

# АЛГОРИТМЫ ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Юнусова Л.Р.<sup>1</sup>, Магсумова А.Р.<sup>2</sup>

<sup>1</sup>Юнусова Лилия Рафиковна – магистрант;

<sup>2</sup>Магсумова Алия Рафиковна – магистрант,

направление: информатика и вычислительная техника,  
магистерская программа: технология разработки программного обеспечения,

кафедра информационных систем,

отделение информационных технологий и энергетических систем,

Высшая инженерная школа

Набережночелнинский институт

Казанский федеральный университет,

г. Набережные Челны

**Аннотация:** в статье рассматривается тема нейросетевых связей, иначе говоря, нейронных сетей. Использование нейронных сетей за последние несколько лет и на сегодняшний день является актуальной и амбициозной темой в сфере компьютерных технологий. Все больше и больше начинают применять в различных областях, в особенности компьютерные программные обеспечения активно пользуются нейронными сетями, как примеры можно привести распознавание образов или обработка фотографий в современных телефонах прямо во время съемки кадров.

**Ключевые слова:** нейронные сети, обучение нейронных сетей, алгоритмы, алгоритм Хэбба, корреляция.

Как и в случае с большинством нейронных сетей, цель состоит в обучении сети таким образом, чтобы достичь баланса между способностью сети давать верный отклик на входные данные, использовавшиеся в процессе обучения (запоминания), и способностью выдавать правильные результаты в ответ на входные данные, схожие, но неидентичные тем, что были использованы при обучении (принцип обобщения). Обучение сети методом обратного распространения ошибки включает в себя три этапа: подачу на вход данных, с последующим распространением данных в направлении выходов, вычисление и обратное распространение соответствующей ошибки и корректировку весов. После обучения предполагается лишь подача на вход сети данных и распространение их в направлении выходов. При этом, если обучение сети может являться довольно длительным процессом, то непосредственное вычисление результатов обученной сетью происходит очень быстро. Кроме того, существуют многочисленные вариации метода обратного распространения ошибки, разработанные с целью увеличения скорости протекания процесса обучения. Также стоит отметить, что однослойная нейронная сеть существенно ограничена в том, обучению каким шаблонам входных данных она подлежит, в то время, как многослойная сеть (с одним или более скрытым слоем) не имеет такого недостатка.

Общепринятой практикой в задачах нейросетевого моделирования справедливо считается применение процедур исключения повторяющихся и противоречивых примеров с целью уменьшения размера обучающей выборки и, как следствие, ускорения самого процесса обучения. Однако практический опыт решения таких задач показывает, что данный подход подтверждает свою эффективность всегда для малых и средних объемов, обучающих данных (до нескольких сотен примеров) и отнюдь не постоянно для выборок больших размеров, насчитывающих тысячи, десятки и сотни тысяч примеров. Связано это с тем, что для небольших выборок почти всегда можно подобрать достаточно простую структуру нейронной сети и осуществить процесс её обучения за приемлемый интервал времени таким образом, что она, если не обучится обобщать все возможные входные наборы с выдачей правильного отклика, то, по крайней мере, запомнит обучающие данные. Для

больших же выборок, требующих к тому же более сложных структур нейронных сетей, временной фактор, обусловленный огромным объёмом вычислений, вступает в противоречие с достижением достаточного уровня ошибки обучения. И если для выборок размера в несколько тысяч примеров можно найти определённый баланс между временем обучения и уровнем ошибки, то в случаях, когда размер выборки достигает десятков и сотен тысяч примеров, не помогает даже вычислительных

Последнее расчётное соотношение отличается от аналогов учётом частоты  $f_k$ , с которой  $k$ -й пример встречается во время моделируемого процесса или явления. Таким образом, производимая с его помощью оценка будет соответствовать существующей в действительности картине. Следует заметить, что в данном алгоритме не используется традиционное обучение нейронной сети по эпохам – циклам однократного предъявления в случайном порядке всех примеров обучающей выборки. Рассмотрим примеры использования предложенного алгоритма для обучения двухслойного персептрона (рис. 1) на примере двух серий вычислительных экспериментов с двумя выборками, содержащими, соответственно, 500 и 5000 уникальных обучающих примеров, для каждого из которых заранее известна частота его реализации во время реального процесса или явления. Используемые структуры двухслойных персептронов включали 5 входных независимых переменных и 2 выходные результирующие переменные (2 искусственных нейрона в выходном слое). Количество скрытых нейронов  $N_{скр}$  варьировалось от 3 до 25. Поскольку инициализация весовых коэффициентов нейронной сети и порядок предъявления примеров обучающей выборки осуществляются случайным образом, ход процесса обучения для одинаковых настроек и структуры на практике не может повториться, а его фактическая скорость немного меняется. В этой связи для каждого опыта была проведена серия из 5 попыток обучения. Все приводимые далее на графиках данные – это средние арифметические значения по 5 попыткам. Для коррекции весовых коэффициентов на каждом такте обучения использовались расчётные соотношения метода обратного распространения ошибки [2,505].

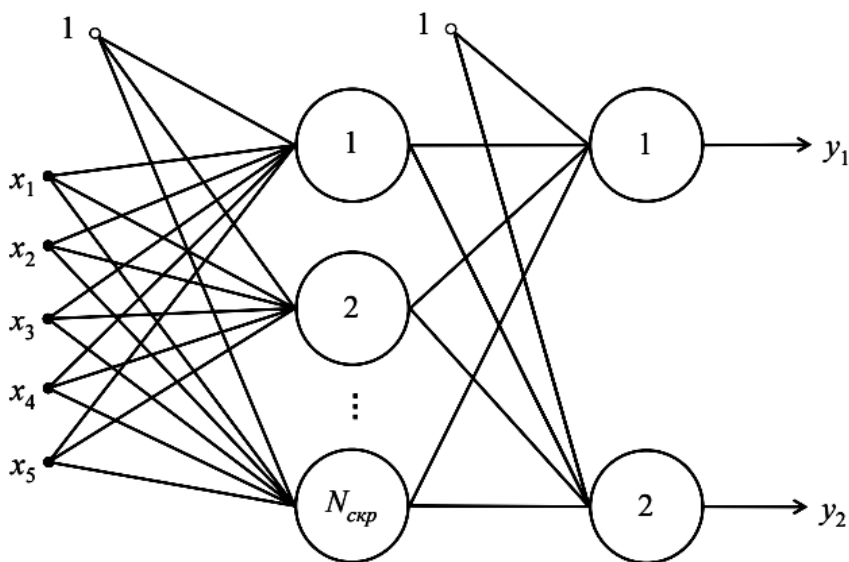


Рис. 1. Структура двухслойного персептрона

На рисунке 2 показана зависимость ошибки обучения от сложности структуры персептрона, изменявшейся за счёт добавления или сокращения количества скрытых нейронов для выборки в 500 примеров после 10000 тактов обучения. На данном

графике сравниваются структуры сетей, обучавшихся по предложенному алгоритму с учётом различной частоты использования каждого примера, и аналогичные им структуры сетей, обучавшихся на выборках с одинаковой частотой использования каждого примера (эквивалент традиционного подхода). Из рисунка видно, что подтверждается общеизвестная тенденция к уменьшению ошибки с усложнением структуры сети. Однако главное наблюдение – это наличие в среднем 0,5% разницы в ошибке в пользу предложенного подхода для двухслойных персептронов практически любой сложности. В некоторых случаях в зависимости от размера выборки, сложности структуры сети и сложности воспроизводимой ею функциональной зависимости разница может достигать 1–2%, что весьма существенно для не до конца обученных персептронов.

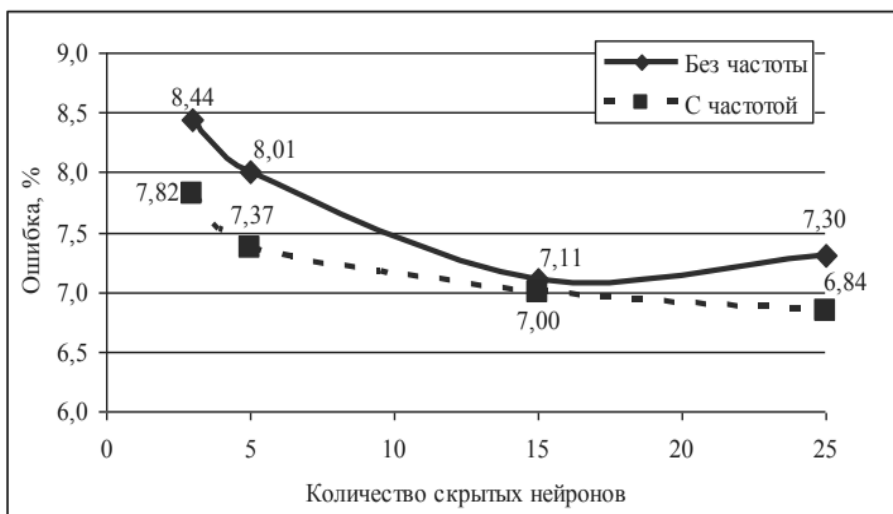
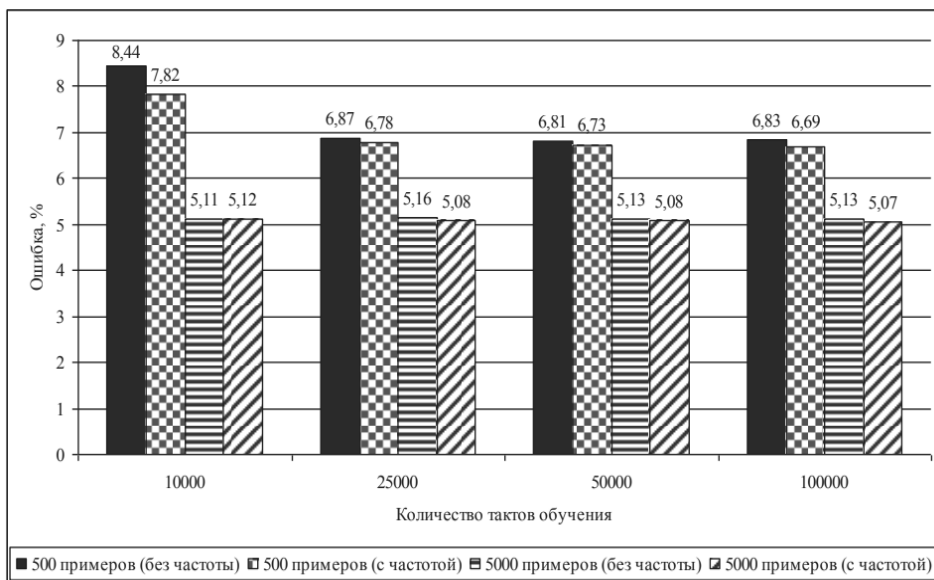


Рис. 2. Зависимость ошибки обучения от количества нейронов в скрытом слое

На рисунке 3 показано распределение ошибок для нейронных сетей, обучавшихся с использованием предложенного и традиционного подходов на 500 и 5000 примерах, в зависимости от количества тактов обучения. Из представленной диаграммы можно сделать следующие выводы: – увеличение продолжительности обучения способствует снижению ошибки для всех случаев; – учёт частоты использования примеров позволяет снизить ошибку, причём для рассмотренных массивов данных это снижение более заметно для выборок среднего размера, чем для больших, что объясняется не очень большой сложностью описываемых функциональных зависимостей; – та же самая причина вкупе с большей репрезентативностью обучающих данных обуславливает, на наш взгляд, меньшую ошибку (на 2–3%) для выборки размера 5000 примеров в сравнении с выборкой, содержащей 500 примеров.



*Рис. 3. Распределение ошибок для выборок из 500 и 5000 примеров при различной продолжительности обучения*

Полученные результаты позволяют сделать общий вывод о том, что использование частоты повторяемости примеров в процессе обучения позволяет быстрее достигать требуемого уровня ошибки.

- Алгоритм обучения Хэбба. По существу, Хэбб предположил, что синаптическое соединение двух нейронов усиливается, если оба эти нейрона возбуждены. Это можно представить, как усиление синапса в соответствии с корреляцией уровней возбужденных нейронов, соединяемых данным синапсом. Поэтому алгоритм обучения Хэбба иногда называется корреляционным алгоритмом.

- Метод сигнального обучения Хэбба предполагает вычисление свертки предыдущих изменений выходов для определения изменения весов.

Много общих идей, используемых в искусственных нейронных сетях прослеживаются в работах С. Гроссберга; в качестве примера можно указать конфигурации входных и выходных звезд, используемые во многих сетевых парадигмах. Входная звезда, как показано на рис.4, состоит из нейрона, на который подается группа входов через синаптические веса. Выходная звезда, показанная на рис.5, является нейроном, управляющим группой весов. Входные и выходные звезды могут быть взаимно соединены в сети любой сложности; Гроссберг рассматривает их как модель определенных биологических функций. Вид звезды определяет ее название, однако, звезды обычно изображаются в сети несколько иначе[3,154].

Входная звезда выполняет распознавание образов, т. е. она обучается реагировать на определенный входной вектор и ни на какой другой. Это обучение реализуется, настраивая веса таким образом, чтобы они соответствовали входному вектору. Выход входной звезды определяется как взвешенная сумма ее входов, это описано в предыдущих разделах. С другой точки зрения, выход можно рассматривать как свертку входного вектора с весовым вектором или меру сходства нормализованных векторов. Следовательно, нейрон должен реагировать наиболее сильно на входной образ, которому был обучен.

В то время как входная звезда возбуждается всякий раз при появлении определенного входного вектора, выходная звезда имеет дополнительную функцию: она вырабатывает требуемый возбуждающий сигнал для других нейронов всякий раз,

когда возбуждается. Для того чтобы обучить нейрон выходной звезды, его веса настраиваются в соответствии с требуемым целевым вектором[1, 45].

### **Метод обучения Уидроу—Хоффа**

Как мы видели, персептрон ограничивается бинарными выходами. Б.Уидроу вместе со студентом университета М.Хоффом расширили алгоритм обучения персептрона для случая непрерывных выходов, используя сигмоидальную функцию. Второй их впечатляющий результат — разработка математического доказательства, что сеть при определенных условиях будет сходиться к любой функции, которую она может представить. Их первая модель — Адалин — имеет один выходной нейрон, более поздняя модель — Мадалин — расширяет ее для случая с многими выходными нейронами.

Выражения, описывающие процесс обучения Адалина, очень схожи с персептронными. Существенные отличия имеются в четвертом шаге, где используются непрерывные сигналы NET вместо бинарных OUT.

### **Список литературы**

1. *Борисов Е.С.* Основные модели и методы теории искусственных нейронных сетей, 2005 г. С.35-47.
  2. *Бенджио, Гудфеллоу, Курвилль:* Глубокое обучение. Издательство: ДМК-Пресс, 2018 г. С.492-568.
  3. *Осовский С.* Нейронные сети для обработки информации, 2002г, С.128-264.
-