

$$S_d(I_i, q) = \frac{|K_{i,q}|}{|K_q|}, \quad (8)$$

де  $|K_{i,q}|$  – кількість однакових ключових слів для зображення  $Q$  та зображень  $I_i$ ;  $|K_q|$  – загальна кількість ключових слів ключових слів у зображенні  $Q$ .

Інтегроване значення подібності визначається наступним чином:

$$S_{final} = r * S_q + (1 - r) * V_d \quad S_{final} = r * S_q + (1 - r) * V_d, \quad (9)$$

де  $r$  – ваговий коефіцієнт.

При тестуванні розглянутої вище системи значення  $r$  приймалися від 0,1 до 0,8 для різних типів текстово-графічних фрагментів електронних документів: зображення з розширеними підписами:  $r = 0,8$ ; зображення зі короткими підписами:  $r = 0,1$ . При цьому було використано колекцію авторефератів дисертаційних робіт. Результати тестування запропонованого метода підтверджують доцільність його використання в системах аналізу електронних текстів. Перспективним розвитком системи є дослідження можливості урахування додаткових метаданих зображення (наприклад, інформації про розташування об'єктів, дату, час формування зображень).

#### ЛІТЕРАТУРА

1. Лізунов, П. П. Автоматичний аналіз подібностей схем та діаграм в електронних текстових документах / П. П. Лізунов, А. О. Білощицький, Л. Е. Чала, С.В. Білощицька, О. Ю. Кучанський, С. Г. Удовенко // Управління розвитком складних систем. – 2017. – № 28. – С. 147 – 156
2. Чала, Л.Э. Оценивание пертинентности лингвистических дескрипторов в системах информационного поиска документов / Л.Э. Чала, Ю.Ю. Харитонов// Восточно-европейский журнал передовых технологий. – 2015. – № 1/9(73). – С. 46–53

### ПРО ОДИН З ПІДХОДІВ ДО МЕДИЧНОГО ДІАГНОСТУВАННЯ НА ОСНОВІ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ

Узундай М.Р., Перова І.Г.

*Херсонський Фізико-Технічний ліцей, м. Херсон, вул. Залаегерсег, 39,  
Харківський національний університет радіоелектроніки, 61166, м. Харків, пр. Науки, 14, email:  
[yzundai1@gmail.com](mailto:yzundai1@gmail.com), [rikywenok@gmail.com](mailto:rikywenok@gmail.com)*

Нейронні мережі – прогресивна технологія, для якої в сучасному світі можна знайти безліч застосувань майже у всіх сферах життя. Одною з найнеобхідніших сфер є медицина, де використовується можливість нейронних мереж мінімізувати відсоток помилок і досягти точності більшої за багатьох спеціалістів. За статистикою лікар діагностує інфаркт міокарда неправильно в 29% випадків, так як використовуючи звичайну діаграму неможливо визначити стовідсотково хвору людину. З цією задачею можуть чудово впоратися нейронні мережі, завдяки яким вдалося зменшити відсоток помилок у діагнозі всього до 3%. Зараз ця технологія присутня в лікарських програмах більшості країн світу.

Об'єктом нашого дослідження є медична вибірка даних dermatology.data [1], в якій за результатами 34 аналізів треба визначити який з 6 недугів мають пацієнти. Таким чином, вхідною інформацією для обробки є таблиця Microsoft Excel в якій зберігаються результати 34 аналізів для 367 пацієнтів. Тобто кожний пацієнт характеризується вектором  $X(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ .

Перше на що ми звернули увагу – це відсутність частини значень аналізів у деяких пацієнтів, тобто наявність пропущених значень. Є багато алгоритмів заповнення пропущених значень, але одним із найточніших є метод нечіткої просторової екстраполяції [2]. Він полягає в розрахунку відстаней між пацієнтами по тих ознаках, що не містять пропусків і на основі цих відстаней провести заповнення прощених значень. Після позбавлення від пропущених даних для правильної роботи нейронної мережі потрібне опрацювання і кодування інформації, яке зазвичай проводиться в інтервалі від -1 до 1, або від 0 до 1. На наступному етапі ми повинні вирішити який вид нейронних мереж слід використовувати. В нашому дослідженні приймали участь нейронна мережа Адаліну, а саме її збільшена версія – Мадаліна і двохаровий перцептрон.

Адаліна – це однашарова мережа, що має кількість входів, що дорівнює кількості ознак кожного пацієнта і 1 вихід (Рис.1).

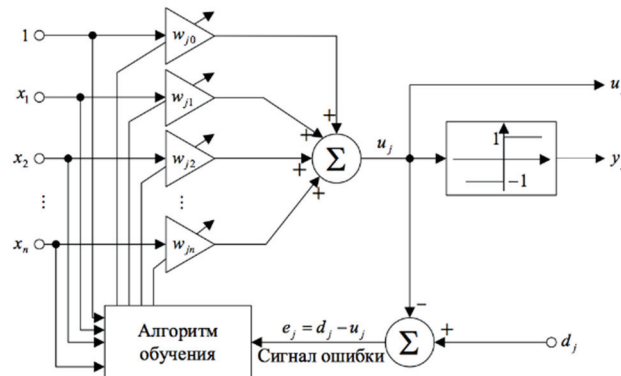


Рисунок 1 – Нейронна мережа Адаліна

Початкові ваги задаються випадково і перераховуються за алгоритмом Качмажа-Уїдрю-Хоффа

$$w_j(k+1) = w_j(k) + \frac{e_j(k)x(k)}{\|x(k)\|^2}.$$

Мадаліна представляє собою декілька одношарових Адалін, паралельно з'єднаних в багатошарову мережу. Для вибірок з багатьма ознаками кожна Адаліна відповідає за свій вхід, після чого сигнали всіх виходів Адалін надходять в опрацювувачий елемент, де ваги перераховуються за формулою:

$$W(k+1) = W(k) + \frac{d(k) - W(k)x(k)}{\|x(k)\|^2} x^T(k)$$

Кількість виходів Мадаліни дорівнює кількості діагнозів (в нашому випадку 6). Мадаліна демонструє точність при навчанні 96,2%, при тестуванні 83-93% (87,6%)

Персептрон – нейронна мережа з нелінійним розподілом, що має декілька входів, кілька проміжних шарів та декілька виходів для кожного пацієнта (Рис. 2). Завдяки перерахуванню ваг і вихідного і прихованих шарів можна досягти майже стовідсоткової точності, але через складну структуру персептрона деякі параметри, такі як кількість нейронів прихованого шару і значення констант, слід підбирати емпірично, через що для кожної вибірки даних треба вводити нові значення, або використовувати максимально наближені до середніх, але при цьому точність може бути гірше.

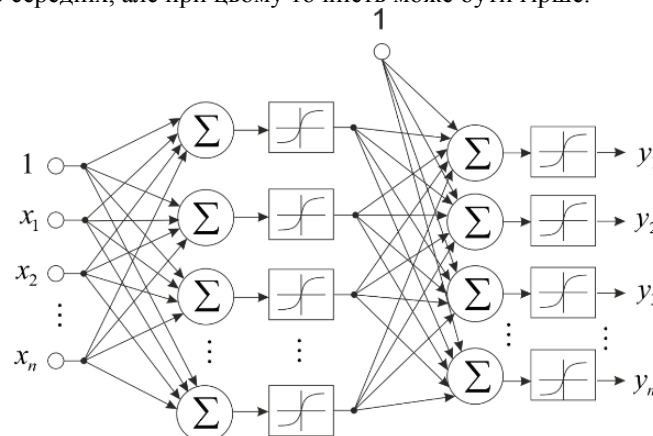


Рисунок 2 – Двошаровий персептрон

Похибки вихідного та прихованого шару персептрона обчислюються за формулами

$$\delta_{out} = (y - d) \cdot y \cdot (1 - y), \quad \delta_{hid} = (W_2 \cdot \delta_{out}) \cdot hid\_out \cdot (1 - hid\_out),$$

де  $y = \frac{1}{1 + e^{-W_2 \cdot hid\_out(k)}}$  – вектор виходів нейронної мережі,  $d$  – вектор відомих діагнозів,  $W_2$  – вектор

синаптичних ваг другого шару,  $hid\_out = \frac{1}{1 + e^{-W_1 \cdot X(k)}}$  – значення виходу першого прихованого шару.

Формули перерахування ваг:

$$W_1(k) = W_1(k-1) - \alpha(x(k)\delta_{hid}^T)$$

$$W_2(k) = W_2(k-1) - \alpha(x(k)\delta_{out}^T)$$

Значення констант у нашій програмі:  $\alpha = 0.3$ ,  $\gamma = 0.5$ , кількість нейронів прихованого шару = 40, кількість епох навчання = 100

При цих значеннях двошаровий перцептрон показує точність при навчанні 99,6%, при тестуванні 98,3%.

З цього можна зробити висновок, що перцептрон при правильних значеннях констант показує точніші результати ніж Мадаліна і може опрацювати більшу кількість даних, але якщо задати константи не так, як цього потребує вибірка даних, наприклад, якщо ми піднімемо кількість прихованих шарів до 100, не змінивши вхідну інформацію, точність одразу ж впаде в 3-4 рази, або взагалі опуститься до 0.

#### ЛИТЕРАТУРА:

1. "Dermatology dataset," May 2008. [Online]. Available: <http://archive.ics.uci.edu/ml/machine-learning-databases/dermatology/dermatology.data>.
2. Бодянский Е.В., Перова И.Г. Восстановление пропусков в таблицах данных на основе метода нечеткой пространственной экстраполяции с использованием манхэттенской метрики // Интеллектуальные системы принятия решения и проблемы вычислительного интеллекта ISDMCI'2015. – Железный порт, Украина. – 2015, с. 306-308

### РЕШЕНИЕ ЗАДАЧИ ФОЛДИНГА БЕЛКА С ИСПОЛЬЗОВАНИЕМ ГИБРИДНОЙ ИСКУССТВЕННОЙ ИММУННОЙ СИСТЕМЫ

Фефелова И.А., Фефелов А.А., Игнатенко І.

*Херсонский национальный технический университет*

Поскольку процессы свертывания протеинов изучены не в полной мере, исследователями предложен ряд упрощенных моделей, базирующихся на физических свойствах молекул и приводящих к задачам комбинаторной оптимизации. Одним из примеров таких моделей являются модели дискретных решеток [1]. При использовании решеток задача прогнозирования третичной структуры состоит в поиске свёртки с минимальной энергией, которая подсчитывается согласно определённому правилу. Доказано, что в такой постановке задача фолдинга протеина NP-полная. Поэтому, в общем случае точные методы не способны справиться с задачей за приемлемое время.

Цель настоящего исследования – разработка нового метода и алгоритма поиска, объединяющего преимущества популяционного подхода и технологии гибридизации. Для поиска оптимальной конформации протеина на выбранной модели предложен гибрид алгоритма дифференциальной эволюции (ДЭ) и искусственной иммунной системы (ИИС) в форме алгоритма клонального отбора. В данной работе проводились исследования с использованием модели двумерной квадратной решетки, которая является частным случаем известной гидрофобно-полярной (HP-модели) Дилла [2].

При разработке популяционных алгоритмов необходимо определить способ кодирования индивидуумов и вид целевой функции. В классическом случае для представления пути в решетке используется символическое кодирование индивидуумов с алфавитом  $A_{rel} = \{F, L, R\}$ , а путь задается последовательностью перемещений: вперед (F), влево (L) и вправо (R). Однако мутация в алгоритме ДЭ работает с вещественными числами. Таким образом, в данной работе потребовалось создание вещественной схемы кодирования. Каждый элемент вектора решений  $x_{ij}$  кодируется вещественным числом в промежутке от 0 до 1, т.е.  $x_{ij} \in [0.0, 1.0]$ . В процессе декодирования индивидуума этот промежуток делится на столько частей, сколько существует допустимых направлений из текущего узла решетки  $N_i$ . Допустимым считается направление, не приводящее к самопересечению, т.е. когда в результате перемещения в одном из направлений  $\{F, L, R\}$  новый узел  $N_j$  окажется не занятым элементом HP-последовательности, представляющей молекулу протеина. Каждая часть промежутка закрепляется за одним из допустимых направлений (рис. 1).

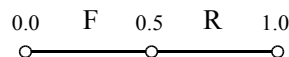


Рис. 1. Пример схемы отображения вещественных чисел в символы направлений (при наличии допустимых направлений  $\{F, R\}$ )