

Добровська Людмила Миколаївна
кандидат педагогічних наук,
доцент кафедри біомедичної кібернетики
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Добровская Людмила Николаевна
кандидат педагогических наук,
доцент кафедры биомедицинской кибернетики
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»

Dobrovska Liudmyla
Candidate of Pedagogical Sciences,
Associate Professor of the Department of Biomedical Cybernetics
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»

Сітченко Олексій Євгенович
студент факультету біомедичної інженерії (ФБМІ)
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»

Ситченко Алексей Евгеньевич
студент факультета биомедицинской инженерии (ФБМИ)
Национального технического университета Украины
«Киевский политехнический институт имени Игоря Сикорского»

Sitchenko Oleksii
Student of the Faculty of Biomedical Engineering (FBME)
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»

КЛАСИФІКАТОР РАКУ МОЛОЧНОЇ ЗАЛОЗИ З ВИКОРИСТАННЯМ ПЕРСЕПТРОНУ ТА ГЕНЕТИЧНОГО АЛГОРИТМУ

КЛАССИФИКАТОР РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ С ИСПОЛЬЗОВАНИЕМ ПЕРСЕПТРОНА И ГЕНЕТИЧЕСКОГО АЛГОРИТМА

BREAST CANCER CLASSIFIER USING PERSEPTRON AND GENETIC ALGORITHM

Анотація. У даній статті розглянуто метод навчання багатошарового персептрону на основі генетичного алгоритму з використанням бази даних про захворювання на рак молочної залози.

Ключові слова: нейронна мережа, персептрон, генетичний алгоритм, рак молочної залози.

Аннотация. В данной статье рассмотрен метод обучения многослойного персептрона на основе генетического алгоритма с использованием базы данных о заболевании раком молочной железы.

Ключевые слова: нейронная сеть, персептрон, генетический алгоритм, рак молочной железы.

Summary. This article discusses a method for training a multilayer perceptron based on a genetic algorithm using a breast cancer database.

Key words: neural network, perceptron, genetic algorithm, breast cancer.

Вступ. На даний час нейронні мережі широко використовуються у повсякденному житті, їх використання відкриває нескінченну кількість можливостей, а результати дають рішення на задачі, які раніше вважалися б нерозв’язними. Нейронні мережі здатні визначати діагнози пацієнтам, прогнозувати хвороби, класифікувати їх ступені або види та приймати рішення щодо лікування пацієнтів, але досі, величезна кількість лікарень віддає перевагу людським робочим кадрам.

Зрозуміло, що машина не може повністю замінити людину, але допомогти і доповнити її у вирішенні проблем чи питань завжди зможе. Коли людину беруть на роботу у неї завжди цікавляться її наявним досвідом чи навичками. Це дуже важливий момент, який також присутній у контексті нейронних мереж. Важливим питанням при роботі із нейронними мережами є якість їх навчання. Існують різні алгоритми навчання мереж. Важливим є пошук нових і покращення існуючих методів навчання нейронних мереж.

Основна частина. У рамках даного дослідження для нейронної мережі буде запропоновано модель багатопередового перцептрон. Для його навчання буде використовуватись генетичний алгоритм.

Генетичний алгоритм — це еволюційний алгоритм пошуку рішень. Він, користуючись деякими правилами та операторами, виконує підбір оптимальних параметрів із згенерованої та непостійної вибірки можливих рішень поставленої задачі. Ключовим об’єктом у генетичному алгоритмі є «особа» — об’єкт, який зображує хромосому. Хромосома складається із генотипів, які у свою чергу складаються із генів. Зазвичай, один ген є одним бітом і може мати лише два значення, «0» або «1». Загальна кількість існуючих осіб у генетичному алгоритмі називається популяцією.

У загальному вигляді генетичний алгоритм можна подати у вигляді рисунку 1.

1. Формування популяції. Початковий етап генетичного алгоритму, на якому відбувається генерація визначеної користувачем кількості осіб, в хромосоми яких записані випадкові значення генів (бітів).

2. Розрахунок пристосованості. За деяким правилом для кожної особи, за її хромосомою, визначається значення пристосованості. За цим значенням

алгоритм судитиме, наскільки дана особа важлива у пошуку оптимального рішення задачі і чи вона загине, чи, пристосувавшись, зможе перейти до наступної популяції.

3. Селекція. Відбір найбільш пристосованих осіб до наступної популяції (наступного циклу алгоритму). У даній роботі було обрано класичний метод турнірної селекції, де уся популяція ділиться по парам, і у кожній парі за пристосованістю осіб опонентів робиться вибір.

4. Виконання умов закінчення роботи алгоритму. Зазвичай цими умовами виступає цільова пристосованість, при отриманні якої особа обирається як оптимальне рішення задачі, або обмеження по кількості пройдених циклів алгоритму.

5. Репродукція. Важливий етап алгоритму, який виступає рушієм у пошуку оптимального рішення. Після селекції гинуть багато осіб, отже необхідно створювати нові. Постійно генерувати нові особи випадковим чином не правильно і це вбиває увесь пошук рішень, роблячи його банальним перебором. У генетичному алгоритмі існують генетичні оператори, і зазвичай їх два:

Оператор кросоверу. Обираються дві особи, а також точка поділу їх хромосом. Поділивши хромосоми цих осіб у визначеній точці, із них (хрестнахрест поєднавши частини хромосом) отримуються дві особи-нащадки, які несуть в собі по дві частини хромосом своїх предків.

Оператор мутації. У обраній хромосомі змінюється значення одного випадкового гену.

6. Отримання найкращої особи. Визначення особи із найкращим значенням пристосованості, як оптимального рішення для поставленої задачі [1–2].

Розглянувши базові принципи перцептрон та генетичного алгоритму постає ключове питання: Як навчати перцептрон за допомогою генетичного алгоритму? Це стає можливим, якщо його набір вагових коефіцієнтів подати як одну хромосому. Те як саме одна хромосома інтерпретує цілу нейронну мережу називається її кодуванням. Для того щоб пояснити принцип кодування хромосоми, необхідно представити структуру перцептрон (рис. 2).

Як видно із рисунку 2, даний перцептрон має структуру із двох входів, одного прихованого шару

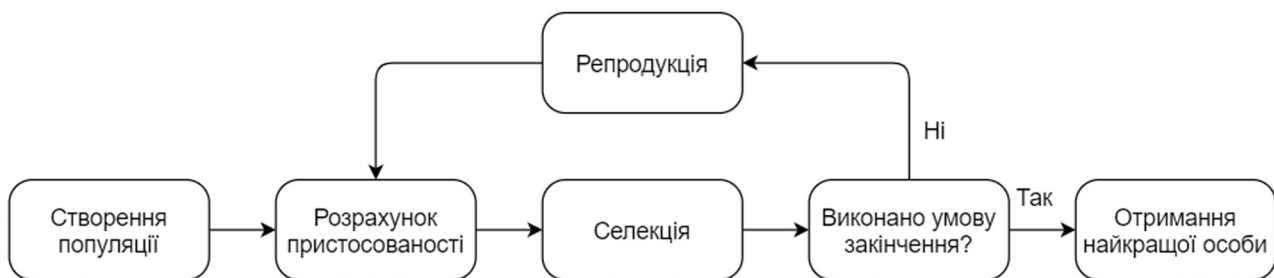


Рис. 1. Загальний вигляд генетичного алгоритму
Джерело: згідно з [1]

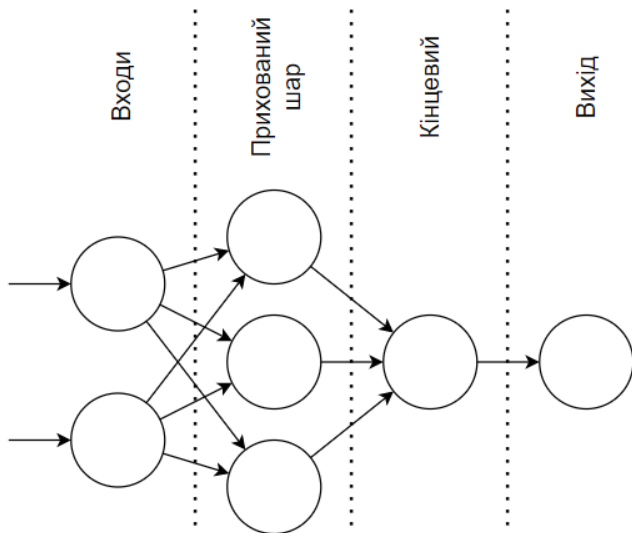


Рис. 2. Приклад структури перцептрон

у якому знаходяться три нейрони, і одного виходу. Знаючи це можна порахувати кількість зв'язаних нейронних зв'язків (а також зсувів) у нейронах. Маємо 9 зв'язків і 4 зсуви, що у сумі дає 13 чисел, які повинні кодувати одна хромосома.

Нехай для двійкового відображення одного числа було встановлено число 3. Перший біт відповідатиме за знак цього числа, а решта за його модуль. Отже маємо, що виділяючи по три біти на число, для кодування перцептроні із такою структурою, хромосома має бути довжиною у 39 генів. Приклад декодування хромосоми на рисунку 3.

У дослідженні розглядається задача бінарної класифікації тож значенням пристосованості для кожної такої особи буде середньоквадратична помилка моделі визначення класів для тренувальної вибірки спостережень наявного набору даних про захворювання на рак молочної залози. І чим вона менша, тим пристосованішою буде особа [3–4].

Для цього дослідження був взятий набір даних про захворювання на рак молочної залози, який надано Центром клінічних наук Університету Вісконсіна, і який у вільному доступі знаходиться в мережі Інтернет. В ньому налічується 569 спостережень та 12 атрибутів. З цих атрибутів буде видалено ID, залишаючи лише 10 атрибутів що несуть у собі середні значення показників ракових клітин, а також один бінарний атрибут для класифікації, в якому зафіксовано чи пухлина доброякісна, чи злоякісна. Всього у базі даних 357 спостережень із доброякісними пухлинами і 212 — із злоякісними. Її було поділено на тренувальну та експериментальну вибірки у відношенні 0,75/0,25.

Визначивши алгоритм навчання багат шарового перцептроні за допомогою генетичного алгоритму, і маючи на руках базу даних захворювань на рак молочної залози, на мові програмування C++ було розроблено програмний додаток для класифікації

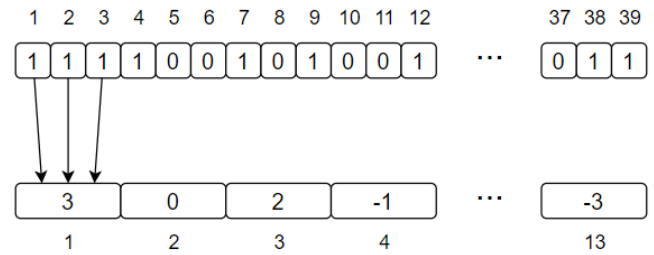


Рис. 3. Приклад отримання інформації із хромосоми

злаякісних пухлин, із графічним користувацьким інтерфейсом створеним за допомогою Qt Creator.

Для тестувань у розробленому програмному додатку було встановлено параметри, представлені у таблиці 1.

Оскільки у обраному наборі даних, окрім бінарного атрибуту класу, присутні ще десять атрибутів, то кількість входів у перцептроні було встановлено також десять. Кількість прихованих шарів, нейронів у них, а також кількість бітів на число було обрано після проведення десятків тестувань, як оптимальні значення для цього дослідження.

Роботу програми та результат проведення сотні циклів алгоритму представлено на рисунках 4 та 5.

Надані програмою дані дозволили сформувавши матриці невідповідностей для тренувальної та експериментальної вибірок, після однієї сотні ітерацій алгоритму. Вони представлені у таблицях 2 і 3.

Маючи дані матриці можна розрахувати точність.

Для тренувальної вибірки точність склала 88.028%.

А для експериментальної — 84.615%.

Висновки. Розглянувши базові принципи роботи багат шарового перцептроні і генетичного алгоритму, було розроблено програмне забезпечення для класифікації раку молочної залози, за допомогою перцептроні та генетичного алгоритму. Проведено його тестування та отримано 84.615% точність на експериментальній вибірці, що є чудовим результатом. Використання даних цього дослідження у клініках може покращити лікування пацієнтів, що хворі на рак молочної залози.

Таблиця 1

Встановлені для тестування параметри

| № | Назва | Значення |
|---|------------------|----------|
| 1 | Вхід | 10 |
| 2 | Вихід | 1 |
| 3 | Прихованих шарів | 3 |
| 4 | Нейронів/шар | 20 |
| 5 | Популяція | 1000 |
| 6 | Бітів/число | 7 |
| 7 | Шанс кроссоверу | 100 |
| 8 | Шанс мутації | 10 |
| 9 | Ліміт популяцій | 100 |

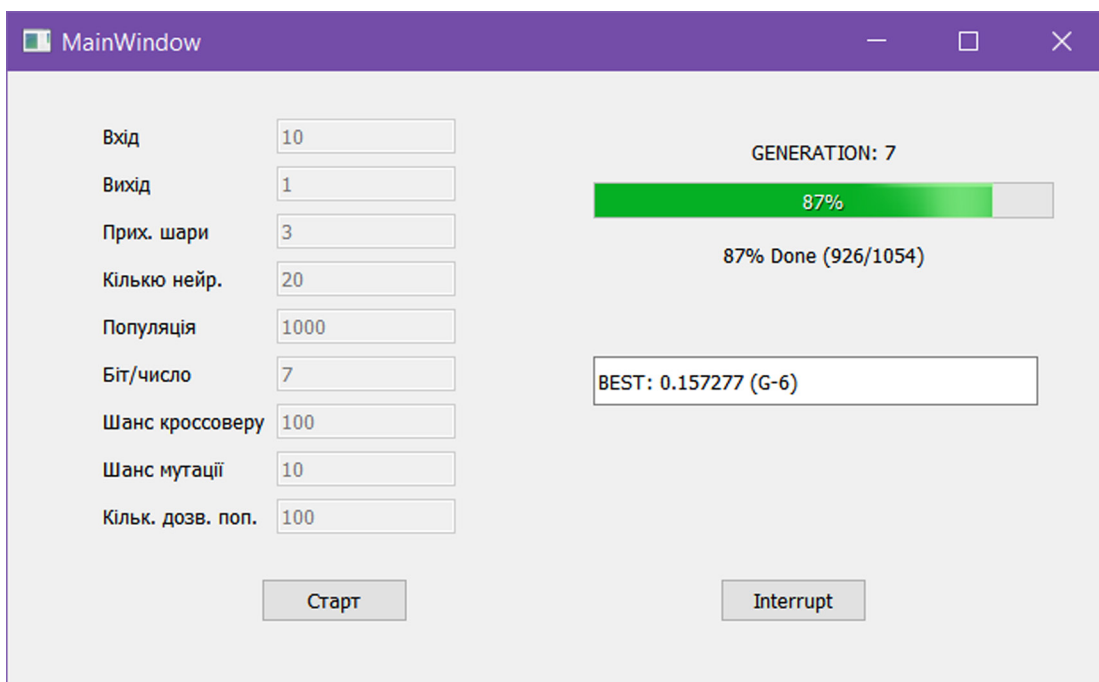


Рис. 4. Процес навчання нейронної мережі

Джерело: розробка автора

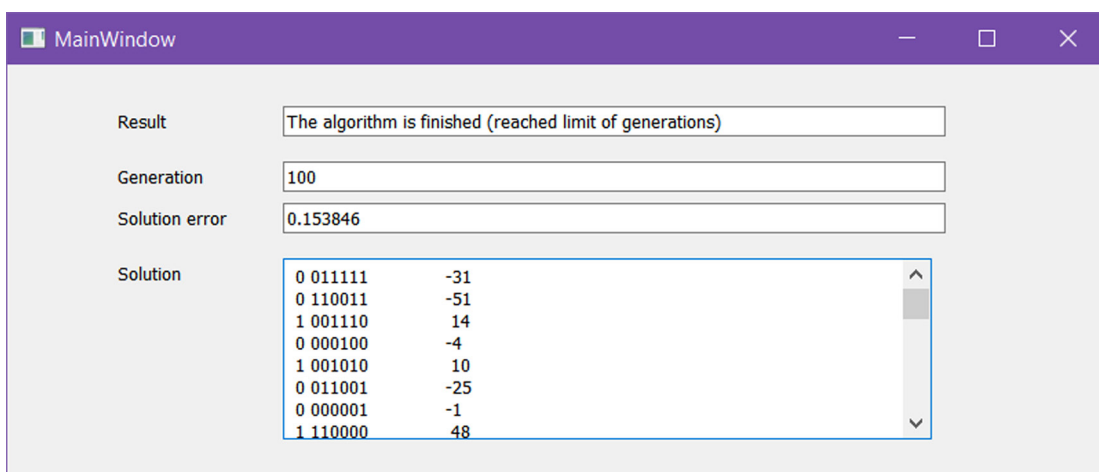


Рис. 5. Помилка найкращої моделі на експериментальній вибірці після навчання

Джерело: розробка автора

Таблиця 2

Матриця невідповідності для тренувальної вибірки

| | | Фактично | |
|---------------|-------------|-------------|----------|
| | | Доброякісні | Злякісні |
| Класифіковано | Доброякісні | 236 | 20 |
| | Злякісні | 31 | 139 |

Таблиця 3

Матриця невідповідності для експериментальної вибірки

| | | Фактично | |
|---------------|-------------|-------------|----------|
| | | Доброякісні | Злякісні |
| Класифіковано | Доброякісні | 76 | 8 |
| | Злякісні | 14 | 45 |

Література

1. Добровська Л. М., Добровська І. А. Теорія та практика нейронних мереж. Київ: НТУУ «КПІ», 2015. 395 с.
2. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. Москва: Горячая линия — Телеком, 2013. 383 с.
3. Цой Ю. Р. Эволюционный подход к настройке и обучению искусственных нейронных сетей — Нейроинформатика, 2006. С. 34–61.
4. Ding S. F., Xu L., Su C. Y. Using Genetic Algorithm to Optimize Artificial Neural Networks // Journal of Convergence Information Technology. 2010. Vol. 5. С. 54–62.