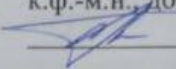


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КубГУ»)

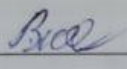
Факультет компьютерных технологий и прикладной математики
Кафедра прикладной математики

Допустить к защите
И.о. заведующего кафедрой
к.ф.-м.н., доцент
 А.В. Письменский

_____ 2023г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(БАКАЛАВРСКАЯ РАБОТА)**

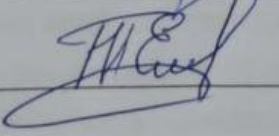
**НЕЙРОСЕТЕВОЕ ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ
СПОРТИВНЫХ СОСТЯЗАНИЙ**

Работу выполнил  _____ В.Е. Жалдаков

Направление подготовки 02.03.02 «Фундаментальная информатика и
информационные технологии»

Направленность (профиль) Математическое и программное обеспечение
компьютерных технологий

Научный руководитель
канд. ф.-м. наук, доц.  _____ А. В. Письменский

Нормоконтролер
преподаватель  _____ Е. С. Троценко

Краснодар
2023

РЕФЕРАТ

Выпускная квалификационная работа 57 стр., 3 ч., 16 рис., 12 источников.

НЕЙРОННАЯ СЕТЬ, ПРОГНОЗИРОВАНИЕ, СПОРТ, ТЕННИС, МНОГОСЛОЙНЫЙ ПЕРСЕПТРОН

Объектом исследования в данной работе является модель нейронной сети.

Предметом исследования является способность нейронной сети анализировать входную информацию, устанавливать зависимости между параметрами и таким образом осуществлять прогноз.

Цель работы: изучение возможностей применения нейронной сети в прогнозировании исходов спортивных состязаний и разработка модели, способной предсказывать результат.

Методологическая основа исследования включает в себя эмпирический метод, сбор и обработка статистической информации, анализ полученных результатов, синтез выводов, аналогию.

В результате работы проведено исследование способности модели нейронной сети многослойного персептрона на основе определенных статистических данных предсказывать результаты теннисных матчей, а также спроектирована модель нейронной сети, способная предсказывать исходы состязаний.

В результате решения поставленных задач определены возможности применения нейронных сетей для прогнозирования результатов спортивных состязаний, на основе ряда экспериментов проведена проверка эффективности работы модели и её способности выполнять поставленные задачи на реально существующих показателях.

СОДЕРЖАНИЕ

Введение	4
1 Общие сведения о нейронной сети и её применении в прогнозировании	7
1.1 Понятие нейронной сети	7
1.2 Функция активации	10
1.3 Проблемы прогнозирования высших спортивных достижений	11
1.3.1 Оценка возможности прогнозирования индивидуальных спортивных результатов	11
1.3.2 Методы прогнозирования в спорте.....	13
1.3.3 Методы экспертных оценок	14
1.3.4 Формализованные методы прогнозирования.....	15
1.3.5 Примеры и оценка прогнозирования достижений в легкой атлетике.....	17
1.4 Применение нейронных сетей для прогнозирования спортивных состязаний.....	21
1.4.1 Основные этапы построения нейросетевой прогнозирующей системы	21
1.4.2 Примеры прогнозирования спортивных матчей.....	22
2 Этапы разработки модели нейронной сети для прогнозирования спортивных состязаний.....	27
2.1 Сбор исходных данных	27
2.2 Формирование и структуризация обучающей выборки	29
2.3 Нормализация данных.....	30
2.4 Выбор подходящей архитектуры нейронной сети	32
2.5 Обучение модели.....	33
2.6 Настройка алгоритма обучения	35
2.7 Проверка разработанной модели	36
2.8 Практическое использование нейронной сети для прогнозирования вероятностей исходов матчей.....	37
3 Построение модели нейронной сети для прогнозирования исходов теннисных матчей	39

3.1 Структуризация показателей теннисистов и формирование обучающей выборки.....	39
3.2 Выбор архитектуры нейронной сети.....	42
3.3 Построение и обучение модели нейронной сети.....	44
3.4 Анализ результатов обучения модели нейронной сети.....	47
Заключение	49
Список использованных источников	50
Приложение А Файл создания модели в пакете Matlab.....	52
Приложение Б Файл создания модели средствами языка Python.....	54

ВВЕДЕНИЕ

Прогноз – это научно обоснованное суждение о возможных состояниях объекта или предмета будущих исследований и (или) об альтернативных событиях и сроках их становления. Прогнозирование – это процесс разработки (формирования) прогноза, специальное научное исследование конкретных перспектив дальнейшего развития какого-либо процесса. Сегодня прогнозирование используется во многих сферах человеческой деятельности: в науке, экономике, медицине и многих других областях. Прогнозирование в общем случае связано с задачами анализа данных, поскольку нет четких границ. Этот процесс является важным фактором во время принятия решений относительно определения направления стратегии развития, тенденций в поведенческой динамике конкретного прогнозируемого объекта или события на основе анализа его состояния в прошлом и настоящем.

На протяжении всей истории спорта людей всегда восхищала способность точно прогнозировать исход спортивных событий. Это человеческое увлечение предсказанием будущего в рамках спорта в последнее время набирает все больше оборотов с появлением сервисов, принимающих ставки на различные состязания не только в парадигме победитель-проигравший, но и на вероятные моменты, возникающие во время матча (число сетов в теннисе, количество мячей, забитых в футбольном противостоянии и другие)

Спортивные прогнозы являются неотъемлемой частью профессиональных ставок (букмекерских игр). Большое количество поклонников спортивных игр, таких как футбол, хоккей и других, где несколько команд соревнуются друг с другом, рано или поздно проявляли интерес к прогнозированию своих результатов, делая это систематически или нерегулярно, стараясь при этом не только доверять своей интуиции, но и разработать алгоритм, позволяющий им удачно предсказать исход

спортивного матча или состязания. Часто игроки используют прогнозы такие, какие они есть, ориентируясь лишь на свои предположения, стремительно получая доход от ставок на спорт, а затем, выигрывая крупные суммы денег, начинают повышать сумму ставки и проигрывают. Профессионалы работают более комплексно и тщательно, снижая единовременную прибыль от каждой ставки, но выравнивая доход с течением времени, обеспечивая стабильность своих выигрышей. Среди них гораздо меньше людей, которые занимаются методами искусственного интеллекта и понимают, как их можно использовать для решения проблемы прогнозирования результатов матчей в спортивных состязаниях.

Неопределенность в исходе соревновательных игр, с одной стороны, повышает привлекательность и азарт спортивных состязаний, с другой же стороны, делает их сложными для науки прогнозирования. Предсказанию результатов спортивных соревнований уже давно уделяется пристальное и обширное внимание. Исследование этого вопроса стало частью математической науки, основным направлением которой является точность прогноза.

До сих пор теннису не уделялось должного внимания, которое ученые уделяли футболу, баскетболу или хоккею. Большинство исследователей использовали статистические подходы, основанные на вычислениях вероятностей того или иного события, лишь немногие обратили внимание на инструменты машинного обучения

Основная цель данной работы – изучение возможностей применения нейронной сети в прогнозировании результатов спортивных состязаний, создание нейросетевой модели, способной предсказывать исход теннисных матчей. Для реализации поставленной цели предполагается выполнить следующие задачи:

- исследование теории нейронных сетей и способов их использования в прогнозировании;

- поиск и сбор необходимой информации о статистических показателях спортсменов и результатах матчей за определенный период, на основе которой будет осуществляться прогноз;
- выбор подходящей архитектуры нейронной сети;
- разработка модели и её реализация;
- создание выборок данных для тренировки и тестирования модели и непосредственно её обучение;
- настройка алгоритма обучения;
- анализ эффективности работы системы путем проведения испытаний и сравнения их результатов с реальными.

Актуальность работы заключается в популярности различных схем заработка на спортивных событиях.

Результатом работы будет являться прогнозная система, позволяющая достаточно точно предсказать исход теннисного матча, которую в дальнейшем можно будет использовать в различных программных системах букмекеров для формирования коэффициентов на то или иное событие или с целью возможного заработка на ставках. Присутствует возможность использования выводов, проведенных в данном исследовании, для создания модели нейронной сети для прогнозирования результатов более комплексных командных видов спорта.

1 Общие сведения о нейронной сети и её применении в прогнозировании

1.1 Понятие нейронной сети

Нейронная сеть (искусственная нейронная сеть, ИНС) – это математическая модель, а также ее программная или аппаратная реализация, которая построена по принципу организации и деятельности биологических нейронных сетей – сетей нервных клеток живого организма. Такое понятие появилось при изучении процессов, протекающих в мозге, и попытке смоделировать нейронные взаимодействия. Первой такой попыткой были нейронные сети У. Маккаллоха и У. Питтса. После разработки алгоритмов обучения полученная модель стала использоваться в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления и т.д. [3].

Существует несколько взглядов на аппарат нейронных сетей в зависимости от области его применения:

- с точки зрения машинного обучения, нейронная сеть – это частный случай методов распознавания образов, дифференциального анализа, методов кластеризации и т.д.;
- в математике обучение нейросетевой модели представляется в качестве многопараметрической задачи нелинейной оптимизации;
- в области кибернетики нейронная сеть используется в задачах адаптивного управления, а также в качестве алгоритмов для робототехники;
- с точки зрения развития вычислительной техники и программирования ИНС – это метод решения проблемы эффективного подобию и инструмент, способный находить взаимосвязь между большим числом переменных;
- сфера искусственного интеллекта определяет нейронную сеть как основу философского течения коннекционизма и базовое направление в

структурном подходе к изучению возможности моделирования естественного интеллекта с использованием компьютерных алгоритмов [1].

Любая архитектура ИНС состоит из искусственных нейронов, каждый из которых представляет собой вычислительный блок, который, в свою очередь, получает информацию, выполняет над ней простые вычисления и передает её следующему нейрону. Они делятся на три основных типа: входные; те, которые включают в себя один или несколько скрытых слоев, и выходные. Пример простейшей архитектуры нейронной сети представлен на рисунке 1.

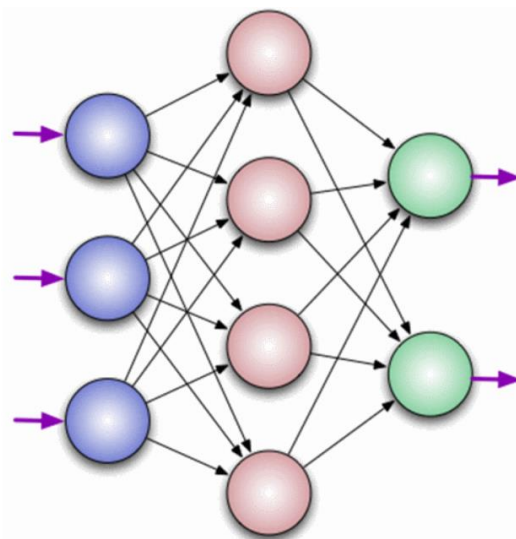


Рисунок 1 – Пример простейшей нейронной сети

Смещенный и контекстный нейроны могут присутствовать в архитектуре. В случае, когда нейронная сеть состоит из большого количества таких структурных единиц, вводится понятие слоя. Соответственно, существует входной слой, который получает информацию, N скрытых слоев (обычно не более трех), которые её обрабатывают, и выходной слой, который создает конечный результат и выводит его пользователю. Каждый из нейронов имеет два основных параметра: входные и выходные данные. В случае входного нейрона: входные данные равны выходным. Остальные получают суммарную информацию всех нейронов из предыдущего слоя,

после чего она преобразуется с помощью функции активации и поступает в поле выходных данных.

Нейроны работают с числами в диапазоне от нуля до единицы или от минус единицы до единицы, в зависимости от представления обучающей выборки. Для обработки чисел, выходящих за пределы этого диапазона, необходимо провести процесс нормализации и стандартизации подходящим для исследуемой выборки методом.

Синапс – это связь между двумя нейронами, так называемый регулируемый параметр, который преобразует нейронную сеть в параметризованную систему. Синапс имеет единственный параметр – вес. Благодаря ему входная информация изменяется при её передаче от одного нейрона к другому. Текущее состояние нейрона определяется как взвешенная сумма его входных данных. Выход нейрона – это функция активации его состояния. Характеристика нейрона представлена на рисунке 2.

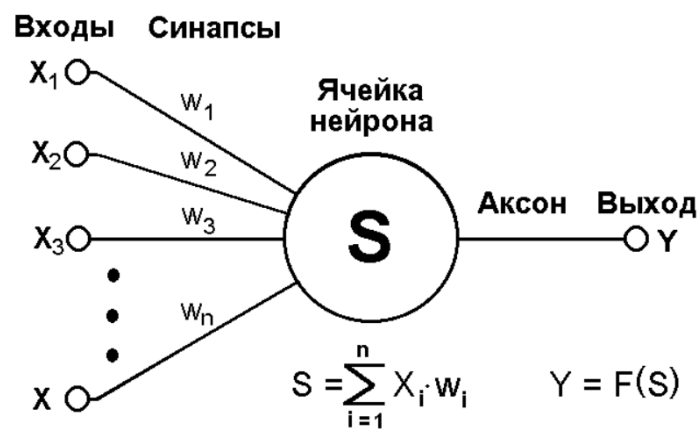


Рисунок 2 – Характеристика ячейки нейрона

На самом деле, совокупность весов нейронной сети или матрица весов – это своеобразный мозг всей системы. Именно благодаря этим весам, входная информация обрабатывается и превращается в результат [3].

1.2 Функция активации

Взвешенная сумма со входов проходит через функцию активации для вывода данных из нейрона. Она определяет выходное значение каждого вычислительного блока в зависимости от результата суммы весов входов и порогового значения. Тип функции активации во многом определяет функциональность нейронной сети и метод её обучения. Существует несколько видов функций активации: линейная, ступенчатая, сигмоидная, тангенциальная, выпрямительная. Выбор того или иного типа зависит от функциональных особенностей: необходимо использовать такую функцию активации, которая лучше аппроксимирует желаемую, это приведет к более быстрому обучению. В данной работе использовались выпрямительная ReLu (1) и сигмоидная (2) функции активации.

Первая является хорошим аппроксиматором, так как любая функция может быть приближена комбинацией ReLu, а также она менее требовательна к вычислительным ресурсам в отличие от гиперболического тангенса, потому что производит простые математические операции, что, очевидно, ускоряет обучение сложной нейронной сети.

$$F(x) = \max(0, x) \quad (1)$$

Сигмоида нелинейна по своей природе и лучше подходит для задач классификации, так как стремится привести значения к одной из сторон кривой, что позволяет находить явные границы при прогнозировании.

$$F(x) = \frac{1}{1+e^{-x}} \quad (2)$$

Лучше всего использовать функцию, которая ускорит процесс обучения и сходимости. Также возможно использовать пользовательскую функцию, которая не принадлежит к определенному типу.

Нейронные сети не программируются в обычном смысле этого слова, они обучаются. Способность к обучению – одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение состоит в нахождении коэффициентов связей между нейронами. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными и выполнять обобщение. Это означает, что в случае успеха сеть сможет вернуть правильный результат на основе данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных [2].

1.3 Проблемы прогнозирования высших спортивных достижений

1.3.1 Оценка возможности прогнозирования индивидуальных спортивных результатов

Прогнозирование понимается как определение свойств или состояния объекта в будущий промежуток времени. Основная характеристика прогнозирования – направленность на предсказание грядущих событий; другая важная особенность – это учет неопределенности, которая обусловлена отсутствием понимания о точном значении, сроках возникновения и способах развития прогнозируемых характеристик и признаков. Суждение о возможном состоянии исследуемого объекта в будущем или путях и сроках его изменения играет важнейшую роль в каждом суждении [3].

Принято считать, что в видах спорта с объективно измеримыми результатами, получаемыми при относительно постоянных внешних

условиях (легкая атлетика, плавание, метание диска и т.д.) это можно сделать очень легко. Но в других видах спорта, например, в гонках на парусниках, биатлоне, в соревнованиях, которые проводятся в изменчивых и трудно сопоставимых условиях, разработка прогноза доставляет определенные трудности, несмотря на то, что в них исход измеряется в метрических единицах. Прогнозирование здесь является не самоцелью, а важным инструментом оптимального планирования и управления подготовкой спортсменов к основным соревнованиям [4].

В гимнастике, прыжках в воду и фигурном катании предсказать результаты проблематичнее из-за оценки сложности программы и точности ее выполнения. Однако прогноз возможен и в таких видах, хотя он будет очень приблизительным и с наибольшей вероятностью неточным. Например, в фигурном катании прогнозирование уровня будущих спортивных достижений осуществляется на основе определения количественных характеристик отдельных программ произвольного катания по показателю средней сложности координации того или иного элемента во время контрольного или соревновательного проката [6]. Менее развита техника прогнозирования в боевых искусствах и спортивных играх за счет огромного количества влияющих на исход параметров. Между тем, в этих видах спорта в последние годы был достигнут определенный успех в прогнозировании результатов выступления спортсмена или команды на крупных соревнованиях [7].

Процесс разработки прогнозов в спорте предполагает совокупность определенных операций, представленных на рисунке 3.



Рисунок 3 – Блок-схема основных операций при разработке прогнозной модели в спорте

Из схемы видно, что последовательность представленных операций направлена на выбор методов и приемов прогнозирования, позволяющих разработать модель объекта прогнозирования. В каждом конкретном случае выбор методов предсказания спортивных результатов зависит от цели прогноза, времени, на которое он разрабатывается, специфики вида спорта, достоверности и полноты исходной информации.

1.3.2 Методы прогнозирования в спорте

М. Кендал [5] пишет, что «осторожный предсказатель (а в прогнозировании необходимо быть осторожным и запоминать гораздо больше ошибок, чем успехов) взвешивает все имеющиеся в его распоряжении методы, подобно тому, как мастер изучает свой набор

инструментов, чтобы использовать тот, который наилучшим образом соответствует данным обстоятельствам».

В литературе существует большое количество схем классификации методов прогнозирования, С. А. Саркисян [7] делит все методы прогнозирования на интуитивные (или экспертные оценки) и формализованные (или фактографические) в зависимости от степени формализации.

В зависимости от общих принципов работы методы интуитивного прогнозирования делятся на две группы: индивидуальные экспертные оценки и коллективные экспертные оценки. С другой стороны, класс формализованных методов делится на группы экстраполяционных, структурно-системных методов, ассоциативных и расширенных информационных методов.

1.3.3 Методы экспертных оценок

В настоящее время методы экспертных оценок широко используются при прогнозировании спортивных результатов и расстановки сил на Олимпийских играх, чемпионатах мира, Европы, надежности выступления спортсменов на ответственных соревнованиях, исхода конкретных матчей и встреч. С их помощью был разработан ряд прогнозов мировых рекордов в нескольких видах спорта. В качестве экспертов привлекаются теоретики, аналитики и практики, имеющие опыт работы в этой области познания.

Существует две формы работы с экспертами: индивидуальная и групповая. При индивидуальной форме прогноз разрабатывается тренером, спортивным аналитиком, независимо от других специалистов. В этом случае для получения прогностической информации используются следующие методы: метод интервью, аналитические записки, написание сценария. При групповой (коллективной) форме работы прогноз проводится специалистами

путем обмена мнениями (устно или письменно, лично или заочно) о прогнозируемом объекте.

Методы экспертных оценок имеют определенные преимущества и недостатки. Пожалуй, наиболее полно положительные и отрицательные аспекты каждого из этих методов, а также вопросы организации и методологии экспертизы, повышения достоверности экспертных оценок по обсуждаемой проблеме рассмотрены в работе В. И. Баландина, Ю.М. Блудова и В. А. Плахтиенко. «прогнозирование в спорте» [9]. Конечно, эксперты могут дать правильную и неправильную оценку. Ниже будет проведен критический анализ рекордных прогнозов, сделанных с привлечением экспертов на примере некоторых видов спорта.

1.3.4 Формализованные методы прогнозирования

Среди формализованных методов, как показывает анализ научной литературы, методы экстраполяции широко используются для прогнозирования спортивных результатов и рекордов в измеряемых видах спорта. Как известно, экстраполяция – это продолжение в будущем тенденции, наблюдавшейся в прошлом. Экстраполяция основана на следующих допущениях:

- развитие явления можно обоснованно охарактеризовать плавной (эволюционной) траекторией – трендом;
- общие условия, определявшие тенденцию развития в прошлом, не претерпят существенных изменений в будущем.

Основой методов экстраполяционного прогнозирования является изучение динамических временных рядов, которые представляют собой значения спортивных результатов, отсортированные по времени. Временной ряд Y_t может быть представлен в следующем виде (3):

$$Y_t = X_t + E_t, \quad (3)$$

где

X_t – детерминированная неслучайная компонента (временной дрейф) процесса роста результатов;

E_t – случайная компонента этого процесса, значения которой в разные моменты времени изменяются независимо друг от друга.

Если детерминированная компонента (тренд) X_t характеризует существующую динамику процесса развития спортивных результатов в целом, то стохастическая компонента E_t отражает случайные колебания или шумы этого процесса. Обе составляющие процесса повышения спортивных достижений обусловлены влиянием определенных факторов, механизм действия которых влияет на поведение результатов с течением времени. Задача прогнозирования состоит в определении типа экстраполяционных функций X_t и E_t на основе исходных эмпирических данных.

Логика подхода к прогнозированию роста данных схематично выглядит следующим образом. На первом этапе проводится анализ эмпирической динамики наилучших результатов в конкретном виде спорта за определенный период времени (скажем, 10-30 лет и более) и строится их временной ряд. На втором этапе, на основе предварительной обработки и преобразования исходных данных, выбирается оптимальный тип функции, описывающей эмпирический ряд. Эти функции можно разделить на две основные группы: линейные и нелинейные. Для экстраполяции роста рекордов в определенных видах спорта предлагается использовать: параболу; экспоненциальную, линейную функцию.

И, наконец, на третьем этапе, используя соответствующие математические уравнения, формулы, вычисляется результат некоторого отдаленного будущего. Прогноз в данном случае заключается в том, что полученная тем или иным способом зависимость «результат-время (годы)» заменяется на интересующие нас годы, и находятся соответствующие значения результата, который принимается за прогноз на этот год. В то же

время, чтобы обосновать прогноз, необходимо доказать, что закон (тенденция), обнаруженный в известном интервале, не изменится за его пределами в определенных границах, что сами параметры качественно изменяться не будут.

1.3.5 Примеры и оценка прогнозирования достижений в легкой атлетике

Сводные данные об ожидаемых результатах в легкой атлетике на 2000 год и их текущем уровне на данный момент приведены на рисунке 4. Видно, что лишь в некоторых случаях теоретические расчеты были подтверждены спортивной практикой. В абсолютном большинстве дисциплин мировые рекорды либо лучше, либо хуже, чем ожидалось. Например, известный американский специалист Е. Jokl [9] разработал свой прогноз, основанный на самых высоких зарегистрированных достижениях с 1 января 1984 года. Он предположил, что в метании копья мировой рекорд составит 103,0 м. Между тем прошло меньше года, и спортсмен из ГДР Уве Хон показал рекордный результат в 104,80 м. Следовательно, Уве Хон отправил снаряд в XXI век 16 лет назад.

Усилия Юрия Седых, который в 1986 году метнул молот на высоту 86,74 м, превзошли рубеж, установленный рядом ученых двадцатого века. Некоторые авторы «запланировали» на 2000 год результат в прыжках в высоту в 6,0 метра, а Сергей Бубка в 1988 году установил рекорд в 6,06 м и за шесть спортивных сезонов увеличил результат до 6,14 м.

Дисциплина	Мировой рекорд на 30.06.2000г.	прогноз мировых рекордов									
		Литературный источник**									
		1	2	3	4	5	6	7	8	9	10
100 м	9,79 М. Грин (США) 16.06.1999	9,55	9,65	9,63	9,47	9,73	9,74	9,45	9,75	9,80	9,70
200 м	1932 М.Джонсон (США) 01.08.1996	18,60	19,42		19,06	19,42	18,96	19,40	19,40	19,50	19,30
400 м	43,18 М.Джонсон (США)	41,80	42,75		42,40	42,65	41,09	43,00	42,70	43,20	42,80
800 м	1,41,11 У.Кипкетер (Дания) 24.08.1997	1,40,0	1,41,00		1,40,00	1,39,10	1,38,40	1,40,0	1,40,0	1,39,00	1,39,5
1500 м	3,26,00 И. Герруж (Мар.) 14.07.1998	3,23,0	3,29,2	3,27,0	3,28,9	3,28,00	3,26,10	3,20,0	3,28,0	3,27,00	3,26,0
5000 м	12,39,36 Х.Гебреселаси (Эфиопия) 13.06.1998	12,15	12,45,1		12,49,0	12,40,0	12,51,4	12,30,0	12,45,0	12,48,00	12,45,
10000 м	26,22,75 Х. Гебреселаси (Эфиопия) 1.06.1998	26,00	26,38,0	26,23	26,14,0	26,30,0	26,34,0	27,00,0	26,55,0	26,50,00	26,40,
110 м с/б	12,91 К.Джексон (Вбр.) 20.08.1993	12,30	12,63		12,66	12,80		12,70	12,80	12,70	12,80
400 м с/б	46,78 К. Янг (США) 6.08.1992	46,20	45,88		44,89	44,90		46,50	45,50	46,30	46,40
3000 м с/пр	7,55,72 Б. Бармасан (Кения) 24.08.1997	7,32	8,02,00	7,44	7,42,00	7,41,00	7,52,40	7,55,0	7,50,0	7,55,00	7,50,0
Марафон*	2:05,42 Х.Ханнучи (Мар.) 24.10.1999					-	2:04,20				
4х100м	37,40 (США) 8.08.1992						37,08			37,00	
4х400 м	2.54,20 (США) 22.07.1998						2.55,00			2,54,50	
Высота	2,45 Х. Сотомайор (Куба) 27.07.1993	2,54	2,33	2,49		2,48	2,45	2,45	2,50	2,52	2,54
Шест	6,14 С. Бубка (Укр.) 31.07.1994	6,40	5,91	6,88	6,03	6,20		5,85	6,00	6,35	6,25
Тройной	18,29 Д. Эдварде (Вбр.) 7.08.1995	18,60	18,14	18,77	18,65	18,41		18,25	18,40	18,40	18,50
Длина	8,95 М.Пауэлл (США) 30.08.1991	9,15	8,68		8,80	9,15	9,09	9,00	9,00	9,10	9,0
Ядро	23,12 Р. Варне (США) 20.05.1990	25,25	22,78	26,38	24,67	24,36	24,80	23,50	23,10	23,40	24,00
Диск	74,08 Ю.Шульт (ГДР) 6.06.1986	84,0	72,58	84,40	78,55	78,90	75,90	77,00	77,60	76,0	75,00
Молот	86,74 Ю. Седых (СССР) 10.08.1986	90,0	81,24	93,30	94,82	94,04	85,61	85,00	88,00	95,00	91,00
Копье	104,80 У. Хон (ГДР) 20.06.1984; 98,48 Я. Железны (Чехия) 25.05.1990	110,0	100,02	103,0	110,00	110,12	103,00	100,0	101,00	105,00	92,00
Десяти-борье	8994 Т. Дворжак (Чехия) 4.07.1999	-				-				9400	

Рисунок 4 – Прогноз мировых рекордов и фактические абсолютные спортивные достижения в легкой атлетике у мужчин на 2000 год

** 1. V. Visek (1970). 2. Я. Брогли, К. Наков (1973). 3. Z. Wazny (1975). 4. P. Tillinger, V. Visek (1977). 5. Е. Разумовский (1983). 6. Е. Jokl (1984). 7. P. Loffer, W. Lohmann (1986). 8. P. Loffer, W. Lohmann (1986). 9. Н.Г. Озолин, Л.С. Хоменков 1987). 10. V. Visek (1990).

По расчетам многих экспертов, результат в беге на 5000 м за 13-16 лет должен был улучшиться на 10-11 с, а на 10000 м – на 20-40 с. Однако выдающийся спортсмен из Эфиопии Хайле Гебреселаси за пять спортивных сезонов, с 1994 по 1998 гг., в беге на 5000 метров отстал от своего лучшего результата на 17,6 секунды и установил мировой рекорд, равный 12, 39, 36. В беге на 10 000 м он трижды устанавливал мировые рекорды – в 1995, 1997 и 1998 годах – и отстал примерно на 21 секунду от своего первого мирового рекорда на этой дистанции.

Почему истинные рекорды превышают ожидания? Основная причина заключается в действии так называемых ускоряющих факторов. Это, прежде всего, появление новых, более современных тренажеров и методик, совершенствование системы отбора спортивных талантов, увеличение числа спортсменов высокой квалификации и оптимизация системы питания спортсменов, улучшение условий жизни и материально-технической базы, рост престижа ряда спортивных дисциплин и повышение конкурентоспособности спортсменов.

Большим недостатком многих прогнозов является то, что невозможно предсказать все спортивные моменты. Прогнозирующие системы, к примеру, не способны учитывать внезапные изменения в результатах, которые, вероятно, произойдут в ближайшем или отдаленном будущем. Действительно, кто сегодня может с полной уверенностью сказать, каковы революционные изменения произошли в 2010 или 2020 годах в спортивном оборудовании, методах тренировок и конкретных видах спорта? Какие высококачественные материалы используются при строительстве спортивных сооружений, производстве спортивного инвентаря и оборудования для спортсменов и самое главное, как они влияют на улучшение результата конкретного спортсмена?

Кроме того, предположения о том, как будет выглядеть лучший мировой результат, не могут основываться исключительно на оценках существующих или лучших результатов сезона. Поскольку первоклассные

рекорды в определенной степени отражают роль выдающейся личности в спорте, как специалист осмелился бы с уверенностью назначить спортсмена, способного показать ожидаемый результат? Ведь многие индивидуальные особенности спортсменов будущим тренерам и ученым сегодня неизвестны.

Таким образом, на основе материалов, рассмотренных в этой главе, можно сделать следующие выводы:

1) прогнозирование крупных спортивных достижений в спорте с помощью метрической измерительной системы, несмотря на ее визуальную простоту, является очень сложной и недостаточно изученной областью даже при условии наличия большого количества публикаций, посвященных этому вопросу. Причиной этого является высокая неопределенность в динамике результатов совершенствования спортивных дисциплин и обусловленность их развития множеством факторов, влияние которых не всегда поддается выявлению и формализации;

2) оценивая в целом позитивные прогнозы последних трех десятилетий в отношении мировых результатов победителей и призеров Олимпийских игр, мы должны отметить, что в большинстве случаев они оказались нереалистичными. Их авторам не удалось избежать серьезных ошибок и просчетов в прогнозировании спортивных исходов в будущем;

3) ретроспективный обзор сделанных нами прогнозов показывает, что их точность, т.е. степень соответствия прогнозируемых и фактических результатов, зависит от особенностей спортивной дисциплины и характера динамики мирового роста и олимпийских рекордов (прямолинейного, скачкообразного, изогнутого), пола спортсменов, продолжительности исторического периода, на основе которого делается прогноз, и наличия информации об ускоряющих и сдерживающих факторах, влияющих на развитие области высших спортивных достижений.

1.4 Применение нейронных сетей для прогнозирования спортивных состязаний

1.4.1 Основные этапы построения нейросетевой прогнозирующей системы

Нейронные сети – это очень мощный и гибкий механизм прогнозирования. При определении того, что прогнозировать, необходимо указать переменные, которые анализируются и могут влиять на будущее состояние объекта прогнозирования. На используемый уровень детализации влияют многие факторы: доступность и точность данных, стоимость анализа и пользовательские предпочтения в отношении результатов прогнозирования. В ситуациях, когда наилучший набор переменных неясен, вы можете попробовать различные альтернативы и выбрать один из вариантов, который дает наилучшие результаты. Обычно именно так делается выбор при разработке прогностических систем на основе анализа исторических данных.

Вторым важным этапом построения нейросетевой прогностической системы является определение следующих трех параметров:

- прогнозируемый период;
- горизонт прогнозирования;
- интервал прогнозирования.

Прогнозируемый период – это основная единица времени, на которую составляется прогноз. Горизонт прогноза – это количество периодов в будущем, которые охватывает прогноз. Это означает, что вам может понадобиться прогноз на 10 дней вперед с данными за каждый день. В данном случае период составляет один день, а горизонт – 10 дней. Наконец, интервал предсказания – это частота, с которой выполняется новое предсказание. Интервал прогнозирования часто совпадает с прогнозируемым периодом.

Выбор периода и горизонта прогнозирования обычно диктуется условиями принятия решений в той области, для которой составляется прогноз. Выбор этих двух параметров – едва ли не самое сложное в прогнозировании работы нейронной сети. Для того чтобы прогноз имел смысл, горизонт прогноза должен быть не меньше времени, необходимого для реализации решения, принятого на основе прогноза. Таким образом, прогноз очень сильно зависит от характера принимаемого решения. В некоторых случаях время, необходимое для внедрения решения, не определено, например, как в случае поставки запасных частей для пополнения запасов ремонтных компаний. Существуют методы работы в условиях такой неопределенности, но они увеличивают вариацию ошибки прогнозирования. Поскольку точность прогнозирования обычно снижается с увеличением горизонта прогнозирования, часто можно улучшить процесс принятия решений за счет сокращения времени, необходимого для реализации решения, и, следовательно, уменьшения горизонта прогнозирования и погрешности.

В некоторых случаях прогнозирование конкретных значений прогнозируемой переменной не так важно, как прогнозирование существенных изменений в ее поведении. Такая задача может возникнуть в случае, когда текущее направление движения рынка изменит свое направление на противоположное.

Точность прогноза, требуемая для решения конкретной задачи, оказывает существенное влияние на систему прогнозирования. Обучающая выборка также оказывает большое влияние на прогноз [9].

1.4.2 Примеры прогнозирования спортивных матчей

Принято считать, что групповые виды спорта, такие как баскетбол, футбол, хоккей и другие, предсказать гораздо сложнее, чем отдельные разновидности состязаний, к которым можно отнести бадминтон, теннис,

бокс и т.д. Командные виды спорта включают в себя больше факторов, влияющих на исход, которые необходимо учитывать, что усложняет их статистический анализ. По этой причине данная сфера является важной областью применения искусственного интеллекта. К числу основных факторов, влияющих на исход игры в теннис, относятся: удача, психология, тип покрытия, судейство, сила игрока и его прошлые результаты и другие. Основная проблема заключается в определении степени воздействия того или иного фактора на исход состязания.

Оценка спортивных результатов в командных видах спорта с точки зрения основных направлений и двигательных шаблонов, стратегии и тактики является предметом уже хорошо развитой дисциплины – символического (нотационного) анализа в спорте. При анализе стратегии и тактики используются различные подходы. Многие авторы предлагают эффективные методы обнаружения сходных перемещений в потоках позиционных данных, чтобы обеспечить основу для анализа часто встречающихся перемещений и тактических схем. Например, они используют записи матчей, которые немецкая Бундеслига проводит с помощью специальных камер с высоты птичьего полета, открывающих вид на целое поле, чтобы лучше анализировать движения игроков и игровую тактику. 90-минутная запись футбольного матча преобразуется в поток позиционных данных. При стандартной скорости записи 25 кадров в секунду совпадение представляет собой последовательность из 135 000 снимков. Каждый удар состоит из 23 позиций (22 позиции игроков двух команд и положение мяча). В целом, игра описывается более чем тремя миллионами координат. Чтобы эффективно обрабатывать такой большой объем данных, исследователи предлагают алгоритм, использующий методы Angle/Arc-Lengths, Dynamic Time Warping Locality Sensitive Hashing [10].

Бедфорд вместе с единомышленниками разработали три уровня стратегии в качестве возможных вариантов повышения эффективности бадминтона. Третья, более продвинутая стратегия использует байесовский

анализ для обновления первоначальных оценок на основе текущей информации о развитии матча и оптимизации оценки рисков при выполнении начальных ударов. Эти стратегии могут быть реализованы в живой игре, поскольку правила игры в бадминтон допускает вмешательство тренера во время состязания.

Глокнер предлагает использовать эвристические методы распознавания при разработке стратегий для спортивных игр. В статье рассматриваются закономерности выбора игрока в гандболе, основанные на «поведении взгляда». Согласно лабораторному эксперименту, авторы протестировали два класса моделей задач принятия решений: модель нейронной сети с параллельным ограничением удовольствия (Parallel constraint satisfaction – PCS) и модель батареи. Оба класса моделей могут быть реализованы как детерминированные, так и вероятностные. Модели предсказывают действия в задаче выбора с точки зрения гандболиста – отдать мяч другому игроку или забросить его в ворота. Авторы использовали данные о поведении зрителей и вариант, созданный с помощью 74 участников (гандболистов), набранных из Государственного тренировочного центра и клубов на севере Германии. Задача участников состояла в том, чтобы перечислить возможные варианты в реальных ситуациях игры в гандбол. Используемый метод подробно описан в работе. В целом, данные включали информацию о продолжительности фиксации взгляда плэймейкера на левой, средней или правой стороне игрового поля и предположения других участников о направлении действий плэймейкера – то, что должно быть предсказано моделями. Результаты показали, что оба класса моделей одинаково хорошо предсказывают изначально сгенерированные варианты участников на основе данных о поведении при просмотре. Таким образом, фиксация зрения имеет прогностическое значение. Авторы приходят к выводу: сетевые модели могут быть успешно применены при принятии решений экспертами в области спорта и предлагают использовать полученные результаты для разработки тренировочных программ при подготовке спортсменов [10].

Группа исследователей в команде с Пфайффером отмечают, что использование нелинейных методов, в частности нейронных сетей, для анализа данных в области соревновательных видов спорта представляется особенно перспективным. На примере трех конкретных задач они показали, как сетевые подходы могут быть успешно применены для решения сложных задач в области спортивной науки. Первое исследование было проведено в области выявления талантов, где самоорганизующаяся карта признаков Кохонена (Self-Organizing Kohonen Feature Map) использовалась для выявления различных паттернов характеристик качеств подростков, затем успешных или неуспешных пловцов. Во втором примере динамически управляемая сеть (Dynamically Controlled Network – DyCoN) – дальнейшее развитие самоорганизующейся карты Кохонена – использовалась для определения тактического поведения в гандбольной команде. Наконец, искусственная нейронная сеть, многослойный персептрон (Multi-Layer-Perceptron), была использована для прогнозирования соревнований элитных пловцов на Олимпийских играх 2004 года в Афинах на основе характеристик их тренировочных нагрузок.

Гранц использовал динамически управляемую сеть (Dynamically Controlled Network – DyCoN) для выявления тактических закономерностей в данных о положении футболистов. Классификация с использованием сети была сопоставлена с результатами, полученными экспертами при классификации вручную. Показано, что выбранная архитектура нейронной сети в основном способна обнаруживать категории тактических моделей. Предварительные результаты показали высокую точность классификации. Две нейронные сети – многослойные персептроны (MLP) были использованы для классификации тактического поведения волейбольных команд в зависимости от формы их защитных позиций. Исследователи пришли к выводу, что схемы защиты в командных видах спорта сугубо индивидуальны и различаются даже в стандартных ситуациях. Искусственные нейронные сети могут быть использованы для распознавания команд на основе формы

схемы расположения игрока. В работе используется байесовская сеть для выявления взаимосвязей между 22 психологическими характеристиками игроков в полупрофессиональном футболе и их влиянием на эффективность коллективной игры [10].

2 Этапы разработки модели нейронной сети для прогнозирования спортивных состязаний

Теннис – сложный вид спорта, в котором на исход матчей может влиять множество факторов, включая уровень мастерства игроков, их физическое состояние, предыдущие результаты в личных встречах и расписание турниров. Однако благодаря достижениям в области машинного обучения и искусственного интеллекта стало возможным разрабатывать прогностические модели, которые могут помочь любителям тенниса и игрокам, делающим ставки, принимать обоснованные решения о потенциальных победителях матчей.

Решение же конкретной задачи прогнозирования исхода матча можно представить в виде последовательности следующих этапов:

- сбор начальных данных об имевших место спортивных событиях того же вида спорта и за определенный период;
- формирование обучающей выборки и ее структуризация под требования используемого программного обеспечения;
- нормализация данных;
- выбор архитектуры нейронной сети для решения задачи;
- обучение нейронной сети;
- настройка алгоритма обучения;
- проверка модели;
- практическое использование обученной нейронной сети для прогнозирования вероятностей возможных исходов [11].

2.1 Сбор исходных данных

Первым шагом в разработке нейросетевой модели для прогнозирования теннисных матчей является сбор данных. Теннисные данные могут поступать

из различных источников, включая официальные источники, такие как веб-сайты АТР, WTA и турниров Большого шлема, сторонних поставщиков, таких как Tennis Abstract и Tennis Explorer, и платформы социальных сетей. Данные должны быть всеобъемлющими и точными, охватывающими различные аспекты игроков, турниров и матчей.

Наличие большого количества исходных данных для обучения искусственной нейронной сети является важным фактором, определяющим возможность эффективного решения поставленной задачи прогнозирования [4]. Однако, как показывают проведенные нами исследования, важным показателем является не только объем, но и происхождение данных. Немаловажным является качественный состав данных в обучающих векторах. Чем больше различных факторов будет учтено, тем глубже можно выполнить кластеризацию, вплоть до получения очень неожиданных, но реально работающих прогнозов. Для прогнозирования исходов теннисных матчей использовались следующие параметры:

- результаты предыдущих противостояний;
- скорость игры;
- ранг и рейтинг противостоящих игроков;
- коэффициенты выигрышных и проигрышных партий;
- физические показатели игроков;
- коэффициенты решающий и затяжных партий;
- процент побед;
- коэффициент партий с победой в дополнительном тайме.

Отдельного обсуждения требуют факторы, которые трудно оценить количественно:

- резкие изменения состояния: травмы, потеря физической формы, смена тренера;
- психологические факторы, которые могут быть связаны с турнирной ситуацией, потерей ведущего игрока или капитана, статусом

турнира (если рассматривать матчи не только основной лиги), предыдущими матчами (как победными, так и проигранными).

Теоретически учитывать их, хоть и сложно, но можно. Однако на практике самым сложным окажется собрать и представить их в обучающей выборке. Именно по этой причине есть смысл учитывать их уже после полученного решения на заключительной стадии, корректируя вероятности исходов на основе собственного опыта прогнозиста.

Собранные данные должны быть надлежащим образом отформатированы и очищены, чтобы удалить любую неуместную или ошибочную информацию. Это также может потребовать преобразования данных в структурированный формат, такой как электронная таблица или база данных. Кроме того, может потребоваться определить недостающие точки данных и решить, как с ними обращаться (например, заполнить их на основе среднего значения или вообще удалить точку данных). Наконец, крайне важно удалить все дубликаты и подготовить набор данных для извлечения признаков – следующего шага в разработке модели нейронной сети. Поэтому, сбор данных – это всеобъемлющий и важный процесс, который требует большого внимания и тщательности для обеспечения качества, полноты и надежности данных, необходимых для разработки мощной модели прогнозирования тенниса.

2.2 Формирование и структуризация обучающей выборки

Процесс формирования обучающей выборки заключается в выборе правильных входных переменных, также известных как характеристики, из собранных данных, которые будут использоваться для обучения модели нейронной сети. Извлечение самых влиятельных характеристик из предложенного набора данных важен, поскольку они оказывают воздействие на точность модели, вычислительную эффективность и обобщение.

В теннисе важные функции для прогнозирования игроков включают такие исторические данные, как результаты встреч, важность турнира, а также рейтинг, возраст, стиль игры и форма каждого спортсмена. Другие характеристики могут включать тип поверхности, местоположение, расстояние в пути и погодные условия. Личные рекорды игроков – еще один важный аспект, принимаемый во внимание при прогнозировании вероятного исхода матча. Расписание турнира, в котором матчи запланированы последовательно, также является важной особенностью.

Структурирование выборки может потребовать использования статистических методов или методов машинного обучения, таких как анализ основных компонентов (РСА), корреляционный анализ или деревья принятия решений. Эти методы помогут определить важность каждой функции при построении точной модели. Кроме того, разработка функций, создание новых функций, производных от других функций, может быть использовано для максимального повышения производительности модели.

Таким образом, этап извлечения признаков, включающий в себя выбор соответствующих признаков, которые точно и лаконично описывают данные, играет существенную роль в прогнозировании спортивных состязаний. Этот шаг требует знания теннисных матчей и методов машинного обучения, чтобы определить параметры, которые внесут наиболее ценный вклад в обучение нейронной сети.

2.3 Нормализация данных

Следующий шаг включает в себя нормализацию данных, что означает стандартизацию значений параметров набора данных. Этот шаг необходим перед обучением модели, поскольку он гарантирует, что входные данные обрабатываются с одинаковой скоростью, и устраняет любые искажения, вызванные изменениями в диапазоне характеристик.

Нормализация обычно достигается путем применения метода масштабирования к извлеченным объектам в наборе данных. Вероятность каждого объекта уменьшается таким образом, чтобы все входные значения попадали в один и тот же интервал, чтобы устранить расхождения в значениях атрибутов. Наиболее распространенные методы нормализации включают масштабирование от минимума к максимуму и нормализацию по Z-оценке. Минимальное-максимальное масштабирование изменяет данные от нуля до единицы, в то время как нормализация Z-оценки стандартизирует их таким образом, что среднее значение равно нулю, а стандартное отклонение равно единице.

Выбор правильного метода нормализации имеет важное значение для производительности нейронной сети. Объекты должны быть нормализованы таким образом, чтобы все они имели равное среднее значение и дисперсию и были примерно распределены по Гауссу. Нейронная сеть будет функционировать более эффективно, когда входные данные будут иметь аналогичный масштаб. Этот шаг гарантирует, что обученная модель не придаст больший вес входным данным просто потому, что им присвоено большее значение, что приведет к расхождениям в результатах модели.

Нормализация данных является одним из важных шагов в построении прогнозирующей модели нейронной сети для теннисных матчей, поскольку она обеспечивает системный подход к масштабированию всех входных характеристик и уменьшению предвзятости. Нормализация делает диапазон входных данных схожим, что снижает риск доминирования или игнорирования функций на этапе обучения. Нормализация в сочетании с извлечением признаков создает надежную модель, которая учитывает все аспекты данных.

2.4 Выбор подходящей архитектуры нейронной сети

Четвертым этапом, связанным с созданием нейросетевой модели для прогнозирования теннисных матчей, является выбор подходящей архитектуры нейронной сети. Архитектура нейронной сети является основой всей модели, так как она определяет структуру нейронной сети и способ обработки входных данных для получения прогнозируемого результата. Этот шаг включает в себя выбор количества и типов слоев, числа нейронов и функций активации, участвующих в преобразовании каждого из них.

Выбранная архитектура нейронной сети должна отражать сложность решаемой задачи. Простая архитектура с несколькими уровнями может подойти для данных с небольшим количеством переменных или неглубокими прогнозами, в то время как при работе с большими наборами данных и многоступенчатой обработкой может потребоваться более сложная структура. Модель должна быть спроектирована таким образом, чтобы свести к минимуму переоснащение сети или ее недостаточное оснащение при максимальном увеличении мощности обобщения.

Существуют различные типы архитектур нейронных сетей, включающих нейронные сети прямого действия, сверточные нейронные сети и рекуррентные нейронные сети. В моделях прогнозирования тенниса часто используются нейронные сети прямого действия, которые обучаются путем распространения входных данных вперед по скрытым слоям. Количество скрытых слоев в типичной нейронной сети прямого действия варьируется от одного до нескольких слоев с десятками или сотнями узлов в каждом слое. Функция активации, применяемая к каждому нейрону, определяет выходное значение этого вычислительного блока, на которое влияют входные значения и вес соединения с нейроном.

Выбор функций активации также влияет на производительность модели, при этом обычно используются функции sigmoid, rectified linear unit (ReLU) и softmax. Сигмоидная функция отображает входные значения в

диапазоне от 0 до 1, в то время как ReLU отображает входные значения от нуля или ниже до нуля, а значения выше нуля остаются неизменными. Softmax используется на выходном уровне и преобразует выходные значения нейронной сети в вероятности.

Выбор подходящей архитектуры нейронной сети имеет решающее значение для оптимальной производительности и желаемых результатов. Архитектура нейронной сети должна основываться на знании предметной области рассматриваемой проблемы, размере и качестве набора данных, а также доступных вычислительных ресурсах. Правильный выбор сетевой архитектуры обеспечит обобщение и масштабируемость модели, что приведет к точному и надежному прогнозированию теннисных матчей.

2.5 Обучение модели

Обучение модели – следующий важный этап разработки нейросетевой модели. После выбора архитектуры нейронной сети и нормализации данных мы теперь готовы к обучению модели. Цель этого шага состоит в том, чтобы скорректировать веса и смещения нейронной сети путем многократного представления входных данных в сеть и сравнения прогнозируемого результата с фактическим результатом.

Процесс обучения включает в себя следующие этапы:

- инициализация весов и смещений: веса смещения нейронной сети должны быть инициализированы случайным образом. Это важно, потому что начало работы с нулевыми весами может привести к нарушению симметрии и снижению производительности;
- прямое распространение: на этом этапе нейронная сеть получает входные данные и обрабатывает их, умножая входные данные на веса, а затем добавляя компенсации. Результирующий вывод передается на следующий уровень путем применения функции активации;

- обратное распространение: на этом шаге вычисляется ошибка между прогнозируемым результатом и фактическим результатом. Затем ошибка передается обратно через слои для корректировки весов и компенсаций нейронной сети;

- обновление весов и смещений: веса и смещения корректируются в процессе обратного распространения, чтобы свести к минимуму разницу между прогнозируемым результатом и фактическим результатом. Степень обучения определяет степень, в которой корректируются значения веса и компенсации.

Процесс обучения должен повторяться много раз или эпох, чтобы нейронная сеть могла изучить корреляцию между входными характеристиками и предполагаемым результатом. Чем больше обучается нейронная сеть, тем ниже становится частота ошибок во время обучения и тем выше точность модели.

Показатели производительности, такие как ошибка обучения, ошибка проверки и ошибка тестирования, используются для отслеживания прогресса модели. Эти показатели используются для определения того, является ли модель неадекватной, завышенной или хорошо обобщенной. Процесс обучения продолжается до тех пор, пока модель не достигнет приемлемого уровня ошибок и наилучшей возможной производительности обобщения.

Таким образом, обучение модели является таким же неотъемлемым шагом в разработке нейросетевой модели для прогнозирования теннисных матчей. Именно в процессе обучения нейронная сеть учится распознавать закономерности во входящих данных и корректировать веса и компенсации, чтобы точно прогнозировать будущие результаты. Этот процесс необходимо повторять много раз с сертификацией и тестированием, чтобы получить наиболее эффективную модель.

2.6 Настройка алгоритма обучения

Этот и следующие шаги обычно повторяются много раз, но в строгом порядке, пока прогнозист не будет доволен результатом или пока он не потеряет терпение.

Настройка алгоритма включает в себя прямой или косвенный контроль размера сети и скорости обучения, манипуляция с функциями активации в определенных слоях и, возможно, в зависимости от выбранной архитектуры нейронной сети, ряда других параметров. Если сеть с фиксированной структурой является самоучкой (сеть Кохонена или основана на радиальных функциях), то мы говорим о показателе размера сети (количества кластеров). В результате некоторые кластеры могут быть вообще не задействованы, а некоторые могут быть переполнены. Неиспользуемые кластеры только мешают процессу обучения в сети, поскольку замедляют работу алгоритма. Переполненные, в свою очередь, гораздо более опасны, поскольку их присутствие означает, что нейронная сеть не может должным образом разделить обучающие примеры по характеристикам – данные становятся средними, а точность прогнозирования для примеров, попадающих в такие кластеры, очень мала.

Если нейронная сеть не только обучается без учителя, но и самоорганизована, что возможно для всех ранее перечисленных архитектур, то в итоге пустых кластеров просто не будет, а избыточными группами в принципе можно управлять, разбивая их в процессе обучения на более мелкие по количеству включенных в них примеров. Обычно порог расстояния от обучающего примера до центра ближайшего кластера указывается в качестве одной из основных настроек алгоритма. При превышении этого порога новый класс с одним центром (весовым коэффициентом) создается в соответствии с реализованным алгоритмом, что соответствует текущему обучающему примеру. Однако для регулирования размера кластера и решения проблемы переплаты можно задать другой

параметр – максимальное количество примеров, образующих данную группу. Когда выясняется, что новый пример назначен определенному классу, для которого предел примеров уже исчерпан, принимается решение о создании нового кластера, центр которого соответствует примеру, наиболее удаленному от центра существующего.

С одной стороны, выбор другой настройки – коэффициента скорости обучения (коэффициента адаптации) влияет на скорость обучения всей сети, а с другой – на миграцию примеров между близлежащими кластерами. Желательно поддерживать баланс, чтобы он не был слишком большим или слишком маленьким. Часто для решения проблем кластеризации используется переменный, постепенно снижающийся коэффициент скорости.

2.7 Проверка разработанной модели

Седьмым шагом является валидация модели. Валидация модели – это процесс проверки точности и надежности выходных данных обученной модели. Оценивая производительность модели по сравнению с ранее невидимыми данными, мы оцениваем способность модели обобщать и делать точные прогнозы.

Процесс валидации включает в себя тестирование модели нейронной сети на тестовом наборе данных, который никогда не был замечен на этапе обучения модели. Тестовый набор данных должен быть репрезентативным для входных данных и содержать различные входные значения для функций, чтобы оценить, насколько хорошо работает нейронная сеть. Тестовый набор данных также должен предоставлять целевые выходные значения для прогнозирования выходных данных нейронной сети, чтобы сравнить фактические значения с прогнозируемыми.

Для оценки точности модели можно использовать такие показатели производительности, как точность, прецизионность, отзывчивость и f1-оценка. Также может быть сгенерирована матрица путаницы для

определения уровня точности нейронной сети в каждом классе или исходе, где истинными предсказаниями являются диагональные ячейки. Кроме того, ROC-кривые и результаты AUC могут быть использованы для оценки способности модели различать классы, где значение AUC, равное 1, представляет собой идеальную классификацию.

Показатели эффективности модели сравниваются с показателями точности модели на этапе обучения, чтобы убедиться в способности модели к обобщению. Если нейронная сеть хорошо работает как с обучающим, так и с тестовым наборами данных, то ее можно развернуть для составления прогнозов по целевому набору данных.

В процессе валидации мы завершаем работу над показателями эффективности модели и оцениваем, может ли модель прогнозировать целевые данные без ошибок. Хорошо работающая модель нейронной сети не только предсказывает исходы теннисных матчей с высокой точностью, но и гарантирует, что модель сможет адаптироваться к новым данным.

2.8 Практическое использование нейронной сети для прогнозирования вероятностей исходов матчей

Заключительный шаг включает в себя использование обученной модели для составления прогнозов на основе новых данных, таких как предстоящие теннисные матчи.

Расширенная модель может быть использована различными способами для прогнозирования исходов матчей и оказания помощи игрокам, занимающимся ставками на теннис. Например, прогнозы модели могут быть представлены в виде процентной вероятности того, что игрок выиграет матч, или определять его шансы на победу. Предсказания также могут быть использованы для создания стратегии ставок, например, для размещения ставок на матчи, вероятности которых, спрогнозированные нейросетью, выше коэффициентов, установленных рынком. Варианты использования

установленной модели зависят от предполагаемого применения и целей заинтересованных сторон.

Стоит отметить, что набор данных для обучения модели нейронной сети и последующий набор данных для проверки будут отличаться от данных турнира, и поэтому некоторые расхождения могут повлиять на результаты. Эффективность модели может постоянно оцениваться, и модель может быть переобучена путем включения новых данных и обновления параметров модели. Такое непрерывное переобучение модели улучшит ее производительность и устраним деградацию.

В заключение, описанные выше этапы являются надежным руководством по разработке нейросетевой модели для прогнозирования теннисных матчей. Нейронная сеть потенциально может помочь игрокам, делающим ставки на теннис, принимать обоснованные решения в этом высококонкурентном и неопределенном виде спорта. Кроме того, по мере того, как методы и алгоритмы глубокого обучения развиваются и становятся более мощными, производительность нейронной сети со временем может быть адаптирована и улучшена. Важно отметить, что прогнозы модели являются не абсолютными, а скорее вероятностными результатами, основанными на прошлых выступлениях теннисистов. Следовательно, её следует использовать как вспомогательный инструмент, а не как единственный фактор при принятии решений [12].

3 Построение модели нейронной сети для прогнозирования исходов теннисных матчей

3.1 Структуризация показателей теннисистов и формирование обучающей выборки

Исходные данные для обучения нейронной сети собирались с веб-ресурса <https://www.ultimatetennisstatistics.com>. На рисунке 5 представлен фрагмент статистики теннисных матчей турнира Australian Open, на котором отражаются противостоящие игроки, покрытие поля, счет и результат матча.

Rank	Player	Opponent	Date	Tournament	Surface	Round	W/L	Score
1	 Ivo Karlovic	 Horacio Zeballos 🏆	16-01-2017	Australian Open	Hard	R128	W	6-7 ⁽⁶⁾ 3-6 7-5 6-2 22-20
2	 Reilly Opelka 🏆	 Thomas Fabbiano	14-01-2019	Australian Open	Hard	R64	L	6-7 ⁽¹⁵⁾ 6-2 6-4 3-6 7-6 ⁽⁵⁾
3	 Ivo Karlovic	 Kei Nishikori	14-01-2019	Australian Open	Hard	R64	L	6-3 7-6 ⁽⁶⁾ 5-7 5-7 7-6 ⁽⁷⁾
4	 Ivo Karlovic	 Yuichi Sugita	15-01-2018	Australian Open	Hard	R64	W	7-6 ⁽³⁾ 6-7 ⁽³⁾ 7-5 4-6 12-10
5	 Ivo Karlovic	 Andreas Seppi 🏆	15-01-2018	Australian Open	Hard	R32	L	6-3 7-6 ⁽⁴⁾ 6-7 ⁽³⁾ 6-7 ⁽⁵⁾ 9-7
6	 Joachim Johansson	 Andre Agassi	17-01-2005	Australian Open	Hard	R16	L	6-7 ⁽⁴⁾ 7-6 ⁽⁵⁾ 7-6 ⁽³⁾ 6-4
7	 Ivo Karlovic	 Ivan Dodig	17-01-2011	Australian Open	Hard	R128	L	6-4 3-6 6-7 ⁽⁵⁾ 6-4 6-4
8	 John Isner 🏆	 Reilly Opelka 🏆	14-01-2019	Australian Open	Hard	R128	L	7-6 ⁽⁴⁾ 7-6 ⁽⁶⁾ 6-7 ⁽⁴⁾ 7-6 ⁽⁵⁾

Рисунок 5 – Статистика теннисных матчей

Помимо общих показателей состязаний также использовалась информация по каждому игроку. На рисунке 6 приведена таблица, включающая такие показатели игроков как ранг и рейтинг, число победных и потерянных очков, коэффициенты решающий и затяжных партий, коэффициент партий с победой в дополнительном тайме.

Rank	Player	Rating▼	Pts. Won	Pts. Lost	Dec. S.	5-th S.	Finals	Tie-Br.	DS TB
1	 Novak Djokovic 🏆	2.340	990	423	72.3%	78.3%	70.2%	65.4%	71.9%
2	 Bjorn Borg	2.278	540	237	73.8%	81.8%	71.7%	54.7%	70.0%
3	 Carlos Alcaraz Garfia 🏆	2.254	151	67	68.8%	88.9%	76.9%	60.0%	87.5%
4	 Rod Laver	2.152	568	264	69.3%	73.5%	72.7%	56.5%	55.6%
5	 Kei Nishikori	2.031	522	257	72.4%	78.8%	46.2%	59.1%	65.2%
6	 Pete Sampras	2.000	930	465	68.2%	68.8%	72.7%	62.8%	63.8%
7	 John McEnroe	1.995	732	367	68.8%	65.8%	71.0%	61.4%	52.4%

Рисунок 6 – Индивидуальные показатели теннисистов

На рисунках 7 и 8 показана таблица теннисистов, отсортированных по рейтингу, которая может дополнить предыдущие данные с рисунка 6 следующими параметрами: общее число очков, число турнирных, ранговых очков, рейтинговые очки на пике карьеры, процент побед, а также количество очков различных турнирных состязаний.

Rank	Name	Pts ▼	Tourn P	Rank P	Ach P	GS
1	 Novak Djokovic 🏆	1024	562	226	236	22
2	 Roger Federer	927	566	200	161	20
3	 Rafael Nadal 🏆	891	526	186	179	22
4	 Jimmy Connors	616	402	151	63	8
5	 Ivan Lendl	605	395	141	69	8
6	 Pete Sampras	522	316	140	66	14
7	 John McEnroe	512	321	128	63	7

Рисунок 7 – Турнирные показатели теннисистов (часть 1)

TF	AF	M	O	BT	T	W@1	W%	Elo
6	0	38	-	66	93	387	83.5%	2629
6	0	28	-	54	103	310	81.9%	2550
0	0	36	1	59	92	209	83.0%	2552
1	2	20	-	31	109	268	81.7%	2521
5	2	18	-	33	94	270	81.5%	2518
5	2	11	-	32	64	286	77.4%	2407
3	5	17	-	32	77	170	81.6%	2583

Рисунок 8 – Турнирные показатели теннисистов (часть 2)

Для сбора и структуризации представленных данных использовались средства языка программирования R, с помощью которых сформирована обучающая выборка.

Вследствие небольшого количества начальных данных для обучения модели нейронной сети использовалась вся доступная информация по

теннисным матчам, проводившимся в рамках общемировых турниров за период с 2005 года по текущий.

Нормализация показателей проводилась по методу Минимакс (Min-Max Scaling) – техника преобразования значений признака, масштабирующая значения таким образом, что они располагаются в диапазоне от 0 до 1. Каждый нормализованный показатель вычисляется с помощью следующей формулы (4):

$$x_{i,norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \quad (4)$$

где

$x_{i,norm}$ – нормализованный элемент признака;

x_{min} – наименьший элемент признака;

x_i – i -й непреобразованный элемент;

x_{max} – наибольший элемент.

Цель такого преобразования – изменить значения числовых столбцов в наборе данных так, чтобы сохранить различия их диапазонов. В машинном обучении набор данных требует нормализации, когда параметры имеют разные диапазоны и тем самым способствуют искажению восприятия взаимоотношений между переменными.

Фрагмент обучающей выборки, используемой при прогнозировании исходов теннисных матчей, приведен на рисунках 9 и 10.

speed	rank	rating	pointsWon	pointsLost	decidingSetsPct	fifthSetsPct	tieBreaksPct	rank.1
1.00000000	0.90702948	0.07272727	0.07446809	0.55263158	0.000	0.000	0.73937153	0.18872017
1.00000000	0.65306122	0.19028213	0.29787234	0.97368421	0.375	0.375	0.52865065	0.03253796
0.59090909	0.65306122	0.19028213	0.29787234	0.97368421	0.375	0.375	0.52865065	0.40780911
0.27272727	0.00000000	1.00000000	0.08510638	0.00000000	1.000	1.000	0.86321627	0.07809111
0.36363636	0.65306122	0.19028213	0.29787234	0.97368421	0.375	0.375	0.52865065	0.06290672
1.00000000	0.48072562	0.27868339	0.31914894	0.73684211	0.333	0.333	0.76709797	0.86767896
0.72727273	0.48072562	0.27868339	0.31914894	0.73684211	0.333	0.333	0.76709797	0.88503254
0.36363636	0.90702948	0.07272727	0.01063830	0.13157895	0.000	0.000	0.42329020	0.75054230

Рисунок 9 – Фрагмент обучающей выборки (часть 1)

rating.1	pointsWon.1	pointsLost.1	decidingSetsPct.1	fifthSetsPct.1	tieBreaksPct.1	result
0.51693622	0.10526316	0.02631579	1.000	1.000	0.30779392	0
0.87671651	0.70175439	0.28947368	0.857	0.857	0.66050198	0
0.33262130	0.80701754	1.00000000	0.571	0.571	0.48745046	0
0.76502899	0.70175439	0.34210526	0.714	0.714	0.68560106	0
0.79676533	0.26315789	0.07894737	1.000	1.000	0.39630119	0
0.09734513	0.14035088	0.55263158	0.000	0.000	0.66050198	0
0.08452853	0.03508772	0.15789474	0.000	0.000	0.36327609	1
0.16081782	0.22807018	0.55263158	0.429	0.429	0.13210040	0

Рисунок 10 – Фрагмент обучающей выборки (часть 2)

3.2 Выбор архитектуры нейронной сети

Для прогнозирования исходов теннисных матчей выбрана архитектура многослойного персептрона. Есть несколько аргументов, которые можно привести относительно того, почему эта архитектура лучше используется для решения поставленной задачи:

1) нелинейность: модель многослойной нейронной сети персептрона (MLP) очень полезна, поскольку она может моделировать нелинейные функции. Это делает нейронную сеть MLP мощным инструментом для работы с очень сложными наборами данных, требующими большого количества входных и выходных данных;

2) адаптивное обучение: модель нейронной сети MLP является адаптивной в том смысле, что она приспосабливается к изменяющейся среде.

Эта способность модели делает ее оптимальной для обработки динамических наборов данных, где распределение данных меняется со временем;

3) обучаемость: модель нейронной сети MLP позволяет эффективно изучать наиболее важные особенности больших наборов данных. Изучая, какие функции наиболее важны, модель MLP можно использовать для создания мощных классификационных моделей, которые обнаруживают закономерности в данных, которые трудно увидеть без алгоритмов машинного обучения;

4) масштабируемость: модель нейронной сети MLP может быть увеличена или уменьшена в зависимости от набора данных и доступных вычислительных ресурсов. Это делает его оптимальным для больших наборов данных, требующих больших вычислительных ресурсов, и для небольших наборов данных, требующих простой, но мощной модели;

5) мощные возможности классификации: модель нейронной сети MLP обладает мощными возможностями классификации, которые позволяют ей точно классифицировать данные. Это делает его идеальным для таких приложений, как распознавание изображений, речи и обработка естественного языка;

6) отказоустойчивость: нейронная сеть MLP обладает высокой устойчивостью к ошибкам и шуму в данных. Это делает его эффективным при работе с наборами данных, в которых отсутствуют или повреждены данные, или которые имеют несбалансированное распределение данных;

7) параллельная обработка: нейронные сети MLP могут обучаться на нескольких процессорах одновременно, что позволяет выполнять обработку данных более быстро и эффективно.

Архитектура многослойного персептрона нейронной сети хорошо подходит для прогнозирования результатов теннисных матчей благодаря своей способности улавливать сложные нелинейные взаимосвязи между переменными, обрабатывать многомерные данные, адаптироваться к обновлениям данных в режиме реального времени и демонстрировать более

высокую общую точность прогнозов, когда по сравнению с другими моделями. На рисунке 11 представлена структура подходящей нейронной сети.

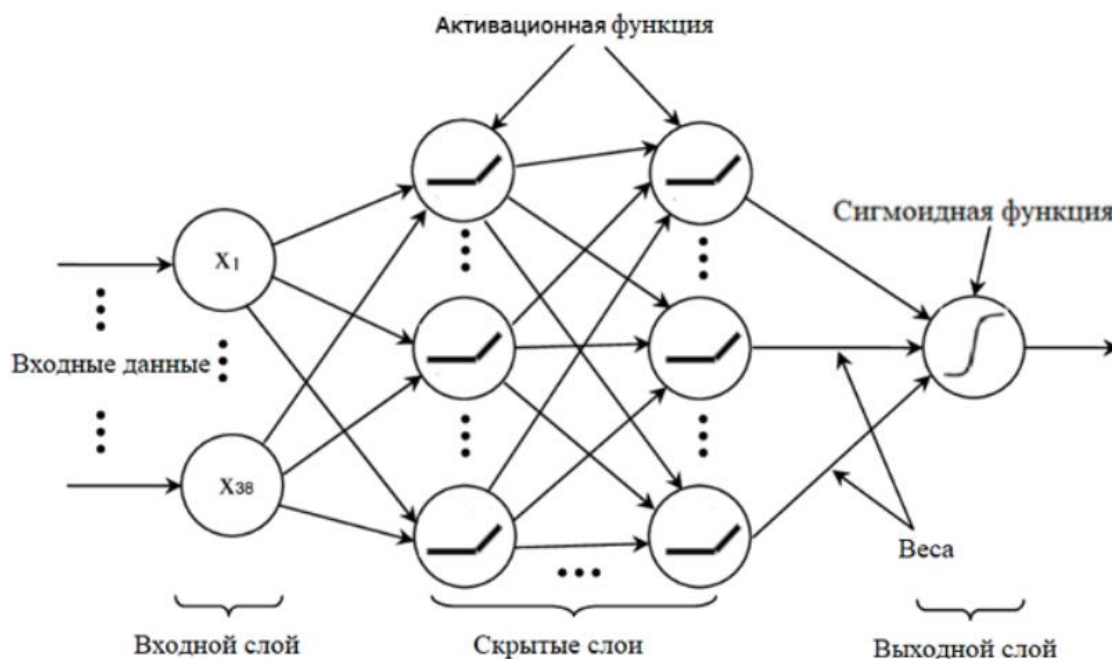


Рисунок 11 – Структура нейронной сети для прогнозирования результатов теннисных матчей

3.3 Построение и обучение модели нейронной сети

В рамках первого эксперимента модель представленной нейронной сети построена средствами пакета Matlab, код представлен в приложении 1. В качестве обучающей выборки использовались данные, представленные на рисунках 9-10. Количество элементов выборки – 225, число параметров – 15.

Обучающая выборка разделена на два фрагмента: 85% – данные для обучения, 15% – данные для валидации. Модель тренировалась 30 эпох с размером пакета равным 16 элементам выборки, после каждого из которых проводилась настройка весов с помощью метода обратного распространения ошибки. На рисунке 12 представлен результат работы модели нейронной сети.

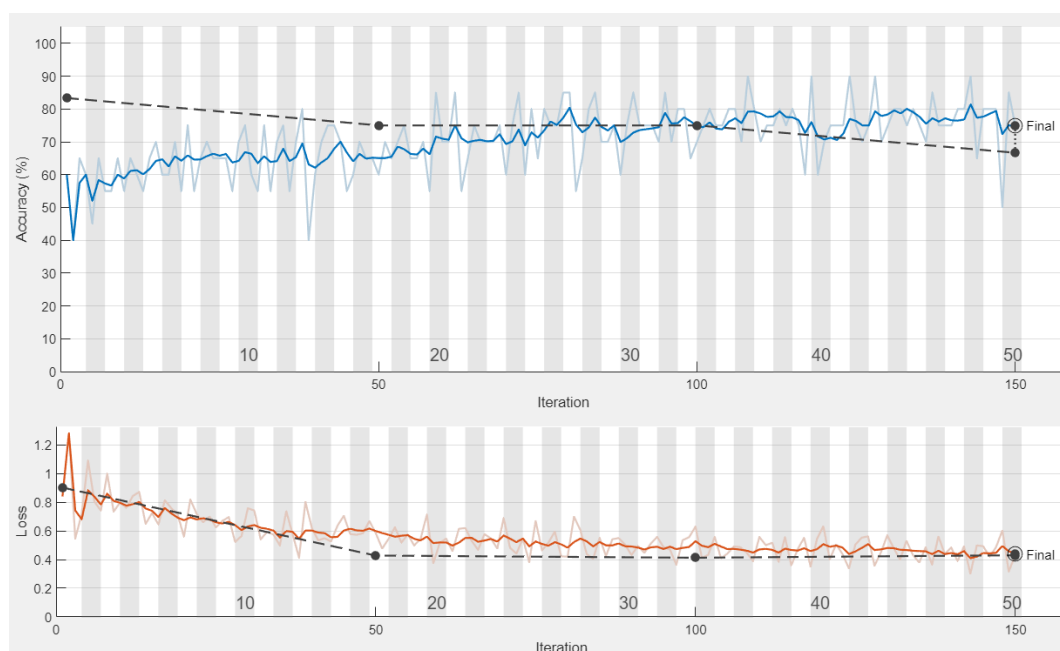


Рисунок 12 – Результат работы нейронной сети

Второй эксперимент выполнялся на языке программирования Python с использованием инструментов для работы с нейронными сетями tensorflow и keras, код представлен в приложении 2.

Обучающая выборка дополнена новыми элементами (752 элемента) и параметрами. Фрагмент новых данных показан на рисунке 13.

rank	totalPoints	tournamentPoints	rankingPoints	achievementsPoints	tGPoints	tFLPoints	tMPoints
0.234920635	0.022678186	0.019469027	0.010	0.04968944	0.004065041	0.00000000	0.005813953
0.117460317	0.051835853	0.072566372	0.020	0.01863354	0.016260163	0.00000000	0.116279070
0.117460317	0.051835853	0.072566372	0.020	0.01863354	0.016260163	0.00000000	0.116279070
0.226984127	0.023758099	0.037168142	0.000	0.00621118	0.012195122	0.00000000	0.058139535
0.426984127	0.006479482	0.010619469	0.000	0.00000000	0.00000000	0.00000000	0.017441860
0.426984127	0.006479482	0.010619469	0.000	0.00000000	0.00000000	0.00000000	0.017441860
0.377777778	0.008639309	0.014159292	0.000	0.00000000	0.004065041	0.00000000	0.011627907
0.030158730	0.181425486	0.219469027	0.165	0.06832298	0.174796748	0.12307692	0.226744186
0.223809524	0.023758099	0.037168142	0.000	0.00621118	0.016260163	0.00000000	0.023255814
0.039682540	0.133909287	0.187610619	0.060	0.03726708	0.134146341	0.24615385	0.139534884
0.234920635	0.022678186	0.019469027	0.010	0.04968944	0.004065041	0.00000000	0.005813953
0.211111111	0.024838013	0.038938053	0.000	0.00621118	0.028455285	0.00000000	0.023255814
0.077777778	0.078833693	0.106194690	0.015	0.06211180	0.077235772	0.04615385	0.098837209

Рисунок 13 – Фрагмент новой обучающей выборки

Обучающая выборка составляет 85% от общего количества данных, тестовая – 15%. В качестве функции ошибки использовалась двоичная

перекрестная энтропия, алгоритмом оптимизации выбран adam. Модель тренировалась 85 эпох с размером пакета равным 16 элементам.

На рисунках 14 и 15 отображены результаты обучения нейронной сети: график функции ошибки и график точности тренировки соответственно.

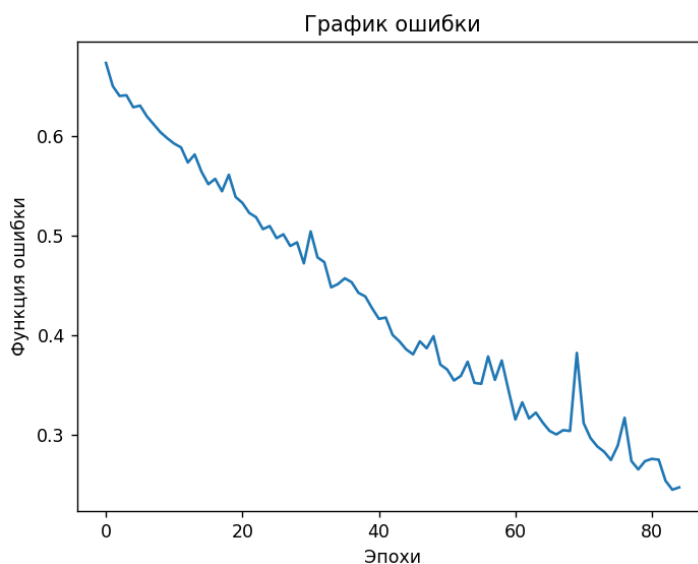


Рисунок 14 – График функции ошибки



Рисунок 15 – График точности

3.4 Анализ результатов обучения модели нейронной сети

Первый эксперимент показал точность прогноза равный 72%, что может свидетельствовать об использовании небольшой выборки для обучения, однако это также доказывает эффективность данного способа прогнозирования. Данный показатель улучшается за счет увеличения обучающей выборки и расширения числа параметров, определяющих теннисный матч.

Во втором эксперименте результат обучения показал точность, достигающую 89%, что может говорить об улучшении нейросетевого прогнозирования по сравнению с первым экспериментом вследствие качественного и количественного увеличения обучающей выборки и более точной настройки алгоритма обучения нейросетевой модели.

Попробуем предсказать результат теннисного матча между Роджером Федерером и Энди Роддиком, завершившийся победой первого. Результат представлен на рисунке 16.

A screenshot of a prediction result displayed in a dark-themed interface. The text 'Прогноз: [0.79715097]' is shown in a light-colored font, with the numerical value enclosed in brackets.

Рисунок 16 – Результат прогноза модели

Нейронная сеть предсказала победу Роджера Федерера со степенью уверенности равной 79%. На основе этого результата можно с уверенностью сказать об эффективности нейросетевого прогнозирования спортивных состязаний.

При наличии результатов кластерного анализа можно использовать обученную нейронную сеть для решения изначально поставленной задачи прогнозирования. Важным преимуществом методов нейросетевой кластеризации является возможность дообучения сетей параллельно в рамках их практического использования. Это значит, что, узнав фактический

результат матча, для которого ранее был сделан прогноз, можно пересчитать вероятности исходов для соответствующего кластера на будущее или продолжить процедуру обучения нейронной сети. Таким образом, постепенно наращивая базу данных по спортивной статистике, обеспечивается повышение качества и точности прогноза.

ЗАКЛЮЧЕНИЕ

Использование нейронных сетей для анализа и прогнозирования спортивных состязаний является перспективным аналогом или дополнением для традиционных методов исследования. В силу своей адаптивности одни и те же нейронные сети могут использоваться для анализа различных видов спорта и прогнозирования различного рода показателей, в то же время найденные для конкретного инструмента закономерности с помощью методов технического анализа могут работать хуже или не работать вообще для каких-либо других инструментов.

Специфика объекта исследования накладывает некоторые особенности на использование нейронных сетей для анализа данных. Такой особенностью является выбор функции активации или функции ошибки нейронной сети, отличной от традиционной среднеквадратичной. В дополнение ко этому архитектура модели может существенно влиять на способы обучения и работу нейросети. Необходимо отметить, что одной из важных составляющих анализа данных с помощью нейронных сетей является предобработка данных, направленная на сокращение размерности входов сети, повышение совместной энтропии входных переменных и нормировку входных и выходных данных.

В ходе работы проведено исследование способов применения нейронных сетей в прогнозировании результатов теннисных матчей, а также продемонстрирована работа модели нейронной сети многослойного персептрона в прогнозировании теннисных матчей.

Нейросетевая модель реализована в виде компьютерной программы, которая может быть использована для получения результатов новых теннисных матчей.

Методика нейросетевого моделирования может быть использована для создания аналогичных программ, позволяющих получать результаты спортивных состязаний.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Гафаров, Ф.М. Искусственные нейронные сети и их приложения: учебное пособие / Ф.М. Гафаров, А.Ф. Галимянов ; Издательство Казанского университета. – Казань, 2018. – 121 с. – УДК 004.032.26, ББК 32.973.2-018+32.813.
2. Описание нейронных сетей, используемых для прогнозирования // Studbooks.net : сайт. – 2015. – URL: https://studbooks.net/2273202/informatika/opisanie_neyronnyh_setey_ispolzuemyh_prognozirovaniya (дата обращения 10.02.2023).
3. Нейронные сети для начинающих // Habr.com : сайт. – 2014. – URL: <https://habr.com/ru/post/312450/> (дата обращения 20.02.2023).
4. Грачев, О.К. Физическая культура : Учебное пособие / О.К. Грачев. – Ростов-на-Дону : ИКЦ «МарТ», 2011. – 464 с. – ISBN 5-241-00576-5.
5. Майнберг, Э. Основные проблемы педагогики спорта: вводный курс / [пер. с нем. под ред. М. Я. Виленского, О. С. Метлушко] – Москва : АО «Аспект-пресс», 1995. – 318 с. – ISBN 5-7567-0014-5.
6. Семченко, О. Пермские ученые занялись спортивными прогнозами / О. Семченко // Поиск. – 2014. – №7. – URL: <http://www.poisknews.ru/theme/infosphere/8916/> (дата обращения: 19.04.2023).
7. Пермская научная школа искусственного интеллекта и ее инновационные проекты 2-е изд / Л.Н. Ясницкий, В.В. Бондарь, С.Н. Бурдин [и др.] ; НИЦ «Регулярная и хаотическая динамика». – Ижевск, 2008. 75 с.
8. Баландин, В. И. Прогнозирование в спорте / В.И. Баландин, Ю.М. Блудов, В.А. Плахтиенко. — Москва : Физкультура и спорт, 1986. – 191,[2] с. : граф.; 20 см. – (Наука – спорту. Основы тренировки).
9. Круглов, В.В. Искусственные нейронные сети. Теория и практика. – 2-е издание, стереотип / В.В. Круглов, В.В. Борисов. – Москва : Горячая линия–Телеком, 2002. – 382 с. : ил. – ISBN 5-93517-031-0 – URL:

<https://studizba.com/files/show/djvu/1761-1-kruglov-v-v-borisov-v-v--iskusstvennye.html> (дата обращения 20.02.2023).

10. Бурева, В. К. Применение методов искусственного интеллекта в спорте / В. К. Бурева, Е. И. Стоянов // Актуальные вопросы технических наук : материалы III Междунар. науч. конф. (г. Пермь, апрель 2015 г.). – Пермь : Зебра, 2015. – 1-12. с. – URL: <https://moluch.ru/conf/tech/archive/125/7831/> (дата обращения: 15.04.2023).

11. Прогнозирование исходов спортивных событий методами нейросетевой кластеризации // Neuronus.com : сайт. – 2010. – URL: <https://neuronus.com/stat/205-prognosis-sport-neural-claster.html> (дата обращения 05.03.2023).

12. Neural Networks Football Result Prediction // Medium.com : сайт. – URL: <https://medium.com/@rolandshum.shc/neural-networks-football-result-prediction-d8b0f933118b> (дата обращения 05.03.2023).

ПРИЛОЖЕНИЕ А

Файл создания модели в пакете Matlab

```
input = readtable("result_table.csv", "VariableNamingRule", "preserve", "TextType", "string");
labelName = "result";
tbl = convertvars(input, labelName, 'categorical');
head(tbl);
tbl = splitvars(tbl);
classNames = categories(tbl{:, labelName});

numObservations = size(tbl, 1);
numObservationsTrain = floor(0.7 * numObservations);
numObservationsValidation = floor(0.15 * numObservations);
numObservationsTest = numObservations - numObservationsTrain -
numObservationsValidation;

idx = randperm(numObservations);
idxTrain = idx(1:numObservationsTrain);
idxValidation =
idx(numObservationsTrain + 1 : numObservationsTrain + numObservationsValidation);
idxTest = idx(numObservationsTrain + numObservationsValidation + 1 : end);

tblTrain = tbl(idxTrain, :);
tblValidation = tbl(idxValidation, :);
tblTest = tbl(idxTest, :);

numFeatures = size(tbl, 2) - 1;
numClasses = numel(classNames);

layers = [
    featureInputLayer(numFeatures)
    fullyConnectedLayer(50)
    batchNormalizationLayer
    reluLayer
```

```

fullyConnectedLayer(numClasses)
softmaxLayer
classificationLayer];

miniBatchSize = 20;

options = trainingOptions('adam', ...
    'MaxEpochs', 30, ...
    'MiniBatchSize',miniBatchSize, ...
    'Shuffle','every-epoch', ...
    'ValidationData',tblValidation, ...
    'Plots','training-progress', ...
    'Verbose',false);

net = trainNetwork(tblTrain,labelName,layers,options);

YPred = classify(net,tblTest(:,1:end-1),'MiniBatchSize',miniBatchSize);

YTest = tblTest{:,labelName};
accuracy = sum(YPred == YTest)/numel(YTest);

figure
confusionchart(YTest,YPred)

```

ПРИЛОЖЕНИЕ Б

Файл создания модели средствами языка Python

```
import numpy as np
import pandas as pd
import tensorflow as tf
from tensorflow import keras
import matplotlib.pyplot as plt

# Load the CSV file using Pandas
data = pd.read_csv("result_table.csv")

# Split the data into input (X) and output (y) variables
X = data.iloc[:, :-1].values
y = data.iloc[:, -1].values

# Normalize the input data
#X = (X - np.mean(X, axis=0)) / np.std(X, axis=0)

# Split the data into training and test sets (e.g. 85/15 split)
train_size = int(len(X) * 0.85)
X_train, X_test = X[:train_size], X[train_size:]
y_train, y_test = y[:train_size], y[train_size:]

# Verify the shape of the training data
print("Shape of training data:")
print("X_train:", X_train.shape)
print("y_train:", y_train.shape)

model = keras.Sequential([
    keras.layers.Dense(38, activation='relu', input_shape=(X_train.shape[1],)),
    keras.layers.Dense(256, activation='relu'),
    keras.layers.Dense(128, activation='relu'),
    keras.layers.Dense(1, activation='sigmoid')
])
```

```
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

```
history = model.fit(X_train, y_train, epochs=85, batch_size=16, verbose=2)
```

```
test_loss, test_acc = model.evaluate(X_test, y_test, verbose=2)
```

```
print("Test accuracy:", test_acc)
```

```
plt.figure()
```

```
plt.plot(history.history['loss'])
```

```
plt.title('График ошибки')
```

```
plt.ylabel('Функция ошибки')
```

```
plt.xlabel('Эпохи')
```

```
plt.show()
```

```
plt.figure()
```

```
plt.plot(history.history['accuracy'])
```

```
plt.title('График точности')
```

```
plt.ylabel('Точность')
```

```
plt.xlabel('Эпохи')
```

```
plt.show()
```

```
new_match =
```

```
[[0.174552429667519,0.566695727986051,1,0.535802469135802,0.31,0.376,0.6048158640226  
63,0.421602787456446,0.434782608695652,0.758620689655172,0.68547418967587,0.730215  
82733813,0.234920634920635,0.0226781857451404,0.0194690265486726,0.01,0.0496894409  
937888,0.0040650406504065,0,0.00581395348837209,0,0.0918367346938776,0.1666666666  
6667,0.275555555555556,0.280952380952381,0.0841269841269841,0.0755939524838013,0.1  
16814159292035,0.005,0.0167597765363128,0.0934959349593496,0.0615384615384615,0.08  
45771144278607,0,0.224489795918367,0.125,0.554269175108538,0.65]]
```

```
prediction = model.predict(new_match)
```

```
print('Прогноз:', prediction[0])
```