алгоритму а не задачі вцілому. Таким чином пошук екстремуму функції оптимізації відбувається на частковій множині розв'язків. З процесом розвитку популяції генетичного алгоритму частковий найкращий розв'язок буде наближатись до точного, якщо такий взагалі існує.

Нехай P – множина особин поточної популяції. Значення найкращого проміжного розв'язку визначається наступним чином:

$$C'_i = \min\{C_i(p) \mid p \in P\}, \quad i = 1 \dots k. \quad (6)$$

Підставивши вираз (6) у формулу (5) одержимо значення відхилення особи p :

$$r(p) = \left[\sum_{j=1}^k \delta_j^h |C_j(p) - C'_j|^h \right]^{1/h}. \quad (7)$$

Оскільки для “найкращих” особин значення відхилення буде мінімальним, слід перетворити значення відхилення у значення функції пристосованості. Позначимо найменше та найбільше відхилення у поточній популяції r_{\min} та r_{\max} відповідно. Значення функції пристосованості особи p визначимо як

$$F(p) = \frac{r_{\max} - r(p)}{r_{\max} - r_{\min}}. \quad (8)$$

Оскільки для операції відбору використовується турнірний метод немає необхідності обрахунку функції пристосованості вцілому, що значно спрощує саму процедуру відбору. Для порівняння особин в групі достатньо порівняти значення відхилення кожної особи і залишити одну хромосому, відхилення якої в групі є мінімальним.

Збіжність та критерії виходу

Застосування методу турнірного відбору дозволяє позбутися потенційної збіжності алгоритму до локального екстремуму функції оптимізації. Проте такий підхід значно ускладнює теоретичний обрахунок збіжності популяції.

Збіжністю будемо вважати такий стан популяції, коли всі особи покоління є майже однаковими і знаходяться в області деякого екстремуму. В такій ситуації кросовер практично ніяк не змінює популяцію, оскільки створені під час цієї операції потомки є копіями батьківських хромосом, а особи, що з'являються внаслідок мутації схильні до вимирання.

Визначимо середнє значення відхилення від оптимального у популяції P :

$$\bar{r} = \frac{\sum_p r(p)}{N}, \quad p \in P, \quad N = |P|. \quad (9)$$

Критерієм щільності значень функції пристосованості популяції алгоритму є відношення:

$$S = \frac{r_{\max} - r_{\min}}{\bar{r}} \quad (10)$$

Іншим важливим питанням є критерій зупинки генетичного алгоритму вцілому. Зазвичай використовують два підходи:

- обмеження на максимальну кількість епох функціонування генетичного алгоритму;
- порівняння пристосованості популяції кількох епох і зупинки при стабілізації цього параметра.

Використання першого підходу вимагає експериментальних досліджень і в загальному випадку не приводить до одержання достатньо хорошого результату.

Для застосування другого підходу слід порівнювати не лише середнє чи найкраще значення пристосованості популяцій генетичного алгоритму але й їх щільність. Тому критерій завершення алгоритму для задачі в запропонованій моделі можна сформулювати наступним чином:

$$S_N < S_{crit} \ll 1, S_{N+1} < S_{crit} \ll 1, \quad (11)$$

$$\left| \frac{\bar{r}_N}{\bar{r}_{N+1}} - 1 \right| < k_{crit} \ll 1. \quad (12)$$

Тут \bar{r}_N та \bar{r}_{N+1} – значення середнього відхилення функції пристосування поточного та наступного поколінь відповідно.

Умова (11) забезпечує невеликий розкид розв'язків задачі в межах одного покоління генетичного алгоритму, а умова (12) – незначну відмінність середнього значення розв'язків двох послідовних епох. Точні значення критеріїв закінчення алгоритму S_{crit} та k_{crit} залежать від кількості параметрів задачі та необхідної точності отриманого результату і вимагають експериментального корегування.

Висновки

В роботі узагальнено задачу пошуку найкоротших шляхів на графі з кількома критеріями та сформовано підходи до її формалізації. Запропоновано генетичний алгоритм оптимізації як найбільш зручний та доцільний підхід до розв'язання цієї задачі за багатьма параметрами.

Для запропонованої моделі генетичного алгоритму сформовано та обгрунтовано метод представлення розв'язку задачі та особливості генетичних операцій (кросовер, мутація). Створено теоретичні засади відбору та критеріїв зупинки генетичного алгоритму.

Описані концепції можуть бути використанні для створення сучасних протоколів маршрутизації, метрика яких враховує як характеристики мережевих з'єднань, так і мережевого обладнання. Час збіжності такого алгоритму може змінюватись залежно від необхідної точності та динаміки зміни мережі. Сформовані підходи дозволяють значно спростити (а для деяких окремих випадків є єдиним варіантом) розв'язання задачі маршрутизації у складних комп'ютерних та транспортних системах.

1. Седжвик Р. Фундаментальные алгоритмы на C++. Алгоритмы на графах / Пер. с англ. – СПб : ООО «ДиаСофтЮП», 2002. – 496 с.
2. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы: построение и анализ. – М.: Центр непрерыв. математического образования, 2000. – 960 с.
3. Погорілий С.Д., Бойко Ю.В., Білоус Р.В. Формування та аналіз паралельних схем алгоритму. Дейкстри. Математичні машини і системи. 2008. Т. 4, – С. 61–71.
4. Goldberg D.E., Genetic Algorithms in Search Optimization and Machine Learning. MA.: – Addison-Wesley, 2000.
5. Sateesh Kumar P., Ramachandram S. Genetic zone routing protocol. // J. of Theoretical and Applied Information Technology. 2008.
6. Gen M., Cheng R. Genetic Algorithms and Engineering Optimization. – New York : Wiley, 2000.
7. Gonen, B. Genetic Algorithm Finding the Shortest Path in Networks. – Reno: University of Nevada, 2006.
8. Погорілий С.Д., Пантелєєва І.В. Методика масштабування алгоритмів маршрутизації. // Наукові праці ДНТУ. Серія Інформатика, кібернетика і обчислювальна техніка. 2007. – Вип. 8 (120), Донецьк, С. 213 – 218.
9. Погорілий С.Д., Мар'яновський В.А., Бойко Ю.В., Верещинський О.А. Дослідження паралельних схем алгоритму Данцига для обчислювальних систем зі спільною пам'яттю. // Математичні машини і системи. – 2009, № 4. – С. 27–37.