

operation – despite the obvious practical constraints. While unlikely to be a mainstream technology, the tantalizing opportunities offered by low temperature CMOS operation, combined with recent changes in the international security environment, make this a technology worth considering for certain situations.

**Acknowledgements.** The authors would like to acknowledge the influence of numerous colleagues over the years in helping to shape his thinking on this subject; here, they would particularly like to thank Peter Bendix, David Binkley, Colin McAndrew, Edward Nowak, and Yannis Tsividis for numerous enlightening discussions over a long and interesting period of time.

1. J. Lilienfeld, U.S. Patent Nos. 1,745,175 (1930), 1,877,140 (1932), and 1,900,018 (1933). 2. Quoted in G. Gilder, "The Soul of Silicon," *Forbes ASAP*, 1 June 1998. 3. C. Mead, "Fundamental Limitations in Microelectronics – I. MOS Technology," *Sol. St. Elec.* vol. 15, pp. 819 – 829 (1972). 4. R. Dennard et al., "Design of Ion Implanted MOSFETs with Very Small Physical Dimensions," *IEEE J. Sol. St. Circ.* vol. SC-9, pp. 256 – 268 (1974). 5. E. Nowak, "Ultimate CMOS ULSI Performance," 1993 IEDM Tech. Dig., pp. 115 – 118. 6. D. Foty and E. Nowak, "MOSFET Technology for Low Voltage/Low Power Applications," *IEEE Micro*, June 1994, pp. 68 – 77. 7. D. Foty, "The Design of Deep Submicron FETs for Low Temperature Operation," *Proceedings of the Symposium on Low Temperature Electronics and High Temperature Superconductors* (ed. by S. Raider), pp. 63 – 77 (1993). 8. D. Foty and E. Nowak, "Performance, Reliability, and Supply Voltage Reduction, with the Addition of Temperature as a Design Variable," *Proceedings of the 1993 European Solid State Device Research Conference*, pp. 943 - 948. 9. Y. Tsividis, "Moderate Inversion in MOS Devices," *Sol. St. Elec.* Vol. 25, pp. 1099 – 1104 (1982). 10. E. Vittoz, "Micropower Techniques," in *Design of MOS VLSI Circuits for Telecommunications* (ed. By J. Franca and Y. Tsividis), Prentice Hall, 1994. 11. D. Foty, "Re-Interpreting the MOS Transistor for the 21<sup>st</sup> Century: Generalized Methods and Their Extension to Nanotechnology," *Proceedings of the 21st Nordic VLSI Design Conference (NorChip)*, pp. 8 – 15 (2003). 12. D. Binkley et al., "A CAD Methodology for Optimizing Transistor Current and Sizing in Analog CMOS Design," *IEEE Transactions on Computer-Aided Design of Circuits and Systems* vol. CAD-22, pp. 225 – 237 (2003). 13. C. Enz, F. Krummenacher, and E. Vittoz, "An Analytical MOS Transistor Model Valid in All Regions of Operation and Dedicated to Low Voltage and Low Current Applications," *Analog Int. Circ. and Signal Proc.* vol. 8, pp 83 – 114 (1995). 14. G. Gildenblat et al., "SP: An Advanced Surface-Potential-Based Compact MOSFET Model," *IEEE J. Sol. St. Circ.* vol. SC-39, pp. 1394 – 1406 (2004). 15. Y. Yi et al., "Temperature-Scaling Theory for Low-Temperature-Operated MOSFET with Deep-Submicron Channel," *Jpn. J. Appl. Phys.* vol. 27, pp. L1958 – L1961 (1988). 16. D. Foty and E. Nowak, "Performance/Reliability Trade-Off and Optimized nMOSFET Design for 77K CMOS Logic," *Proceedings of the Symposium on Low Temperature Electronics and High Temperature Superconductors* (ed. by S. Raider), pp. 89 - 95 (1993).

УДК 638.235.231

Райд Фатхи Сахавне

Національний університет "Львівська політехніка",  
кафедра САПР

## МОДЕЛЮВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РЕАЛІЗАЦІЇ АЛГОРИТМІВ РОЗПІЗНАВАННЯ

© Райд Фатхи Сахавне, 2004

Розглянуті методи опису та зображення нейронних мереж для реалізації алгоритмів розпізнавання. Розроблено послідовні, паралельні та послідовно-паралельні схеми реалізації обчислень у нейронних елементах.

Methods of the description and representation of neural networks of realization of algorithms of recognition are considered. Algorithms of emulation of such networks are developed and realized. Examples of results of work of the program of emulation are resulted.

Завжди існував і існує великий інтерес до проблемно-орієнтованих та спеціалізованих систем обробки і розпізнавання зображень. Це передусім пов'язано з вимогами реального часу, обмежених

розмірів обчислювача, специфічних умов експлуатації та автономної роботи [2]. На сьогодні багато таких систем використовується в гідро - та радіолокації, системах контролю та ідентифікації [3].

Для реалізації таких систем використовують як існуючу елементну базу (наприклад, програмовані логічні матриці), так і спеціалізовану – однорідні обчислювальні середовища [ 2,4 ], нейронні мережі.

Для побудови нейронної мережі розпізнавання зображень з навчанням використовуємо алгоритми розпізнавання, що базуються на моделях обчислення оцінок. Це достатньо відомі і вивчені алгоритми [1]. Вони були базою для відпрацювання математичної теорії побудови та обґрунтування алгоритмів розв'язання погано формалізованих задач обробки і аналізу даних. Для цього класу алгоритмів сформульовано поняття моделі алгоритму розпізнавання, яка дозволяє перейти до системного вибору алгоритму для розв'язання конкретної задачі розпізнавання (алгоритм вибору алгоритму).

Основою цих моделей є принцип прецедентності (збігу). Він полягає в аналізі близькості між частинами описів раніше класифікованих об'єктів та об'єкта, який необхідно розпізнати. Наявність близькості є частковим випадком і оцінюється деяким заданим правилом (числовою оцінкою), із набору оцінок близькості. При цьому виробляється загальна оцінка об'єкта для класу, яка і є значенням функції входження об'єкта до класу.

Будемо розглядати нейронний елемент (НЕ) з функцією активації такого вигляду:

$$\sum_{i=1}^n w_i x_i \geq p \quad (1)$$

яка приймає значення 1 або 0 залежно від того перевищує значення зваженої суми величину порогу  $p$  чи ні. Це найпростіша порогова функція від векторів  $X = (x_1, \dots, x_n)$  – вхідні сигнали,  $W = (w_1, \dots, w_n)$  – вагові коефіцієнти та порогу  $p$ . У загальному випадку, функцією активації нейрона може бути довільна монотонна дійсна функція. Трактувати (1) можна і так: нейронний елемент приймає рішення відносно подальшого проходження сигналу на основі зважених вхідних сигналів. Вагові коефіцієнти вказують на важливість вхідних сигналів (інформативність вхідних сигналів). Якщо інформація про важливість вхідних сигналів відсутня або вважається, що всі елементи вхідного вектора інформативно рівноцінні, тоді приймають  $w_i = 1, i = \overline{1, n}$  і вираз (1) матиме вигляд

$$\sum_{i=1}^n x_i \geq p. \quad (2)$$

Будемо позначати  $NE(x, n, p, w)$  та  $NE(x, n, p)$  нейронні елементи з функціями активації відповідно (1) та (2).

Коли не буде двозначності і з контексту буде зрозуміло, про що йдеться, будемо також розуміти під виразами  $NE(x, n, p, w)$  та  $NE(x, n, p)$  числа 1 або 0 залежно від виконання нерівностей у виразах (1) та (2). Таку ж інтерпретацію дамо виразам

$$\left( \sum_{i=1}^n w_i x_i \geq p \right) \text{ та } \left( \sum_{i=1}^n x_i \geq p \right). \quad (3)$$

В наведених нейронних елементах “виходами нейрона” (значеннями функції активації) були 1 або 0. Розглядають також НЕ, в яких виходами є цілі числа від 0 до  $d (d > 1)$ . Якщо значення зваженої суми більше або дорівнює  $p$ , тоді інтервал  $[0, p]$  розбивається на  $d$  однакових інтервалів і значення функції активації приймається таким, що дорівнює номеру інтервалу, до якого належить зважена часткова сума. Такі нейронні елементи будемо позначати  $NE_d(x, n, p, w)$  та  $NE_d(x, n, p)$  і відповідно ці вирази прийматимуть значення з інтервалу  $[0, d]$ . Нейронний елемент  $NE_1(x, n, p)$  буде відповідати НЕ  $NE(x, n, p)$  (в наших позначеннях).

Мовою програмування Pascal будемо описувати наведені нейронні елементи такими типами даних:

$NE_d(x, n, p, w)$ Type_NE_d= record d,n,out : word; p,A : real X: array [1..n] of real; W: array [q..n] of real; end;	$NE_d(x, n, p)$ Type_WE= record d,n,out : word; p,A : real; X: array [1..n] of real; end;
--	--

Out – значення “виходу нейрона” (функції активації); A – значення акумулятора нейрона (накопичувача, часткова або повна зважена сума); n – кількість входів НЕ; p – поріг; X – масив вектор вхідних даних; W – масив вектора вагових коефіцієнтів; d – кількість градацій вихідного значення нейрона.

Традиційні схеми обчислень під час функціонування нейронних мереж та нейронних елементів не передбачають фіксованого порядку обробки. Тому обчислення задаються в непроцедурній формі. При програмній реалізації НЕ на традиційних послідовних обчислювачах, обчислення виконуються послідовно.

Нехай це буде Pascal-подібна конструкція:

```

A1: A:=0
for i:=1 to n do
    A: = A + X [i] * W [i];
Out: = F (A,p,d);
    
```

Функція  $F$  визначає значення виходу НЕ залежно від зваженої суми  $A$ , порогу  $p$  та кількості градацій виходу нейрона  $d$ . У циклі послідовно формується зважена сума  $A$ . Для виконання цього фрагменту необхідно  $n$  операцій додавання,  $n$  операцій множення та виконання порогової функції  $F_i$ . На рис. 1 показано граф такого нейронного елемента та його опис.

Під час реалізації НЕ на паралельних системах, обчислення (1) та (2) оптимально (максимально паралельно, мінімально за часом) виконуються у вигляді паралельно-конвеєрної схеми на неповному двійковому дереві

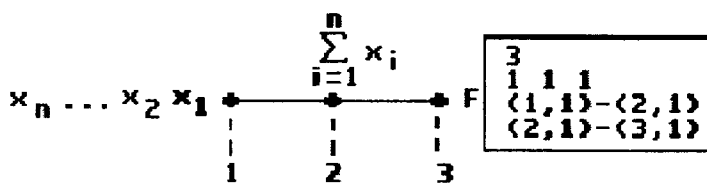


Рис. 1. Граф та опис нейронного елемента під час послідовної обробки вхідного вектора

В  $A_i$  доступ до елементів вектора є послідовним, тобто на обробку вектор надходить покомпонентно  $(x_1, \dots, x_n)$ , у випадку паралельної реалізації алгоритму вектор надходить паралельно, тобто є доступ до всіх елементів вхідного вектора одночасно.

Алгоритми НЕ (послідовні, паралельні та послідовно-паралельні) будемо зображати у вигляді графа  $G(V, E)$ , де  $V$  – вершини графа,  $E$  – дуги. У нашому випадку  $V$  – це вхідні дані або операції,  $G$  – інформаційні зв’язки між операціями або вхідними даними та операціями. Такі графи будемо описувати багаторівневою структурою вигляду

$$\langle m, (n_1, \dots, n_m), \{(i_1, j_1) - (i_2, j_2)\} \rangle, \quad (4)$$

де  $m$  – кількість рівнів;  $n_i$  – кількість вершин на  $i$ -му рівні ( $i = \overline{1, m}$ );  $(i_1, j_1) - (i_2, j_2)$  – дуга від вершини графа, що знаходиться на  $i_1$ -му рівні та має на цьому рівні порядковий номер  $j_1$  до вершини з порядковим номером  $j_2$   $i_2$ -го рівня.

На рис. 2 показано граф та опис нейронного елемента при паралельній обробці вхідного вектора.

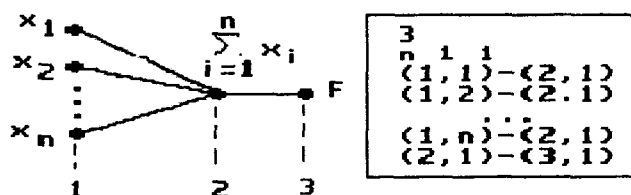


Рис. 2. Граф та опис нейронного елемента під час паралельної обробки вхідного вектора

На рис. 3 зображено граф і його опис паралельно-послідовних (конверсних) обчислень в НЕ.

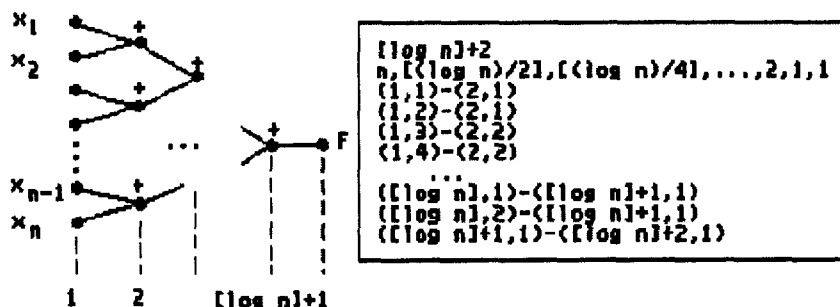


Рис. 3. Граф та опис нейронного елемента під час паралельно-послідовних (мультиконверсних) обчислень

Під нейронною мережею (НМ) будемо розуміти систему з'єднаних між собою певним чином нейронних елементів. Вони з'єднуються так, що входами їх є або вхідні дані, або виходи НЕ. Відповідно виходи НЕ з'єднуються з входами інших НЕ або вони є виходами нейронної мережі. Більше формально, як і у вже наведеному вище подані НЕ, будемо використовувати багаторівневу графову структуру  $G(V, E)$ , де  $V$  – вершини графів (у нас це нейронні елементи),  $E$  – дуги графа (це входи-виходи нейронних елементів, з'єднання НЕ).

На рис. 4 показано фрагмент графа нейронної мережі, який описується структурою

Фрагмент графа	Пояснення
$m$	Кількість рівнів (шарів) НМ
$n_i, i = \overline{1, m}$	Кількість НЕ на $i$ -му рівні НМ
$(i_1, j_1) - (i_2, j_2)$	Зв'язки між НЕ, вхідними даними та виходом НМ

Перший рівень – це вхідні дані (вектор вхідних даних, який надходить на обробку в НМ), останній рівень – це вихід НМ.

Нехай на рівні  $k \in m$ -входовий нейронний елемент  $NE(x, m, p)$  з порогом  $p$ . Вектор  $X$  формується на інших рівнях нейронної мережі, його елементи  $x_i$  – виходи нейронних елементів. Нехай  $x_i = NE(Y_i, n_i, p_i)$  значення виходу нейронного елемента, входом якого є вектор  $Y_i = (y_{i1}, \dots, y_{in_i})$  довжиною  $n_i$ , порогом  $p_i, i = \overline{1, m}$ . Тоді вектор  $X$  має вигляд  $X = (NE(Y_1, n_1, p_1), \dots, NE(Y_m, n_m, p_m))$ , а нейронний елемент  $NE(x, m, p)$  матиме вигляд

$$NE(x, m, p) = NE(\{NE(Y_i, n_i, p_i), i = \overline{1, m}\}, m, p) \quad (5)$$

Запис  $\{NE(Y_i, n_i, p_i), i = \overline{1, m}\}$  означає вектор довжиною  $m$ , значення якого дорівнюють  $NE(Y_i, n_i, p_i)$

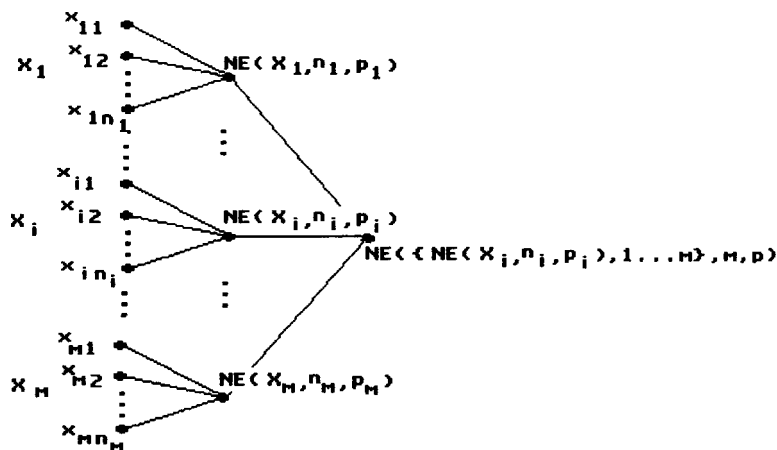


Рис. 4. Граф нейронної мережі

Нейронну мережу можна описати виразом (5) багаторівневого входження запису NE. На найвищому рівні опису (зовнішній запис NE) фігурує опис NE, виходом якого є вихід нейронної мережі. На найнижчому рівні входами NE є вхідні дані нейронної мережі. На рис. 3 зображено фрагмент графічного зображення такого рекурсивного опису нейронної мережі. Вхідний вектор NE ( $NE(Y, n, p)$ ) можна формулювати на різних рівнях і відповідно його вихід може слугувати входами для інших NE, які розміщені на різних рівнях. На основі запропонованих моделей розроблено емулятор, вікно роботи якого показано на рис. 5.

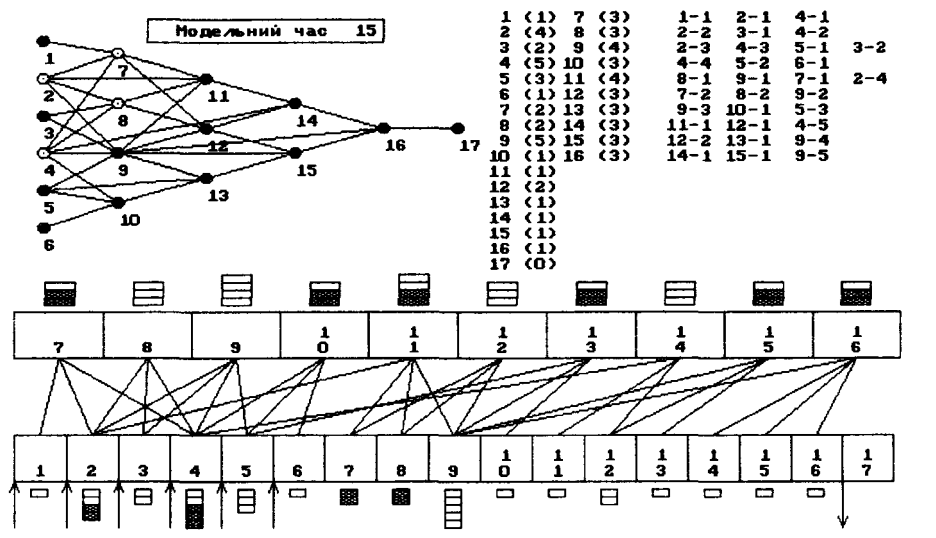


Рис. 5. Вікно роботи емулятора нейронної мережі

**Висновки.** Наведено розроблені моделі нейронних елементів з послідовною, паралельною та послідовно-паралельною організацією обчислень. На їх основі розроблено моделі нейронних мереж та принципи їх функціонування. Використовуючи ці моделі, реалізовано емулятор, який дозволяє проводити верифікацію та налагодження нейронних мереж, наведено вікно роботи програми емулятора.

1. Горелик А.Л. и Скрипкин В.А. Методы распознавания. – М., Высш. школа, 1977. – 222 с. 2. Параллельная обработка информации. Вычислительные системы, структуры и среды для решения задач большой размерности. Т. 3 / Под ред. В.В.Грицика, – К. Наукова думка, 1986, – 288 с. 3. Применение цифровой обработки сигналов. / Под ред. Э.Оппенгейма. – М.: Мир, 1980. – 550 с. 4. Kisil B.V. Computing media: A Promising Parallel Computational Tool. // Pattern Recognition And Image Analysis, V4, № 3 1994, – p. 233–237.