



## АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ ВСТРЕЧНОГО РАСПРОСТРАНЕНИЯ С КОНТРОЛИРУЕМЫМ ОБУЧЕНИЕМ В ЗАДАЧАХ КЛАССИФИКАЦИИ ТЕКСТОВЫХ ДОКУМЕНТОВ

*Золотухин О.В.*

*Харьковский национальный университет радиоэлектроники*

В настоящее время интеллектуальная обработка текстовых документов привлекает все большее внимание, а среди таких задач в качестве одной из наиболее актуальных можно отметить задачу классификации, которую удобно рассматривать с позиций теории распознавания образов. Для решения такого типа задач в качестве весьма эффективного аппарата в настоящее время широко используются искусственные нейронные сети (ИНС), благодаря своим универсальным аппроксимирующим свойствам и способности к обучению. С помощью ИНС были получены вполне приемлемые результаты, все они не лишены некоторых недостатков, ограничивающих их применимость. Архитектуры ИНС, основанные на обучаемом векторном квантовании [1], в том числе с нечеткой модификацией [2], являются по сути процедурами стохастической аппроксимации, характеризуются низкой скоростью сходимости, что требует больших объемов обучающих выборок. Использована нечеткая вероятностная нейронная сеть [3]. Вероятностные и радиально-базисные ИНС подвержены «проклятию размерности» в случае высокой размерности входных векторов-образов, что существенно затрудняет их использование в режиме реального времени. Таким образом, является целесообразным разработка быстродействующей ИНС, предназначенной для решения задач классификации текстовых документов в режиме реального времени при последовательном поступлении данных на вход системы.

Специфика задачи классификации текстовых документов требует существенной модификации как архитектуры ИНС встречного распространения, т.к. в режиме обучения на вход сети подаются классифицированные образы, целесообразно в первом скрытом слое использовать не традиционную самоорганизующуюся карту, обучаемую без учителя, а нейронную сеть векторного квантования, обучаемую с учителем, что позволяет повысить быстродействие, и в выходном слое вместо звезд Гроссберга целесообразно использовать элементарные персептроны Розенблатта с нелинейной функцией активации (релейной), принимающей только два значения: 1, если предъявляемый образ относится к данному конкретному классу, и 0 – в противном случае.

Исходной информацией для обучения является последовательность векторов-образов  $x(1), x(2), \dots, x(k), x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T \in R$  с известной классификацией, которая в реальном времени, что характерно для задач Web Mining, подается на нулевой (рецепторный) слой сети. Нейроны первого скрытого слоя  $N_j^{LV} (j=1, 2, \dots, m)$ ,  $m$  – количество возможных классов, задаваемые априорно, являясь по сути адаптивными линейными ассоциаторами, предназначены для нахождения центроидов и границ классов, при этом  $(n \times 1)$ -



векторы, описывающие эти центроиды  $c_j(k) = (c_{j1}(k), c_{j2}(k), \dots, c_{jn}(k))^T$ , являются синаптическими настраиваемыми весами каждого из нейронов  $N_j^{LV}$ . При этом все входные сигналы перед подачей на нулевой слой нормируются так, что  $\|x(k)\| = 1$ .

Выходной слой сети образован  $m$ -элементарными перцептронами Розенблатта [2]  $N_j^R$  с сигмоидальной функцией активации, при этом

$$y_j(k) = \psi(\gamma u_j(k)) = \psi\left(\sum_{i=1}^n \gamma w_{ji} c_{ji}(k)\right) = \psi(\gamma w_j^T c_j(k)) = \frac{1}{1 + e^{-\gamma u_j(k)}} = \frac{1}{1 + e^{-\gamma w_j^T c_j(k)}}$$

где  $\gamma$  – параметр крутизны активационной функции,  $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})^T$  – вектор синаптических весов  $N_j^R$ .

На рисунке показан вид сигмоидальной активационной функции  $\psi(\gamma u_j)$  в зависимости от параметра крутизны  $\gamma$ . При этом, чем больше значение  $\gamma$ , тем ближе  $\psi(\gamma u_j)$  к релейной функции

$$\psi(\gamma u_j) = \begin{cases} 1, & \text{при } u_j \geq 0, \\ 0, & \text{при } u_j < 0, \end{cases} \quad (1)$$

обычно используемой в задачах распознавания образов.

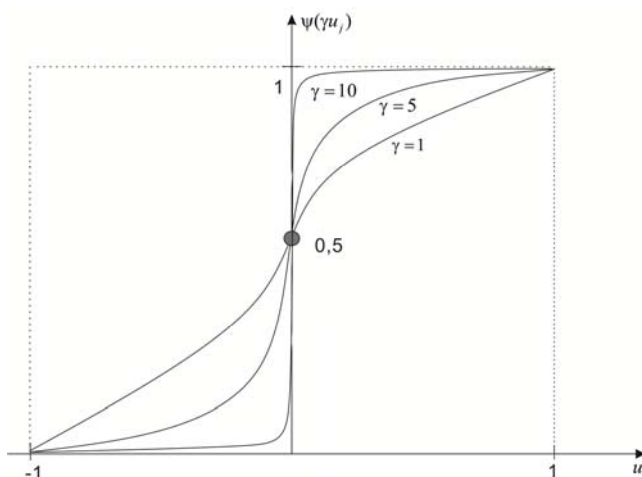
Понятно, что при  $\gamma \rightarrow \infty$ ,  $\psi(\gamma u_j)$  совпадает с (1), не претерпевая при этом разрыва производной в точке  $u_j = 0$ .

Здесь интересно также заметить, что векторы  $c_j(k)$ , являясь синаптическими весами  $N_j^{LV}$ , подаются в качестве входных сигналов на выходные нейроны  $N_j^R$ .

Выходные сигналы сети  $y_j(k)$  принимают два значения

$$y_j(k) \approx \begin{cases} 1, & \text{если } x(k) \text{ относится к } j\text{-ому классу,} \\ 0, & \text{в противном случае,} \end{cases}$$

при этом точные значения 1 или 0 никогда не достигаются.



Зависимость сигмоидальной функции от параметра крутизны

1. Umer M.F., Khiyal M.S.H. Classification of textual documents using learning vector quantization // Information Technology Journal. - 2007. - 6(1). - P.154-159.  
 2. Бодянский Е.В., Рябова Н.В., Золотухин О.В. Обработка текстовых документов с помощью адаптивного нечеткого обучаемого векторного квантования // Вісник Національного технічного університету «ХПІ». – 2011. – №53. – С.109-115.  
 3. Бодянский Е.В., Рябова Н.В., Золотухин О.В. Классификация текстовых документов с помощью нечеткой вероятностной нейронной сети // Восточно-европейский журнал передовых технологий – 2011. – №6/2 (54). – С.16-18.