**ЛЕКЦІЯ 10. ГЕНЕТИЧНІ АЛГОРИТМИ**

* [Історія появи еволюційних алгоритмів](http://www.victoria.lviv.ua/html/oio/html/theme10.htm" \l "10_1)
  + [Еволюційна теорія](http://www.victoria.lviv.ua/html/oio/html/theme10.htm#10_1_1)
  + [Природний відбір і генетичне спадкування](http://www.victoria.lviv.ua/html/oio/html/theme10.htm#10_1_2)
* [Задачі оптимізації](http://www.victoria.lviv.ua/html/oio/html/theme10.htm#10_2)
* [Робота генетичного алгоритму](http://www.victoria.lviv.ua/html/oio/html/theme10.htm#10_3)
* [Застосовування генетичних алгоритмів](http://www.victoria.lviv.ua/html/oio/html/theme10.htm#10_4)

**Історія появи еволюційних алгоритмів**

Природа вражає своєю складністю і багатством всіх своїх проявів. Серед прикладів можна назвати складні соціальні системи, імунні і нейронні системи, складні взаємозв'язки між видами. Вони - всього лише деякі з чудес, що стали більш очевидні, коли ми стали глибше досліджувати себе самих і світ навколо нас. Наука - це одна із систем віри, що змінює інші, якими ми намагаємось пояснити те, що спостерігаємо, цим самим, змінюючи себе, щоб пристосуватися до нової інформації, одержуваної з зовнішнього середовища. Багато чого з того, що ми бачимо і спостерігаємо, можна пояснити єдиною теорією: теорією еволюції через спадковість, змінність і відбір.

Теорія еволюції вплинула на зміну світогляду людей із самої своєї появи. Теорія, що Чарльз Дарвін представив у роботі, відомої як "Походження Видів", у 1859 році, стала початком цієї зміни. Багато областей наукового знання в даний час насолоджуються волею думки в атмосфері, що багатьом зобов'язана революції, викликаною теорією еволюції і розвитку. Але Дарвін, подібно багатьом своїм сучасникам, хто припускав, що в основі розвитку лежить природний відбір, не міг не помилятися. Наприклад, він не зміг показати механізм спадкування, при якому підтримується змінність. Його гіпотеза про пангенезис виявилася неправильною. Це було за п'ятдесят років до того, як теорія спадковості почала поширюватися по світу, і за тридцять років до того, як "еволюційний синтез" зміцнив зв'язок між теорією еволюції і відносно молодою наукою генетикою. Однак Дарвін виявив головний механізм розвитку: відбір у сполученні зі змінністю або, як він його називав, "спуск із модифікацією". У багатьох випадках, специфічні особливості розвитку через змінність і відбір все ще не безперечні, однак, основні механізми пояснюють неймовірно широкий спектр явищ, що спостерігаються в Природі.

Тому не дивно, що вчені, які займаються комп'ютерними дослідженнями, звернулися до теорії еволюції в пошуках натхнення. Можливість того, що обчислювальна система, наділена простими механізмами змінності і відбору, могла б функціонувати за аналогією з законами еволюції в природних системах, була дуже приваблива. Ця надія стала причиною появи ряду обчислювальних систем, побудованих на принципах природного відбору.   
Історія еволюційних обчислень почалася з розробки ряду різних незалежних моделей. Основними з них були генетичні алгоритми і класифікаційні системи Голанда (Holland), опубліковані на початку 60-х років і які отримали загальне визнання після виходу у світ книги, що стала класикою в цій області, - "Адаптація в природних і штучних системах" ("Adaptation in Natural and Artifical Systems", 1975).

Головні труднощі з можливістю побудови обчислювальних систем, заснованих на принципах природного відбору і застосуванням цих систем у прикладних задачах, полягає в тому, що природні системи досить хаотичні, а всі наші дії, фактично, носять чітку спрямованість. Ми використовуємо комп'ютер як інструмент для рішення визначених задач, що ми самі і формулюємо та акцентуємо увагу на максимально швидкому виконанні при мінімальних витратах. Природні системи не мають ніяких таких цілей чи обмежень, у всякому разі, нам вони не відомі. Виживання в природі не спрямовано до деякої фіксованої мети, замість цього еволюція робить крок вперед у будь-якому доступному напрямку. Можливо це велике узагальнення, але я думаю, що зусилля, спрямовані на моделювання еволюції за аналогією з природними системами, до дійсного часу можна розбити на дві великі категорії:

1. системи, що змодельовані на біологічних принципах. Вони успішно використовуються для задач функціональної оптимізації і можуть легко бути описані небіологічною мовою;
2. системи, що є біологічно більш реалістичними, але які не виявилися особливо корисними в прикладному змісті. Вони більше схожі на біологічні системи і менш спрямовані (чи не спрямовані зовсім). Вони мають складну і цікаву поведінку, і, скоріше, незабаром набудуть практичного застосування.

Звичайно, на практиці ми не можемо розділяти ці речі так строго. Ці категорії - просто два полюси, між якими лежать різні обчислювальні системи. Ближче до першого полюса - еволюційні алгоритми, такі як Еволюційне Програмування (Evolutionary Programming), Генетичні Алгоритми (Genetic Algorithms) і Еволюційні Стратегії (Evolution Strategies). Ближче до другого полюса - системи, що можуть бути класифіковані як Штучне Життя (Artificial Life).

Звичайно, еволюція біологічних систем не єдине "джерело натхнення" творців нових методів, що моделюють природні процеси.

**Еволюційна теорія**

Як відомо, еволюційна теорія затверджує, що життя на нашій планеті виникло спочатку лише в найпростіших формах - у вигляді одноклітинних організмів. Ці форми поступово ускладнювалися, пристосовуючись до навколишнього середовища, породжуючи нові види, і тільки через багато мільйонів років з'явилися перші тварини і люди. Можна сказати, що кожен біологічний вид з часом вдосконалює свої якості так, щоб найбільш ефективно справлятися з найважливішими задачами виживання, самозахисту, розмноження і т.д. Таким шляхом виникло захисне забарвлення в багатьох риб і комах, панцир у черепахи, отрута в скорпіона і багато інших корисних застосувань.

За допомогою еволюції природа постійно оптимізує все живе, знаходячи часом неординарніші рішення. Неясно, за рахунок чого відбувається цей прогрес, однак йому можна дати наукове пояснення, ґрунтуючись всього на двох біологічних механізмах - природного відбору і генетичного спадкування.

**Природний відбір і генетичне спадкування**

Ключову роль в еволюційній теорії грає природний відбір. Його суть полягає в тому, що найбільш пристосовані особини краще виживають і приносять більше нащадків, ніж менш пристосовані. Помітимо, що сам по собі природний відбір ще не забезпечує розвиток біологічного виду. Тому дуже важливо вивчити, яким чином відбувається спадкування, тобто як властивості нащадку залежать від властивостей батьків.

Основний закон спадкування інтуїтивно зрозумілий кожному - він полягає в тому, що нащадки схожі на батьків. Зокрема, нащадки більш пристосованих батьків будуть, швидше за все, одними з найбільш пристосованих у своєму поколінні. Щоб зрозуміти, на чому заснована ця подібність, нам буде потрібно небагато поглибитися в побудову природної клітини - у світ генів і хромосом.

Майже в кожній клітині будь-якої особини є набір хромосом, що несуть інформацію про цю особину. Основна частина хромосоми - нитка ДНК, що визначає, які хімічні реакції будуть відбуватися в даній клітині, як вона буде розвиватися і які функції виконувати.

Ген - це відрізок ланцюга ДНК, відповідальний за визначену властивість особини, наприклад за колір очей, тип волосся, колір шкіри і т.д. Вся сукупність генетичних ознак людини кодується за допомогою приблизно 60 тис. генів, довжина яких складає більш 90 млн. нуклеотидів.

Розрізняють два види клітин: статеві (такі, як сперматозоїд і яйцеклітина) і соматичні. В кожній соматичній клітині людини міститься 46 хромосом. Ці 46 хромосом - насправді 23 пари, причому в кожній парі одна з хромосом отримана від батька, а друга - від матері. Парні хромосоми відповідають за однакові ознаки - наприклад, батьківська хромосома може містити ген чорного кольору око, а парна їй материнська - ген голубого кольору. Існують визначені закони, що керують участю тих чи інших генів у розвитку особини. Зокрема, у нашому прикладі нащадок буде чорнооким, оскільки ген блакитних очей є "слабким" (рецесивним) і подавляється геном будь-якого іншого кольору.

В статевих клітинах хромосом тільки 23, і вони непарні. При заплідненні відбувається злиття чоловічої і жіночої статевих клітин і утворюється клітина зародка, що містить саме 46 хромосом. Які властивості нащадок одержить від батька, а які - від матері? Це залежить від того, які саме статеві клітини брали участь у заплідненні. Справа в тім, що процес вироблення статевих клітин (так званий мейоз) в організмі піддається випадкам, завдяки яким нащадки все-таки багато в чому відрізняються від своїх батьків. При мейозі, зокрема, відбуваються наступне: парні хромосоми соматичної клітини зближаються впритул, потім їхні нитки ДНК розриваються в декількох випадкових місцях і хромосоми обмінюються своїми частинами (рис. 1).

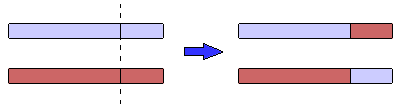


Рис.1. Умовна схема кросінговеру

Цей процес забезпечує появу нових варіантів хромосом і зветься "кросинговер". Кожна з нових хромосом з'явиться потім всередині однієї з статевих клітин, і її генетична інформація може реалізуватись в нащадках даної особини.

Другий важливий фактор, що впливає на спадковість, - це мутації, що виражаються в зміні деяких ділянок ДНК. Мутації також випадкові і можуть бути викликані різними зовнішніми факторами, такими, як радіоактивне випромінювання. Якщо мутація відбулася в статевій клітині, то змінений ген може передатися нащадку і проявитися у вигляді успадкованої хвороби або в інших нових властивостях нащадка. Вважається, що саме мутації є причиною появи нових біологічних видів, а кросинговер визначає вже змінність всередині виду (наприклад, генетичні розходження між людьми).

**Задачі оптимізації**

Як уже було відзначено вище, еволюція - це процес постійної оптимізації біологічних видів. Тепер ми в стані зрозуміти, як відбувається цей процес. Природний відбір гарантує, що найбільш пристосовані особини дадуть досить велику кількість нащадків, а завдяки генетичному спадкуванню ми можемо бути упевнені, що частина нащадків не тільки збереже високу пристосованість батьків, але буде мати і деякі нові властивості. Якщо ці нові властивості виявляються корисними, то з великою імовірністю вони перейдуть і в наступне покоління. Таким чином, відбувається нагромадження корисних якостей і поступове підвищення пристосованості біологічного виду в цілому. Знаючи, як вирішується задача оптимізації видів у природі, ми тепер застосуємо схожий метод для рішення різних реальних задач.

Задачі оптимізації - найбільш розповсюджений і важливий для практики клас задач. Їх приходиться вирішувати кожному з нас або в побуті, розподіляючи свій час між різними справами, або на роботі, домагаючись максимальної швидкості роботи програми чи максимальної прибутковості компанії - у залежності від посади. Серед цих задач є розв'язувані простим шляхом, але є і такі, точне рішення яких знайти практично неможливо.

Уведемо позначення і приведемо кілька класичних прикладів. Як правило, у задачі оптимізації ми можемо керувати декількома параметрами (позначимо їх значення через x1, x2, ..., xn, а нашою метою є максимізація (чи мінімізація) деякої функції, f(x1, x2, ..., xn), що залежить від цих параметрів. Функція f називається цільовою функцією. Наприклад, якщо потрібно максимізувати цільову функцію "дохід компанії", тоді керованими параметрами будуть число співробітників компанії, обсяг виробництва, витрати на рекламу, ціни на кінцеві продукти і т.д. Важливо відзначити, що ці параметри зв'язані між собою - зокрема, при зменшенні числа співробітників швидше за все впаде й обсяг виробництва.

Генетичний алгоритм - новітній, але не єдино можливий спосіб рішення задач оптимізації.

Відомо два основні шляхи рішення таких задач - переборний та градієнтний. Розглянемо класичну задачу комівояжера. Суть задачі полягає у знаходженні короткого шляху проходження всіх міст.

* Переборний метод є найпростішим. Для пошуку оптимального рішення (максимум цільової функції) потрібно послідовно обчислити значення функції у всіх точках. Недоліком є велика кількість обчислень.
* Іншим способом є градієнтний спуск. Обираємо випадкові значення параметрів, а потім значення поступово змінюють, досягаючи найбільшої швидкості зросту цільової функції. Алгоритм може зупинитись, досягнувши локального максимуму. Градієнтні методи швидкі, але не гарантують оптимального рішення (оскільки цільова функція має декілька максимумів).

Генетичний алгоритм уявляє собою комбінацію переборного та градієнтного методів. Механізми кросоверу (схрещування) та мутації реалізують переборну частину, а відбір кращих рішень - градієнтний спуск.

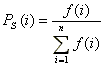
Тобто, якщо на деякій множині задана складна функція від декількох змінних, тоді генетичний алгоритм є програмою, яка за зрозумілий час знаходить точку, де значення функції знаходиться достатньо близько до максимально можливого значення. Обираючи прийнятний час розрахунку, отримуємо одне з кращих рішень, які можна отримати за цей час.

**Робота генетичного алгоритму**

Уявимо собі штучний світ, населений множиною істот (особин), причому кожна особина - це деяке рішення нашої задачі. Будемо вважати особину тим більше пристосованої, чим краще відповідне рішення (чим більше значення цільової функції воно дає). Тоді задача максимізації цільової функції зводиться до пошуку найбільш пристосованої істоти. Звичайно, ми не можемо поселити в наш віртуальний світ всі істоти відразу, тому що їх дуже багато. Замість цього ми будемо розглядати багато поколінь, що змінюють один одного. Тепер, якщо ми зуміємо ввести в дію природний відбір і генетичне спадкування, тоді отримане середовище буде підкорятися законам еволюції. Помітимо, що, відповідно до нашого визначення пристосованості, метою цієї штучної еволюції буде саме створення найкращих рішень. Очевидно, еволюція - нескінченний процес, у ході якого пристосованість особин поступово підвищується. Примусово зупинивши цей процес через досить довгий час після його початку і вибравши найбільш пристосовану особину у поточному поколінні, ми одержимо не абсолютно точну, але близьку до оптимальної відповідь. Така, коротенько, ідея генетичного алгоритму. Перейдемо тепер до точних визначень і опишемо роботу генетичного алгоритму більш детально.

Для того щоб говорити про генетичне спадкування, потрібно наділити наші особини хромосомами. У генетичному алгоритмі хромосома - це деякий числовий вектор, що відповідає параметру, який підбирається, а набір хромосом даної особини визначає рішення задачі. Які саме вектори варто розглядати в конкретній задачі, вирішує сам користувач. Кожна з позицій вектора хромосоми називається ген.

Простий генетичний алгоритм випадковим образом генерує початкову популяцію структур. Робота генетичного алгоритму уявляє собою ітераційний процес, що продовжується доти, поки не виконаються задане число поколінь або будь-який інший критерій зупинки. В кожному поколінні генетичного алгоритму реалізується відбір пропорційно пристосованості, одноточковий кросинговер і мутація. Спочатку, пропорційний відбір призначає кожній структурі імовірність Ps(i) рівну відношенню її пристосованості до сумарної пристосованості популяції:



Потім відбувається відбір (із заміщенням) усіх n особин для подальшої генетичної обробки, відповідно до величини Ps(і).

При такому відборі члени популяції з більш високою пристосованістю з більшою імовірністю будуть частіше вибиратися, ніж особини з низькою пристосованістю. Після відбору, n обраних особин випадковим чином розбиваються на n/2 пари. Для кожної пари з імовірністю Pc може застосовуватися кросинговер. Відповідно з імовірністю 1-Pc кросинговер не відбувається і незмінені особини переходять на стадію мутації. Якщо кросинговер відбувається, отримані нащадки заміняють собою батьків і переходять до мутації.

Визначимо тепер поняття, що відповідають мутації і кросинговеру в генетичному алгоритмі.

**Мутація** - це перетворення хромосоми, що випадково змінює одну чи декілька її позицій (генів). Найбільш розповсюджений вид мутацій - випадкова зміна тільки одного з генів хромосоми.

**Кросинговер** (у літературі по генетичних алгоритмах також вживається назва кросовер або схрещування) - це операція, при якій із двох хромосом породжується одна чи декілька нових хромосом. Одноточковий кросинговер працює в такий спосіб. Спочатку, випадковим образом вибирається одна з l-1 точок розриву. (Точка розриву - ділянка між сусідніми бітами в рядку.) Обидві батьківські структури розриваються на два сегменти по цій точці. Потім, відповідні сегменти різних батьків склеюються і виходять два генотипи нащадків.

Наприклад, припустимо, один батько складається з 10 нулів, а інший - з 10 одиниць. Нехай з 9 можливих точок розриву обрана точка 3. Батьки і їхні нащадки показані нижче.

Кросинговер

Батько 1 0000000000 000~0000000--> 111~0000000 1110000000 Нащадок 1

Батько 2 1111111111 111~1111111 --> 000~1111111 0001111111 Нащадок 2

Після того як закінчується стадія кросинговеру, виконуються оператори мутації. У кожному рядку, що піддається мутації, кожен біт з імовірністю Pm змінюється на протилежний.

Популяція, отримана після мутації записує поверх старої і цим цикл одного покоління завершується. Наступні покоління обробляються подібним чином: відбір, кросинговер і мутація.

В даний час дослідники ген пропонують багато інших операторів відбору, кросинговеру і мутації. От лише найбільш розповсюджені з них.

Елітні методи відбору гарантують, що при відборі обов'язково будуть виживати кращий чи кращі члени популяції сукупності. Найбільш поширена процедура обов'язкового збереження тільки одної кращої особини, якщо вона не пройшла як інші через процес відбору, кросинговеру і мутації. Елітизм може бути впроваджений практично в будь-який стандартний метод відбору.

Двоточковий кросинговер і рівномірний кросинговер - цілком гідні альтернативи одноточковому оператору. В двоточковому кросинговері вибираються дві точки розриву, і батьківські хромосоми обмінюються сегментом, що знаходиться між двома цими точками. У рівномірному кросинговері, кожен біт першого батька успадковується першим нащадком із заданою імовірністю; у противному випадку цей біт передається другому нащадку. І навпаки.

Блок-схема генетичного алгоритму зображена на рис. 2. Спочатку генерується початкова популяція особин (індивідуумів), тобто деякий набір рішень задачі. Як правило, це робиться випадковим образом. Потім ми повинні змоделювати розмноження всередині цієї популяції. Для цього випадково відбирається декілька пар індивідуумів, відбувається схрещування між хромосомами в кожній парі, а отримані нові хромосоми втілюються в популяцію нового покоління. У генетичному алгоритмі зберігається основний принцип природного відбору - чим пристосованіше індивідуум (чим більше відповідне йому значення цільової функції), тим з більшою імовірністю він буде брати участь у схрещуванні. Тепер 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 ж особин, скільки початкова, але в силу відбору пристосованість у ній у середньому вище. Тепер описані процеси відбору, схрещування й мутації повторюються вже для цієї популяції і т.д.

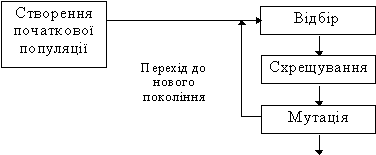


Рис. 2. Блок-схема генетичного алгоритму

В кожному наступному поколінні ми будемо спостерігати виникнення зовсім нових рішень нашої задачі. Серед них будуть як погані, так і хороші, але завдяки відбору число прийнятних рішень буде зростати. Помітимо, що в природі не буває абсолютних гарантій, і найпристосованіший тигр може загинути від рушничного пострілу, не залишивши нащадків. Імітуючи еволюцію на комп'ютері, ми можемо уникати подібних небажаних подій і завжди зберігати життя кращому з індивідуумів поточного покоління - така методика називається "стратегією елітизму".

**Застосовування генетичних алгоритмів**

Генетичні алгоритми в різних формах застосовуються до 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 областях, включаючи біологічні (екологія, імунологія і популяційна генетика) та соціальні (такі як економіка і політичні системи) системи.

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. У деяких випадках, потрібно знайти ті значення параметрів, при яких досягається найкраще точне значення функції. В інших випадках, точний оптимум не потрібний - рішенням може вважатися будь-яке значення, краще за певну задану величину. У цьому випадку, генетичні алгоритми - часто найбільш прийнятний метод для пошуку "прийнятних" значень. Сила генетичного алгоритму полягає в його здатності маніпулювати одночасно багатьма параметрами, що використовується в сотнях прикладних програм, включаючи проектування літаків, налаштування параметрів алгоритмів і пошуку стійких станів систем нелінійних диференціальних рівнянь.

Наприкінці можна підсумувати, що генетичні алгоритми є ефективною процедурою пошуку, що конкурує з іншими процедурами. Ефективність генетичних алгоритмів сильно залежить від таких деталей, як метод кодування рішень, операторів, налаштування параметрів, окремих критеріїв успіху. Теоретична робота, відбита в літературі, присвяченої генетичним алгоритмам, не дає підстав говорити про вироблення певних строгих механізмів для чітких передбачень.