

– *адаптивність* – здатність до адаптації (*адаптація* – сукупність пристосувальних, реакцій на зміну зовнішнього середовища, що протікають на різних ієрархічних рівнях, та організації сигналів зовнішнього середовища, що впливають на організм. НМ є адаптивними системами, оскільки можуть пристосовуватися до змін зовнішнього середовища за допомогою зміни своєї структури і значень параметрів);

– *здатність до навчання*: НМ здатні удосконалювати свою роботу (навчатися шляхом адаптації), використовуючи приклади для налаштування на вирішення певної задачі;

– *узагальнення* – здатність інтегрувати окремі дані для визначення закономірностей та пролонгації результатів, що дозволяє після навчання на одних даних застосовувати отримані знання для інших даних. Для НМ під властивістю узагальнення розуміється здатність генерувати правильні виходи для вхідних сигналів, які не були враховані в процесі навчання. Однак НМ не завжди може узагальнювати дані. Якщо НМ була перетренована, тобто вона запам'ятала тренувальні дані як таблицю і видає коректні дані лише при точній вказівці «адреси» в ній, то в цьому випадку мережа не зможе коректно інтерполювати вхідні дані між тренувальними. При низькому рівні узагальнення перетворення, здійснюване НМ, недостатньо гладко. *Гладкість* перетворення в НМ безпосередньо пов'язана з необхідністю вибрати найпростішу модель в умовах відсутності апріорної інформації. У даному випадку під найпростішою мається на увазі найбільш гладка функція, що найкращим чином апроксимує перетворення, яке розглядається. Вимога гладкості забезпечує гарні інтерполяційні властивості НМ і гарантує їм мінімальну складність структури, що прямо впливає на обсяг обчислень, які виконуються при навчанні. Іншою вимогою є достатня *репрезентативність* тренувального набору (навчальної вибірки) даних: НМ тим краще узагальнює, чим щільніше і рівномірніше розташовані тренувальні дані у вхідному просторі. Якщо тестові дані завжди будуть надаватися між близько розташованими тренувальними шаблонами, то мережа зможе проводити коректне узагальнення, інтерполюючи вхідні дані;

– *універсальна апроксимація*: НМ здійснюють наближення функцій багатьох змінних за допомогою лінійних операцій і суперпозицій функцій однієї змінної. Радянські вчені

О. М. Колмогоров і В. І. Арнольд показали, що будь-яку неперервну функцію N змінних можна отримати за допомогою операцій додавання, множення і суперпозиції з неперервних функцій однієї змінної. На основі цього доведено ряд теорем про апроксимації неперервних функцій багатьох змінних НМ з використанням довільної неперервної функції однієї змінної. НМ дозволяють з будь-якою точністю обчислювати довільну неперервну функцію. Отже, з їх допомогою можна як завгодно точно апроксимувати функціонування будь-якої неперервної системи. У 1987 році Р. Хехт-Нільсеном була запропонована теорема, яка доводить зображуваність функції багатьох змінних досить загального вигляду за допомогою НМ з прямими повними зв'язками, N нейронами вхідного шару, $(2N+1)$ нейронами прихованого шару з наперед відомими обмеженими функціями активації (наприклад, сигмоїдальними) і N_M нейронами вихідного шару з невідомими функціями активації. З теореми Хехт-Нільсена впливає зображуваність будь-якої багатовимірної функції декількох змінних за допомогою НМ фіксованого розмірності. Таким чином, НМ є універсальними структурами, що дозволяють реалізувати будь-який обчислювальний метод;

– *самоорганізація* являє собою процес динамічної перебудови системи з метою адаптації до зовнішнього середовища. Самоорганізація на рівні організму відбувається за допомогою навчання, в результаті якого динамічно перебудовуються нейронні структури головного мозку. Самоорганізація НМ називається також здатністю до навчання: НМ можуть автономно «вивчати» статичні і динамічні властивості модельованого об'єкта на основі результатів вимірювань, які проводилися в минулому, а потім діяти таким чином, щоб прийняти найкраще рішення при невідомому стані зовнішнього середовища. Звичайні комп'ютери повинні бути попередньо запрограмовані, щоб мати можливість обробляти дані, вони не можуть працювати за межами рішень, що задаються програмою. Таким чином, інженерія знань не може бути повною мірою реалізована на звичайних комп'ютерах, так як вони не можуть приймати рішення в нових зовнішніх умовах;

– *асоціативна пам'ять*: НМ розглядається як мультистабільна система, стійкий стан якої (атрактор) відповідає запам'ятованим

станам. Якщо зовнішній вплив встановлює мережу у стан, близький до одного з атракторів, то, будучи надана сама собі, мережа за одну або декілька ітерацій конвергенції перейде у стан атрактора. Таким чином, якщо початковий стан відображає неповний або спотворений набір раніше запам'ятованих даних, то після конвергенції весь цей набір даних буде відновлений. На цьому базується властивість асоціативної пам'яті, що дозволяє застосовувати нейрокомп'ютери для розпізнавання образів, прогнозування складних процесів і т. п. Обсяг асоціативної пам'яті НМ залежить від кількості стійких станів, положення яких визначається значенням матриці зв'язків, отриманих в результаті навчання. Число таких станів не може бути як завгодно великим. Якщо число запам'ятованих станів перевищує деяку межу, відбувається насичення пам'яті і з'являються неправдиві стійкі стани, що спотворюють вміст асоціативної пам'яті;

– *розподіленість пам'яті* – інформація зберігається за багатьма адресами, розподілених таким чином, що кожен елемент даних представляється шаблоном активності, розподіленим по багатьом обчислювальним елементам, і кожен обчислювальний елемент бере участь в уявленні багатьох різних елементів даних. У звичайних комп'ютерах реалізується локальна пам'ять, або локальне подання, в якому використовується один обчислювальний елемент для кожного елемента даних. На основі розподіленої архітектури подання інформація в НМ може дробитися і оброблятися по частинах;

– *ізоморфізм НМ топології навчальних даних і структурній інформації про модельовану залежність*. В основі принципів побудови штучних НМ закладені: *наївний ізоморфізм Кьолера*, який вимагає, щоб подання події в НМ, мало просторову структуру, що нагадує структуру відображуваного реального об'єкта, а також *ізоморфізм Гріна*, що допускає, щоб подання події в НМ мало таку логічну структуру (не обов'язково фізично реалізовану в просторовій формі), яка дозволяла б розбивати це подання, розчленовувати його і знову відновлювати за допомогою відповідних процедур або процесів направленої уваги так, що встановлюється зв'язок між частинами, поверхні або аспектами явищ реального світу. Відомо, що існують ділянки мозку –

«сенсорні проекційні поля», – в яких справді зберігається просторова структура образів. Однак залишається невідомим, наскільки далеко має поширюватися код уявлень в сенсі просторового ізоморфізму, щоб у ньому враховувалися організуючі властивості досвіду. Показано, що в ряді завдань необхідне структурне подання може бути отримано за допомогою адаптивних процесів і не потрібно, щоб уявлення були повністю ізоморфні піднаглядним явищам;

– *паралельна архітектура, масовий паралелізм і розподіленість обчислень* – обробка інформації в НМ виконується кількома процесорними елементами, в той час як у звичайних комп'ютерах, що мають тільки один центральний процесор, інформація обробляється послідовно, крок за кроком;

– *ієрархічна організація структури, цілісність і дробимість елементів НМ*: багато завдань штучного інтелекту неможливо вирішити без використання ієрархічних структур, що дозволяють будувати моделі складних об'єктів з простіших; робота ієрархічної структури вимагає, щоб інформаційний елемент у кожному ієрархічному рівні вів себе як єдине ціле, але при переході з рівня на рівень допускав дроблення, причому при переході з верхнього ієрархічного рівня на нижній це дроблення відповідає виділенню складових його елементів, а при переході з нижнього рівня на верхній воно відповідає включенню певної частини цього елемента в більш складний об'єкт;

– *попередня організація в навчанні*: незважаючи на те, що НМ мають більш широкі можливості до адаптації і навчання на прикладах порівняно з традиційними методами штучного інтелекту, вирішити на їх основі складні задачі штучного інтелекту за рахунок одного лише навчання, мабуть, неможливо. Необхідно, щоб вихідна структура НМ була попередньо організована. Для виконання цієї умови необхідно мати достатньо розвинені інструментальні засоби, що дозволяють формувати структуру НМ до початку її навчання;

– *функціональна блочність* – полягає у побудові архітектури НМ на стандартизованих функціональних блоках. Для вирішення задач штучного інтелекту необхідно будувати НМ, які мають велику кількість нейронів, тому базовими елементами для інструментальних засобів передорганізації НМ повинні бути

великі функціональні блоки, внутрішня організація і властивості яких визначені задалегідь і відомі розробнику;

– *поліалгоритмічність* – властивість, що дозволяє одній і тій самій НМ одночасно обробляти вхідну інформацію за різними методами. Поліалгоритмічність пов'язана з наявністю в мережі різних вихідних нейронів, з'єднаних з нейронами інших ієрархічних рівнів. При цьому ланцюжки, по яких збудження передається від входу до вихідних нейронів, мають різну структурну конфігурацію, яка залежно від сполучень схем проміжних нейронів, що включають збуджуючі або гальмують синапси, відповідає різним математичним формулам;

– *логічна прозорість*: логічно прозорою називають НМ, яка вирішує задачу зрозумілим для людини способом, для якої легко сформулювати словесний опис у вигляді явного методу. Така мережа повинна мати мінімальну структурну складність і при цьому задовольняти вимогам (перевагам) користувача і (або) вимогам методу автоматизованого здобуття знань до виду результуючої мережі;

– *здатність отримувати знання з даних*: в процесі навчання НМ засвоює (апроксимує) найбільш загальні закономірності, присутні в навчальних даних, витягуючи тим самим неявні для людини знання з даних. Труднощі використання таких мереж складаються в нашій обмеженій здатності пояснити, як НМ вирішує завдання: внутрішні уявлення НМ, що виходять в результаті навчання, настільки складні, що їх неможливо проаналізувати, за винятком найпростіших випадків. Спрощення НМ, що досягається контрастуванням, дозволяє отримувати логічно прозорі НМ, на основі структури яких можна побудувати вербальний опис методу отримання відповіді за наявною нейромережевою моделі – витягти явні знання з даних;

– *точність вирішення задач* НМ визначається помилкою розпізнавання (апроксимації залежності) для навчальних даних, а також помилкою розпізнавання (апроксимації залежності) для тестових даних;

– *ефективність (якість) вирішення завдань* НМ визначається точністю (помилкою) рішення задачі для навчальних і тестових

даних, простотою, логічної прозорістю і швидкодією отриманої нейромережевої моделі, а також витратами на побудову нейромережевої моделі (вимоги до апаратних засобів, ітераційність і витрати часу методу навчання);

– *варіативність моделей апроксимованої залежності*: для однієї і тієї ж таблиці даних і різних мереж або однієї мережі, але з різною початковою генерацією вихідних значень набору параметрів, що настроюються, після навчання, спрощення за єдиною схемою та вербалізацією можуть вийти різні НМ (нейромережеві моделі) і, відповідно, кілька методів вирішення однієї і тієї ж задачі. За кінцевою таблицею даних завжди будується кілька методів (моделей) рішення. Далі методи починають перевірятися і конкурувати між собою. Комбінуючи фрагменти декількох методів (моделей), можна сконструювати нову теорію. У силу цього неєдиність одержуваного знання (варіативність, неєдиність нейромережевої моделі) не є недоліком;

– *складність НМ* визначається кількістю і топологією міжнейронних зв'язків, складністю виконуваних перетворень і кількістю нейроелементів;

– *надійність та стійкість мережі до відмов окремих елементів*, складових НМ, проявляється в тому, що відмова одного або декількох нейроелементів мережі не призводить до відмови всієї НМ і не може істотно впливати на роботу мережі в цілому.

1.5 Загальне уявлення про синтез нейромереж

Нехай задана вибірка даних $\langle x, y \rangle$, $x = \{x^s\}$, $y = \{y^s\}$, $s=1, 2, \dots, S$, $x^s = \{x_j^s\}$, $j=1, 2, \dots, N$, $y^s = \{y_p^s\}$, $p = 1, 2, \dots, M$, де $\langle x^s, y^s \rangle$ – s -й екземпляр, x^s – набір значень *вхідних ознак*, що характеризує s -й екземпляр, y^s – набір значень *вихідних ознак*, зіставлений s -му екземпляру, s – номер екземпляра, j – номер вхідної ознаки, p – номер вихідної ознаки, S – кількість екземплярів, N – кількість вхідних (описових) ознак, M – кількість вихідних (цільових) ознак.

Тоді *задача синтезу нейромережевої моделі* за заданою вибіркою полягає в *ідентифікації її структури і параметрів*: $НМ = \langle C, W, DF, TF \rangle$, таким чином, щоб значення критерію

оптимальності ξ нейромоделі НМ було мінімальним: $\xi(\text{НМ}, X, Y) \rightarrow \min$, де C – матриця, що визначає наявність синаптичних зв'язків між елементами мережі (рецепторами, нейронами); $W = \{B, W(C)\}$; $W(C)$ – матриця вагових коефіцієнтів, відповідних зв'язків, присутніх в мережі НМ; $B = B(C)$ – вектор зміщень (порогів) нейронів мережі; $DF = DF(C)$ – вектор дискримінантних функцій нейроелементів; $TF = TF(C)$ – вектор функцій активації нейронів мережі; $\xi(\text{НМ}, X, Y)$ – критерій, що визначає ефективність використання нейромережевої моделі НМ для апроксимації залежності між набором вхідних параметрів X і відповідним йому вектором значень вихідного параметра Y .

Як правило, в якості критерію оптимальності нейромоделі використовується квадратичний критерій:

$$\xi = \frac{1}{2} \sum_{s=1}^S \sum_{p=1}^M \left(y_p^s - \text{НМ}_p^s(x^s) \right)^2,$$

де НМ_p^s – розрахункове значення на p -му виходу нейромоделі НМ при подані на її входи екземпляра x^s .

Процес побудови нейромереж на основі відомої вибірки навчальних даних може бути поданий у вигляді послідовності етапів:

- вибір системи інформативних ознак;
- структурний синтез;
- параметричний синтез;
- оптимізація і спрощення побудованої нейромоделі.

Розглянемо основні задачі етапів процесу побудови НМ.

Вибір системи інформативних ознак: з заданого набору ознак виділяють піднабір ознак меншого обсягу, які несуть у собі достатню для побудови нейромоделі інформацію. Виконання даного етапу забезпечує зменшення кількості входів мережі, зменшення розмірності навчальних і тестових даних, зменшення кількості ваг мережі, спрощення структури мережі, підвищення її узагальнювальних властивостей та інтерпретабельності у порівнянні з використанням великого за розміром первісного набору ознак;

Структурний синтез – етап, на якому формується топологія зв'язків, вибираються нейрони, що надалі визначають принцип

функціонування мережі та її ефективність для розв'язання досліджуваної задачі.

Так, нейромережі, що мають невелику кількість нейронів і лінійні функції активації, як правило, через свої обмежені апроксимаційні здібності не дозволяють вирішувати реальні практичні завдання. У той же час вибір надлишкової кількості нейронів у мережі призводить до проблеми перенавчання та втрати апроксимаційних властивостей нейромоделі.

Оскільки для вирішення більшості практичних задач розпізнавання і оцінювання можуть бути використані нейромережі прямого поширення з нейроелементами, що використовують зважену суму як дискримінантну функцію, вибір виду вагових функцій при структурному синтезі нейромоделей, як правило, не здійснюють.

Головною метою дослідника на етапі структурного синтезу є визначення загального вигляду структури шуканого зв'язку, представленого у вигляді нейромережевої моделі, вихідного параметра і вектора керуючих змінних.

Задача структурного синтезу нейромоделі полягає в пошуку структури моделі виду $HM = \langle C, TF \rangle$, для якої $\xi(HM, X, Y) \rightarrow \min$, де $C = C(L, A)$ – матриця, що визначає наявність синаптичних зв'язків між елементами мережі (рецепторами, нейронами); A – максимально допустима кількість нейронів у мережі.

На даний час не існує стандартних методів, які утворювали б сувору теоретичну базу для вирішення задачі структурного синтезу нейромоделей. Будучи вузловим у процесі синтезу нейромоделей реальних об'єктів, цей етап не має суворих і закінчених математичних рекомендацій щодо його реалізації. Тому його реалізація вимагає спільної роботи фахівця у відповідній предметній області та аналітика, спрямованої на якомога більш глибоке проникнення у фізичний механізм досліджуваного зв'язку.

У ряді випадків структуру нейромоделі можна вибрати апіорі, виходячи з фізичних, хімічних та емпіричних міркувань про характер взаємодії досліджуваного об'єкта з зовнішнім середовищем. Однак для складних об'єктів, процесів і систем, що характеризуються багатомірністю і нелінійністю, загальних апіорних міркувань може виявитися недостатньо.

У цьому випадку завдання вибору структури зводиться до пошуку математичними методами оптимізації за апостеріорними даними з використанням наявної додаткової інформації. Використання таких методів передбачає формування певних гіпотез про топологію мережі, які, як правило, перевіряють на основі критеріїв, що базуються на наступних ідеях:

- досягнення компромісу між складністю моделі і точністю її оцінювання;

- пошук моделі, найбільш стійкої до варіювання складу вибіркового даних, на основі яких вона оцінюється.

Існуючі *методи автоматичного пошуку оптимальної структури нейромережових моделей*, як правило, використовують *жадібну стратегію пошуку*. Так конструктивні методи починають пошук з мінімально можливою архітектурою мережі (нейромережа з мінімальною кількістю шарів, нейронів і міжнейронних зв'язків) і послідовно на кожній ітерації додають нові шари, нейрони і міжнейронні зв'язки. При використанні деструктивних методів на початковій ітерації оцінюється ефективність нейромоделі, що містить максимально допустиму кількість шарів, нейронів і міжнейронних зв'язків, потім у процесі пошуку структура такої моделі скорочується до найбільш прийнятної.

Однак такі методи внаслідок застосування жадібною стратегії досліджують незначну частину простору всіх можливих структур нейромоделей і схильні до потрапляння в локальні оптимуми.

Параметричний синтез: відбувається навчання нейромережової моделі, тобто підбираються такі значення вагових коефіцієнтів мережі, а іноді і параметрів функцій активації, при яких мережа найбільш ефективним чином дозволяє вирішувати поставлену задачу.

Задача параметричного синтезу нейромоделі заданої структури полягає в пошуку такого набору значень вагових коефіцієнтів і зміщень (порогів), при якому досягається мінімум критерію помилки нейромоделі. Навчання НМ полягає в зміні значень синаптичних ваг в результаті послідовного пред'явлення екземплярів навчальної вибірки.

Розрізняють дві **концепції навчання нейромереж**: навчання з учителем і навчання без учителя.

Навчання без учителя передбачає використання для налаштування вагових коефіцієнтів мережі тільки значення вхідних ознак екземплярів навчальної вибірки. При цьому реальні значення вихідного параметра невідомі, а процес навчання полягає в об'єднанні близько розташованих екземплярів у кластери. Таким чином, екземпляри, розташовані досить близько одне до одного в просторі ознак, будуть віднесені до одного класу при їх класифікації за допомогою налаштованої мережі.

Навчання з учителем передбачає, що навчальна вибірка містить значення вхідних ознак, що характеризують даний об'єкт, і значення вихідного параметра. При такому вигляді навчання мінімізується сума квадратів різниці між модельними та реальними виходами досліджуваного об'єкта.

У загальному випадку навчання нейронних мереж зводиться до задачі *багатовимірної нелінійної оптимізації*, розв'язуваної, як правило, за допомогою *градієнтних* або *стохастичних* (включаючи еволюційні) методів.

Оптимізація побудованої нейромоделі. На можливість застосування нейромережеских моделей на практиці істотний вплив роблять складність побудованої нейромережі та швидкість обчислення значення цільового параметра по набору даних, що не входить в навчальну вибірку. Тому актуальним є спрощення структури синтезованої нейромоделі.

Завдання оптимізації побудованої нейромоделі полягає в пошуку таких нових значень $C' \subseteq C, W', DF', TF'$, при яких досягаються оптимальні значення заданих критеріїв оптимальності $\xi_1, \xi_2, \dots, \xi_K$, враховують основні характеристики нейромоделі, де K – кількість цільових критеріїв.

В даний час існує два основні *підходи до оптимізації структури* синтезованою нейромережевою моделі:

- методи видалення зв'язків;
- методи видалення нейронів.

Методи видалення зв'язків зменшують кількість елементів матриці ваг нейромережі, спрощуючи таким чином її структуру. При використанні таких методів видаляються зв'язки, які мають

найменші значення синаптичних ваг або надають найменший вплив на ефективність функціонування нейромережевої моделі. Виділяють такі методи видалення зв'язків: метод обнуління найменших ваг, метод оптимального руйнування нейромережі, метод оптимальної розбудови мережі.

Методи видалення нейронів спрощують структуру нейромоделі шляхом виключення з неї нейронів, видалення яких не призводить до значного збільшення помилки мережі. При їх використанні додатково необхідно оцінювати значимість нейронів шляхом обчислення помилки спрощеної моделі. До таких методів відносять: метод видалення нейронів з урахуванням їх важливості і метод видалення нейронів з використанням вартісної функції.

Проте існуючі підходи спрощення нейромоделей, як правило, передбачають використання штрафних функцій, що в багатьох випадках не дозволяє отримувати оптимальну структуру нейромоделі, а іноді призводить до неможливості збіжності процесу спрощення нейромоделі або до видалення значущих зв'язків або нейронів.

1.6 Методи оптимізації в задачах синтезу нейромереж

Для побудови інтелектуальних моделей, як правило, застосовуються оптимізаційні методи, сутність яких полягає в пошуку екстремуму *цільового критерію*, в якості якого звичайно використовується середньоквадратична похибка.

Градiєнтні методи – це методи *багатовимірної нелінійної безумовної оптимізації*, що використовують обчислення похідних цільового функціоналу при корекції керованих параметрів для знаходження екстремуму цільової функції.

Нехай задано правило обчислення деякої функції $y = f(w, x)$, де w та x – складені вектори-стовпці змінних і відомі реалізації функції y при відомих значеннях x . Потрібно оцінити невідомі параметри w , таким чином, щоб значення функції y при цих параметрах незначно відрізнялися від відомих реалізацій y^* при однакових заданих значеннях x , тобто необхідно мінімізувати деяку цільову функцію $F(w)$, де w – складений вектор-стовпець змінних, що керуються, за кінцеву кількість ітерацій.

Представимо функцію $F(w)$ у вигляді інтегрального критерію якості

$$F(w) = \frac{1}{k} \sum_{m=1}^k Q(\varepsilon(w, m)),$$

де k – номер поточної ітерації, $m=1, 2, \dots, k$ – номери попередніх ітерацій, а $Q(\varepsilon(w, k))$ – деякий миттєвий критерій якості, що залежить від вектора помилки $Q(\varepsilon(w, m))$: $\varepsilon(w, m) = y(m) - y^*(m)$ і який має вигляд квадратичної форми: $Q(\varepsilon(w, m)) = \varepsilon^T(w, m) R \varepsilon(w, m)$, де R – позитивно визначена матриця.

Гradientні методи мінімізації інтегрального критерію якості засновані на використанні градієнта цільової функції $F(w)$. Ці методи носять ітеративний характер, оскільки компоненти градієнта виявляються нелінійними функціями.

Усі далі розглянуті gradientні методи засновані на ітераційній процедурі, що реалізована відповідно до формули:

$$w_{k+1} = w_k + \alpha_k p(w_k),$$

де w_k, w_{k+1} – поточне і нове наближення значень керованих змінних до оптимального рішення, відповідно, α_k – крок збіжності, $p(w_k)$ – напрямок пошуку в N -вимірному просторі керованих змінних. Спосіб визначення $p(w_k)$ і α_k на кожній ітерації залежить від особливостей конкретного методу.

Метод Коші (метод найшвидшого спуску) полягає в реалізації правила:

$$w_k = w_{k-1} - \gamma \nabla_w Q(\varepsilon(w_{k-1}, k)) = w_{k-1} - \gamma \frac{\partial Q(\varepsilon(w_{k-1}, k))}{\partial w}.$$

Позначивши поточний gradient $g = \frac{\partial Q}{\partial w}$, одержимо:

$w_{k+1} = w_k - \gamma_k g_k$, де γ_k – швидкість навчання.

Величина γ або покладається постійною, і тоді звичайно послідовність w_k сходиться в околицю оптимального значення w , або вона є спадаючою функцією часу так, як це робиться в стохастичних методах оптимізації й адаптації.

Дана процедура може виконуватися доти, доки значення керованих змінних не стабілізуються, або доки помилка не зменшиться до прийняттого значення.

Слід зазначити, що дану процедуру характеризує повільна швидкість збіжності і можливість потрапляння в локальні мінімуми цільового функціонала.

Метод Ньютона відносно до задачі мінімізації цільової функції $F(w)$ має вигляд:

$$w_k = w_{k-1} - \left(\sum_{m=1}^k \frac{\partial}{\partial w} \left(\frac{\partial Q(m)}{\partial w} \right)^T \right)^{-1} \gamma \nabla_w Q(\varepsilon(w_{k-1}, k)).$$

Обчислення градієнта $\nabla_w Q(\varepsilon(w_{k-1}, k))$ може бути виконано за допомогою методу Коші.

У результаті задача зводиться до обчислення матриці других похідних $\sum_{m=1}^k \frac{\partial}{\partial w} \left(\frac{\partial Q(m)}{\partial w} \right)^T$ мінімізованого функціонала.

Метод Ньютона за певних умов має істотно більшу швидкість збіжності, ніж метод Коші.

Методи спряжених градієнтів утворюють підклас методів, що збігаються квадратично. Методи спряжених градієнтів використовують метод Коші для обчислення похідних цільової функції щодо керувальних змінних.

Всі методи спряжених градієнтів на першій ітерації починають роботу з пошуку у найкрутішому напрямку спуску – протилежному градієнту цільової функції: $p_0 = -g_0$. Це напрямок, у якому цільова функція зменшується найбільш швидко. Однак зазначимо, що це не обов'язково забезпечує найшвидшу збіжність.

Після цього виконується лінійний пошук, щоб визначити оптимальну відстань для руху у поточному напрямку пошуку: $w_{k+1} = w_k + \alpha_k p_k$.

Наступний напрям пошуку визначається так, щоб він був спряженим з попередніми напрямками. Пошук відбувається за спряженим напрямком градієнта, щоб визначити розмір кроку, що мінімізує цільову функцію. Загальна процедура для визначення нового напрямку пошуку поєднує новий найкрутіший напрямок спуску з попереднім напрямком пошуку:

$$p_{k+1} = -g_k + \beta_k p_{k-1}.$$

У більшості методів спряжених градієнтів розмір кроку корегується при кожній ітерації, на відміну від інших методів, де

швидкість навчання використовується для визначення розміру кроку.

Різні версії методів спряжених градієнтів відрізняються способом, за яким обчислюється константа β :

– у методі Флетчера-Рівса це правило має вигляд:

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}},$$

де β – відношення квадрата норми поточного градієнта до квадрата норми попереднього градієнта;

– у методі Полака-Ріб'єра це правило має вигляд:

$$\beta_k = \frac{\Delta \mathbf{g}_{k-1}^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}},$$

де β – внутрішній добуток попередньої зміни в градієнті і поточного градієнта, розділений на квадрат норми попереднього градієнта.

Квазіньютонівські методи подібні методам спряжених градієнтів, оскільки також засновані на властивостях квадратичних функцій. Відповідно до викладених вище методів пошуку рішення здійснюється за системою спряжених напрямків, тоді як квазіньютонівські методи мають позитивні риси методу Ньютона, однак використовують тільки перші похідні. В усіх методах зазначеного класу побудова векторів напрямків пошуку здійснюється за допомогою формули:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha_k A_k \mathbf{g}_k,$$

де A_k – матриця порядку $N \times N$, що називається *метрикою*. Методи пошуку уздовж напрямків, обумовлених цією формулою, називаються *методами змінної метрики*, оскільки матриця A змінюється на кожній ітерації.

Градієнт \mathbf{g}_k може бути знайдений за допомогою методу Коші і, отже, квазіньютонівські методи полягають у знаходженні матриці A_k .

Метод Левенберга-Марквардта є одним з найпопулярніших та ефективних квазіньютонівських методів. В ньому використовується метод Коші, щоб обчислити якобіан J цільової функції щодо керованих змінних. Керовані змінні змінюються відповідно до коригувального правила, що у матричній формі має вигляд:

$$\mathbf{H} = J^T J, \quad \mathbf{g} = J^T \mathbf{e},$$

$$w_{k+1} = w_k - [H_k + \eta I]^{-1} g_k,$$

де J – якобіан, e – вектор помилок, η – скаляр, I – одинична матриця.

На початковій стадії пошуку параметру η_0 приписується велике значення. Якщо після першого кроку отримана точка з меншим значенням цільової функції, то слід встановити: $\eta_{k+1} = \eta^{(-)} \eta_k$, в іншому випадку – встановити: $\eta_{k+1} = \eta^{(+)} \eta_k$ і знову реалізувати попередній крок. Тут $\eta^{(-)}$ та $\eta^{(+)}$ деякі константи, такі що: $\eta^{(-)} < 1$, $\eta^{(+)} > 1$.

Гradientні методи оптимізації передбачають обчислення значень цільової функції, що ускладнює, а в деяких випадках і унеможлиблює їх застосування на практиці. Еволюційні методи, на відміну від gradientних, не вимагають диференційованості цільової функції та не накладають обмежень на її вигляд.

Еволюційний (генетичний) пошук – група багатовимірних, стохастичних, евристичних оптимізаційних методів, заснованих на ідеї еволюції за допомогою природного відбору. Методи генетичного пошуку отримані в процесі узагальнення та імітації в штучних системах таких властивостей живої природи, як природний відбір, пристосовність до змінюваних умов середовища, спадкоємність нащадками життєво важливих властивостей від батьків і т. ін.

Формально методи генетичного пошуку можуть бути описані у вигляді такої функції: $GM = GM(P_0, N, L, f, \Omega, \Psi, \Theta, T)$, де $P_0 = \{H_1^0, H_2^0, \dots, H_N^0\}$ – початкова популяція – множина рішень задачі, поданих у вигляді хромосом; $H_j^0 = \{h_{1j}^0, h_{2j}^0, \dots, h_{lj}^0\}$ – j -та хромосома (рішення) популяції P_0 – набір значень незалежних змінних, поданих у вигляді генів; h_{ij}^0 – i -й ген j -ї хромосоми популяції P_0 – значення i -го оптимізованого параметру задачі, що входить в j -те рішення; N – кількість хромосом в популяції (розмір популяції); L – довжина хромосом (кількість генів); f – цільова функція (фітнес-функція, функція пристосованості, функція здоров'я, функція придатності); Ω – оператор відбору; Ψ – оператор схрещування; Θ – оператор мутації; T – критерій зупинення.

На кожній ітерації генетичного пошуку метод працює не з єдиним рішенням, а із деякою множиною рішень (сукупністю хромосом), за рахунок чого забезпечується паралельність пошуку.

При цьому кожна нова множина рішень залежить лише від попередньої і, в загальному випадку, є кращою за попередню.

Оскільки генетичні методи в процесі пошуку використовують деяке *кодування множини параметрів* замість самих параметрів, то вони можуть ефективно застосовуватися для вирішення задач дискретної оптимізації, визначених як на числових множинах, так і на скінчених множинах довільної природи.

Для роботи генетичних методів як інформація про функцію, що оптимізується, використовуються її значення в точках простору пошуку і не потрібно обчислювати похідні або інші характеристики. Тому дані методи можуть бути застосовані до широкого класу функцій, зокрема до тих, що не мають аналітичного опису. Таким чином, методи генетичного пошуку є достатньо гнучкими і можуть бути застосовані до широкого кола задач, в тому числі до задач, для розв'язування яких не існує загальновідомих методів.

Сутність генетичного пошуку полягає в циклічній заміні однієї популяції наступною, більш пристосованою. Таким чином, популяція існує не тільки в просторі, але і в часі. Часто можна вважати, що вся *популяція* складається в просторі і в часі з дискретних поколінь (генерацій, епох) $P_0, P_1, P_2, \dots, P_T$.

Покоління P_{t+1} – це сукупність *особин* (хромосом, рішень), *батьки* яких належать поколінню P_t . Покоління P_0 є початковою популяцією. Процес формування покоління P_0 називається ініціалізацією. Кожне покоління є результатом циклу роботи генетичного методу.

Кожна хромосома (особина, точка в просторі пошуку) оцінюється *мірою її пристосованості* відповідно до того, наскільки є гарним відповідне їй рішення задачі. Пристосованість визначається як обчислена цільова функція (фітнес-функція) для кожної з хромосом.

Правила *відбору (селекції)* прагнуть залишити лише ті точки-рішення, де досягається оптимум цільової функції.

Найбільш пристосовані особини дістають можливість відтворювати нащадків за допомогою перехресного *схрещування* з іншими особинами популяції. Це призводить до появи нових особин, які поєднують в собі деякі характеристики, успадковані ними від батьків. Найменш пристосовані особини з меншою

Ймовірно зможуть відтворити нащадків, внаслідок чого ті властивості, якими вони володіли, поступово зникатимуть з популяції в процесі еволюції.

Таким чином, з покоління в покоління, гарні характеристики розповсюджуються по всій популяції. Комбінація гарних характеристик від різних батьків іноді може призводити до появи суперпристосованого нащадка (або *мутанта*), чия пристосованість більша, ніж пристосованість будь-якого з його батьків. Схрещування найбільш пристосованих особин призводить до того, що досліджуються найбільш перспективні ділянки простору пошуку. Зрештою, популяція збігатиметься до оптимального рішення задачі.

Після схрещування інколи відбуваються *мутації* – спонтанні зміни в генах, які випадковим чином розкидають точки по всій множині пошуку.

Процес збіжності за рахунок відбору повинен бути більш виражений порівняно з розкидом точок за рахунок мутації і інвертування, інакше збіжність до екстремумів не матиме місця. Збіжність за рахунок відбору не повинна бути занадто швидкою, інакше всі точки можуть зібратися поблизу локального екстремуму, а інший, можливо глобальний, так і не буде знайдено.

Після отримання нащадків за допомогою схрещування та мутації розмір популяції збільшується. Для подальших перетворень число хромосом поточної популяції зменшується до заданого розміру популяції.

Подальша робота генетичного методу є ітераційним процесом застосування *генетичних операторів* до особин наступного покоління. Генетичні оператори необхідні, щоб застосувати принципи спадковості і мінливості до популяції. Такі оператори володіють властивістю ймовірності, тобто вони не обов'язково застосовуються до всіх особин, що вносить додатковий елемент невизначеності в процес пошуку рішення. Невизначеність не має на увазі негативного чинника, а є своєрідною мірою свободи роботи генетичного методу.

Схему роботи узагальненого генетичного методу наведено на рис. 1.4.

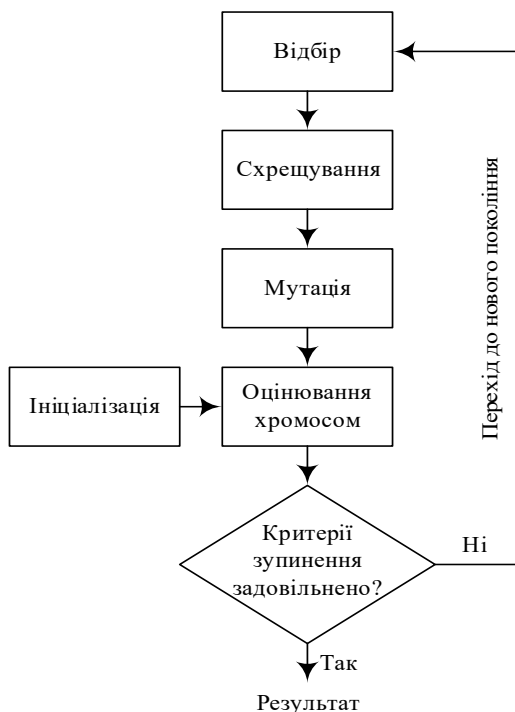


Рисунок 1.4 – Схема роботи узагальненого генетичного метода

Узагальнений метод генетичного пошуку можна записати таким чином.

Крок 1. Встановити лічильник ітерацій (часу): $t = 0$. Виконати ініціалізацію (initialization) початкової популяції особин: $P_t = \{H_1, H_2, \dots, H_N\}$.

Крок 2. Оцінити особини поточної популяції (evaluating) шляхом обчислення їхніх фітнес-функцій $f(H_j), j = 1, 2, \dots, N$.

Крок 3. Перевірити умови закінчення пошуку (termination criteria). Як такі умови можуть бути використані: досягнення максимально допустимого часу функціонування методу, числа ітерацій, значення функції пристосованості і т. ін. Якщо критерії закінчення пошуку задовільнено, тоді виконати перехід до кроку 12.

Крок 4. Збільшити лічильник ітерацій (часу): $t = t + 1$.

Крок 5. Вибрати частину популяції (батьківські особини) для схрещування (selection of parents) P' .

Крок 6. Сформувати батьківські пари (mating) з особин, що відібрані на попередньому кроці.

Крок 7. Схрестити (crossover) вибрані батьківські особини.

Крок 8. Застосувати оператор мутації (mutation) до особин P' .

Крок 9. Обчислити нову функцію пристосованості $f(H_j)$ особин, отриманих в результаті схрещування та мутації.

Крок 10. Сформувати нове покоління шляхом вибору особин, що вижили, виходячи з рівня їх пристосованості (replacing, selection of survivors).

Крок 11. Перейти до кроку 3.

Крок 12. Зупинення.

Генетичні методи є більш ефективним інструментом пошуку в порівнянні з класичними методами оптимізації за таких умов:

– досліджуваний простір пошуку є великим, негладким (існують точки розриву) і неунімодальним (є декілька оптимумів);

– цільова функція пошуку може мати шуми;

– задача не вимагає знаходження надточного глобального оптимуму. Тобто необхідно достатньо швидко знайти прийнятне рішення, що досить часто має місце в реальних задачах.

Таким чином, генетичний пошук може успішно використовуватися для вирішення комбінаторних задач, а також для пошуку оптимальних значень полімодальних функцій.

До переваг генетичних методів відносять:

– відсутність необхідності в специфічних знаннях про вирішувану задачу. Проте у випадку, якщо додаткова інформація про досліджувану систему, об'єкт або процес є відомою, то вона може бути використана в процесі пошуку;

– концептуальна простота та прозорість реалізації;

– можливість розпаралелювання;

– простота кодування вхідної і вихідної інформації.

Некритичність до виду параметрів досліджуваних систем (можливість використання експертної, емпіричної, довідкової та іншої інформації про об'єкт, поданої різними типами даних);

– можливість застосування до великого кола задач без внесення серйозних змін у внутрішню структуру методу;

– можливість адаптивності параметрів генетичного пошуку до особливостей вирішуваної задачі;

– менша ймовірність попадання і зациклення в локальному оптимумі, яка досягається за рахунок використання популяційного підходу;

– можливість застосування в методі інших пошукових процедур.

До *недоліків генетичного пошуку* відносять:

– високу ітеративність;

– сильну залежність ефективності генетичного пошуку від його параметрів (розмір популяції, початкова точка пошуку, імовірнісні характеристики генетичних операторів і т. ін.);

– *епістазис* – внутрішню залежність між змінними (генами), закодованими в хромосомі. Якщо гени не пов'язані один з одним, то вважається, що епістазис малий або не існує. У випадку, якщо гени є взаємозалежними, то епістазис може створювати проблеми для генетичного пошуку, спричинені тим, що при схрещуванні ланцюжки взаємозалежних генів будуть зруйновані, що призводить до появи низько пристосованих нащадків. Вирішення проблеми епістазису полягає в тому, щоб зберігати в хромосомі взаємозалежні гени, розташовуючи їх близько один до одного. При групуванні залежних генів істотно знижується ймовірність того, що вони будуть зруйновані при застосуванні схрещування;

– *передчасна збіжність*, пов'язана з недостатньою різноманітністю хромосом в популяції. Найпоширенішою причиною передчасної збіжності є недостатній розмір популяції. Таким чином, вирішенням такої проблеми може бути збільшення кількості хромосом в популяції.

Враховуючи особливості, переваги і недоліки оптимізаційних методів, можна надати такі *рекомендації щодо вибору пошукового методу* при вирішенні практичних задач:

– якщо простір пошуку є дискретним і невеликим за розміром, то можна скористатися методом повного перебору, який знайде найкраще рішення. Генетичний метод, на відміну від методу повного перебору, може з більшою ймовірністю зійтися до локального оптимуму, а не до глобального. Проте генетичний пошук швидше знайде субоптимальне рішення, що знаходиться недалеко від дійсного оптимуму;

– якщо цільова функція в пошуковому просторі є гладкою і унімодальною, то будь-який градієнтний метод (наприклад, метод найшвидшого спуску) буде ефективнішим, ніж генетичний пошук;

– якщо про простір пошуку відома деяка додаткова інформація (наприклад, для задачі комівояжера), то методи пошуку, що використовують апріорні відомості про пошуковий простір, часто

перевершують будь-який універсальний метод, у тому числі й генетичний метод;

– при достатньо складному рельєфі цільової функції методи пошуку, що працюють із єдиним рішенням на кожній ітерації (наприклад, простий метод спуску), можуть зациклитися в локальному рішенні. Генетичні методи працюють з набором із декількох рішень, тому вони мають менше шансів зійтися до локального оптимуму і надійно функціонують на багатоекстремальних поверхнях.

? 1.7 Контрольні питання

1. Біологічні нейрони. Математичні моделі нейроелементів.
2. У чому полягають переваги еволюційних методів?
3. Властивості штучних нейромереж.
4. Для чого призначений оператор мутації?
5. Загальна характеристика та принципи побудови нейромереж.
6. Загальне уявлення про навчання нейромереж.
7. Структурний синтез нейромереж.
8. Параметричний синтез нейромереж.
9. Класифікація та види моделей нейромереж.
10. Наведіть особливості методів спряжених градієнтів.
11. Наведіть особливості, переваги та недоліки градієнтних методів оптимізації.
12. Наведіть послідовність виконання узагальненого еволюційного пошуку.
13. Назвіть особливості еволюційних методів.
14. Особливості багатопараметричного пошуку.
15. Поняття: нейрон, нейромережа, нейрокомп'ютер.
16. Поняття: синапс, вага (ваговий коефіцієнт), поріг, дискримінантна (вагова) функція, функція активації, бажаний і реальний вихід нейромережі.
17. Опишіть квазіньютонівські методи.
18. Порівняйте методи еволюційного пошуку з іншими методами оптимізації.
19. Які методи відносять до еволюційних?
20. Порівняйте методи спряжених градієнтів.
21. Порівняйте стратегії створення початкової популяції.

22. Поясніть призначення методу Левенберга-Марквардта. Порівняйте його з іншими градієнтними методами.
23. Поясніть схему роботи методу Ньютона.
24. Проаналізуйте метод Коші.
25. Проаналізуйте узагальнену схему роботи еволюційних методів.
26. Проаналізуйте умови ефективного використання методів еволюційного пошуку.
27. У чому подібність і відмінність біологічних і формальних нейронів?
28. Характеристики нейромереж.
29. Що таке фітнес-функція?
30. Яким чином відбувається формування нового покоління при еволюційному пошуку?
31. Які критерії зупинення використовуються при еволюційному пошуку?
32. Які недоліки еволюційного пошуку та в чому вони полягають?

1.8 Практичні завдання



Завдання 1. Побудувати за графіки функцій активації:

- а) логістичний сигмоїд (функція logsig);
- б) тангенційний сигмоїд (функція tansig);
- в) лінійна (функція purelin);
- г) лінійна уніполярна з насиченням (функція satlin);
- д) радіально-базисна (функція radbas);
- е) порогова (функція hardlim).



Завдання 2. Побудувати графіки перших похідних функцій активації:


- а) логістичний сигмоїд (функція dlogsig);
- б) тангенційний сигмоїд (функція dtansig);
- в) лінійна (функція dpurelin);
- г) лінійна уніполярна з насиченням (функція dsatlin);
- д) радіально-базисна (функція dradbas);
- е) порогова (функція dhardlim).



Завдання 3. Проаналізуйте, чи пов'язані (і якщо так, то яким чином, а якщо ні, то чому) такі властивості нейромереж, як:

- а) адаптивність і функціональна блочність;
- б) асоціативна пам'ять і універсальна апроксимація;

- в) варіативність і узагальнення;
- г) ефективність і точність;
- д) здатність до навчання і складність;
- е) здатність отримувати знання з даних і самоорганізація;
- ж) ієрархічність і розподіленість пам'яті;
- з) ізоморфізм і попередня організація;
- и) логічна прозорість і поліалгоритмічність;
- к) надійність і пластичність;
- л) нелінійність і паралельна архітектура;
- м) однорідність і надійність;
- н) паралельна архітектура і адаптивність;
- о) пластичність і асоціативна пам'ять;
- п) поліалгоритмічність і варіативність ;
- р) попередня організація і ефективність;
- с) розподіленість пам'яті і здатність до навчання;
- т) самоорганізація і здатність отримувати знання з даних;
- у) складність і ієрархічність ;
- ф) точність і ізоморфізм;
- х) узагальнення і логічна прозорість;
- ц) універсальна апроксимація і надійність;
- ч) функціональна блочність і нелінійність.

 Завдання 4. Підготувати реферат на одну з таких тем.

1. Історія розвитку нейроінформатики.
2. Фізичні моделі нейронів.
3. Фізичні моделі нейромереж.
4. Сучасні програмні засоби для моделювання нейромереж.
5. Сучасні апаратні засоби для моделювання нейромереж.
6. Еволюційні методи в задачах синтезу нейромереж.