

1. Philip J. Gill. *Enabling the Age of Network Computing* // Oracle magazine. November/December, 1997. 2. Philip J. Gill. *Network Computing Architecture* // Oracle magazine. November/December, 1997. 3. Philip J. Gill. *The Network Computing Rule Book: Playing to Win* // Oracle magazine. November/December, 1997. 4. *Архитектура распределенных приложений. По материалам Gupta Corp* // Компьютеры+Программы. 1995. № 5. 5. Голощук Р.О. *Огляд інформаційних технологій розподілених обчислень для розробки мережевоцентричних застосувань* // Вісн. ДУ "Львівська політехніка". 1998. № 383. С.62-68. 6. Пелецишин А. М. *Розробка інформаційних систем у Web-середовищах* // Вісн. ДУ "Львівська політехніка". 1997. № 315. С.193-207

УДК 681.32:681.5

Дубровін В., Субботін С.

Запорізький державний технічний університет

ВИБІР ФУНКЦІЙ АКТИВАЦІЇ ФОРМАЛЬНОГО НЕЙРОНА ТА ДОСЛІДЖЕННЯ ЇХ ВПЛИВУ НА ЯКІСТЬ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ

© Дубровін В., Субботін С.

Розглянуто модель формального нейрона. Досліджено загальні та особисті вимоги до функцій активації і вироблено критерії їх порівняння. Наведено дані про результати експериментів із вибору найкращих функцій активації.

Вступ

Необхідність усунення не самих невідповідностей, а їх причин – і насамперед потенційних – завжди підкреслювалася засновниками Total Quality Management. Значно краще – знайти потенційні проблеми якомога раніше та знищити на самому початку, до того, як вони зможуть створити реальні труднощі. Тому при управлінні якістю акцент зміщується з перевірки продукції на діагностику виробничих процесів [1].

При будівництві адаптивних систем діагностики перспективним є використання нейронних мереж (НМ). Однією з важливих проблем у галузі штучних НМ є розробка ефективних методів та алгоритмів навчання НМ. На жаль, сьогодні ця проблема залишається невирішеною через те, що більшість алгоритмів навчання (особливо, для багатопараметричних НМ) повільно збігаються. Тому особливу значимість для практичної реалізації систем на основі НМ можуть мати дослідження оптимізації окремих параметрів НМ. Одним з таких параметрів, що впливає як на швидкість, так і на точність навчання, є тип функції активації формального нейрона НМ, що досліджується в цій роботі.

Модель формального нейрона

Формальний нейрон є основним елементом при будівництві НМ. Він виконує параметричне нелінійне перетворення вхідного вектора x в скалярну величину y . Це перетворення складається з двох етапів: спочатку обчислюється дискримінантна функція, що є відрізком багатовимірного ряду Тейлора, яка далі перетворюється у вихідну величину y . Коефіцієнти розкладання відрізка багатовимірного ряду Тейлора утворюють вектор вагових коефіцієнтів w , або пам'ять нейрона.

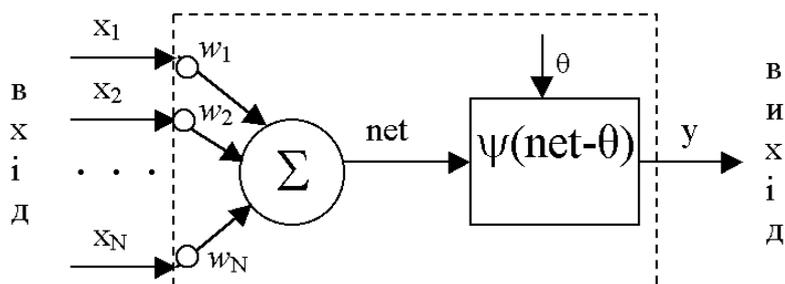


Схема формального нейрона

Дискримінантна функція нейрона першого порядку має вигляд [2]:

$$\text{net} = \sum_{i=1}^N w_i x_i ,$$

де N – кількість входів x_i ; w_i – ваги нейрона.

Функції активації

Нелінійне перетворення $y = \psi(\text{net} - \theta)$, де θ – порог нейрона, задається функцією активації ψ . Такою функцією, зокрема, можуть бути: для дискретного нейрона – порогова функція (1) або (2), а для нейрона із реальним виходом – сигмоїдна функція (3), що стали вже класичними:

$$\psi(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (1)$$

$$\psi(x) = \begin{cases} +1, & x \geq 0, \\ -1, & x < 0. \end{cases} \quad (2)$$

$$\psi(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

У цій “класиці” є свої “плюси” (є стандарт для порівняння моделей НМ) та “мінуси” (із розгляду в більшості робіт із НМ вилучаються інші функції активації, що призводить до створення менш оптимальних систем на їх основі).

Тому необхідно дослідити, які функції можуть застосовуватися як функції активації і які з них краще використовувати на практиці.

Функція активації повинна задовольняти такі вимоги [2,3]:

(Т1) Вона повинна бути обмеженою, тобто на інтервалі значень x $(-\infty, a]$ набувати значення y^0 , на інтервалі $[b, +\infty)$ – значення y^1 , на інтервалі (a, b) – значення y : $y^0 \leq y \leq y^1$, де y^0 и y^1 – деякі константні мінімальне та максимальне значення (звичайно вважають, що вони дорівнюють 0 та 1 або -1 та +1, відповідно), а a та b – деякі константи, причому $a \leq b$.

(Т2) Вона повинна бути монотонною на інтервалі (a, b) , тобто $\Delta\psi(x) = \psi(x + \Delta x) - \psi(x)$ не повинна змінювати знак при $\Delta x > 0$ та $x \in (a, b)$, $x + \Delta x \in (a, b)$.

Очевидно, що функцій, що задовольняють (Т1) и (Т2), може бути нескінченна множина. Для зручності реалізації і використання функцій активації можна сформулювати таку додаткову вимогу:

(Т3) Функція активації повинна бути легко і швидко обчислюваною, тобто містити якомога менше арифметичних дій, повинна складатися тільки з констант та швидкообчислюваних елементарних функцій, що є в наборі у розробника (при програмній реалізації – тих, що є в стандартних бібліотеках).

Обмеження (Т1)-(Т3) суттєво скорочують видову різноманітність можливих функцій активації, з яких можна виділити основні типи (1)-(9):

$$\vartheta(x) = \begin{cases} 1, x > 1 \\ \frac{x^n + 1}{2}, -1 \leq x \leq 1 \\ 0, x < -1 \end{cases}, \text{ де } n - \text{ парне число, } n > 1 \quad (4)$$

$$\vartheta(x) = \begin{cases} 1, x > \frac{\pi}{2} \\ \frac{\sin(x) + 1}{2}, -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0, x < -\frac{\pi}{2} \end{cases} \quad (5)$$

$$\psi(x) = \begin{cases} 1 - \frac{e^{-xx}}{2}, x \geq 0 \\ \frac{e^{-xx}}{2}, x < 0 \end{cases} \quad (6)$$

$$\vartheta(x) = \frac{\tanh(x) + 1}{2}, \text{ де } \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

$$\vartheta(x) = \begin{cases} 1, x > 1 \\ \frac{\arcsin(x) + 1}{2}, -1 \leq x \leq 1 \\ 0, x < -1 \end{cases} \quad (8)$$

$$\vartheta(x) = \begin{cases} 1, x > \frac{\pi}{4} \\ \frac{\operatorname{tg}(x) + 1}{2}, -\frac{\pi}{4} \leq x \leq \frac{\pi}{4} \\ 0, x < -\frac{\pi}{4} \end{cases} \quad (9)$$

Крім загальних обмежень (Т1)-(Т3), що накладаються на функції активації, окремі моделі НМ та алгоритми їх навчання можуть додатково накладати свої особисті обмеження, які розглянуто нижче.

Одношаровий перцептрон

Одношаровий перцептрон являє собою формальний нейрон (див. вище). Єдиний вихід перцептрона може бути дискретним або реальним (див. схему).

(Т4) Дискретний одношаровий перцептрон (тобто перцептрон із дискретним виходом) очевидно вимагає, щоб функція активації набувала тільки значення y^0 або y^1 і може використовувати тільки порогові функції активації (1) або (2).

(Т5) Одношаровий перцептрон з реальним виходом може використовувати будь-яку з функцій (3)-(9), використання ж функцій (1) або (2) перетворить його у дискретний перцептрон.

Багатошарова нейронна мережа (БНМ)

БНМ (багатошаровий перцептрон) являє собою мережу, що складається з декількох шарів формальних нейронів, причому нейрони одного шару між собою не зв'язані, а нейрони сусідніх шарів – зв'язані [2-5].

БНМ з дискретними виходами накладає на функцію активації ті ж вимоги, що й одношаровий дискретний перцептрон (Т4), а БНМ із реальними виходами – ті ж вимоги, що й одношаровий перцептрон з реальним виходом (Т5).

Алгоритми навчання БНМ накладають додаткові обмеження на функції активації:

(Т6) Градієнтний метод навчання, або метод зворотного розповсюдження помилки [2-6], що використовує корегуюче правило ваг $w_{k+1} = w_k - \alpha_k g_k$ (де w_k – поточний вектор ваг і порогів НМ; g_k – поточний градієнт; α_k – швидкість навчання) накладає на функцію активації ті самі вимоги, що й (Т5), а також вимагає, щоб функція активації була диференційованою та перша похідна функції активації $\psi(x)$ не була б тотожною нулю, бо інакше функція активації була би постійною.

(Т7) Метод Ньютона, або метод зворотного розповсюдження помилки другого порядку [2,7], корегуюче правило ваг якого має вигляд $w_{k+1} = w_k - A_k^{-1} g_k$, де A_k^{-1} – матриця других похідних (гесіан), накладає на функцію активації ті ж вимоги, що й (Т6), а крім того, вимагає, щоб у функції активації була похідна другого порядку, яка не була б тотожною нулю.

Алгоритм спряжених градієнтів Флетчера-Рівса [7], що використовує корегуюче правило ваг $p_0 = -g_0$, $w_{k+1} = w_k + \alpha_k p_k$, $p_{k+1} = -g_k + \beta_k p_{k-1}$, $\hat{a}_k = (g_k^T g_k) / (g_{k-1}^T g_{k-1})$, накладає на функцію активації ті ж вимоги, що й (Т7).

Алгоритм спряжених градієнтів Полака-Рибьера [7], корегуюче правило ваг якого має вигляд $p_k = -g_k + \beta_k p_{k-1}$, $\hat{a}_k = (\Delta g_{k-1}^T g_k) / (g_{k-1}^T g_{k-1})$, також накладає вимоги (Т7).

Алгоритм Левенберга-Марквардта [7], що використовує корегуюче правило ваг $H = J^T J$, $g = J^T e$, $w_{k+1} = w_k - [H_k + \mu I]^{-1} g_k$, де J – якобіан, e – вектор помилок, μ – скаляр, I – одинична матриця, накладає на функцію активації ті ж вимоги, що й (Т7).

Методика порівняння функцій активації

З'ясувавши, які вимоги до функцій активації висуваються конкретними моделями НМ та алгоритмами їх навчання, можна зробити висновок, що для дискретних моделей НМ є придатними тільки функції активації (1) або (2), а для мереж із реальними виходами – ті з функцій (3)-(9), які задовольняють вимоги окремих моделей і алгоритмів навчання цих мереж.

Далі будемо розглядати тільки НМ з реальними виходами: одношаровий перцептрон та БНМ. Через те, що для цих НМ може бути використана будь-яка з функцій активації (3)-(9), виникають питання:

1. Яку з функцій (3)-(9) краще використовувати на практиці?
2. Які критерії вибору тієї чи іншої функції активації (3)-(9)?

З опису розглянутих алгоритмів навчання НМ випливає, що збіжність алгоритмів навчання залежить, крім навчальних даних та початкових ваг, ще й від максимально допустимої загальної помилки навчання (критерій якості навчання) та кількості допустимих циклів навчання (критерій швидкості навчання) для всієї вибірки [7]. Ці два критерії доцільно використовувати для порівняння функцій (3)-(9). Зафіксувавши почергово кожний з них, можна прослідкувати, як для кожної моделі НМ змінюються значення іншого критерію при однакових наборах різних навчальних вибірок та однакових початкових ваг. Результати експериментів слід оцінювати за такими критеріями, як час роботи алгоритма навчання і витрачена кількість циклів навчання.

Експерименти

Для проведення експериментів було створено програмний комплекс, що реалізував наведені моделі та алгоритми навчання НМ та дозволяв збирати і документувати результати експериментів.

Експерименти проводилися для одношарових перцептронів, дво- та три- шарових БНМ (в різних серіях експериментів варіювалася також кількість входів НМ) .

Як вхідні дані для навчання НМ використовувалися вибірки із псевдовипадкових чисел і вибірки значень параметрів реальних виробів.

Порівняльна характеристика функцій активації

Функція	Середній час обчислення, 10^{-3} с	Середній час навчання при фіксованій максимальній кількості циклів навчання (1000 циклів), с		Середня кількість циклів навчання БНМ
		Одношаровий перцептрон, алгоритм Уїдроу-Хоффа	БНМ, алгоритм Левенберга-Марквардта	
3	2,09	0,77	4,61	4,7
4	0,61	0,67	3,5	3,3
5	1,0	0,78	3,57	3,3
6	1,94	0,78	3,71	3,7
7	3,66	0,93	3,53	3
8	1,24	0,31	3,91	4,7
9	0,71	0,84	3,79	4

Фрагменти результатів експериментів наведено в таблиці.

Аналіз результатів та висновки

Експерименти показали, що:

1. Функції (3)–(5), (8) для одношарового перцептрона в більшості випадків дозволяли значно скорше ніж інші (іноді в 1.5–2 рази) досягати збіжності при фіксованій мінімальній загальній помилці.

2. Функції (3), (6), (7) через дуже швидке зростання експоненти іноді призводили до збоїв внаслідок переповнення розрядної сітки.

3. Функції (3), (5), (6), (8) при фіксованих максимальних кількостях циклів навчання (10,100,1000) давали значно меншу загальну помилку, ніж інші функції.

4. Функції (4)–(6), (7) дозволяли за меншу кількість циклів навчання досягати бажаної збіжності.

На підставі отриманих експериментальних даних (які, як показали результати експериментів, сильно залежали від вхідних даних та початкових ваг мережі) можна зробити висновки про те, які функції, для якої моделі та методу навчання виявилися кращими, а які гіршими: для одношарового перцептрона кращими виявилися функції (3), (4), (5), (8); для БНМ – (4), (5), (7).

Отже, в результаті даної роботи досліджено вимоги до функцій активації, розроблено критерії їх порівняння та проведено експерименти, що дозволяють рекомендувати використовувати в моделях НМ крім класичної функції сигмоїда (3) також інші функції активації (4), (5), (8).

1. Дубровін В., Степаненко О., Луценко Д. *Діагностика процесів та керування якістю // Досвід розробки і застосування САПР в мікроелектроніці / Матеріали п'ятої міжнар. наук.-техн. конф. CADSM'99. Львів, 1999. С.22–24.* 2. Аведьян Э.Д. *Алгоритмы настройки многослойных нейронных сетей // Автоматика и телемеханика. 1995. № 4. С.106–118.* 3. Бовель Е.И., Паршин В.В. *Нейронные сети в системах автоматического распознавания речи // Зарубежная радиоэлектроника. 1998. № 4. С.50–57.* 4. Корнеев В.В. *Параллельные вычислительные системы. М., 1999.* 5. Boseniuk T., van der Meer M., Poschel T. *A Multiprocessor system for high speed simulation of neural networks // Journal of New Generation Computer Systems. 1990. № 3. P.65–71.* 6. *Нейрокомпьютеры и интеллектуальные роботы / Н.М. Амосов, Т.Н. Байдык, А.Д. Гольцев и др.; под ред. Н.М. Амосова. К., 1991.* 7. *Neural Network Toolbox User's Guide by Howard Demuth & Mark Beale. Natick: MathWorks Inc, 1997.*

УДК 621.396.6.001.63

Керницький А.Б., Лобур М.В.

НУ “Львівська політехніка”, кафедра САПР

ПРОБЛЕМИ ПОБУДОВИ СИСТЕМ АВТОМАТИЗАЦІЇ ТЕХНІЧНОГО ДОКУМЕНТООБІГУ ДЛЯ МАЛИХ ПІДПРИЄМСТВ

© Керницький А.Б., Лобур М.В., 2000

Розглянуто проблеми, що виникають у малих підприємств при розробці, побудові та впровадженні систем автоматизації технічного документообігу.

Система електронного документообігу є однією з найважливіших складових інформаційної системи організацій та підприємств різного рівня, яка забезпечує введення документів у систему, зберігання, пошук, маршрутизацію, збирання й аналіз інформації про поточний стан виконання виробничих і адміністративних процедур.

Увага більшості вітчизняних та закордонних дослідників у даний час прикута до реалізації та впровадження систем електронного документообігу для банків, великих промислових гігантів, корпорацій та офісів [1,2,3]. Це відображає реальний стан справ на ринку продуктів цього типу. Потреби та вимоги користувачів до системи, яка б автоматизувала технічний документообіг для виробничих підприємств невеликого розміру, ніхто здебільшого не аналізував. Відгукуючись на зростаючі потреби малих підприємств у стратегічному виборі такої системи, вважаємо за необхідне проаналізувати потреби, особливості технічного документообігу та сформулювати рекомендації та вимоги до реалізації та впровадження такої системи.

Раніше проблема вибору системи вирішувалася досить просто. Існувало кілька поставальників, сильні та слабкі сторони яких були більш або менш відомі. Після ретельного визначання передбачуваних потреб і порівняння їх із списками обладнання та програмного забезпечення та їх специфікаціями як простий наслідок виводилася остаточна конфігурація системи.