

Використання генетичних алгоритмів в моделюванні міграційних процесів

Описані основні теоретичні аспекти генетичних алгоритмів як методу комп'ютерного моделювання, показано можливість застосування його в дослідженні міграційних процесів, розраховано приклад моделювання міграції та, в порівнянні з отриманою звичайною лінійною регресією, доведено якість використання генетичних алгоритмів.

Ключові слова: міграційні процеси, комп'ютерне моделювання, генетичні алгоритми, популяція, хромосоми, мутація, еволюція, регресія.

Постановка проблеми. Проблеми міграції сьогодні досить актуальні, в працездатному віці значна частина населення задля покращення власного добробуту та добробуту своєї родини вирушає на пошуки джерел отримання прибутків за кордон. Трудова міграція, як відомо, виконує важливу роль економічного та соціального амортизатора, що пом'якшує негативні наслідки циклічних коливань в економіці, частково компенсує падіння доходів, дозволяє значній частині населення уникнути зубожіння в умовах падіння виробництва і зростання безробіття, формує альтернативну сферу зайнятості. Крім того, залишається достатньо потужним і вплив міграції на демографічні проблеми суспільства.

Виклад основного матеріалу. Дослідження міграційних процесів більш якісно можна провести з використанням економіко-математичних методів і моделей. Ці методи дозволяють не тільки зробити адекватний аналіз стану міграції населення, знайти та оцінити ступінь впливу міграційних чинників, але й створити прогнози щодо міграційних процесів на майбутнє, що дозволить певним чином використовувати їх при регулюванні процесів міграції та проведенні міграційної політики.

Серед методів комп'ютерного моделювання останнім часом набувають поширення у використанні методи нечіткої логіки, штучного інтелекту, генетичних алгоритмів тощо. Сьогодні генетичні алгоритми довели свою конкурентоспроможність під час вирішення багатьох завдань і особливо у практичних економічних ситуаціях, де математичні моделі мають складну структуру й застосування їх стандартних методів програмування край ускладнено. У машинному навчанні вони використовуються при проектуванні нейронних мереж або керуванні роботами. Вони також застосовуються при моделюванні процесу розвитку в різних предметних областях, включаючи біологічні (екологія, імунологія і популяційна генетика) та соціальні (такі як економіка і державна політика) системи [1].

© О. Р. Овчиннікова, 2013.

Проте, можливо найбільш популярне застосування генетичних алгоритмів – оптимізація багатопараметричних функцій. Багато реальних задач можуть бути сформульовані як пошук оптимального значення, де значення – складна функція, що залежить від певних вхідних параметрів. Так, відомо, що на чисельність мігрантів впливають багато різноманітних чинників. У деяких випадках потрібно знайти ті значення параметрів, при яких досягається найкраще точне значення функції (оптимальний міграційний потік).

Ідея генетичних алгоритмів запозичена в живій природі й полягає в організації еволюційного процесу, кінцевою метою якого є отримання оптимального рішення. Тобто, генетичний алгоритм – це проста модель еволюції в природі, реалізована у вигляді комп'ютерної програми. У ньому використовуються як аналог механізму генетичного успадкування, так і аналог природного відбору. При цьому зберігається біологічна термінологія в спрощеному вигляді:

Хромосома – вектор (послідовність) з нулів та одиниць. Кожна позиція (біт) називається геном. Індивідуум – генетичний код. Набір хромосом – варіант вирішення задачі. Кросовер – операція, при якій дві хромосоми обмінюються своїми частинами. Мутація – випадкова зміна однієї або декількох позицій у хромосомі. Життєвий цикл популяції – це кілька випадкових схрещувань (за допомогою кросовера) і мутацій, в результаті яких до популяції додається якась кількість нових індивідуумів.

Щоб змоделювати еволюційний процес, генерують спочатку випадкову популяцію – декілька індивідуумів з випадковим набором хромосом (числових векторів). Генетичний алгоритм імітує еволюцію цієї популяції як циклічний процес схрещування індивідуумів і зміни поколінь.

Відбір в генетичному алгоритмі – це процес формування нової популяції зі старої, після чого стара популяція гине. Після відбору до нової популяції знову застосовуються операції кросовера і мутації, потім знову відбувається відбір, і так далі.

Таким чином, модель відбору визначає, яким чином слід будувати популяцію наступного покоління. Як правило, імовірність участі індивідуума в схрещуванні береться пропорційною його пристосованості. У будь-якому випадку кожне наступне покоління буде в середньому кращим за попереднє. Коли пристосованість індивідуумів перестає помітно збільшуватися, процес зупиняють і як рішення задачі оптимізації беруть найкращого із знайдених індивідуумів.

Розглянемо детальніше концепцію генетичного програмування на прикладі символічної регресії – задачі знаходження формули, яка описує якусь залежність [3]. Більш формально, її можна сформулювати як побудова регресійної моделі у вигляді суперпозиції заданих функцій.

Алгебраїчна формула: $3X^2 + 8X + 1$.

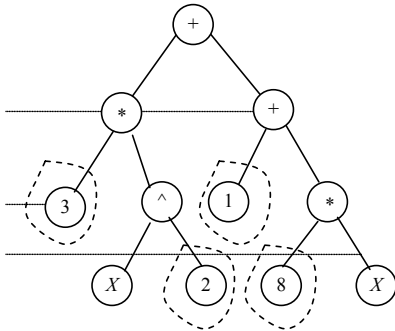


Рис. 1. Приклад синтаксичного дерева (Складено автором)

У даному випадку хромосомою є масив чисел: 3, 2, 1, 8

Досить зручним варіантом представлення програми є синтаксичне дерево. У даному випадку роль «програм» грають алгебраїчні формули (рис. 1). Листові вузли відповідають змінним або числовим константам, а нелистові вузли містять операцію, яка виконується над дочірніми вузлами.

Ми будемо використовувати синтаксичні дерева в якості хромосом. Функція пристосованості синтаксичного дерева обчислює наскільки добре відповідна алгебраїчна формула

апроксимує дані, і, фактично, являє собою суму квадратів відхилень значень згенерованої функції від тренувальних даних.

Топології графа міграційного генетичного алгоритму представлена у [2, с. 32]. В даній роботі модель міграції (Migration) розглядає популяцію як безліч підпопуляцій. Кожна підпопуляція обробляється окремим процесором. Ці підпопуляції розвиваються незалежно одна від одної протягом однакової кількості поколінь T (час ізоляції). Після закінченню часу ізоляції відбувається обмін особами між підпопуляціями (міграція). Кількість осіб, які зазнали обміну (ймовірність міграції), метод відбору осіб для міграції та схема міграції визначає частоту виникнення генетичного різноманіття в підпопуляціях та обмін інформацією між популяціями. Відбір осіб для міграції може відбуватися таким чином:

- випадкова одноманітна вибірка з числа осіб;
- пропорційний відбір: для міграції беруться найбільш придатні особи.

Окремі підпопуляції в паралельних генетичних алгоритмах можна умовно прийняти за вершини деякого графа. У зв'язку з цим можна розглядати топологію графа міграційного генетичного алгоритму. Найбільш поширеною



Рис. 2. Міграція з топологією повної мережі (Складено автором)

топологією міграції є повний граф (рис. 2), при якій особи з будь-якої підпопуляції можуть мігрувати в будь-яку іншу підпопуляцію. Для кожної підпопуляції повна кількість потенційних іммігрантів будувється на основі всіх

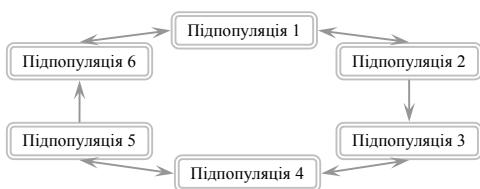


Рис. 3. Міграція з топологією кільця (варіант 1) (Складено автором)

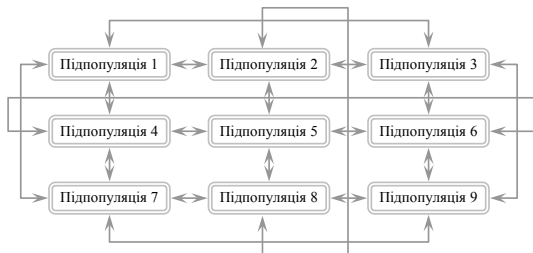


Рис. 4. Міграція з топологією кільця (варіант 2) (Складено автором)

підпопуляцій. Мігруюча особа випадковим чином вибирається з цього загального числа.

Інша основна міграційна схема – це топологія кільця (рис. 3). Тут особи передаються між сусідніми підпопуляціями. Таким чином, особи з однієї підпопуляції можуть мігрувати тільки в одну – сусідню підпопуляцію.

На рис. 4 представлена стратегія міграції, схожа на топологію кільця. Як і при топології кільця, міграція здійснюється тільки між найближчими сусідами.

Однак міграція в цій моделі також можлива між «крайніми» підпопуляціями (тороїдальні крайові міграції).

Побудуємо оптимізаційну модель міграції населення. Для того щоб зрозуміти як працює генетичний алгоритм на практиці, візьмемо дані статистики (табл. 1), тобто чотириох чинників X , Y , Z та R (кількість безробітних, навантаження на одне вільне робоче місце, кількість працевлаштованих осіб, не зайнятих трудовою діяльністю, середня номінальна заробітна плата) та результативного показника F (кількість мігрантів).

Створюємо текстовий документ та використовуємо середовище Java, побудуємо функцію залежності, яка буде задовольняти нашим вихідним даним [5]. На 551 ітерації було згенеровано найкращу функцію залежності, рівняння якої має наступний вигляд:

Best function is:

$$\begin{aligned}
 & f(x, y, z, r) = \\
 & = (((x \times 0,9717058645649455) \times ((x \times 0,0015102156234496622) \times \\
 & \quad \times 0,005559867179321554)) \times ((-0,15141504370393777) \times \\
 & \quad ((x \times (-0,0038014237327761435)) \times \text{sqrt}(\text{abs}(z)))))) / \\
 & \quad / 1445,723497514941)
 \end{aligned}$$

Вихідні дані*

| Дата | Кількість мігрантів, осіб | Кількість безробітних, осіб | Навантаження на одне вільне робоче місце, осіб | Кількість працевлаштованих осіб, не зайнятих трудовою діяльністю, осіб | Середня номінальна заробітна плата, грн. |
|------------------|---------------------------|-----------------------------|--|--|--|
| | <i>F</i> | <i>X</i> | <i>Y</i> | <i>Z</i> | <i>R</i> |
| Січень 2010 р. | 485 | 20404 | 64 | 1835 | 1480 |
| Лютий 2010 р. | 494 | 21288 | 74 | 1938 | 1521 |
| Березень 2010 р. | 474 | 20519 | 37 | 2697 | 1696 |
| Квітень 2010 р. | 468 | 17767 | 38 | 4572 | 1625 |
| Травень 2010 р. | 449 | 15861 | 27 | 4033 | 1799 |
| Червень 2010 р. | 519 | 14544 | 29 | 3554 | 1922 |
| Липень 2010 р. | 682 | 14047 | 31 | 2996 | 1911 |
| Серпень 2010 р. | 589 | 13606 | 23 | 2404 | 1806 |
| Вересень 2010 р. | 2235 | 13608 | 32 | 3144 | 1860 |
| Жовтень 2010 р. | 1629 | 12891 | 44 | 2923 | 1841 |
| Листопад 2010 р. | 1017 | 14417 | 62 | 1848 | 1882 |
| Грудень 2010 р. | 639 | 20503 | 53 | 1229 | 2159 |
| Січень 2011 р. | 404 | 23445 | 63 | 2334 | 1786 |
| Лютий 2011 р. | 589 | 23165 | 52 | 2433 | 1808 |
| Березень 2011 р. | 563 | 21646 | 25 | 3494 | 1990 |
| Квітень 2011 р. | 468 | 19098 | 37 | 5270 | 1932 |
| Травень 2011 р. | 470 | 17372 | 21 | 4095 | 2021 |
| Червень 2011 р. | 525 | 15658 | 28 | 3276 | 2186 |
| Липень 2011 р. | 758 | 14467 | 28 | 3104 | 2182 |
| Серпень 2011 р. | 606 | 13581 | 22 | 2287 | 2143 |
| Вересень 2011 р. | 2229 | 12028 | 29 | 2848 | 2135 |
| Жовтень 2011 р. | 1894 | 10897 | 33 | 2626 | 2213 |
| Листопад 2011 р. | 1098 | 12132 | 34 | 1293 | 2168 |
| Грудень 2011 р. | 664 | 16559 | 51 | 887 | 2390 |
| Січень 2012 р. | 510 | 20404 | 39 | 1997 | 2122 |
| Лютий 2012 р. | 556 | 21288 | 45 | 4301 | 2276 |
| Березень 2012 р. | 550 | 20519 | 26 | 3769 | 2270 |
| Квітень 2012 р. | 434 | 17767 | 29 | 4771 | 2279 |
| Травень 2012 р. | 564 | 15861 | 21 | 3616 | 2408 |
| Червень 2012 р. | 739 | 14544 | 38 | 3113 | 2547 |
| Липень 2012 р. | 612 | 14047 | 27 | 2596 | 2588 |
| Серпень 2012 р. | 630 | 13606 | 27 | 1803 | 2507 |
| Вересень 2012 р. | 2014 | 13608 | 35 | 2474 | 2445 |
| Жовтень 2012 р. | 1672 | 12891 | 33 | 2825 | 2483 |
| Листопад 2012 р. | 1127 | 14417 | 50 | 1434 | 2494 |
| Грудень 2012 р. | 612 | 20503 | 93 | 1066 | 2728 |

* Складено автором на основі джерел: [4].

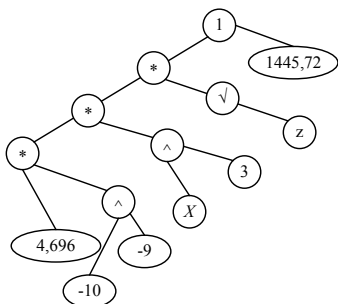


Рис. 5. Синтаксичне дерево отриманої моделі міграції (Складено автором)

Отримана залежність була оптимізована шляхом розрахунку значень листових вузлів, тобто вузлів, які мають числову константу. Після спрощення синтаксичне дерево отримало такий вигляд (рис. 5):

Отож, в нашому випадку хромосомою буде масив чисел:

$$4,696; 10; -9; 3; 1445,72$$

Підставивши отримане рівняння в алгоритм, отримаємо такі значення функції (табл. 2):

Побудуємо модель міграції методом регресійно-кореляційного аналізу.

Для цього скористаємось статистичним додатком в Microsoft Excel. Отримано такі статистичні дані (табл. 3):

Далі отримаємо регресійну модель (1) і за нею розрахуємо емпіричні значення функції за всіма точками.

$$Y = 2391,987804 - 0,115111191 \times x_1 + 9,318352739 \times x_2 + 0,022954379 \times x_3 - 0,03383779 \times x_4 \quad (1)$$

Таблиця 2

Змодельовані значення *

| Z | X | f(x, y, z, k) | Z | X | f(x, y, z, k) |
|------|-------|---------------|------|-------|---------------|
| 3144 | 20404 | 1547,23215 | 1293 | 21646 | 1184,678461 |
| 2923 | 21288 | 1694,288098 | 887 | 19098 | 673,8973439 |
| 1848 | 20519 | 1206,390269 | 1997 | 17372 | 761,0387356 |
| 1229 | 17767 | 638,6847053 | 4301 | 15658 | 817,8277735 |
| 2334 | 15861 | 626,1961439 | 3769 | 14467 | 603,8329123 |
| 2433 | 14544 | 492,9364349 | 4771 | 13581 | 562,0412922 |
| 3494 | 14047 | 532,2068564 | 3616 | 12028 | 339,9090572 |
| 5270 | 13606 | 593,9706357 | 3113 | 10897 | 234,5193494 |
| 4095 | 13608 | 523,8152313 | 2596 | 12132 | 295,5410532 |
| 3276 | 12891 | 398,2905841 | 1803 | 16559 | 626,2798653 |
| 3104 | 14417 | 542,3177503 | 2474 | 20404 | 1372,505394 |
| 2287 | 20503 | 1338,916782 | 2825 | 21288 | 1665,643592 |
| 2848 | 23445 | 2234,029849 | 1434 | 20519 | 1062,701804 |
| 2626 | 23165 | 2069,247991 | 1066 | 17767 | 594,8249504 |

* Складено автором.

Таблиця 3

Регресійна статистика *

| Регресійна статистика | |
|-----------------------|----------------|
| R | 0,6522598479 |
| R-квадрат | 0,4254429092 |
| Нормований R-квадрат | 0,3513065104 |
| Стандартна похибка | 435,4264384665 |
| Спостереження | 36 |

* Складено автором.

Для оцінки того, наскільки регресійна модель відповідає дійсності, відобразимо на графіку модель, що побудована методом генетичного алгоритму, отримані значення за регресійною моделлю та реальну ситуацію (рис. 6).

Як бачимо, регресійна модель значно поступається моделі, яка була побудована методом генетичного алго-

ритму, вона не так чітко відображає тенденцію розвитку, тому й покладатися на адекватне прогнозування майбутніх показників за такої ситуації не варто.

Розглянемо вихідні дані, тобто статистичні показники кількості мігрантів, які спостерігалися впродовж досліджуваних років, та показники, які було згенеровані за допомогою методу генетичного алгоритму. Як бачимо модель досить точно повторює контури кривої реальних статистичних даних, отож можна стверджувати, що вона є адекватною й метод генетичних алгоритмів має право на існування й практичне застосування для побудови прогнозів.

Висновки. Таким чином, в роботі розглянуті теоретичні аспекти та поняття генетичного алгоритму. В процесі дослідження міграції населення методом генетичних алгоритмів було побудовано модель, точність та доцільність якої наглядно представлена на графіку. Отож, модель, що отримана таким нестандартним методом, як генетичний алгоритм, є вартою

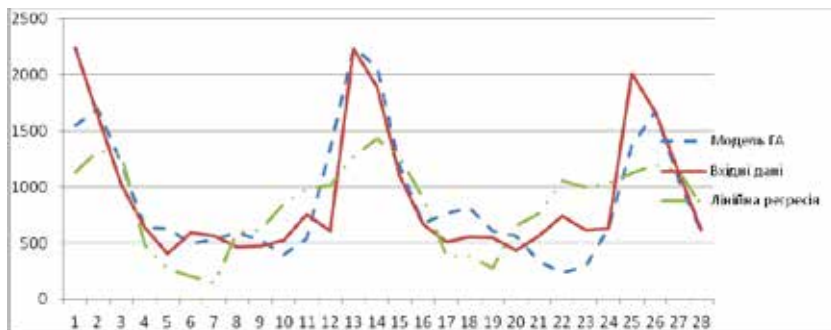


Рис. 6. Реальна, зспрогнозована методом ГА та регресійна модель міграції населення (Складено автором)

уваги. Її можна використовувати для прогнозування міграції на майбутнє, і, відштовхуючись від цього, здійснювати відповідні заходи регулювання міграційних процесів.

Список використаних джерел

1. Багищев Д. И. Генетические алгоритмы решения экстремальных задач / Д. И. Багищев ; Нижегородский госуниверситет. – Нижний-Новгород, 2009. – 62 с.
2. Панченко Т. В. Генетические алгоритмы : уч.-мет. пособие / под ред. Ю. Ю. Тарасевича. – Астрахань : Изд. дом «Астраханский университет», 2007. – 87 с.
3. Символьная регрессия. – Режим доступа : <http://habrahabr.ru/post/163195/>.
4. Статистичний збірник основних показників соціально-економічного розвитку Хмельницької області за 2012 рік / під ред. В. В. Скальського. – Хмельницький, 2013. – 36 с.
5. Gen M., Cheng R. Genetic Algorithms and Engineering design. – John Wiley & Sons. – 1997. – 352 p.

Овчинникова Е.Р. Использование генетических алгоритмов в моделировании миграционных процессов.

Описаны основные теоретические аспекты генетических алгоритмов как метода компьютерного моделирования, показана возможность применения его в исследовании миграционных процессов, рассчитано пример моделирования миграции и, по сравнению с полученной обычной линейной регрессией, доказано качество использования генетических алгоритмов.

Ключевые слова: миграционные процессы, компьютерное моделирование, генетические алгоритмы, популяция, хромосомы, мутация, эволюция, регрессия.

Ovchinnikova O.R. Using of Genetic Algorithms in Modeling of Migration.

This paper describes the main theoretical aspects of genetic algorithms as a method of computer simulation, we show the possibility of applying it to the study of migration, calculated example of modeling and migration, compared with ordinary linear regression obtained, proven quality using genetic algorithms.

Key words: migration, computer modeling, genetic algorithms, population, chromosome mutation, evolution, regression.