



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение эвм и информационные технологии»

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
НА ТЕМУ:**

**«Классификация известных архитектур нейронных
сетей прямого распространения»**

Студент ИУ7-51Б

_____ Кузнецова А. В.

Руководитель

_____ Кострицкий А. С.

2022 г.

РЕФЕРАТ

В научно-исследовательской работе рассматриваются архитектуры нейронных сетей прямого распространения, и проводится сравнительный анализ по критериям.

Научно-исследовательская работа содержит расчетно-пояснительную записку объемом 18 страниц, включая 1 таблицу, 4 иллюстрации, 1 приложение и список использованных источников из 15 наименований.

Ключевые слова: нейронные сети прямого распространения, сверточная нейронная сеть, РБФ-сеть, перцептрон, алгоритм обратного распространения ошибки, классификация.

Содержание

РЕФЕРАТ	3
ВВЕДЕНИЕ	5
1 Анализ предметной области	6
1.1 Основные определения	6
2 Классификация существующих архитектур нейронных сетей прямого распространения	8
2.1 Существующие архитектуры	8
2.1.1 Однослойный перцептрон	8
2.1.2 Сверточные нейронные сети	10
2.1.3 Сети радиально-базисных функций	11
2.2 Критерии оценки архитектур	13
2.3 Сравнение архитектур	14
ЗАКЛЮЧЕНИЕ	15
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	17
ПРИЛОЖЕНИЕ А Презентация научно-исследовательской ра- боты	18

ВВЕДЕНИЕ

Искусственные нейронные сети представляют собой технологию, уходящую корнями во множество дисциплин: нейрофизиологию, математику, статистику, физику, компьютерные науки и технику [1]. Актуальность исследований в этом направлении подтверждается массой различных применений нейронных сетей [2].

К задачам, успешно решаемым нейронными сетями на данном этапе их развития относится распознавание зрительных, слуховых образов, распознавание текста, создание систем голосового управления, синтез речи, формирование моделей и различных нелинейных и трудно описываемых математически систем, прогнозирование развития этих систем во времени, прогнозирование развития циклонов и других природных процессов, прогнозирование изменений курсов валют и других финансовых процессов, разнообразные конечные автоматы: системы массового обслуживания и коммутации, телекоммуникационные системы [3].

Таким образом, можно сказать, что использование нейронных сетей является актуальным направлением, которое непрерывно развивается, по пути устранения недостатков.

Целью данной научно-исследовательской работы является классификация известных архитектур нейронных сетей прямого распространения.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- сформировать основные понятия, касающиеся темы нейронных сетей;
- изучить существующие архитектуры нейронных сетей прямого распространения;
- классифицировать архитектуры;
- сформулировать критерии оценки архитектур;
- провести сравнение архитектур на основании выделенных критериев.

1 Анализ предметной области

В данном разделе вводятся основные определения, касающиеся темы нейронных сетей.

1.1 Основные определения

Искусственная нейронная сеть — сеть устроенных определенным образом элементов — искусственных нейронов. Она обрабатывает входную информацию и в процессе изменения своего состояния во времени формирует совокупность выходных сигналов [4].

Нейронные сети прямого распространения — нейронные сети, в которых все нейроны и входные блоки соединены с передним нейроном и не связаны с нейронами, которые, расположены в том же слое и в предыдущем [5].

Нейронные сети прямого распространения являются наиболее популярным и хорошо зарекомендовавшим себя на различных задачах видом нейронных сетей.

Несмотря на существенные различия, отдельные типы нейронных сетей прямого распространения обладают несколькими общими чертами. Основу каждой нейронной сети прямого распространения составляют относительно простые, в большинстве случаев — однотипные, элементы (ячейки), имитирующие работу нейронов мозга. Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены или заторможены. Он обладает группой синапсов — однонаправленных входных связей, соединенных с выходами других нейронов, а также имеет аксон — выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов. Общий вид нейрона приведен на рисунке 1.1 [2].

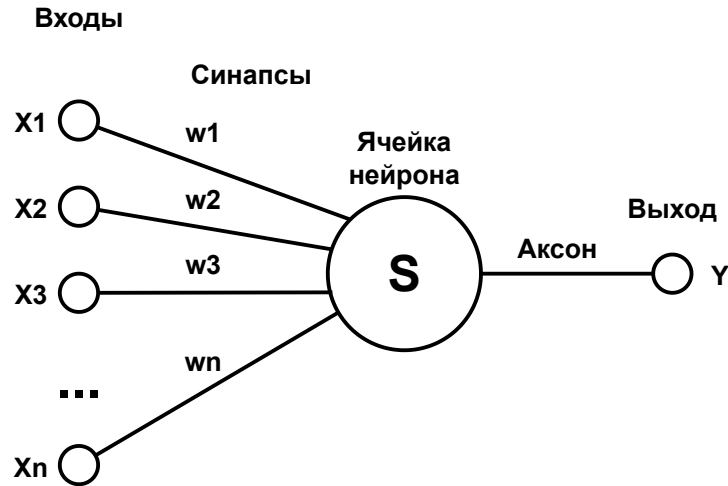


Рисунок 1.1 – Структура нейрона

Каждый i -й синапс характеризуется величиной синаптической связи (ее весом) w_i [2].

Текущее состояние нейрона обычно определяется, как взвешенная сумма его входов (1.1):

$$s = \sum_{i=1}^n x_i \cdot w_i. \quad (1.1)$$

Выход нейрона определяется функцией его состояния (1.2):

$$y = f(s). \quad (1.2)$$

Входной слой искусственных нейронов получает информацию из окружающей среды, а выходной слой передает ответ; между этими уровнями может находиться один или несколько «скрытых» слоев (без прямого контакта с окружающей средой), где происходит большая часть обработки информации. Выход нейронной сети зависит от веса связей между нейронами в разных слоях. Каждый вес указывает на относительную важность определенного соединения. Если сумма всех взвешенных входных сигналов, полученных конкретным нейроном, превышает определенное пороговое значение, нейрон отправит сигнал каждому нейрону, к которому он подключен в следующем слое [4].

2 Классификация существующих архитектур нейронных сетей прямого распространения

В данном разделе описываются существующие архитектуры, выделяются критерии их оценки, и проводится сравнение описанных архитектур по выделенным критериям.

2.1 Существующие архитектуры

Существуют следующие архитектуры нейронных сетей прямого распространения:

- однослойный перцептрон;
- сверточные нейронные сети;
- сети радиально-базисных функций.

2.1.1 Однослойный перцептрон

Однослойная сеть прямого распространения представляет собой простейший случай прямонаправленной многослойной сети. Она включает в себя слой входных нейронов и слой выходных. Нейроны входного слоя просто передают сигналы на выходной слой, не преобразуя их. В выходном слое происходит преобразование сигналов и формирование реакции сети. В отличие от принятой методики обозначения, такие нейронные сети называются однослойными, а не двухслойными. Этим подчеркивается, что вычисления производятся лишь одним слоем сети [6].

Идея однослойного перцептрона была предложена нейрофизиологом Ф. Розенблаттом в 50-е годы XX века, это была одна из первых моделей нейронной сети. В данной модели сигнал, который поступает на вход проходит лишь один слой нейронов и впоследствии формируется выходной сигнал с помощью сигмоидальной функции [7].

Однослойный перцептрон представляет собой простейшую сеть, которая состоит из группы нейронов, образующих слой (рис. 2.1 [7]).

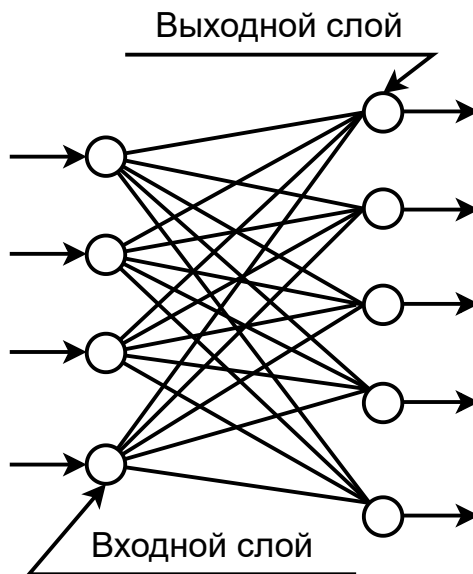


Рисунок 2.1 – Схема однослойного перцептрона

Входные данные кодируются вектором значений, каждый элемент подается на соответствующий вход каждого нейрона в слое. В свою очередь, нейроны вычисляют выход независимо друг от друга. Размерность выхода (то есть количество элементов) равна количеству нейронов, а количество синапсов у всех нейронов должно быть одинаково и совпадать с размерностью входного сигнала [8].

Процесс обучения однослойного перцептрона представляет собой обучение с учителем, при котором сети предъявляется выборка обучающих примеров. Обучение однослойного перцептрона проводится по следующему алгоритму:

1. Каждый образец подается на входы сети.
2. Проходит обработку внутри структуры нейронной сети.
3. Вычисляется выходной сигнал сети, который сравнивается с соответствующим значением целевого вектора, представляющего собой требуемый выход сети.
4. Вычисляется ошибка.

5. Происходит изменение весовых коэффициентов связей внутри сети в зависимости от выбранного алгоритма

Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет приемлемо низкого уровня [8].

Перцептрон, который содержит скрытые слои, называется многослойным перцептроном.

2.1.2 Сверточные нейронные сети

Сверточная нейронная сеть – одна из разновидностей нейронных сетей, предназначенных для эффективного анализа преимущественно двумерных и трехмерных (RGB-изображения) данных (например, распознавания объектов на изображениях) [9]. Данная архитектура входит в состав технологий глубокого обучения и была создана целенаправленно для решения современных проблем, связанных с восприятием, например, таких как зрение или распознавание речи. Под «глубиной» в моделях глубокого обучения подразумевается многослойная структура, где каждый слой выполняет определенную функцию. Уникальность сверточных нейронных сетей состоит в использовании и комбинировании двух типов слоев, а именно сверточных и субдискретизирующих (рис. 2.2 [9]).



Рисунок 2.2 – Модель сверточной нейронной сети

Сверточный слой отвечает за применение определенного набора фильтров к поданному на вход сигналу (изображению), формируя, таким образом кар-

ты признаков. То есть операция свертки подразумевает, что каждый фрагмент входа поэлементно умножается на небольшую матрицу весов (ядро), а результат суммируется. Эта сумма является элементом выхода, который и называется картой признаков. Взвешенная сумма входов пропускается через функцию активации [10]. Слой пулинга (подвыборки, субдискретизации) представляет собой нелинейное уплотнение карты признаков, проходя нелинейное преобразование. Пулинг интерпретируется как разбиение карты признаков на более мелкие матрицы, нахождение их максимальных элементов, т. е. происходит увеличение «глубины» значений [10]. Благодаря тому, что каждый последующий слой (идущий после первого субдискретизирующего слоя) принимает на вход карты признаков меньшей размерности, увеличивается скорость работы сети [11].

Сверточная нейронная сеть способна достигать еще большей точности за счет возможностей настроек глубинной архитектуры, а именно: количество фильтров каждого сверточного слоя, размеры ядер свертки, глубина сети и т. д.

Для обучения сверточной нейронной сети используется специфическая версия алгоритма обратного распространения ошибки, который относится к методам обучения с учителем [9].

2.1.3 Сети радиально-базисных функций

Сеть радиальных базисных функций представляет собой нейронную сеть прямого распространения сигнала, содержащую промежуточный (скрытый) слой радиально симметричных нейронов [12]. Любой нейрон в сети преобразовывает расстояние от входного вектора до соответствующего ему «центра» по некоторому линейному закону.

В сетях с радиальными базисными функциями скрытые нейроны реализуют функции, радиально изменяющиеся вокруг выбранного центра и принимающие ненулевые значения только в окрестности этого центра.

В таких сетях роль скрытого нейрона заключается в отображении радиального пространства вокруг одиночной заданной точки либо вокруг группы точек, образующих кластер. Нейронные сети с радиальными базисными функциями отличаются простотой, так как содержат только один скры-

тый слой (рис. 2.3[13]), поэтому сложный, неформализованный этап подбора структуры сетей может быть исключен [13].

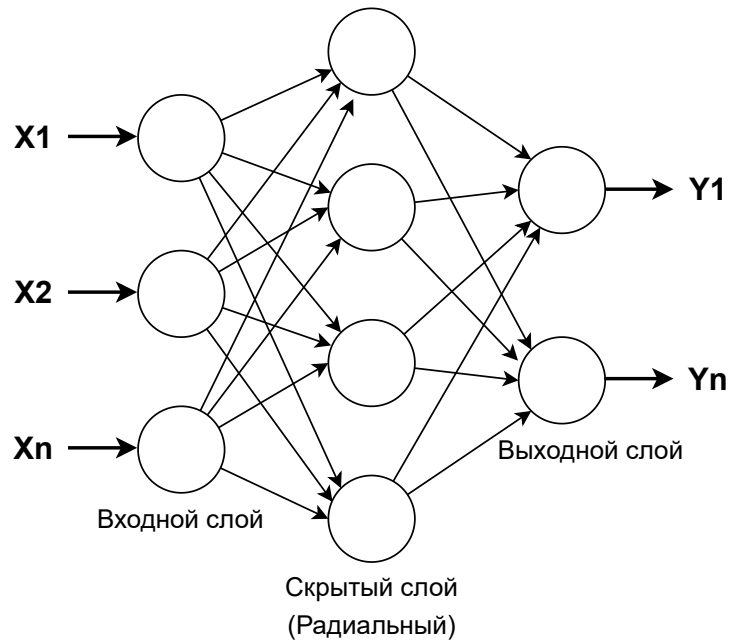


Рисунок 2.3 – Трехслойная модель сети радиально-базисных функций

Скрытый слой — это слой радиально-базисных функций. В качестве активационной функции таких нейронов обычно используется одномерная или многомерная (в зависимости от размерности задачи моделирования) функция Гаусса. Количество скрытых нейронов определяется количеством заданных стационарных узлов интерполирования, а выходы этих нейронов, соответственно, представляют собой меру степени соответствия вектора входных переменных центрам одного из узлов. Нейроны выходного слоя призваны линейно взвесить выходы нейронов скрытого слоя, что позволяет рассчитать выходы сети для поданной на неё входной комбинации значений переменных [14].

РБФ-сети обучаются по принципу обучения с учителем. Целью обучения нейронной сети является определение таких значений весовых коэффициентов, при которых рассчитанные значения выходных переменных будут наиболее близки к выходным значениям обучающих примеров. В нейронных сетях на основе радиально-базисных функций весовые коэффициенты выходного слоя могут быть вычислены однозначно за один расчётный цикл с использованием базовых операций матричной алгебры благодаря линейной связанности с выходными значениями и, как следствие, с погрешностью расчёта для

каждого отдельно взятого примера [14]. Обучение для РБФ-сетей происходит быстро, во многих случаях мгновенно вследствие однократного выполнения расчётного цикла.

При настройке РБФ-сетей можно учесть репрезентативность данных в обучающей выборке, регулируя плотность узлов интерполирования в скрытом слое [15].

Следует заметить, что структура РБФ-сети требует использования значительного размера скрытого слоя. Этому существует формальное объяснение в рамках задачи классификации. РБФ-сеть нелинейно отображает сигналы из входного пространства в пространство выходов нейронов скрытого слоя (скрытое пространство). Выходные нейроны являются линейными, а поэтому, согласно теореме Ковера, все входные шаблоны могут быть линейно разделены лишь при достаточно высокой размерности скрытого пространства. Это условие определяет главный недостаток РБФ-сетей — проклятие размерности, так как с увеличением размерности входного пространства число необходимых нейронов скрытого слоя возрастает экспоненциально [15]. Таким образом, сложность РБФ-сети увеличивается с количеством нейронов в скрытом слое.

Кроме того, классические РБФ-сети имеют некоторые проблемы в своей структуре и алгоритме обучения, поэтому не могут моделировать сильно нелинейные системы.

2.2 Критерии оценки архитектур

Сравнение описанных архитектур нейронных сетей прямого распространения будет проводиться по следующим критериям:

- количество слоев;
- однозначность результата обучения;
- использование общих весов;
- возможность дообучения.

От первого критерия зависит сложность архитектуры нейронной сети, из которой следует длительность ее проектирования. Вторым критерий характеризует продолжительность обучения нейронной сети. Использование общих весов обеспечивает детектирование некоторых признаков в любом месте входных данных и уменьшает количество весовых коэффициентов. Четвертый критерий влияет на время, затраченное при добавлении новых обучающих примеров.

2.3 Сравнение архитектур

Обозначим введенные критерии оценки архитектур следующим образом:

- K1 — количество слоев;
- K2 — однозначность результата обучения [14] [9];
- K3 — использование общих весов [11] [14];
- K4 — возможность дообучения [14] [9].

Результаты сравнения архитектур нейронных сетей прямого распространения представлены в таблице 2.1.

Таблица 2.1 – Сравнение архитектур нейронных сетей прямого распространения

Архитектура	K1	K2	K3	K4
Однослойный перцептрон	1 слой	нет	нет	есть
Сверточная нейронная сеть	более 2 слоев	нет	есть	есть
РБФ-сеть	2 слоя	есть	нет	нет

ЗАКЛЮЧЕНИЕ

В ходе выполнения данной работы были классифицированы известные архитектуры нейронных сетей прямого распространения.

Цель, поставленная в начале работы, была достигнута. В ходе ее выполнения были решены следующие задачи:

- сформированы основные понятия, касающиеся темы нейронных сетей;
- изучены существующие архитектуры нейронных сетей прямого распространения;
- классифицированы архитектуры;
- сформулированы критерии оценки архитектур;
- проведено сравнения архитектур на основании выделенных критериев.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Саймон Хайкин. Нейронные сети: полный курс, 2 издание. — М.: Издательский дом «Вильямс», 2006. — 1104 с.
2. С. Короткий. Нейронные сети: основные положения, 2002.
3. И. В. Заенцев. Нейронные сети: основные модели. — Воронеж: ВГУ, 1999.
4. В. В. Ксенофонтов. Нейронные сети. — М.: Проблемы науки, 2020.
5. О. К. Хужаев, Ш. А. Ядгаров, В. С. Пак. Сравнение моделей нейронной сети для прогнозирования временных рядов // Технические науки – от теории к практике, 2014.
6. А. В. Чижков, С. В. Сеитова. Классификация нейросетевых архитектур // Известия ЮФУ. Технические науки, 2009.
7. Н. А. Алешин. Однослойный перцептрон и алгоритм его обучения. — Пенза: МЦНС «Наука и просвещение», 2019. — 124 с.
8. Ф. М. Гафаров. Искусственные нейронные сети и их приложения: учеб. пособие. — Казань: КФУ, 2018. — 121 с.
9. А. И. Бредихин. Алгоритмы обучения сверточных нейронных сетей // Вестник ЮГУ, 2019.
10. Е. В. Воробьев, Е. В. Пучков. Классификация текстов с помощью сверточных нейронных сетей // Молодой исследователь Дона, 2017.
11. Д. В. Плотников. Вычислительный интеллект в задаче распознавания образов // Актуальные проблемы авиации и космонавтики, 2017.
12. *Е.В. Б.* Нейросетевой метод расширения динамического диапазона аналого-цифрового преобразователя. — —М. : Перспективы развития информационных технологий, 2011.
13. О. В. Аникина, О. М. Гущина, Е. В. Панюкова, Н. Н. Рогова. Табличная реализация искусственной нейронной сети радиальных базисных функций для классификации образов // Современные информационные технологии и ИТ-образование, 2018.

14. П. Л. Папаев, И. В. Макляев, С. В. Дударов, Н. Н. Рогова. Сравнение моделирующей способности нейронных сетей радиально-базисных функций и двухслойных перцептронов // Успехи в химии и химической технологии, 2020.
15. К. В. Махотило. Разработка методик эволюционного синтеза нейросетевых компонентов систем управления: дис. — Харьков: ХГПУ, 1998.

ПРИЛОЖЕНИЕ А

Презентация научно-исследовательской работы

Презентация научно-исследовательской работы содержит 8 слайдов, на которых представлено краткое описание рассмотренных архитектур нейронных сетей и визуализация результатов сравнения.