УДК 681.142.2

ОСОБЛИВОСТІ ВИКОРИСТАННЯ АЛГОРИТМУ ОБЕРНЕНОГО РОЗПОВСЮДЖЕННЯ ПОХИБКИ ДЛЯ НАВЧАННЯ ДИНАМІЧНИХ НЕЙРОННИХ МЕРЕЖ

© Наконечний Ю.М., 2005
Національний університет „Львівська політехніка”

Розглянуто особливості способів навчання нейронних мереж з використанням алгоритму оберненого розповсюдження похибки, приведено співвідношення для обчислення виходів нейронів і градієнтів функціоналу похибки по настроюваних параметрах, обґрунтовано доцільність використання алгоритму оберненого розповсюдження похибки для навчання багатошарових нейронних мереж

Алгоритм оберненого розповсюдження похибки для багатошарової нейронної мережі визначає стратегію обчислення вагових коефіцієнтів нейронів кожного шару мережі з використанням градієнтних методів оптимізації. Його основу, як і для одношарової нейронної мережі, визначає функціонал, який формується у вигляді квадратичної суми різниць між бажаними і фактичними значеннями векторів сигналів вихідного шару нейронної мережі. При використанні алгоритму оберненого розповсюдження похибки обчислюється різниця між бажаними і реальними значеннями вихідної величини і на її основі розраховуються вектори градієнта функціоналу похибки як функції настроюваних параметрів кожного із шарів мережі. Вказані вектори визначають величини і напрямки приростів вагових коефіцієнтів кожного шару нейронної мережі, при яких відбувається переміщення по поверхні функціоналу похибки в напрямку її зменшення. Послідовність таких кроків приводить до знаходження точки, в якій функціонал похибки досягає мінімального значення.

Таким чином, по завершенні процесу навчання мережі вагові коефіцієнти нейронів кожного шару приймають такі значення, при яких для кожного вхідного вектора на виході мережі формується вектор, компоненти якого співпадають з компонентами вектора цільової функції, який підводився до мережі на стадії її навчання в момент подачі цього вхідного вектора.

Навчання нейронної мережі з використанням алгоритму оберненого розповсюдження похибки може здійснюватись двома способами. Перший спосіб полягає в тому, що задану вхідну послідовність тривалістю Q циклів мережа обробляє послідовно, проводячи після кожної наступної вибірки вхідного вектора $a(q)$ корекцію вагових коефіцієнтів нейронів у всіх шарах. Такий спосіб навчання називається адаптацією нейронної мережі.

При реалізації другого способу навчання на кожній вибірці до мережі підводяться значення вхідного і цільового векторів і після проходження певного числа вибірок проводиться корекція вагових коефіцієнтів. Завершення процесу навчання мережі визначається числом епох, які можуть бути заданими, або досягненням певного значення функціоналу похибки, чи будь-якими іншими умовами, які обумовлюють закінчення процесу навчання.

Навчання багатошарової нейронної мережі при використанні алгоритму оберненого розповсюдження похибки відбувається поетапно.

На першому етапі до входу мережі підводиться перша вибірка вхідного вектора $a^{1,0}$, по якій при випадкових значеннях вагових коефіцієнтів нейронів всіх шарів, що встановлюються за допомогою генератора випадкових чисел, обчислюються значення сигналів на виходах нейронів.

Для обчислення сигналу на виході будь-якого нейрона мережі після кожної вибірки використовується співвідношення

$$a_i^{q,M-K} = f_i^{M-K}\left(\sum_{j=1}^{S^{M-(K+1)}} W_{ij}^{M-K} * a_j^{q,M-(K+1)} + W_{i0}^{M-K} * a_{i0}^{q,M-(K+1)}\right), \quad (1)$$

де K = 0, 1, 2, … , M-1; M – кількість шарів; S^M – число нейронів в M-му шарі; q – номер вибірки; $a_i^{q,M-K}$ – сигнал на виході i-го нейрона, який розташований в (M-K)-му шарі і обчислений на q-ій вибірці вхідного вектора $a^{q,0}$, або вихідного вектора $a_i^{q,M-(K+1)}$ попереднього шару; f_i^{M-K} – функція активації і-го нейрона (M-K)-го шару; W_{ij}^{M-K} –

ваговий коефіцієнт по j-му входу i-го нейрона розташованого в (M-K)-му шарі; $S^{M-(K+1)}$ – кількість нейронів M-(K+1)-го шару; $a_j^{q,M-(K+1)}$ – вхідний сигнал j-го нейрона M-(K+1)-го шару на q-ій вибірці; $a_0^{q,M-(K+1)}=1$ – зміщення для всіх нейронів M-(K+1)-го шару на q-ій вибірці.

Вираз для обчислення сигналу a_j^{M-K} на виході i-го нейрона (M-K)-го шару після проходження Q вибірок вхідного сигналу, тобто після завершення однієї епохи має вигляд:

$$a_i^{M-K} = \sum_{q=1}^{Q} f_i^{M-K} \Big(\sum_{j=1}^{S^{M-(K+1)}} W_{ij}^{M-K} * a_j^{q,M-(K+1)} + W_{i0}^{M-K} * a_{i0}^{q,M-(K+1)} \Big) , \qquad (2)$$

де Q – число вибірок.

У випадку навчання динамічної нейронної мережі (рис. 1), коли на один із входів першого шару мережі подається сигнал у вигляді часової послідовності вибірок *p(q)*, обчислення a_j^1 сигналу на виході i-го нейрона першого шару по завершенні однієї епохи проводиться за формулою:

$$a_i^1 = \sum_{q=1}^{Q} f_i \Big(\sum_{j=1}^{S^0} W_{ij}^1 * p[q-(j-1)] + W_{i0}^1 * a_{i0}^0 \Big) , \qquad (3)$$

де, $a_1^{q,0} = p(q)$; $a_2^{q,0} = p(q-1)$; ...; $a_i^{q,0} = p[q-(i-1)]$; ...; $a_{S_0}^{q,0} = p[q-(S_0-1)]$; $a_1^{q,0}$, $a_2^{q,0}$, $a_3^{q,0}$, ..., $a_{S_0}^{q,0}$ – компоненти вектора сигналу на вході першого шару для q-ої вибірки числової послідовності *p(q)*.

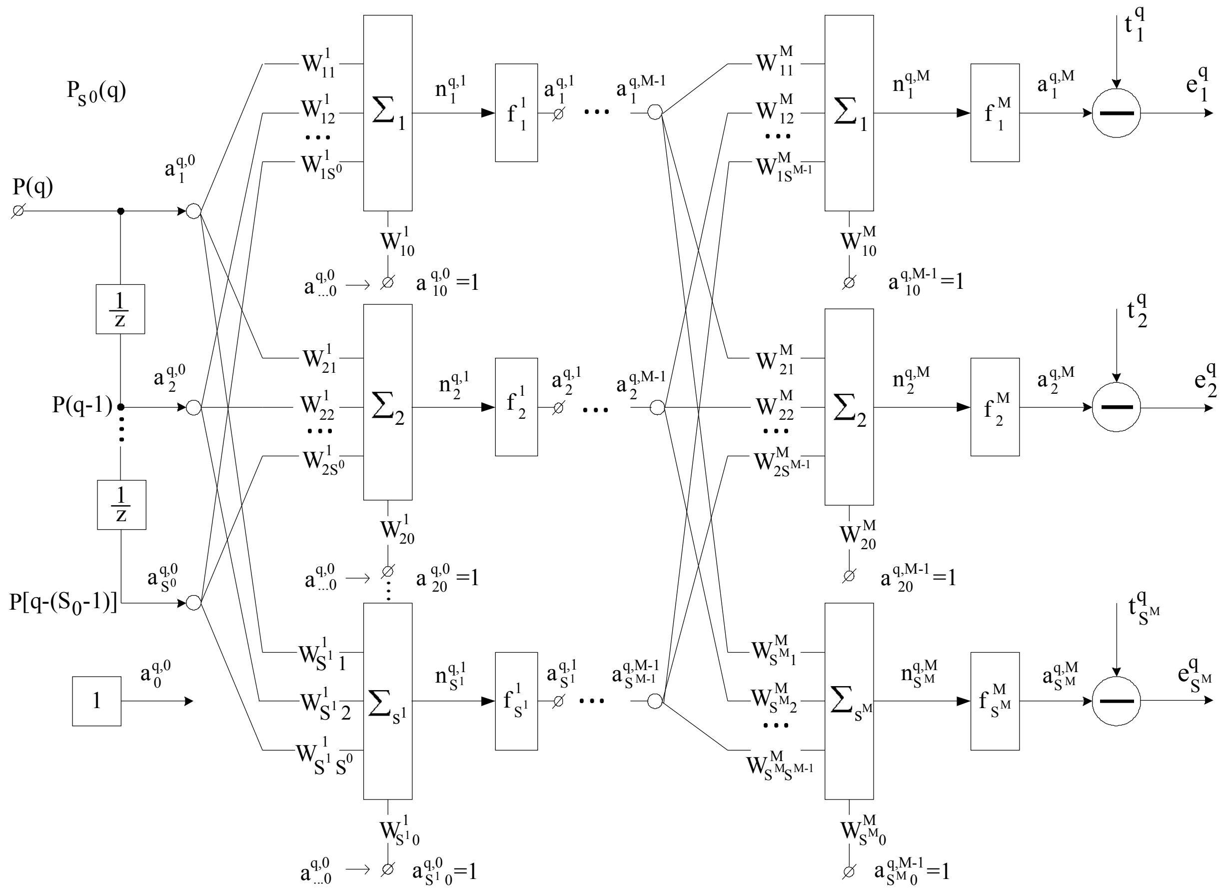

Рис. 1. Структурна схема динамічної нейронної мережі

Для обчислення сигналів на виходах нейронів у всіх наступних шарах при навчанні динамічної нейронної мережі використовується співвідношення (2).

Після розрахунку сигналів на виходах нейронів всіх шарів обчислюється значення функціоналу похибки

$$J = \frac{1}{2} \sum_{q=1}^{Q} \sum_{i=1}^{S^M} (t_i^q - a_i^{qM})^2 . \qquad (4)$$

У випадку, коли в M-му шарі використовується один нейрон, а це, як правило, характерно для динамічних нейронних мереж, вираз (5) набуває вигляду:

$$J=\frac{1}{2}\sum_{q=1}^{Q}(t_1^q-a_1^{qM})^2. \qquad (5)$$

Будемо розглядати загальний випадок, коли в останньому шарі мережі використовується S^M нейронів.

На другому етапі навчання нейронної мережі обчислюються градієнти функціоналу похибки по вагових коефіцієнтах всіх шарів. У зв'язку з тим, що вихідний шар багатошарової нейронної мережі можна розглядати як одношарову нейронну мережу, входами якої є виходи попереднього шару багатошарової мережі, то співвідношення, які одержані для обчислення градієнтів функціоналу похибки по вагових коефіцієнтах нейронів одношарової мережі, можуть бути використані для розрахунку градієнтів функціоналу похибки по вагових коефіцієнтах нейронів вихідного шару багатошарової нейронної мережі.

В цьому випадку вираз для обчислення градієнта функціоналу похибки для i-го нейрона M-го шару по виходу j-го нейрона (M-1)-го шару на q-ій вибірці має вигляд:

$$\frac{\partial J^q}{\partial W_{ij}^M}=-(t_i^{qM}-a_i^{qM})*f_M^{'}(n_i^{qM})*a_j^{q(M-1)}, \qquad (6)$$

де t_i^{qM} – компонента вектора цільової функції, що визначає бажаний вихід i-го нейрона M-го шару на q-ій вибірці; a_i^{qM} – вихід i-го нейрона M-го шару на q-ій вибірці; $f_M^{'}(n_i^{qM})$ – похідна функції активації комбінованого вводу i-го нейрона M-го шару на q-ій вибірці; i = 1, 2, 3, ... , S^M ; j = 1, 2, 3, ... , S^{M-1}.

Якщо ввести позначення $\Delta_i^M=(t_i^{qM}-a_i^{qM})*f_M^{'}(n_i^{qM})$, то вираз (6) можна переписати у вигляді:

$$\frac{\partial J^q}{\partial W_{ij}^M}=-\Delta_i^{qM}*a_j^{q(M-1)}, \qquad (7)$$

а обчислення градієнта функціоналу похибки i-го нейрона по завершенні однієї епохи проводиться за формулою

$$\frac{\partial J}{\partial W_{ij}^M}=-\sum_{j=1}^{Q}\Delta_i^{qM}*a_j^{q(M-1)}. \qquad (8)$$

В матричній формі вираз (8) можна подати у такому вигляді:

$$\frac{\partial J}{\partial W^M}=-\sum_{j=1}^{Q}diag[f_M^{'}(n^{qM})]*e^{qM}*[a^{q(M-1)}]^T, \qquad (9)$$

де $\dfrac{\partial J}{\partial W^M}$ – матриця градієнтів функціоналу похибки по вагових коефіцієнтах M-го шару, розмірність матриці $[S^M\times(S^{M-1}+1)]$; $diag[f_M^{'}(n^{qM})]=[f_M^{'}(n_1^{qM}),\ f_M^{'}(n_2^{qM}),\dots,f_M^{'}(n_{S^M}^{qM})]$ – діагональна матриця похідних функцій активації нейронів M-го шару на q-ій вибірці розмірністю $[S^M\times S^M]$;

$e^{qM}=[(t_1^{qM}-a_1^{qM}),(t_2^{qM}-a_2^{qM}),\dots,(t_{S^M}^{qM}-a_{S^M}^{qM})]$ – вектор різницевих сигналів M-го шару на q-ій вибірці розмірністю $[S^M\times 1]$; $[a^{q(M-1)}]^T=[a_0,a_1^{q(M-1)},a_2^{q(M-1)},\dots,a_{S^{M-1}}^{q(M-1)}]$ – транспонований вектор вхідних сигналів (M-1)-го шару на q-ій вибірці, включно із зміщенням $a_0=1$ розмірністю $[1\times(S^{M-1}+1]$.

В багатошаровій нейронній мережі при розповсюдженні похибок у зворотньому напрямку похибка кожного нейрона наступного шару вносить свій вклад в похибку нейрона попереднього шару і тому при обчисленні градієнтів функціоналу похибки по вагових коефіцієнтах нейронів попереднього шару необхідно враховувати ступінь впливу цих похибок на значення обчислюваних градієнтів. Ступінь такого впливу для кожного нейрона попереднього шару визначається похибкою на виході нейрона наступного шару і значенням вагового коефіцієнту нейрона наступного шару, через який ця похибка подається до нейрона попереднього шару.

Вираз для обчислення градієнта функціоналу похибки k-го нейрона (M-1)-го шару по виходу j-го нейрона (M-2)-го шару на q-ій вибірці з урахуванням ланцюгового правила визначення часткових похідних згідно рис.1 буде мати такий вигляд:

$$\frac{\partial J}{\partial W_{kj}^{q(M-1)}}=\sum_{i=1}^{S^M}\frac{\partial J}{\partial a_i^{qM}}*\frac{\partial a_i^{qM}}{\partial n_i^{qM}}*\frac{\partial n_i^{qM}}{\partial a_k^{q(M-1)}}*$$
$$*\frac{\partial a_k^{q(M-1)}}{\partial n_k^{q(M-1)}}*\frac{\partial n_k^{q(M-1)}}{\partial W_{kj}^{q(M-1)}}. \qquad (10)$$

Підставляючи значення часткових похідних по кожній із змінних в праву частину співвідношення (10) і враховуючи те, що підрахунок градієнта функціоналу похибки відбувається після проходження Q вибірок вхідного і вихідного сигналів, запишемо вираз для обчислення градієнта функціоналу похибки k-го нейрона (M-1)-го шару по j-му входу після завершення однієї епохи:

$$\frac{\partial J}{\partial W_{kj}^{M-1}}=-\sum_{q=1}^{Q}\sum_{i=1}^{S^M}(t_i^{qM}-a_i^{qM})*f_M^{'}(n_i^{qM})*$$
$$*W_{ik}^M*f_{M-1}^{'}(n_k^{q(M-1)})*a_j^{q(M-2)}. \qquad (11)$$

Ввівши позначення

$$\Delta_k^{q(M-1)}=\sum_{i=1}^{S^M}(t_i^{qM}-a_i^{qM})*f_M^{'}(n_i^{qM})*$$
$$*W_{ik}^M*f_{M-1}^{'}(n_k^{q(M-1)}), \qquad (12)$$

приведемо вираз (11) до вигляду:

$$\frac{\partial J}{\partial W_{kj}^{M-1}}=-\sum_{q=1}^{Q}\Delta_k^{q(M-1)}*a_j^{q(M-2)}. \qquad (13)$$

В матричній формі вираз для обчислення

градієнта функціоналу похибки по вагових коефіцієнтах нейронів (M-1)-го шару буде таким:

$$[\frac{\partial J}{\partial W_{kj}^{M-1}}] = -\sum_{q=1}^{Q}[diag[f_{M-1}^{'}(n^{q(M-1)})][W^M]^T * \\ *[diag[f_M^{'}(n^{qM})] * e^{qM}[a^{q(M-2)}]^T]], \quad (14)$$

де $[W^M]^T = \begin{vmatrix} W_{11} & W_{21} & \dots & W_{S^M 1} \\ W_{12} & W_{22} & \dots & W_{S^M 2} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ W_{1S^{M-1}} & W_{2S^{M-1}} & \dots & W_{S^M S^{M-1}} \end{vmatrix}$ – транспонована матриця вагових коефіцієнтів нейронів M-го шару, розмірність матриці $[S^{M-1} \times S^M]$.

Обчислення градієнтів функціоналу похибки l-го нейрона (M-2)-го шару проводиться з використання більш складних виразів у порівнянні з виразами для обчислення градієнтів функціоналу похибки нейронів (M-1)-го шару у зв'язку з тим, що при обчисленні градієнтів функціоналу похибок нейронів цього шару необхідно враховувати вплив похибок нейронів M-го і (M-1)-го шарів.

Вираз для обчислення градієнта функціоналу похибки l-го нейрона (M-2)-го шару по j-му вході буде таким:

$$\frac{\partial J}{\partial W_{lj}^{M-2}} = -\sum_{q=1}^{Q}\sum_{k=1}^{S^{M-1}}\sum_{i=1}^{S^M}(t_i^{qM} - a_i^{qM}) * f_M^{'}(n_i^{qM}) * \\ * W_{ik}^M * f_{M-1}^{'}(n_i^{q(M-1)}) * W_{kl}^{M-1} * f_{M-2}^{'}(n_l^{q(M-2)}) * a_j^{q(M-3)}. \quad (16)$$

Якщо ввести позначення

$$\Delta_l^{q(M-2)} = \sum_{k=1}^{S^{M-1}}\sum_{i=1}^{S^M}(t_i^{qM} - a_i^{qM}) * f_M^{'}(n_i^{qM}) * W_{ik}^M * \\ * f_{M-1}^{'}(n_i^{q(M-1)}) * W_{kl}^{M-1} * f_{M-2}^{'}(n_l^{q(M-2)}), \quad (17)$$

то приходимо до такого виразу, за допомогою якого можна обчислювати значення градієнта функціоналу похибки будь-якого нейрона (M-2)-го шару по одному із компонентів вектора вихідного сигналу (M-3)-го шару після завершення однієї епохи:

$$\frac{\partial J}{\partial W_{lj}^{M-2}} = -\sum_{q=1}^{Q}\Delta_l^{q(M-2)} * a_j^{q(M-3)}. \quad (18)$$

Матриця градієнтів функціоналу похибки по настроюваних параметрах нейронів (M-2)-го шару по завершенні однієї епохи буде такою:

$$[\frac{\partial J}{\partial W^{M-2}}] = -\sum_{q=1}^{Q}[diag(f_{M-2}^{'}(n^{q(M-2)}))][W^{M-1}]^T * \\ *[diag(f_{M-1}^{'}(n^{q(M-1)}))][W^M]^T * \\ *[diag(f_M^{'}(n^{qM}))]e^{qM}[a^{q(M-3)}]^T. \quad (19)$$

Якщо (M-2)-ий шар вважати вхідним шаром динамічної нейронної мережі, то матриця градієнтів функціоналу похибки l-го нейрона цього шару по завершенні однієї епохи буде такою:

$$[\frac{\partial J}{\partial W_{lj}^{M-2}}] = -\sum_{q=1}^{Q}[diag(f_{M-2}^{'}(n^{q(M-2)}))][W^{M-1}]^T * \\ *[diag(f_{M-1}^{'}(n^{q(M-1)}))][W^M]^T * \\ *[diag(f_M^{'}(n^{qM}))]e^{qM}[p^q]^T, \quad (20)$$

де

$$[p^q]^T = [a_o^{M-3}, p(q), p(q-1), p(q-2), \\ \dots, p[q-(S^{M-3}-1)]]^T \quad (21)$$

транспонований вектор вхідного сигналу з урахуванням зміщення.

Як видно з співвідношень (8), (13), (18), вирази для обчислення градієнтів функціоналу похибки по настроюваних параметрах кожного із шарів подібні і записуються у вигляді сум добутків двох співмножників, перший з яких визначається помилкою Δ, що перенесена з виходу мережі на вхід шару, для котрого обчислюються складові градієнта похибки, а друга – компонентою вхідного вектора, або вектора виходів попереднього шару, яка відповідає входу нейрона, для якого обчислюється градієнт функціоналу. Значення градієнтів функціоналу похибки по настроюваних параметрах всіх шарів, які одержані по завершенні кожної наступної епохи, використовуються для корекції вагових коефіцієнтів нейронів кожного шару у напрямку мінімізації середньої квадратичної похибки.

1. Каллан Р. Основные концепции нейронных сетей. – М.: Вильямс, 2001. – 288 с. 2. Осовский С. Нейронные сети для обработки информации. – М.: Вильямс, 2002. – 246 с. 3. Галушкин А. Теория нейронных сетей. – М.: Вильямс, 2000. – 311 с. 4. Терехов В., Ефимов Д., Тюкин Ю. Нейросетевые системы управления. – М.: Вильямс, 2002. – 197 с.