

Лекція 15.
Еволюційні і
генетичні
алгоритми

Еволюційні і генетичні алгоритми

Основні визначення

Еволюційні алгоритми, які моделюють процеси природної еволюції, були запропоновані в 60-х роках минулого сторіччя. Їхньою особливістю є те, що вони опираються на природну еволюцію, використовуючи такі її основні механізми, як селекція, схрещування і мутація.

Моделювання процесу природної еволюції для ефективної оптимізації є першорядним завданням теоретичних і практичних досліджень у галузі еволюційних алгоритмів.

В 70-х роках минулого століття з'явився генетичний алгоритм Холланда, який для своєї реалізації запропонував застосовувати оператори селекції, схрещування і мутації, на відміну від еволюційних алгоритмів Реченберга і Швевела, які пропонували використовувати тільки оператори селекції і мутації. Якщо використати біологічні терміни, то еволюційні алгоритми моделюють природну еволюцію за допомогою непарною репродукції. У зв'язку з цим можна дати означення цих алгоритмів.

Означення 1. Еволюційним алгоритмом називають алгоритм, який моделює непарну репродукцію.

Означення 2. Генетичним алгоритмом називають алгоритм, який моделює парну репродукцію.

Парна репродукція характеризується рекомбінацією двох батьківських стрічок для отримання нащадків, яку називають схрещуванням. Перевага різних генетичних операторів в еволюційних і генетичних алгоритмах визначила відношення до використання розміру популяції. Так Холланд підкреслював можливість рекомбінації у великих популяціях, а Реченберг і Швевел розглядали мутацію в дуже малих популяціях.

Щоб зробити зрозумілими принципи роботи еволюційних і генетичних алгоритмів, пояснимо, як влаштовані механізми спадковості в природі. У кожній клітині будь-якої тварини міститься вся генетична інформація цієї особини. Ця інформація записана у вигляді набору молекул ДНК, кожна з яких – це ланцюжок, що складається з молекул нуклеотидів чотирьох типів, порядок розміщення яких і несе інформацію. Таким чином, генетичний код індивіда – це просто дуже довгий рядок символів, де використовуються всього 4 букви. У тваринній клітині кожна молекула ДНК оточена оболонкою, таке утворення називається *хромосомою*.

Кожна якість особини (колір очей, спадкові хвороби, тип волосся і т.д.) кодується певною частиною хромосоми, яку називають *геном* цієї властивості. Наприклад, ген кольору очей, містить інформацію, що кодує певний колір очей.

При розмноженні тварин зливаються дві батьківські статеві клітини, і їхні ДНК, взаємодіючи, утворюють ДНК нащадка. Основний спосіб взаємодії – *схрещування (кросовер)*. При схрещуванні ДНК предків діляться на дві частини, а потім обмінюються своїми половинками.

При спадковості можливі мутації через радіоактивність чи інше, в результаті чого можуть змінитися деякі гени в статевих клітинах одного з батьків. Змінені гени передаються нащадкові і додають йому нових властивостей.. Якщо ці нові властивості корисні, вони, скоріше за все, збережуться в такому вигляді – при цьому відбудеться стрибкоподібне підвищення пристосованості виду.

Головні труднощі побудови обчислювальних систем, що ґрунтуються на принципах природного відбору і застосуванні їх у природних задачах, полягає в тому, що природні системи досить хаотичні, а всі наші дії фактично чітко спрямовані. Ми використовуємо комп'ютер як інструмент для вирішення певних завдань, ми самі і формулюємо та акцентуємо увагу на максимально швидкому виконанні завдань при мінімальних витратах. Природні системи не мають ніяких таких цілей чи обмежень. Вживання в природі не спрямоване до деякої фіксованої мети, але замість цього еволюція робить рок вперед у будь-якому доступному напрямі. Можливо це велике узагальнення, але можна сказати, що зусилля, спрямовані на моделювання еволюції за аналогією з природними системами, можна розбити на дві великі категорії.

- Системи, що змодельовані на біологічних принципах. Їх успішно використовують для вирішення завдань функціональної оптимізації і можна легко описати небіологічною мовою.

- Системи, що є біологічно більш реалістичними, але вони не виявилися особливо корисними в прикладному сенсі. Вони більш схожі на біологічні системи і менш спрямовані або не спрямовані зовсім у своїй роботі.

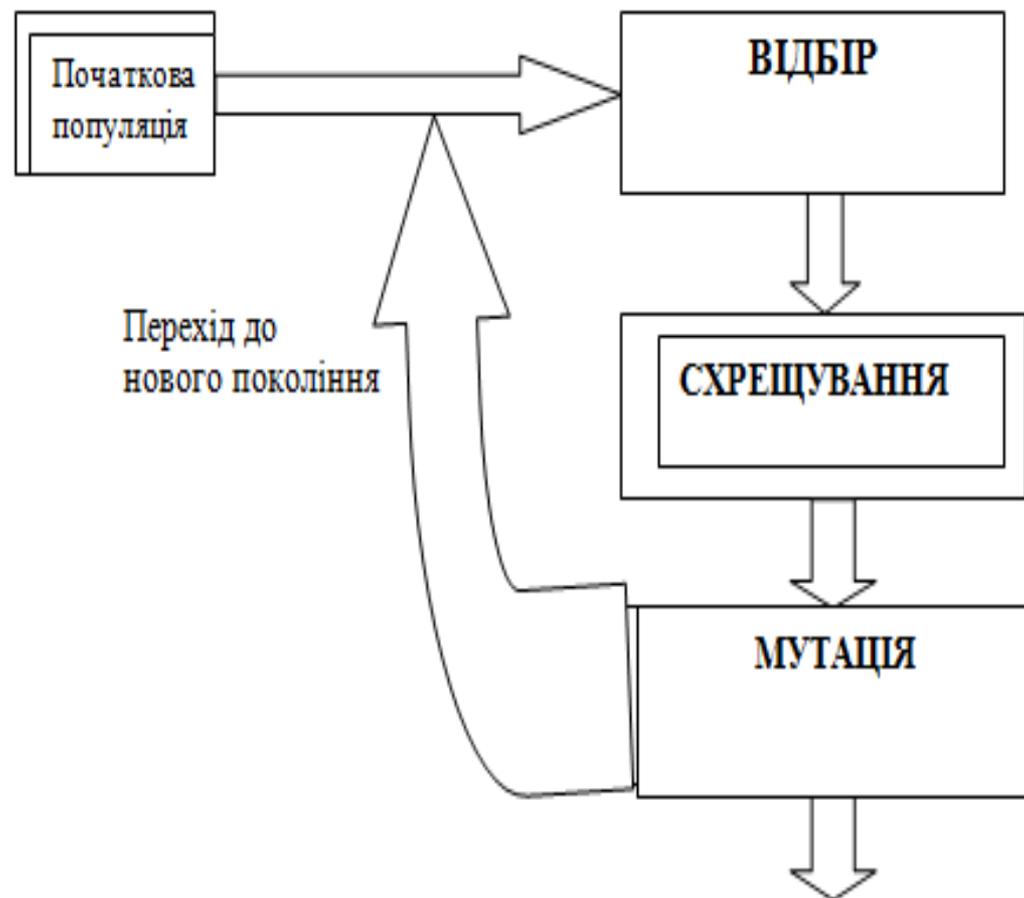
На практиці не можна розділяти ці речі так строго. Ці категорії – два полюси, між якими лежать різні обчислювальні системи. Першому полюсу ближчі еволюційні та генетичні алгоритми, а другому – системи, що можуть бути класифіковані як *«штучне життя»*.

Можна сказати, що кожен біологічний вид з часом удосконалює свої якості так, щоб найбільш ефективно справлятися з виживанням, самозахистом, розмноженням і т.д. За допомогою еволюції природа постійно оптимізує все живе, знаходячи часом неординарні рішення. Неясно, за рахунок чого відбувається цей прогрес, однак йому можна дати наукове пояснення, спираючись усього на два біологічні механізми – природний відбір і генетичне спадкування.

Структура та оператори генетичних алгоритмів

Розглянемо структури і оператори тільки генетичного алгоритму, оскільки вони повністю включають структуру і оператори еволюційного алгоритму.

У загальному вигляді робота генетичного алгоритму являє собою ітераційний процес, який триває до тих пір, поки не виконується задана кількість поколінь, або будь-який інший критерій припинення. На кожному поколінні генетичного алгоритму реалізується відбір, пропорційний пристосуванню, схрещуванню і мутації. Структурну схему роботи генетичного алгоритму представлено на рисунку.



Згідно зі структурною схемою спочатку генерується початкова популяція особини, тобто деякий набір вирішення завдання. Як правило, це робиться випадково. Потім моделюється розмноження всередині цієї популяції. Для цього випадково відбирається декілька пар індивідів, відбувається схрещування між хромосомами в кожній парі, а отримані нові хромосоми втілюються в популяцію нового покоління. В генетичному алгоритмі зберігається основний принцип природного відбору: чим пристосовані ший індивід (чим більше відповідає йому значення цільової функції), тим з більшою ймовірністю він братиме участь в схрещуванні. Далі моделюються мутації – у декількох випадково обраних особинах нового покоління змінюються деякі гени. Потім стара популяція частково або цілком знищується і ми переходимо до розгляду наступного покоління.

Популяція наступного покоління в більшості реалізацій генетичних алгоритмів містить стільки особин, скільки й початкова, але внаслідок відбору пристосованість у ній у середньому вища. Потім описані процеси відбору, схрещування і мутації повторюються вже для цієї популяції і т.д.

Оператор селекції

Означення 3. Оператором селекції називають оператор, який здійснює відбір хромосом відповідно до значення їхніх функцій пристосування.

Для розрахунків є два популярні типи операторів селекції – **рулетка** і **турнір**. Метод рулетки ґрунтується на відборі персон за допомогою N «включень» рулетки. Колесо рулетки має по одному сектору для кожного члена популяції. Розмір i -го сектора, пропорційний відповідній величині $P_c(i)$ обчислюють за формулою ,
$$P_c(i) = \frac{\varphi(i)}{\sum_{i=1}^N \varphi(i)}$$

Де $\varphi(i)$

- функція пристосування кожного члена популяції.

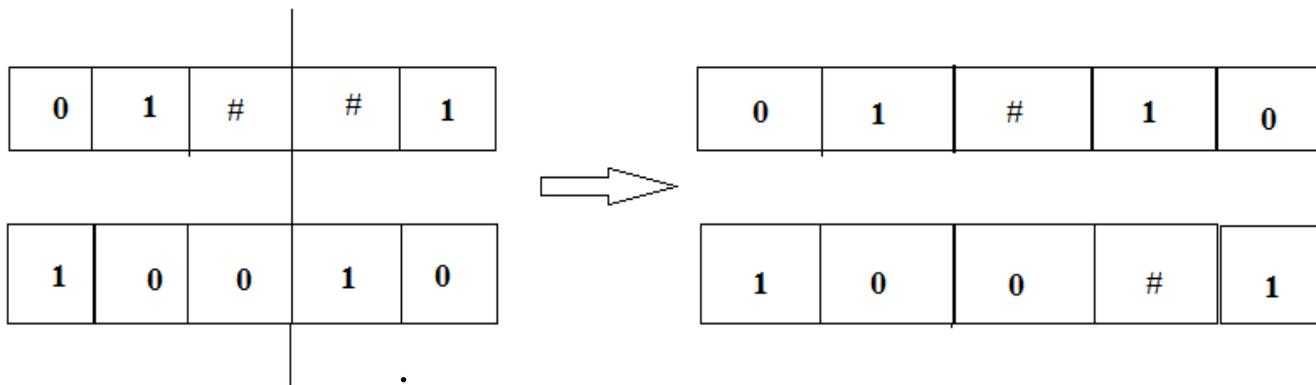
При такому відборі члени популяції з вищим пристосуванням і більшою ймовірністю будуть частіше вибиратися, ніж персони з низьким пристосуванням.

Метод турнірного відбору ґрунтується на проведенні турнірів, щоб відібрати N персон. Кожний турнір побудований на виборі k елементів із популяції та виборі кращої персони серед них. Найбільш поширений турнірний відбір з $k=2$.

Оператор схрещування

Означення 4. Оператором схрещування називають оператор, який здійснює обмін частинами хромосом між двома хромосомами в популяції.

Оператор схрещування може бути одно точковим або багато точковим. Одноточковий оператор схрещування працює так. Спочатку випадково вибирають одну з перших точок розриву. Обидві батьківські структури розривають на два сегменти по цій точці розриву. Потім відповідні сегменти різних батьків склеюють і отримують два генотипи потомків.

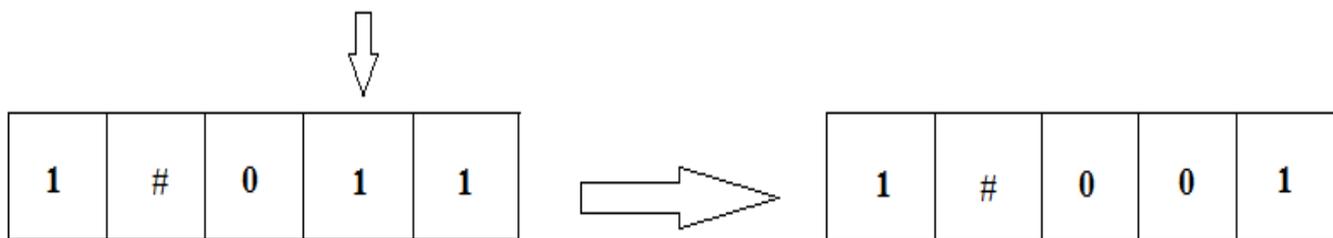


На рисунку показано дію одно точкового оператора схрещування, де точка розриву дорівнює трьом.

Оператор мутації

Означення 5. Оператором мутації називають оператор, який стохастично змінює свої частини хромосом.

Кожний ген стрічки, який підпадає мутації з ймовірністю $P_{\mu}(i)$ (звичайно дуже маленькою), змінюється на другий ген.



На рисунку показано мутацію четвертого гена

Еволюційний алгоритм

Одна з різновидностей еволюційного алгоритму має три кроки:

- 1) Створити первісну популяцію індивідів розміру λ ;
- 2) Обчислити природність функції індивідів $F(x_i), i = 1, 2, \dots, \lambda$
- 3) Виконати селекцію $\mu < \lambda$
кращих індивідів;
- 4) Створити λ / μ
потомків кожного з μ
індивідів з невеликими варіаціями;
- 5) Повернутися до кроку 2 алгоритму.

Базовий генетичний алгоритм

Цей алгоритм, запропонований американським вченим Джоном Генрі Холландом.

Кроки базового генетичного алгоритму:

•Визначити генетичне представлення задачі;

2) Створити первісну популяцію індивідів $P(0) = x_1^0, x_2^0, \dots, x_N^0, t = 0$

3) Обчислити середню придатність індивіда

$$\varphi_{cp}(t) = \sum_{i=0}^N \varphi(x_i) / N$$

4) Обчислити нормалізовані значення ступеня придатності

$$\varphi(x_i) / \varphi_{cp}(t)$$

для кожного індивіда;

5) Призначити кожному індивіду x_i

ймовірності $P(x_i, t)$

пропорційно нормалізованій придатності;

6) Вибрати N векторів з $P(t)$, використовуючи отримані розподілення;

7) Отримати набір відібраних батьків;

8) Сформувати випадково з цього набору, $N/2$ пар;

9) Для формування нової популяції $P(t+1)$ застосувати для кожної пари схрещування, а також інші генетичні оператори;

10) При $t=t+1$ повернутися до кроку 3 алгоритму.

Особливості генетичних алгоритмів, передумови для їхньої адаптації

Дослідження в галузі евристичних та генетичних алгоритмів дає змогу виділити таку їхню важливу особливість, як ефективність, що суттєво залежить від таких взаємозалежних параметрів, як ймовірність застосування генетичних операторів, їхній тип і розмір популяції.

Генетичні алгоритми є **ймовірнісними алгоритмами спрямованого пошуку** і підтримують дані про індивідуальні точки в просторі пошуку, відомому як популяція. Цей пошук в алгоритмі може бути розглянутий як ітераційна програма, яку застосовують для отримання кращих індивідів за допомогою таких операторів, як селекція, схрещування і мутація.

Стратегії мутації і схрещування різні. Мутація ґрунтується на випадку. Результат навіть одного кроку мутації завжди непередбачуваний. Результат схрещування менш випадковий, оскільки при схрещуванні тільки стрічки (якщо розглядати інформаційні потоки в комп'ютері) знаходяться в одній популяції. При цьому пошук за допомогою схрещування призводить до збіжності популяції і може локалізувати оптимум без застосування мутації тільки в разі досить великого розміру популяції. Ці оператори статичні, тобто їхні параметри та ймовірність їхнього використання фіксовані як спочатку, так і залишаються постійними до завершення роботи алгоритму. Однак є докази того, що недостатньо один раз встановити набір операторів і постійно використовувати його в подальшому, тай немає такого набору операторів, які були б оптимальними для всіх задач. Але є докази того, що оптимальний алгоритм операторів для конкретної задачі залежить від ступеня збіжності популяції і буде змінюватися в часі. Спираючись на теоретичні і практичні підходи, деякі автори запропонували різні методи адаптивного управління операторами.

Генетичний алгоритм як метод спрямованого пошуку на кожному кроці генерує нові точки в просторі пошуку для подальшого розвитку популяції. Кожній точці в просторі пошуку відповідає унікальне значення ступеня придатності. Тобто можна говорити про простір пошуку як простір функції ступеня придатності. При цьому популяція дає оцінку ефективності роботи алгоритму, яку можна визначити як **неоднорідну функцію розподілу ймовірності (НФРЙ)**. На відміну від однорідного розподілення ймовірності, який характеризує випадковий пошук, НФРЙ відображає можливі взаємодії між членами популяції.

Генетичний пошук може розглядатися як програма, яка складається з покрокового виконання двох процесів формування проміжної популяції (модифікації популяції): 1) за допомогою селекції та генерації нового набору точок; 2) за допомогою генетичних операторів схрещування та мутації. Таким чином ефективність алгоритму залежить від двох факторів: від підтримання ефективної популяції і від відповідності НФРЙ простору пошуку ступеня природності та ефективності популяції. Перший з цих факторів залежить від розміру популяції та алгоритму селекції, а другий – від дії операторів і пов'язаних з ними параметрів цієї популяції.

Дослідження 6 розмірів популяції, 10 рівнів схрещування, 7 рівнів мутації та 2 типів схрещування показало, що доточкове схрещування не гірше, а інколи краще від одно точкового. У невеликій популяції продуктивність дуже залежить від рівня мутації і менше – від рівня схрещування. Коли розмір популяції збільшується, то чутливість до мутації знижується. При цьому було встановлено такі параметри: розмір популяції 20-30; ймовірність мутації 0,005-0,01; ймовірність схрещування 0,75-0,95. Тим часом теоретичний аналіз оптимальних розмірів популяції дав можливість стверджувати, що розмір популяції залежить від розмірів простоту пошуку. Інтенсивність використання мутації і схрещування повинна бути прив'язана до розміру популяції. Емпіричні дослідження показали, що оптимальне співвідношення між параметрами повинно бути таким $\ln N + 0,93 \ln m + 0,45 \ln n = 0,56$, де N – розмір популяції, m – рівень мутації, n – довжина стратегії. Цей вираз може бути апроксимований, тобто $N \times m \times n = 1,7$. Однак і цей вираз не є кінцевим, оскільки він розглядається без конкретного завдання.

Мутація стає більш ефективною, ніж схрещування, якщо розмір популяції невеликий, і навпаки. Інші фактори, такі як схема представлення, селекція, функції міри готовності, також впливають на ефективність використання мутації та схрещування.

Останні декілька років досліджень привели до появи нових операторів, зокрема однорідного схрещування, яке виконують на $L/2$ точках схрещування стрічки довжиною L і яке є більш ефективним.

За результатами досліджень можна зробити такі два важливі висновки

- Значення НФРЙ є наслідком виконання генетичних операторів (наприклад, схрещування) і залежить від їхнього типу.
- Ймовірність того, що нова точка, згенерована за допомогою генетичних операторів, буде ефективнішою, ніж попередня батьківська, і також залежить від типу генетичного оператора. Значення ймовірності змінюється також від покоління до покоління. Крім того, ймовірність відтворює відповідність між НФРЙ і простором пошуку міри придатності популяції. Ці дослідження в сукупності з дослідженнями еволюційних алгоритмів, де вже використовувалися адаптивні оператори, збільшили інтерес до можливостей алгоритмів, які здатні адаптувати оператори чи параметри. Метою введення механізму адаптації є те, щоб узгодити значення НФРЙ, які формуються алгоритмом і простором пошуку міри придатності.

Застосування генетичних алгоритмів

Генетичні алгоритми в різних формах застосовуються для вирішення багатьох наукових і технічних проблем. Генетичні алгоритми використовуються при створенні інших обчислювальних структур, наприклад, автоматів або мереж сортування. У машинному навчанні вони використовуються при проектуванні нейронних мереж або керуванні роботами. Вони також застосовуються при моделюванні розвитку в різних предметних галузях, включаючи біологічні (екологія, імунологія й опційна генетика) та соціальні (економіка і політика) системи.

Проте найбільше застосовуються генетичні алгоритми для оптимізації багатопараметричних функцій. Багато реальних завдань можуть бути сформульовані як пошук оптимального значення, що залежить від певних вхідних параметрів. У деяких випадках потрібно знайти такі значення параметрів, при яких досягається найкраще точне значення функції. Сила генетичного алгоритму полягає в його здатності маніпулювати одночасно багатьма параметрами, що використовуються в сотнях прикладних програм, включаючи проектування літаків, налаштування параметрів алгоритмів пошуку стійких станів системи нелінійних диференціальних рівнянь.

Останніми роками значно збільшилася кількість праць, особливо зарубіжних вчених, спрямованих на розвиток теорії генетичних алгоритмів і на їхнє практичне застосування. Результати останніх досліджень показують, що генетичні алгоритми можуть ширше застосовуватися в разі поєднання їх з іншими методами і технологіями. З'явилися праці, в яких доводиться ефективність інтеграції генетичних алгоритмів і методів теорії нечітких множин, а також нейронних обчислень і систем.

Ефективність такої інтеграції знайшла практичне підтвердження в розробках відповідних обчислювальних систем. Так, фірма Attar Software включила компоненти генетичних алгоритмів, орієнтованих на вирішення завдань оптимізації в свої обчислювальні системи, які призначені для розроблення експертних систем, а фірма NIBS Inc. Включила компоненти генетичних алгоритмів у свої обчислювальні системи для нейронних мереж, які орієнтовані на прогнозування ринку цінних паперів.