АДАПТАЦИЯ АКТИВАЦИОННЫХ ФУНКЦИЙ В ГЛУБОКИХ НЕЙРОННЫХ СЕТЯХ

Слепанская В.Д.
Научный руководитель – д.т.н., проф. Бодянский Е.В.
Харьковский национальный университет радиоэлектроники
(61166, Харьков, просп. Науки,14, каф. Искусственного интеллекта,
тел. (057) 702-13-37), e-mail: valeriia.slepanska@nure.ua

In this paper the modification of PReLU used in Deep Neural Networks (DNN) is proposed. This modification called Adaptive Parametric Rectifier Linear Unit (AdPReLU) tunes both synaptic weights of neuron and parameters of activation function that permits to accelerate the learning process of DNN in whole and to reduce the amount of network`s hidden layers. The proposed AdPReLU is generalization of activation functions usually used in known DNN.

В настоящее время глубокие нейронные сети (DNN) [1,2] получили широкое распространение для решения задач, возникающих в Data Mining и, прежде всего, распознавания образов, классификации, прогнозирования, эмуляции, интеллектуального управления и т.п., благодаря высоким аппроксимирующим свойствам, достигаемых в процессе обучения.

Нейронные сети в широком смысле являются нелинейными адаптивными динамическими системами, причем нелинейность обеспечивается с помощью активационных функций нейронов, при этом наиболее часто используемый элементарный персептрон Ф. Розенблатта может быть описан с помощью соотношения:

$$\hat{y}_j(k) = \psi_j\left(u_j(k)\right) = \psi_j\left(w_j^T x(k)\right), \quad (1)$$

где $\hat{y}_j(k)$ – выходной сигнал j-ого нейрона на k-ом шаге обработки информации, ψ_j — нелинейная функция активации, $u_j(k)$ сигнал внутренней активации, $x(k) = \left(1,\ x_1(k), \ldots\ x_i(k) \ldots x_n(k)\right)^T$ — вектор входных сигналов, $w_j(k) = (\Theta_{j0}, w_{j1}, \ldots\ w_{ji} \ldots w_{jn})^T$ — вектор синаптических весов, настраиваемых в процессе обучения (δ-правило) может быть описан с помощью рекуррентной процедуры:

$$w_j(k) = w_j(k-1) + \eta_j(k) e_j(k) \psi'\left(u_j(k) x(k)\right) = w_j(k-1) + (k)\sigma_j(k) x(k), \quad (2)$$

где $\eta_j(k)$ – это параметр шага обучения, определяющий скорость сходимости процедуры.

Традиционные сигмоидальные активационные функции мелких нейронных сетей (SNN) порождают ряд существенных вычислительных проблем, возникающих в процессе обучения, в связи с чем в DNN вместо сигмоидальных используются функции типа ReLU или их модификации, например PReLU, имеющая вид:

$$\psi'(u_j) = \begin{cases} \Upsilon_j^R \ if \ u_j \geq 0, \\ \Upsilon_j^L \ \ u_j \ otherwise, \end{cases} \tag{3}$$

где Υ_j^R, Υ_j^L – параметры "крутизны" (gain parameters) кусочно-линейной функции. Алгоритм обучения при этом приобретает простой вид:

$$w_j(k) = \begin{cases} w_j(k-1) + \eta_j(k)\Upsilon_j^R e_j(k)x(k) = w_j(k-1) + \eta_j(k)\Upsilon_j^R \left(y(k) - w_j^T x(k-1)x(k)\right)x(k) \\ if \ u_j(k) \geq 0, \\ w_j(k-1) + \eta_j(k)\Upsilon_j^L e_j(k)x(k) = w_j(k-1) + \eta_j(k)\Upsilon_j^L \left(y(k) - w_j^T x(k-1)x(k)\right)x(k) \\ otherwise, \end{cases} \tag{4}$$

где $y(k)$ – внешний обучающий сигнал.

Вместе с тем, следует помнить, что функции типа ReLU, PReLU или подобные им не отвечают условиям аппроксимационных теорем, лежащих в основе теории нейронных сетей, в связи с чем DNN зачастую содержат слишком большое количество скрытых слоев, чтобы обеспечить требуемое качество кусочно-линейной аппроксимации.

Улучшить это качество можно, введя дополнительный контур для настройки параметров крутизны, при этом сам процесс обучения этих параметров может быть описан с помощью рекуррентных соотношений:

$$\begin{cases} \Upsilon_j^R(k) = \Upsilon_j^R(k-1) + \eta_j(k)\left(y_j(k) - \Upsilon_j^R(k-1)u_j(k)\right)u_j(k) \ if \ u_j \geq 0, \\ \Upsilon_j^L(k) = \Upsilon_j^L(k-1) + \eta_j(k)\left(y_j(k) - \Upsilon_j^L(k-1)u_j(k)\right)u_j(k) \ otherwise. \end{cases} \tag{5}$$

Использование дополнительного контура на основе (5) позволяет сократить время обучения и количество слоев DNN.

Список использованной литературы

1. Goodfellow I., Bengio Y., Courville A. Deep Learning – MIT Press, 2016 – 787 p.

2. Graupe D. Deep Learning Neural Networks: Design and Case Studies – New Jersey: World Scientific, 2016 – 260 p.