



Д.Р. Хугаева, И.В. Лёзина

## ПРЕДВАРИТЕЛЬНАЯ ИНИЦИАЛИЗАЦИЯ ВЕСОВЫХ КОЭФФИЦИЕНТОВ ПРИ РЕШЕНИИ ЗАДАЧИ ИДЕНТИФИКАЦИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ МНОГОСЛОЙНЫМ ПЕРСЕПТРОНОМ

(Самарский университет)

Выбор алгоритма инициализации весов является важным моментом при разработке моделей нейронных сетей глубокого обучения, так как начальные значения весов оказывают большое влияние на результат обучения сети.

При слишком больших значениях весовых коэффициентов процесс обучения может надолго застрять в седловой зоне, когда ошибка обучения будет оставаться постоянной, в то время как начальные значения, близкие к оптимальным, позволяют избежать образования локальных минимумов и ускорить процесс обучения сети.

Существует несколько алгоритмов предварительной оптимизации весовых коэффициентов сети. К ним относятся методы случайной инициализации, метод имитации отжига, генетические алгоритмы и эволюционные алгоритмы.

Чаще всего при обучении глубоких сетей используется случайная инициализация весовых коэффициентов в диапазоне  $[0, 1]$ . Такая инициализация может быть далека от оптимальной, особенно при использовании сигмоидальных функций активации, которые замедляют обучение сети из-за своего ненулевого среднего значения, которое может привести к раннему насыщению нейронов, преимущественно верхнего скрытого слоя [1]. Поэтому при использовании случайной инициализации важно выбрать оптимальный диапазон начального распределения, так как он оказывает большое влияние на результат процедуры оптимизации и на дальнейшую способность сети к обобщению.

В связи с этим были разработаны более конкретные методы случайной инициализации, которые используют информацию об архитектуре сети и типе используемой функции активации [2].

Текущий стандартный подход для инициализации весов нейронной сети, использующих сигмоидальную функцию активации, называется инициализацией "glorot" или "xavier", названной по имени ученого-исследователя Ксавье Глорота.

Метод инициализации "xavier" основан на вычислении случайных чисел с равномерным распределением вероятности в диапазоне от

$$\left[-\frac{1}{\sqrt{n}}; \frac{1}{\sqrt{n}}\right],$$

где  $n$  – количество входных нейронов сети. Но и он имеет один существенный недостаток – градиент обратного распространения затухает во время обучения.

Более усовершенствованный подход – это так называемая нормализованная инициализация Ксавье («normalized xavier»). В этом случае случайные числа рассчитываются в диапазоне:



$$\left[ -\frac{\sqrt{6}}{\sqrt{n+m}}; \frac{\sqrt{6}}{\sqrt{n+m}} \right],$$

где  $n$  – количество нейронов в предыдущем слое, а  $m$  – количество узлов в текущем слое.

В качестве алгоритма предварительной инициализации весов можно также использовать методы глобальной оптимизации, например, метод роя частиц. Метод имитирует роевое поведение, наблюдаемое в децентрализованных самоорганизующихся системах, таких как стаи птиц. Каждая рассматриваемая частица передвигается по пространству решений с некоторой скоростью и представляет собой потенциальное решение задачи оптимизации. Направление движения частицы определяется положением наилучшей точки среди собственных решений, а также положением наилучшей точки в ближайшей окрестности. С каждой итерацией частицы сдвигаются в сторону обоих решений, в результате скапливаясь вблизи одной точки, являющейся, в лучшем случае, глобальным минимумом функции.

Коррекция скорости и положения частиц после каждой итерации происходит по следующим формулам:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1r_1(p_{ij}(t) - x_{ij}(t) + c_2r_2(\tilde{p}_j(t) - x_{ij}(t))),$$
$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1),$$

где  $w$  – коэффициент инерции;  $c_1$  и  $c_2$  – положительные константы, называемые когнитивным и социальным параметрами, соответственно;  $r_1$  и  $r_2$  – случайные числа, равномерно распределенные на интервале  $[0, 1]$ ;  $j$  – индекс частицы, достигшей наилучшего положения среди всех особей роя в предыдущей итерации.

В данной работе были рассмотрены методы случайной инициализации и метод роя частиц. Также была разработана автоматизированная система для оценки эффективности рассматриваемых алгоритмов при решении задачи идентификации плотности распределения вероятности многослойным персептроном.

Оценка проводилась на случайных величинах, распределённых по нормальному, экспоненциальному, равномерному законам и закону Вейбулла.

В качестве пороговой функции выбрана сигмоидальная функция, в качестве основного алгоритма обучения – метод наискорейшего спуска с обратным распространением ошибки.

Основные параметры сети представлены на рисунке 1.

Результаты обучения сети с применением различных алгоритмов предварительной инициализации представлены в таблице 1.

В результате исследования видно, что наилучшие результаты сети на тестовых данных достигаются при использовании метода роя частиц в качестве алгоритма начальной инициализации весовых коэффициентов сети.



Количество нейронов скрытого слоя	5
Коэффициент обучения	0,1
Количество итераций обучения	1000
Количество дифференциальных коридоров	10
Размер обучающих выборок	500
Предварительная инициализация весов	Случайная инициализация [0, 1]

Рис. 1. Окно программы с параметрами сети

Таблица 1 – Зависимость процента верных распознаваний от алгоритма инициализации весовых коэффициентов

Метод предварительной инициализации	Процент верных распознаваний
случайная инициализация [0, 1]	91,78
инициализация xavier	94,56
нормализованная инициализация xavier	97,34
метод роя частиц	99,07

### Литература

1 Glorot, X. Understanding the difficulty of training deep feedforward neural networks [Электронный ресурс]. – URL: <https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>.

2 Weight Initialization for Deep Learning Neural Networks [Электронный ресурс]. – URL: <https://machinelearningmastery.com/weight-initialization-for-deep-learning-neural-networks/>.