

Московский Государственный Университет имени М.В.
Ломоносова
Физический факультет
Кафедра атомной физики, физики плазмы и
микроэлектроники

Курсовая работа студента 2 курса
Семенова М. В.

**Исследование качества прогнозирования геомагнитного
индекса на разных фазах солнечного цикла**

Научные руководители:
Зав. лаб., к.ф.-м.н. Доленко С. А.
Инженер Широкий В. Р.

Москва
2024

Оглавление

1. Введение	3
2. Используемые данные.....	4
3. Выбор архитектуры и параметров моделей	5
3.1 Тривиальная модель	5
3.2 Линейная модель	5
3.3 Многослойный персептрон.....	5
3.4 LSTM.....	7
3.5 Сверточные нейронные сети	8
3.6 Сверточная LSTM	9
3.7 Transformer.....	10
4 Алгоритм оптимизации, функции потерь, методы регуляризации.	11
4.1 Решение задачи оптимизации параметров модели.....	11
4.2 Проблема переобучения.....	12
4.3 Модифицированные функции потерь	12
5 Результаты моделей и выбор наилучшей модели	13
5.1 Поиск моделей-кандидатов.....	14
5.2 Выбор лучшей модели	15
5.3 Более подробный анализ transformer модели.....	16
6 Концепция построения композитной модели.....	16
7 Результаты и их анализ	17
8. Выводы	21

1. Введение

Магнитосфера Земли – это сложная динамическая система, состояние которой зависит от множества факторов, по большей части определяемых воздействием Солнца. Совокупность этих факторов принято называть космической погодой. Эти факторы могут оказывать значительное влияние на функционирование космических и авиационных аппаратов, технических устройств на земле и на самочувствие людей, поэтому прогнозирование космической погоды имеет высокую актуальность.

Для описания состояния магнитосферы и степени её возмущения используются специальные геомагнитные индексы, которые характеризуют развитие возмущений магнитного поля Земли. Сильные возмущения магнитного поля Земли называют магнитными бурями. Они могут вызвать различные проблемы, такие как нарушение радиосвязи, функционирования электросетей, навигационных систем, а также трубопроводов и линий электропередач. Развитие методов прогнозирования и оперативное отслеживание геомагнитных возмущений необходимы для предотвращения серьезных последствий.

В данной работе для прогнозирования значения геомагнитного индекса используются методы машинного обучения (ММО). Это обусловлено высокой сложностью описания процессов в системе "Солнце – солнечный ветер – магнитосфера Земли", что делает традиционные подходы к прогнозированию практически невозможными или недостаточно эффективными.

Для анализа динамики магнитосферы Земли в данной работе используется многомерный временной ряд, включающий параметры межпланетного магнитного поля, скорость и плотность солнечного ветра, а также геомагнитный индекс Dst. Этот индекс представляет собой среднюю в часовом интервале величину возмущения горизонтальной составляющей напряженности магнитного поля Земли, отсчитываемого от спокойного уровня, определенную по данным четырех низкоширотных обсерваторий, равномерно распределенных по долготе [[Подробнее](#)].

Одним из ключевых источников возмущений магнитосферы Земли является солнечная активность (СА). Хорошо известно, что СА подвержена циклическим изменениям, наиболее выраженным из которых является цикл с примерно 11-летним периодом [[Подробнее](#)]. При этом физические процессы в магнитосфере Земли на разных фазах цикла СА могут различаться. Кроме того, различные циклы СА могут значительно отличаться друг от друга, как по своей амплитуде, так и по характеру динамики СА.

Традиционный подход к прогнозированию геомагнитных индексов с помощью ММО состоит в том, что модель строится (обучается) на всех имеющихся исторических данных, за исключением нескольких последних лет, которые используются для тестирования модели. Однако при этом

эффективность прогнозирования на разных циклах СА и на разных фазах цикла СА может различаться.

К настоящему времени накоплен однородный массив данных, измеренных с помощью одной и той же аппаратуры, за период с конца 1997 по конец 2023 года, т.е. более чем за два полных цикла СА. Это позволило поставить вопрос об исследовании того, какой подход является более эффективным – традиционный, основанный на построении единой модели, или композитный, основанный на построении отдельных моделей для различных циклов СА и/или их фаз. В случае успеха такого подхода может быть построена композитная модель, которая вначале автоматически определяет цикл СА и его фазу, в результате чего выбирает наиболее подходящую для прогноза отдельную модель.

Таким образом, **целью** настоящей работы является проверка эффективности композитного подхода при использовании различных ММО. Поставленными в данной работе **задачами** являются построение единой модели и отдельных моделей ММО для разных циклов и разных фаз цикла СА, с использованием различных ММО, а также сравнение результатов применения единой и отдельных моделей в разных ситуациях.

2. Используемые данные

Было произведено так называемое погружение многомерного временного ряда входных параметров на 24 часа, что означает использование 25 среднечасовых значений каждого входного параметра, включая текущее значение.

Для построения моделей ММО для прогноза Dst-индекса использовались следующие входные параметры:

а) Параметры солнечного ветра (СВ) в точке Лагранжа L1 между Землёй и Солнцем:

- Скорость СВ v (измеряется в км/с)
- Плотность протонов в СВ n_p (измеряется в см⁻³)
- Температура плазмы СВ (измеряется в Кельвинах)

б) Параметры межпланетного магнитного поля (ММП) в системе GSM:

- B_x (x-компонента ММП)
- B_y (y-компонента ММП)
- B_z (z-компонента ММП)
- Амплитуда B (модуль ММП)

в) Значения геомагнитного индекса Dst.

г) Временные характеристики, связанные с вращением Земли вокруг Солнца и вокруг своей оси - гармонические функции с периодом один год и одни сутки (для них погружение не применялось ввиду наличия однозначной взаимосвязи между значениями этих характеристик с разными задержками).

Общее количество входных признаков задачи, таким образом, составило 204.

Использовались данные СВ и ММП с космического аппарата ACE (Advanced Composition Explorer) [[подробнее](#)], полученные с приборов SWEPAM и MAG, соответственно.

Ввиду специфики данных, получаемых с космических аппаратов, в них могут быть пропуски. Для их исключения недостающие значения параметров интерполировались, в случае, если число подряд идущих пропусков не превышало 12. После этого проводилось погружение временного ряда, и затем удаление оставшихся примеров с пропусками.

В данной работе для оценки и сравнения качества работы моделей использовались коэффициент множественной детерминации R^2 и среднеквадратичное отклонение (СКО).

3. Выбор архитектуры и параметров моделей

Для адекватного определения эффективности композитного подхода было принято решение выбрать наилучшую модель среди моделей, обученных на одинаковых наборах данных. В качестве обучающей выборки использовался массив данных с ноября 1997 года по декабрь 2021 года. В качестве тестовой выборки – промежуток с января 2022 года по декабрь 2023 года. Валидационный набор был получен путем выбора случайных элементов и тренировочного набора в отношении 1:10.

3.1 Тривиальная модель

Для начальной оценки качества прогнозирования использовалась так называемая тривиальная инерционная модель. Значение прогнозируемой величины у такой модели равно последнему известному значению прогнозируемой переменной на момент прогнозирования.

3.2 Линейная модель

Это модель представляет приближение искомой зависимости линейной функцией входных признаков.

3.3 Многослойный персептрон

Многослойный персептрон (МСП) — это класс искусственных нейронных сетей прямого распространения, состоящих как минимум из трех слоев: входного, скрытого и выходного. За исключением входных, все нейроны использует нелинейную функцию активации.

При обучении МСП используется обучение с учителем и алгоритм обратного распространения ошибки.

Персептрон с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Условиями являются: достаточное количество нейронов скрытого слоя.

[\[Подробнее\]](#)

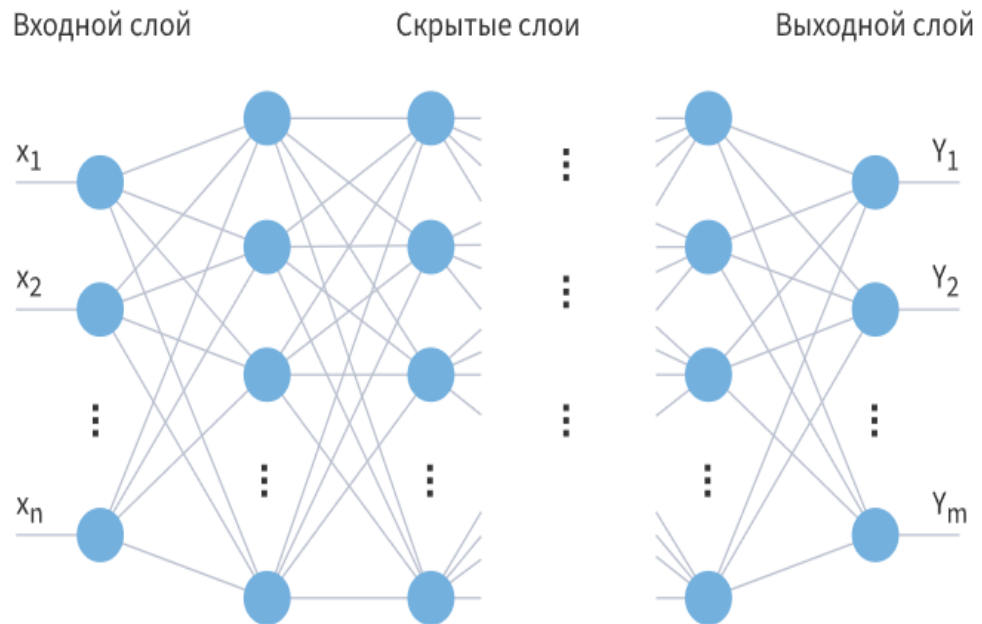


Рис. 1. Пример многослойного персептрона.

Нейроны скрытых слоев содержат функции активации (передаточные функции), это нелинейные функции, благодаря чему персептрон способен аппроксимировать не только линейные зависимости. Примеры функций активаций приведены на Рис. 2.

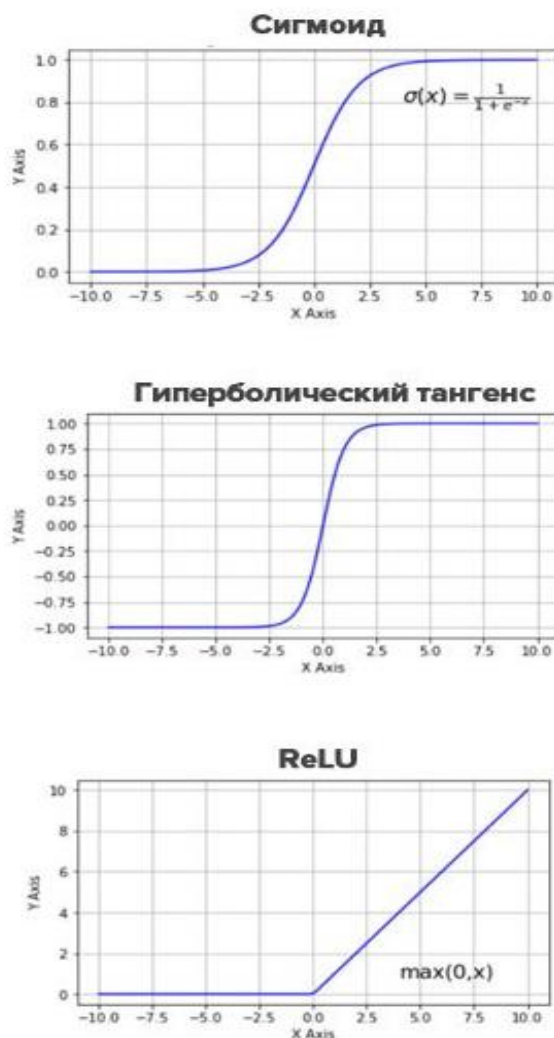


Рис. 2. Распространённые функции активации.

В данной работе использовался персептрон с одним скрытым слоем, содержащим 16 нейронов с функцией активации ReLU, скорость обучения 0,001, гиперпараметры этой и последующих моделей выбраны эмпирически, как наиболее успешные.

3.4 LSTM

LSTM (англ. *Long short-term memory*, *долгая краткосрочная память*) - это разновидность архитектуры рекуррентных нейронных сетей, предназначенная для решения задач классификации, обработки и прогнозирования временных рядов.

Данная архитектура специально разработана для устранения проблемы долгосрочной зависимости, которая свойственна рекуррентным сетям (взрыв и

затухание градиента). Их специализация — запоминание информации в течение длительных периодов времени.

[\[Подробнее\]](#)

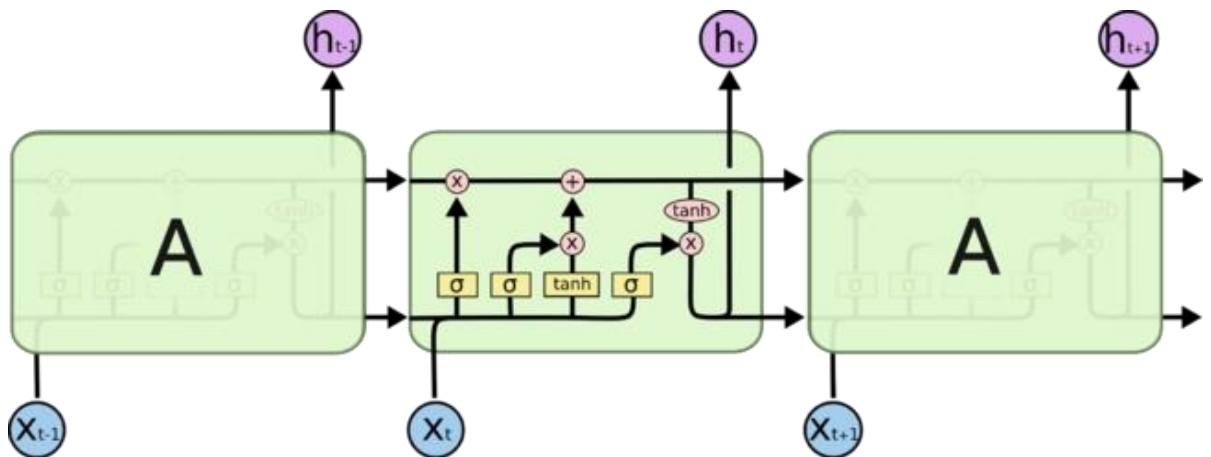


Рис. 3. Рекуррентная нейронная сеть LSTM.

В данной работе применялась архитектура с одним скрытым слоем, содержащим 8 нейронов, со значением dropout, равным 0,2 (описание dropout будет ниже), скорость обучения равна 0,01.

3.5 Сверточные нейронные сети

Сверточная нейронная сеть (англ. convolutional neural network, CNN) — специальная архитектура искусственных нейронных сетей, нацеленная на эффективное распознавание образов.

Идея применения данной сети состоит в поиске потенциальных связей и характерных величин в нескольких соседних временных интервалах, от которых могут зависеть последующие значения Dst индекса.

[\[Подробнее\]](#)

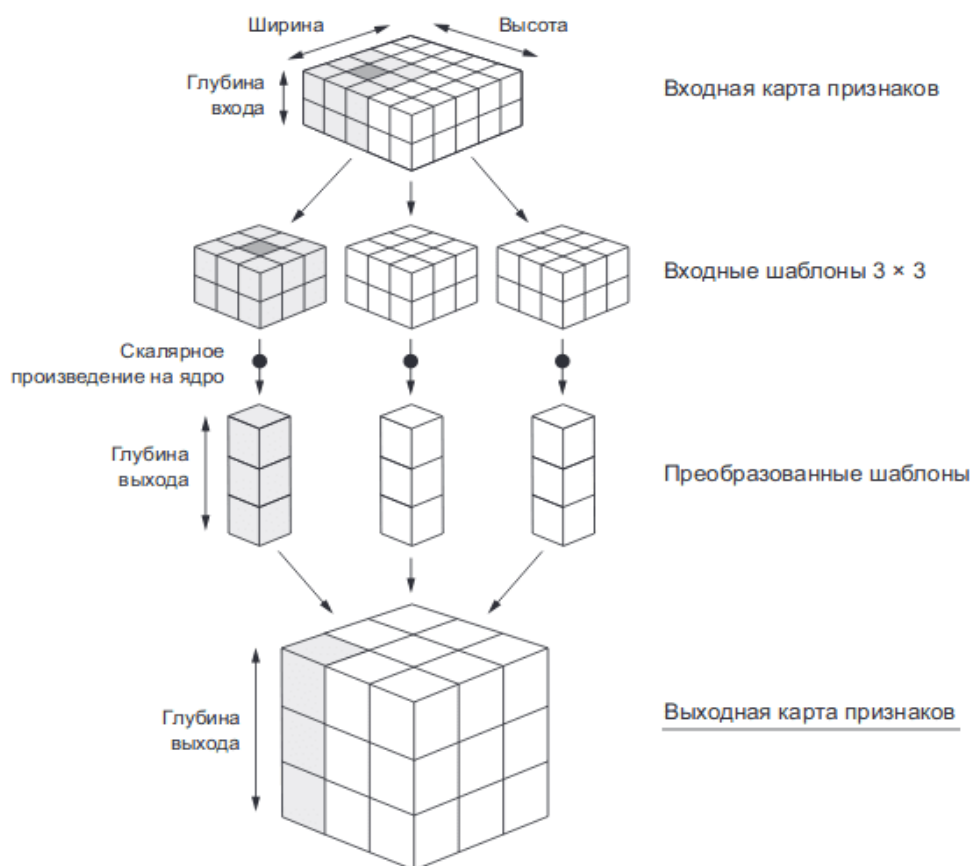


Рис. 4. Свёрточная нейронная сеть.

В данной работе использовалась модель с одним сверточным слоем, с параметрами: размер фильтра – 3, количество каналов – 16, функция активации – ReLU; и одного слоя нейронов с параметрами: количество нейронов – 16, функция активации – ReLU; скорость обучения равна 0,01.

3.6 Сверточная LSTM

Представляет собой объединение сверточного блока и блока LSTM. Данная архитектура объединяет идеи ее составляющих, сверточный блок выявляет наиболее полезные признаки, характерные для соседних временных промежутков, а блок LSTM по их временной зависимости предсказывает значение в определенный момент времени.

Использовалась модель со следующей структурой:

- 1) Сверточный блок:
 - Размер фильтра – 3
 - Количество фильтров – 16
 - Функция активации – ReLU
 - Функция MaxPool
 - Функция BatchNorm (подробнее далее)

2) LSTM блок:

- Размер слоя – 8
- Dropout – 0,2
- Функция BatchNorm

3) Полносвязный блок:

- Размер слоя – 8
- Функция активации – ReLU

3.7 Transformer

Transformer модели реализуют механизм self-attention, который позволяет в каждый момент времени обратиться к произвольному элементу из прошлого без потерь информации.

Архитектура трансформера состоит из кодировщика и декодировщика. Кодировщик получает на вход векторизованную последовательность с позиционной информацией. Декодировщик получает на вход часть этой последовательности и выход кодировщика. Кодировщик и декодировщик состоят из слоев. Слои кодировщика последовательно передают результат следующему слою в качестве его входа. Слои декодировщика последовательно передают результат следующему слою вместе с результатом кодировщика в качестве его входа.

[\[Подробнее\]](#)

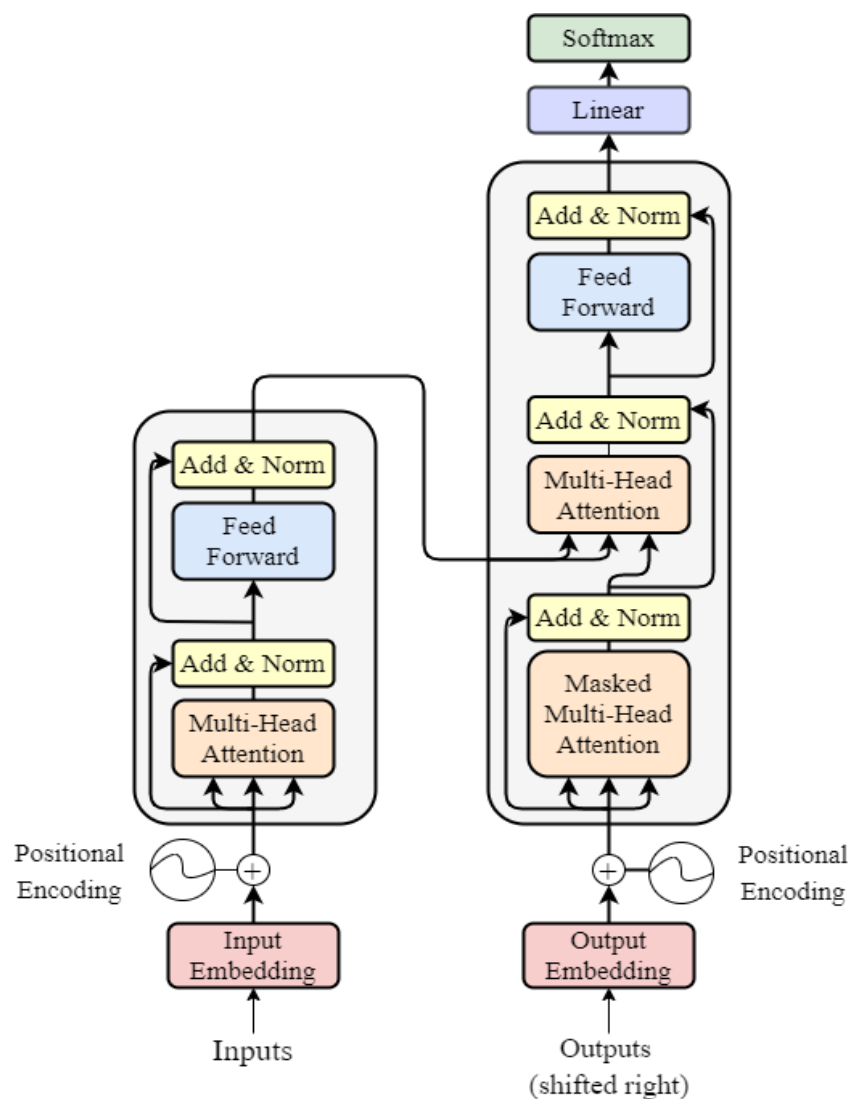


Рис. 5. Архитектура нейронной сети Transformer.

Была использована модель со следующими параметрами:

- Размер скрытого слоя – 8
- Количество голов – 4
- Dropout – 0,025
- Скорость обучения – 0,01

4 Алгоритм оптимизации, функции потерь, методы регуляризации.

4.1 Решение задачи оптимизации параметров модели

Процесс обучения модели ММО представляет собой поиск минимума целевой функции (так называемой функции потерь) в многомерном

пространстве параметров модели. Для решения этой задачи следует определить целевую функцию и алгоритм оптимизации. В данной работе в качестве функции потерь используется средний квадрат ошибки модели, выбранный, как наиболее универсальная функция потерь [\[Подробнее\]](#). При использовании алгоритма обратного распространения ошибки выбран метод оптимизации “Adam”, он показал себя, как достаточно надежный и быстрый алгоритм [\[Подробнее\]](#).

4.2 Проблема переобучения

Данную проблему можно описать, как ухудшение результата на тестовом наборе данных при улучшении результатов на тренировочном наборе данных в процессе обучения. Это возникает из-за того, что данные представляют собой не точные значения, а имеющие определённые погрешности, другими словами, в данных может содержаться некоторый шум. В таких случаях переобученная модель начинает запоминать шум, что негативно сказывается на результатах на тестовом наборе.

Для неглубоких алгоритмов (МСП, LSTM, сверточная сеть) был реализован критерий ранней остановки по валидационному набору данных, согласно которому тренировка модели останавливалась после 20 эпох обучения без улучшения результата на валидационном наборе данных, и фиксировалась модель, показавшая наилучший результат на этом наборе.

Для более глубоких алгоритмов (Transformer, сверточная LSTM), помимо критерия ранней остановки, для борьбы с переобучением были применены методы регуляризации, такие как нормализация пакета и “Dropout”. [\[Подробнее\]](#)

4.3 Модифицированные функции потерь

Ввиду специфики задачи, наиболее интересующими нас результатами являются прогнозы Dst-индекса во время бурь, следовательно, модель должна «уделять больше внимания» примерам, которым соответствуют большие по модулю значения индекса. Одним из способов удовлетворения этого требования могут быть специализированные функции потерь, которые более восприимчивы к модулю значения. В качестве такой функции была выбрана следующая зависимость:

$$\text{MSE} + 0.03D^2$$

где значение D равно величине Dst индекса, а MSE – средний квадрат ошибки, использованный ранее.

Параметры зависимости были выбраны эмпирически. Данная реализация показала неплохие результаты на начальном этапе, но качество

прогнозирования оказалось менее стабильным, поэтому было принято решение подробнее изучить этот метод в дальнейшей работе.

Для визуализации прогноза была также построена корреляционная диаграмма взаимосвязи предсказанного моделью значения и ее истинного значения (Рис. 6):

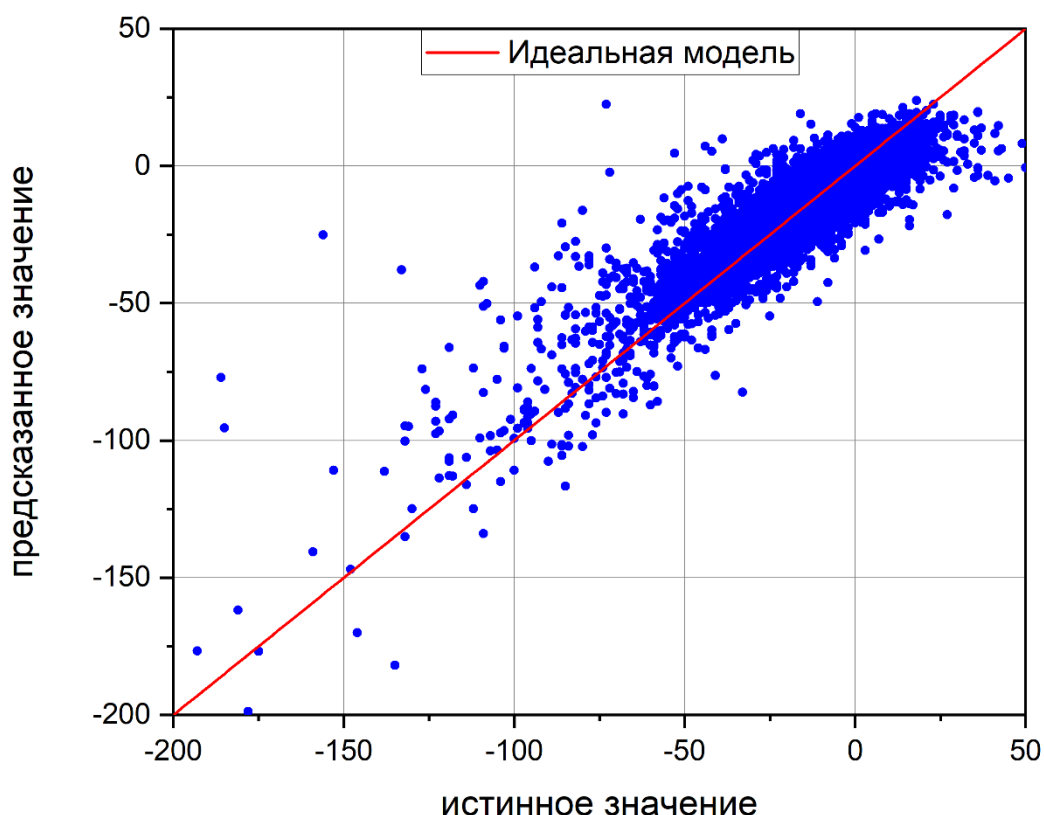


Рис. 6. Корреляционная диаграмма истинного значения индекса Dst и значения, предсказанного многослойным персептроном.

Исследуя эту диаграмму, можно сделать вывод о том, какие значения модель прогнозирует лучше других. Например, из диаграммы видно, что для модели характерен недопрогноз (уменьшение по модулю прогнозируемых значений по сравнению с истинными) для примеров с большими по модулю значениями индекса Dst. После этого можно с помощью модифицированных функций потерь искусственно заставлять модель прогнозировать какие-то значения лучше других, и таким способом построить композитную модель, части которой построены на различных функциях потерь. Данный метод также будет подробнее изучен в дальнейшей работе.

5 Результаты моделей и выбор наилучшей модели

Для сокращения требуемых вычислительных ресурсов и времени расчётов было принято решение оценить качество прогнозирования моделей на единственном наборе данных (с единственным горизонтом прогноза), после чего выбрать наилучшие модели и подробнее проанализировать их результаты. При таком подходе используется предположение о сходной зависимости качества прогноза от его горизонта для разных моделей. Учитывая то обстоятельство, что качество всех моделей достаточно быстро деградирует с увеличением горизонта прогноза, данное предположение представляется разумным.

5.1 Поиск моделей-кандидатов

Для проведения отбора моделей-кандидатов был выбран тренировочный набор с горизонтом прогноза в 5 часов. На этом наборе удалось получить следующие результаты (Рис. 7):

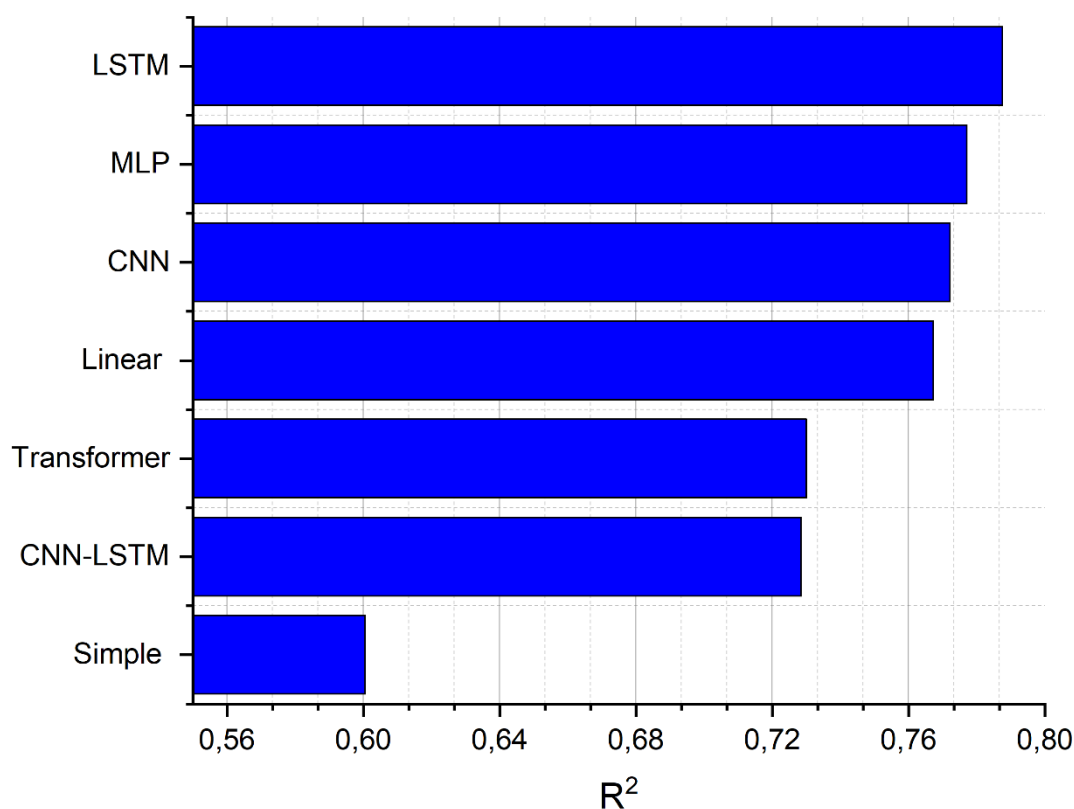


Рис. 7. Сравнение результатов прогнозирования индекса Dst с горизонтом прогноза 5 часов различными методами машинного обучения.

Как видно из диаграммы, лучше всего себя показали персептрон и LSTM. Причиной таких результатов может являться то, что более простая модель

(линейная регрессия) не способна аппроксимировать нелинейные зависимости. Более глубокие архитектуры (сверточная сеть, Transformer, сверточная LSTM) имеют гораздо больше обучаемых параметров; в этой ситуации даже применение методов регуляризации не позволило предотвратить переобучение в достаточной мере.

5.2 Выбор лучшей модели

Для выбора лучшей модели, которая будет использоваться в дальнейших экспериментах, было осуществлено сравнение качества прогнозов МСП и LSTM с горизонтом от 1 до 6 часов, так как этот диапазон представляет наибольший интерес. По окончании измерений были получены следующие результаты (тривиальная модель представлена для сравнения в качестве реперной) (Рис. 8):

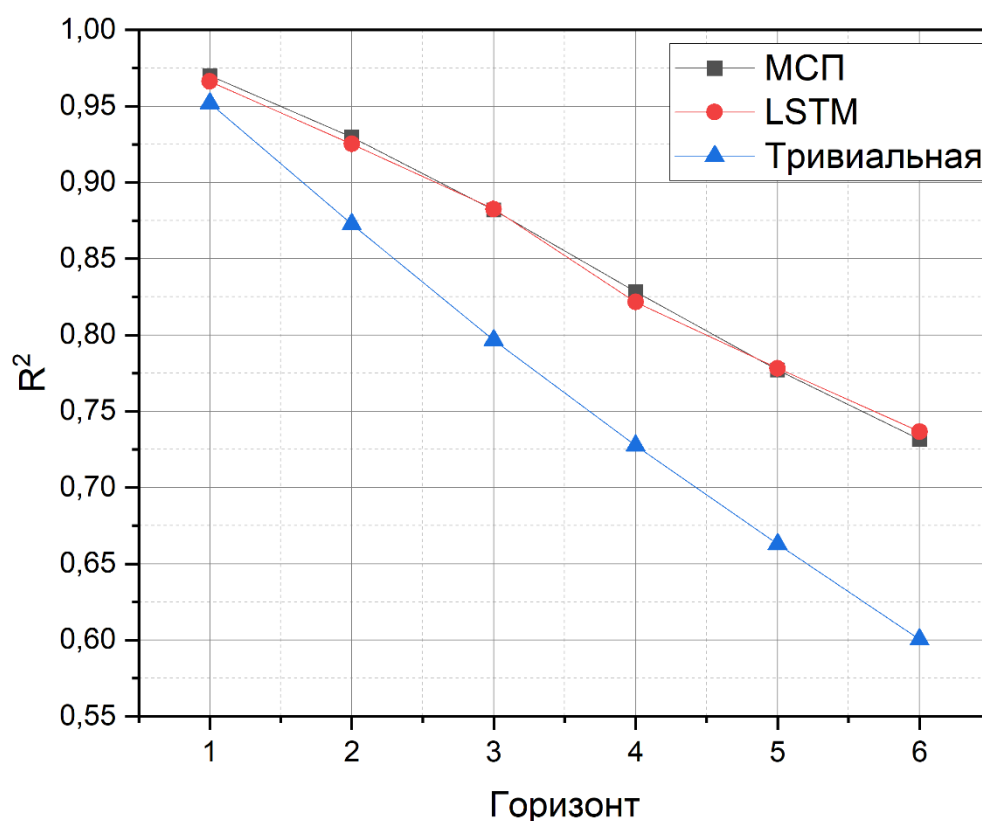


Рис. 8. Сравнение результатов прогнозирования значения индекса Dst с помощью многослойного персептрона, LSTM и тривиальной модели, в зависимости от горизонта прогноза.

Из этих данных можно сделать вывод, что обе модели дают близкие результаты. В результате для дальнейшего использования был выбран персептрон.

5.3 Более подробный анализ transformer модели

Так как данная модель реализует механизм self-attention, она предназначена в основном для последовательностей с большой длиной. В данном случае использованная глубина просмотра в прошлое для временного ряда, по-видимому, оказалась недостаточно велика для реализации потенциала данной архитектуры. Тем не менее, учитывая сложность архитектуры модели и специфику входных признаков, описанную выше, данная модель показала неплохой результат, поэтому возможны эксперименты с этой моделью и большей величиной погружения временного ряда в дальнейших работах.

6 Концепция построения композитной модели

Выше упоминалось построение моделей, основанных на модифицированных функциях потерь, с целью повышения качества прогноза на отдельных участках временного ряда. Другим методом, осуществляющим схожую идею, является построение композитной модели, включающей составляющие модели, обученные на различных участках временного ряда, характеризующихся определёнными свойствами. В данном случае рассматривались составляющие модели, обученные на различных фазах цикла СА или на участках временного ряда, соответствующих возмущённой магнитосфере с заданным уровнем возмущённости. Композитная модель должна включать классификатор, определяющий текущее состояние объекта, и набор составляющих моделей, соответствующих каждому из этих состояний. Далее качество прогноза таких композитных моделей будет сравниваться с моделью, обученной на всех имеющихся данных без деления на участки.

В данной работе такой подход был проверен на следующих группах данных:

- активные фазы – часть данных, соответствующих примерно половине каждого цикла СА с высокой активностью Солнца
- рост активности – часть данных, соответствующих примерно четверти каждого цикла СА на фазе роста СА
- максимум активности – часть данных, соответствующих примерно четверти каждого цикла СА с наиболее высокой активностью Солнца
- бури -100 нТл – отрезки временного ряда, соответствующие геомагнитным возмущениям с индексом Dst, равным -100 нТл, или сильнее

- бури -50 нТл – отрезки временного ряда, соответствующие геомагнитным возмущениям с индексом Dst, равным -50 нТл, или сильнее
- бури -30 нТл – отрезки временного ряда, соответствующие геомагнитным возмущениям с индексом Dst, равным -30 нТл, или сильнее

Для каждой группы в качестве тренировочного набора были взяты отрезки соответствующих типов, принадлежащие 23-му и 24-му циклам СА, а в качестве тестового набора – отрезки соответствующих типов, принадлежащие 25-му циклу СА.

Для дальнейшего обозначения каждой группе был присвоен определенный номер:

- 3 - активная фаза
- 5 - фаза роста активности
- 6 – фаза максимальной активности
- 9 - бури -100 нТл и сильнее
- 10 - бури -50 нТл и сильнее
- 11 - бури -30 нТл и сильнее

7 Результаты и их анализ

На Рисунках 9 и 10 приведено сравнение результатов применения перцептрона, обученного на полном тренировочном наборе данных (23 и 24 циклы СА) с результатами применения перцептронов, обученных на данных каждой из перечисленных выше групп. По оси абсцисс отложена величина горизонта прогноза, по оси ординат – разность между значениями R^2 модели, натренированной на определенной группе данных и R^2 модели, натренированной на всем наборе данных.

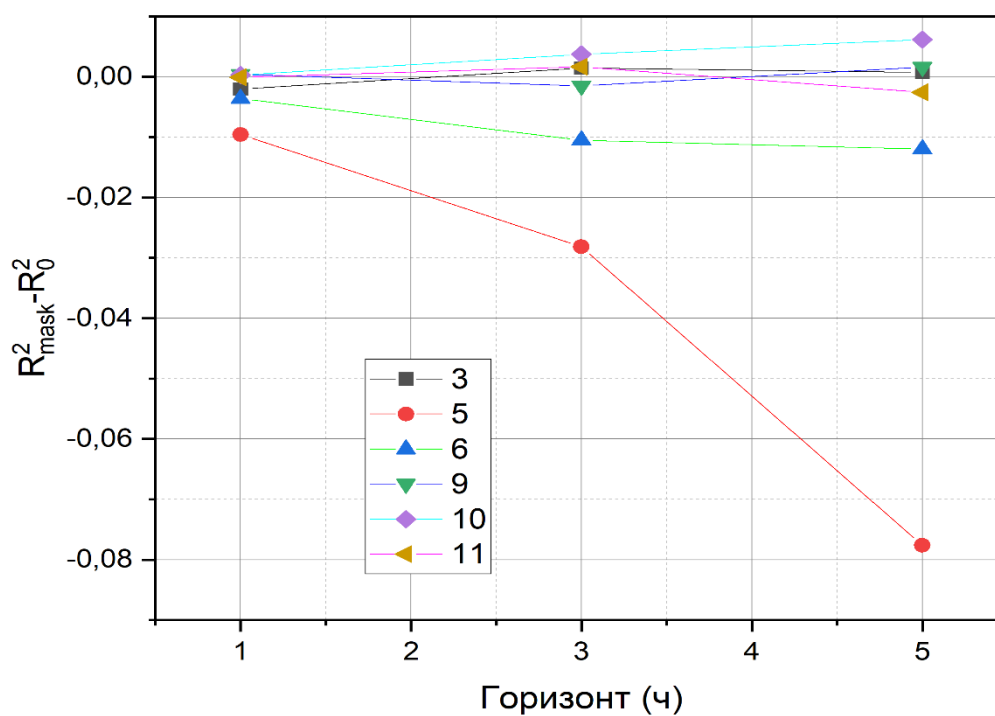


Рис. 9. Разность показателя R^2 на данных указанных типов между персептроном, натренированным на данных того же типа, и персептроном, натренированным на всех данных. Расшифровка типов данных приведена в тексте.

Для большей наглядности построим график без результатов, соответствующих наборам 5 и 6.

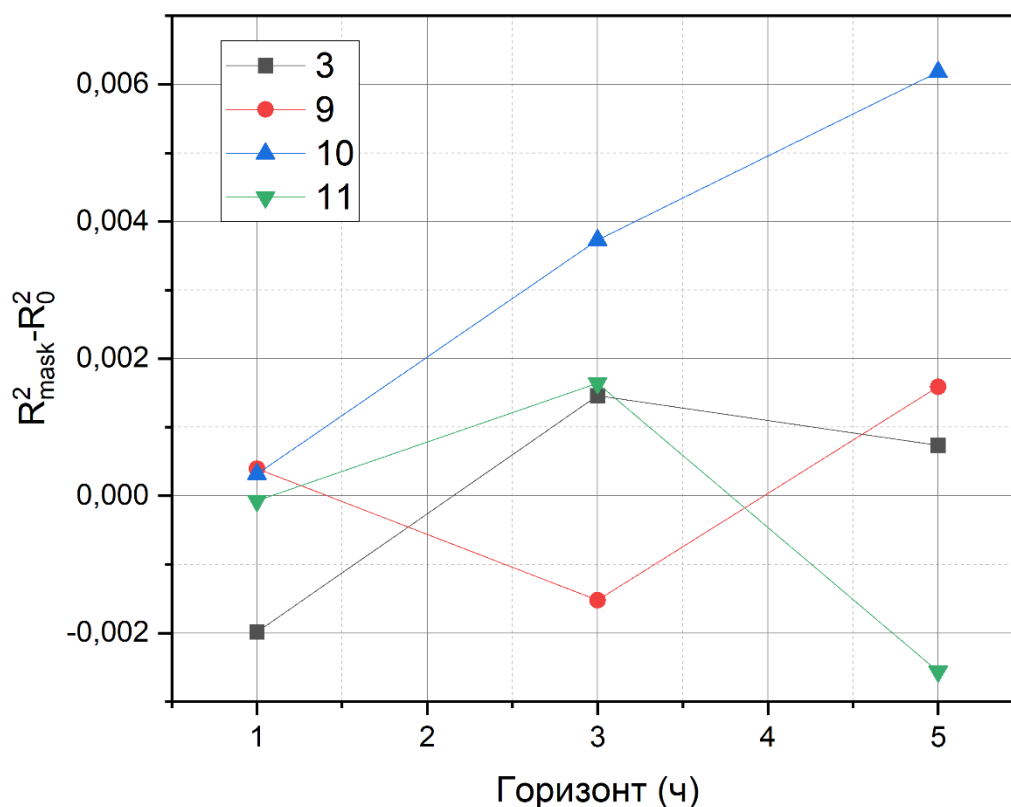


Рис. 10. То же без результатов, соответствующих наборам 5 и 6.

Из данного графика видно, что наибольший интерес представляет набор с номером 10, обеспечивший небольшой, но устойчивый прирост качества прогнозирования для различных горизонтов прогноза.

Для более подробного анализа, были вычислены показатели также для других горизонтов прогноза (2, 4 и 6 часов). Помимо этого, для каждого горизонта были произведены по 5 запусков сети с идентичными гиперпараметрами, но различными начальными весами. После этого по результатам разных запусков для каждого горизонта были вычислены значения среднего и среднеквадратичного отклонения (Рис. 11).

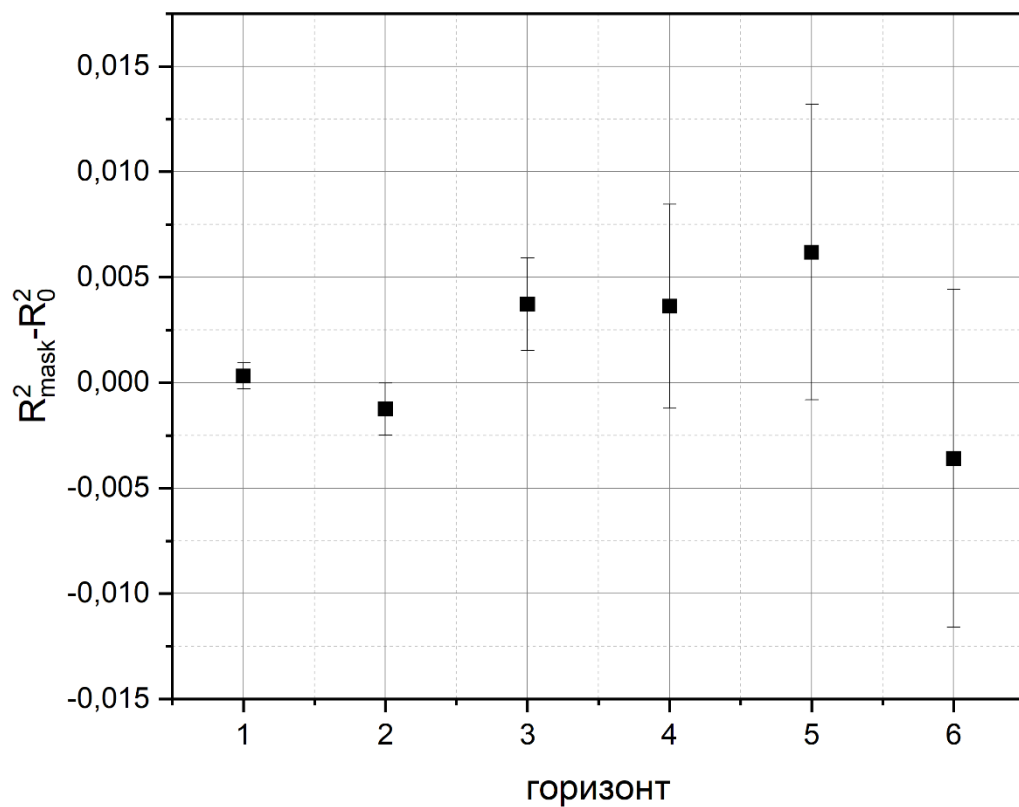


Рис. 11. Разность показателя R^2 на данных типа 10 (бури -50 нТл и сильнее) между перцептроном, натренированным на данных того же типа, и перцептроном, натренированным на всех данных. Планки погрешностей соответствуют стандартному отклонению результатов по 5 идентичным сетям с различными начальными весами.

Как видно из графика, почти все значения отличаются от нуля в пределах погрешности, что означает, что такая модель пока не может дать существенного прироста к качеству прогноза. Однако количество примеров данного набора данных значительно меньше, чем у полного, что может означать, что для эффективного использования такого подхода требуется больше данных. Имеет смысл в дальнейших исследованиях рассмотреть варианты получения дополнительных данных требуемого типа путем различных видов аугментации или с помощью генеративных нейросетевых систем. Из того, что данная модель показывает результат не хуже модели, обученной на полном наборе данных, может следовать, что данные, включённые в эту группу (значения индекса Dst от -50 нТл и ниже), вносят наибольший вклад в качество прогноза моделей.

8. Выводы

В работе исследованы некоторые новые аспекты решения задачи нейросетевого прогнозирования состояния магнитосферы на примере Dst индекса. Получены следующие основные результаты:

1) Произведено сравнение результатов 6 моделей машинного обучения и тривиальной инерционной модели в качестве репера при прогнозировании значения индекса Dst с горизонтом в 5 часов. Выбрана наиболее эффективная модель – многослойный персептрон.

2) Изучены различные методы решения проблемы переобучения моделей, после чего выбраны наиболее подходящие (ранний останов, а также dropout и нормализация пакета для глубоких моделей).

3) Проведен анализ результатов прогноза моделей, на основании которых построена модифицированная функции потерь. Необходимы дальнейшие исследования для определения методики эффективного применения этого подхода.

4) Показано, что при обучении на наборах данных, разделенных на различные группы, не удастся получить нейросетевые модели, которые бы устойчиво показывали заметно лучшее качество прогноза, чем модель, обученная на всех данных. Наиболее перспективными выглядят модели, обученные на наборе данных, соответствующих значениям индекса Dst - 50 нТл и ниже. Это направление работ нуждается в дальнейших исследованиях. Сделано предположение, что для успеха такого подхода требуется большее количество данных.

9. Список цитируемых источников

1. http://www.wdcb.ru/stp/geomag/indices_description.ru.html#dstind
2. http://www.wdcb.ru/stp/solar/solar_cycles.ru.html
3. <https://www.swpc.noaa.gov/products/ace-real-time-solar-wind>
4. <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content394/Content394.htm>
5. <https://neurohive.io/ru/osnovy-data-science/lstm-nejronnaja-set/>
6. [https://ru.wikipedia.org/wiki/%D0%A2%D1%80%D0%B0%D0%BD%D1%81%D1%84%D0%BE%D1%80%D0%BC%D0%B5%D1%80_\(%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B3%D0%BE_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D1%8F\)](https://ru.wikipedia.org/wiki/%D0%A2%D1%80%D0%B0%D0%BD%D1%81%D1%84%D0%BE%D1%80%D0%BC%D0%B5%D1%80_(%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B3%D0%BE_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D1%8F))

7. <https://education.yandex.ru/handbook/ml/article/linear-models>
8. <https://habr.com/ru/companies/mvideo/articles/782360/>