

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: Метод анализа нестационарных сигналов с применением
искусственного интеллекта

Студент гр. 7303

Давыдов А.А.

Преподаватель

Мандрикова Б.С.

Санкт-Петербург

2022

ЗАДАНИЕ НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Студент: Давыдов А.А.

Группа: 7303

Тема НИР: Метод анализа нестационарных сигналов с применением искусственного интеллекта

Задание на НИР: разработать приложение, которое позволит классифицировать состояние сигнала вторичных космических лучей

Сроки выполнения НИР: 01.09.2022 – 31.10.2022

Дата сдачи отчета: 27.10.2022

Дата защиты отчета: 31.10.2022

Студент

Давыдов А.А,

Руководитель

Мандрикова Б.С.

АННОТАЦИЯ

С ростом технических объектов и развитием научно-технической модернизации в настоящее время все сильнее ощущается проблема создания эффективных методов обработки и анализа нестационарных сложно структурированных данных. Наиболее сложной задачей является изучение и анализ природных данных, ввиду недостаточности знаний об исследуемых процессах и огромному числу воздействующих факторов. Целью данной НИР является разработка приложения, которое позволит классифицировать сигналы вторичных космических лучей, выявляя в них такие аномалии, как Форбуш эффект и GLE или определяя сигнал как спокойный.

СОДЕРЖАНИЕ

	Введение	5
1.	Постановка задачи	8
2.	Результаты работы в осеннем семестре	9
3.	План работы на весенний семестр	120
	Заключение	21
	Список использованных источников	22

ВВЕДЕНИЕ

С ростом технических объектов и развитием научно-технической модернизации в настоящее время все сильнее ощущается проблема создания эффективных методов обработки и анализа нестационарных сложно структурированных данных. Наиболее сложной задачей является изучение и анализ природных данных, ввиду недостаточности знаний об исследуемых процессах и огромному числу воздействующих факторов [1]. Анализ природных данных находит применение в разных сферах человеческой деятельности – физика, биология, медицина, экономика и др. Особую актуальность имеют методы, направленные на своевременное обнаружение и идентификацию аномалий. Примерами могут служить задачи оперативного распознавания аномалий в данных геофизического мониторинга – предсказание землетрясений, цунами, обнаружение предикторов магнитных бурь, аномалий геологической среды и других катастрофических явлений природы. Важным свойством таких методов является их способность к адаптации, обеспечивающая возможность обнаружения и идентификации быстрых изменений состояния системы или объекта, свидетельствующих о возникновении аномалий. Объектом научно-исследовательской работы являются космические лучи (КЛ). Основными источниками космических лучей являются звёзды и происходящие в них реакции. Солнце как ближайшая к нам звезда и источник термоядерных реакций [1] является основным объектом исследования природы космических лучей. По своей природе КЛ делятся на две категории - первичные космические лучи, то есть частицы, находящиеся в космическом пространстве и вторичные, являющиеся продуктом взаимодействия первичных КЛ с атмосферой Земли (в результате таких реакций порождаются широкие атмосферные ливни, которые можно наблюдать на расстоянии тысяч метров) [2]. Вариации космических лучей имеют сложную структуру. Анализ представленных данных показывает наличие высокого уровня шума, обусловленного погрешностями аппаратуры, а также помехами

природного происхождения, такими как осадки, ветер и т. д. В качестве примера на Рис.1 ниже показаны данные НМ высокоширотной станции «Инувик» (координаты: 68.36, -133.72, США) и среднеширотной станции «Москва» (координаты: 55.75, 37.6156, Россия) зарегистрированные в период спокойной солнечной активности



Рис. 1 Данные нейтронных мониторов

Частицы космического излучения, приходящие на Землю из космического пространства образуют поток, который непостоянен и зависит как от времени, так и от места наблюдения за космическими лучами. Изменения потока космических лучей во времени и в пространстве называют вариациями космических лучей, и основную роль в их формировании играет Солнце. Различают периодические вариации, к которым относят 22-летние, 11-летние, 27-дневные и солнечно-суточные вариации, которые связаны с солнечной активностью и которые можно наблюдать и регистрировать с заданной периодичностью [2]. Существуют также непериодические вариации, коротко называемые аномальными эффектами. К ним относят эффект Форбуша и солнечные вспышки. Хотя вероятность их наступления и повышается с ростом активности Солнца, но точно предсказать их пока не удаётся, что в свою очередь является проблемой, так как космические лучи влияют как на человека, так и электронику: системы теле-, радио и спутниковой связи, системы GPS и ГЛОНАСС, электрические сети, нефтегазовые комплексы и т.д. Существующие методы прогнозирования интенсивности космических лучей и предотвращения

негативных последствий не являются достаточно эффективными, что говорит об *актуальности* данной работы.

Основным прибором наземного наблюдения за космическими лучами являются нейтронные мониторы [3]. Принцип работы основывается на генерации нейтронов приходящими нуклонами и ядрами в результате ядерной реакции и дальнейшем подсчёте образовавшихся нейтронов. Регистрируемые данные можно интерпретировать как функцию интенсивности космических лучей от времени. Это дает возможность анализировать на предмет выявления аномальных эффектов.

Целью данной работы является разработка метода анализа данных нейтронных мониторов и выявления GLE-событий и эффектов Форбуша на основе применения нейронных сетей.

Объектом исследования являются данные нейтронных мониторов, представляющие собой интенсивность вторичных космических лучей.

Предметом исследования являются методы анализа данных нейтронных мониторов и обнаружения аномалий в космических лучах.

Практическая ценность работы заключается в создании метода, благодаря которому можно будет в оперативном режиме детектировать момент наступления аномальных эффектов в космических лучах, и предотвратить негативные последствия космической погоды, а также способствует развитию фундаментальных исследований в области космоса.

1. ПОСТАНОВКА ЗАДАЧИ

Задачи:

- 1) Создать программные утилиты для формирования размеченного набора данных для обучения НС с учителем.
- 2) Протестировать различные архитектуры нейронных сетей для разделения данных (сигнала КЛ) на три класса: GLE и Форбуш аномалии, спокойный сигнал. Описать математическую модель решаемой задачи. В качестве входных данных использовать минутные “сырые” данные с нейтронных мониторов.

2. РЕЗУЛЬТАТЫ РАБОТЫ В ОСЕННЕМ СЕМЕСТРЕ

2.1. Формирование обучающего набора данных

Для формирования обучающего набора данных использованы следующие общедоступные ресурсы:

- Архивные данные сайта прогноза космической погоды (с индексами степени шторма) URL: <https://www.spaceweatherlive.com/ru/arhiv.html>

Для проведения эксперимента, было выбрано 700 штормовых дней, 700 дней с умеренными аномалиями, 700 спокойных дней. Итого, размер обучающего набора данных – 2100 записей. Для дальнейших экспериментов также были выкачаны все возможные данные космической погоды (с 1996 года по настоящее время). Датасет размером 2100 записей был сформирован из соображений отбора данных по классу с наименьшим числом записей – штормовые дни (700 зарегистрированных дней).

- Архивные данные регистрации вторичных КЛ на станциях с нейтронными мониторами URL: <https://www.nmdb.eu/nest/>

Согласно выбранным датам в пункте выше, были выкачаны 2100 суток с минутными данными регистрации КЛ. Данные брались с нейтронного монитора станции “OULU”. В ряде выгруженных данных имелись дубликаты минутных данных, либо же пропуски. Для устранения этих ошибок была осуществлена предварительная обработка данных. Пропуски в данных заполнялись медианным значением соседних данных с окном в 50 значений, повторы в данных подлежали простой процедуре удаления.

2.2. ИС классификации нестационарных сигналов КЛ

Для первичных экспериментов были взяты “сырые” минутные данные КЛ для попыток выделения признаков в сигналах с условием присутствия в них зашумленных данных. Затем, для решения задачи классификации нестационарных сигналов КЛ были использованы следующие архитектуры:

- Deep Belief Network
- CNN + LSTM

2.2.1. DBN (Deep Belief Network)

Математическая модель

Сеть глубоких убеждений (DBN)[1] – представляет собой генеративную модель, в основе которой лежит глубокая архитектура некоторого числа стеков ограниченных машин Больцмана (RBM)[2]. Архитектура данной сети представлена на Рис.2.

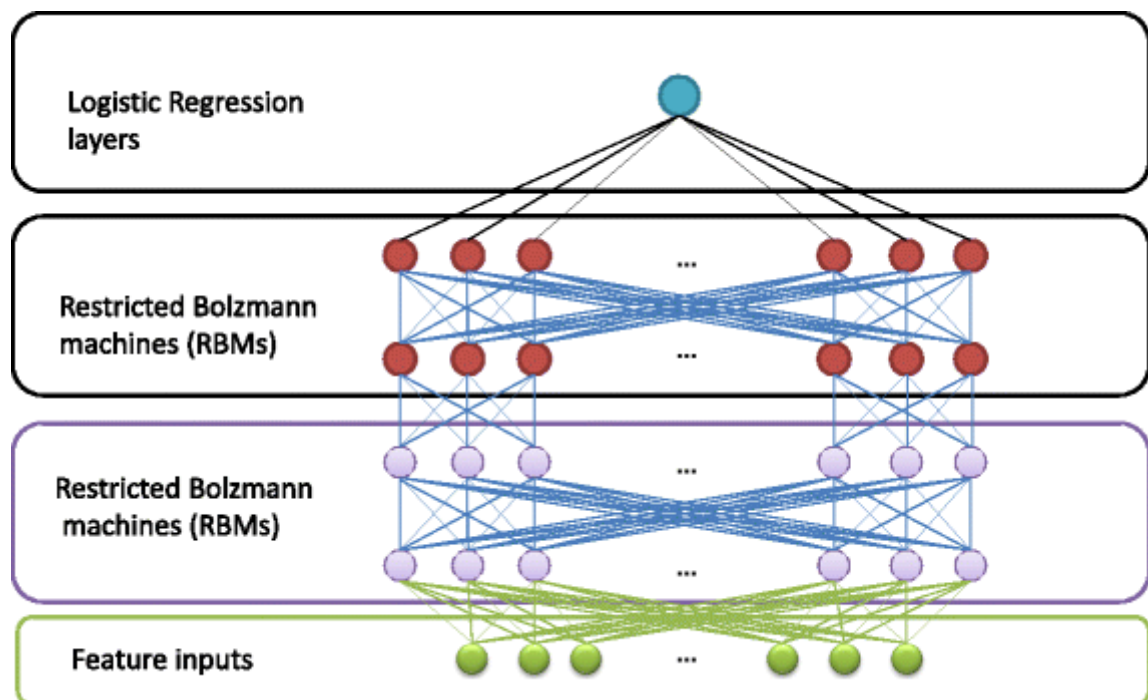


Рис. 2 Сеть глубоких убеждений

Рассмотрим подробнее принцип работы стека RBM. С помощью каждой модели RBM в стеке происходит выполнение нелинейного преобразования по отношению к входным данным. На выходе модели RBM получается набор векторов, которые передаются следующей в последовательности RBM в

качестве входных параметров. Таким образом, достигается гибкость архитектуры сети DBN, упрощая их расширение.

Обучаясь на тренировочной выборке без надзора, DBN учиться восстанавливать входные данные с использованием вероятностного подхода. Данный этап тренировки называется предварительным. После его завершения слои выполняют функцию детектирования объектов. За этапом предварительной тренировки следует обучение DBN с учителем для разрешения задачи классификации входных данных.

Процесс предварительной тренировки DBN

Так как DBN включает в себя стек моделей RBN, для их обучения используется метод контрастивной дивергенции (CD)[3]. Таким образом, для обучения одной RBM используется следующее уравнение в градиентном спуске (одно обновление весов):

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial \log(p(v))}{\partial w_{ij}}$$

- Вероятность видимого вектора:

$$p(v) = \frac{1}{Z} \sum_h e^{-E(v,h)}$$

- Z – функция разделения (необходима для нормализации)
- $E(v, h)$ – энергетическая функция состояния сети

Шаги метода CD:

1. Подача на вход обучающий вектор
2. Обновление скрытых юнитов с учетом видимых:

$$p(h_j = 1 | \mathbf{V}) = \sigma(b_j + \sum_i v_i w_{ij})$$

- \mathbf{b}_j – вектор смещения для \mathbf{h}_j
- σ – сигмовидная функция активации

3. Обновление видимых юнитов с учетом скрытых

$$p(v_i = 1 | \mathbf{H}) = \sigma(a_i + \sum_j h_j w_{ij})$$

- \mathbf{a}_i – вектор смещения для \mathbf{V}_i

4. С учетом восстановленных видимых юнитов, необходимо снова обновить скрытые юниты. Необходимо то же уравнение, что на шаге 2

5. Обновить веса

$$\Delta w_{ij} \propto \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstruction}}$$

Описанный выше алгоритм применяется к первой RBM в стеке. После того как первая RBM обучена, ее выходные данные передаются в видимый слой следующей RBM. Данный процесс повторяется до тех пор, пока не будут обучены все RBM.

Экспериментальный результат

Так как при обучении используются минутные суточные данные, то входной вектор имеет форму: (1440).

Опишем конфигурацию запуска:

1. Размерность стека из RBM: 4
2. Количество скрытых единиц в слое: 512
3. Число эпох предварительной тренировки: 40
4. Число эпох тренировки сети классификации: 50
5. Размер пакета: 32
6. Скорость обучения RBM: 10^{-3}
7. Скорость обучения сети классификации: 10^{-3}

8. Алгоритм оптимизации: SGD[4]

Метрика точности на тестовых данных составила: 40%. Если из обучающего и тестового набора убрать спокойные дни, тогда точность возрастает до 70%. Таким образом, мы понимаем, что при обучении сеть плохо различает спокойные дни от дней с умеренной степенью штормовых возмущений.

2.2.2. CNN + LSTM

Математическая модель

CNN (Convolutional neural network)

CNN – сверточная нейронная сеть [5], в основе которой лежит алгоритм глубокого обучения, нацеленный на присвоение важности(веса и смещения) аспектам во входных данных. Архитектура данной сети представлена на Рис.3.

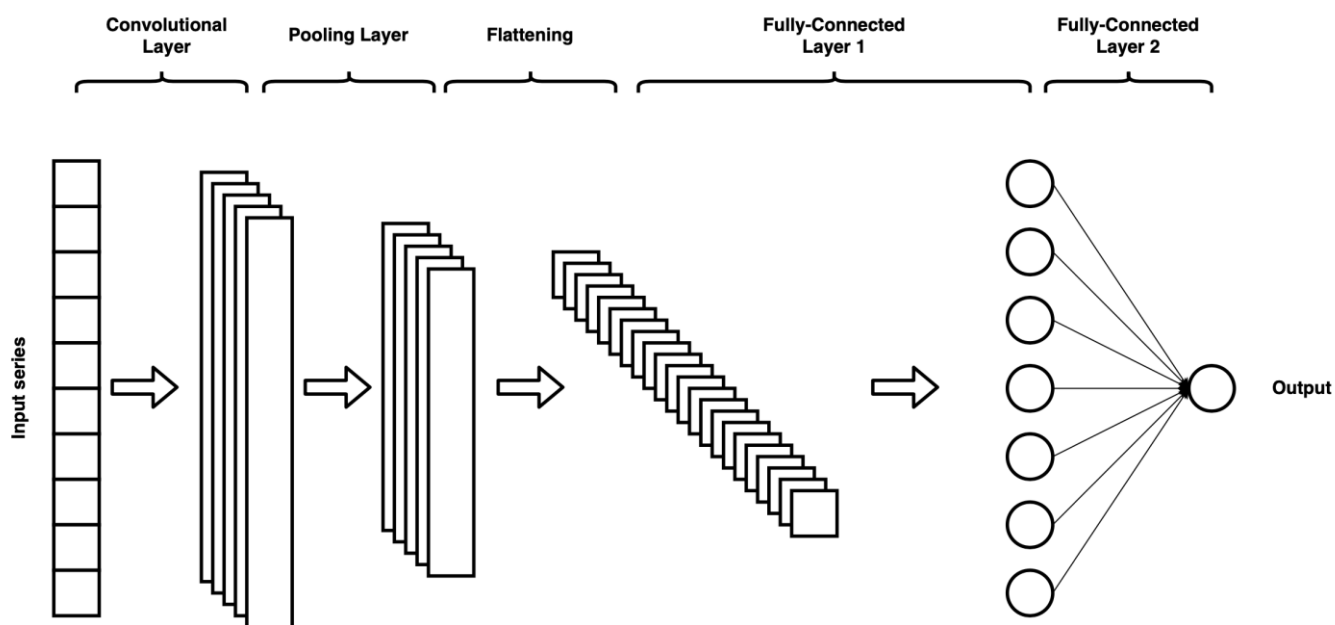


Рис. 3 Сверточная сеть (Conv1D)

Рассмотрим принцип работы сверточного слоя(Convolutional Layer). Слой свёртки имеет ряд фильтров, ядро которых обрабатывает фрагменты предыдущего слоя. Производится сумма результатов поэлементного

произведения для каждого фрагмента. Веса ядра свёртки обновляются в процессе обучения. Более наглядно процедура, которую выполняет сверточный слой, изображена на Рис.4.

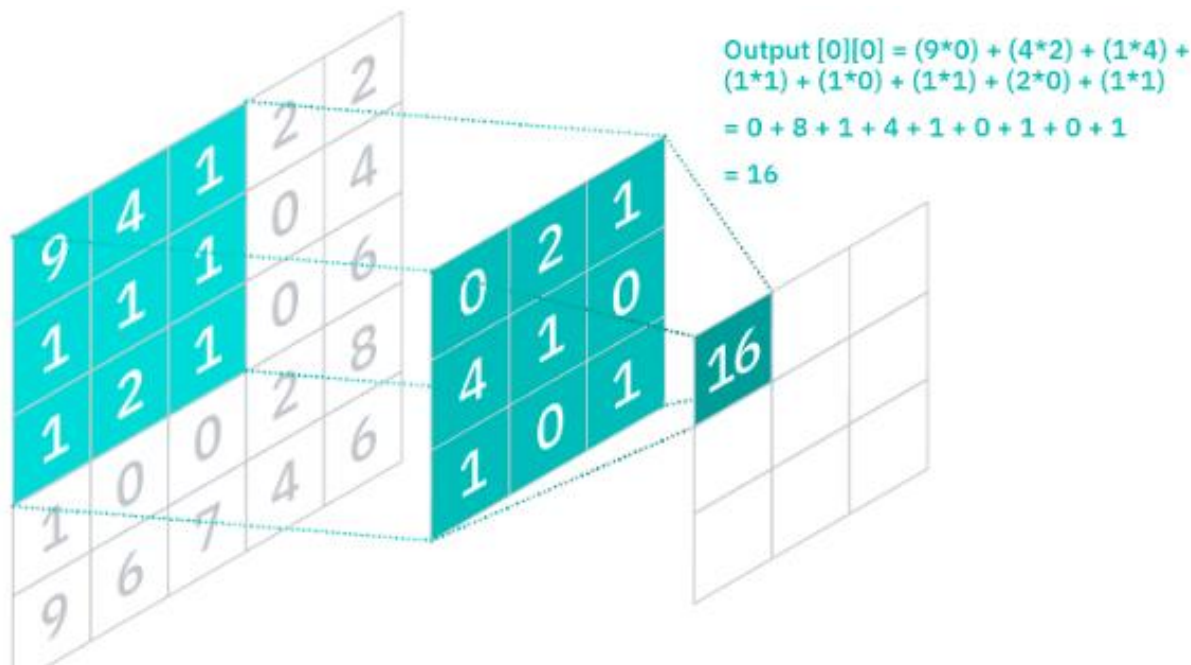


Рис. 4 Принцип работы свёрточного слоя

Важная деталь слоя свертки — это достаточно небольшое число параметров. Например, если размер входных данных 30000, а размер ядра 3x3, также сверточный слой имеет 6 фильтров, тогда число обучаемых параметров данного слоя $9 \times 6 = 54$. Такое количество параметров в сравнении с его числом в полносвязной нейронной сети можно считать существенной разницей.

Стоит также рассмотреть принцип работы слоя субдискретизации (Pulling Layer). Слой субдискретизации выполняет нелинейное уплотнение карты признаков. Карта признаков — это выходной результат чередования слоев свертки и субдискретизации. Как правило, уплотнение данных уплотняет группу единиц данных 2x2 в одну единицу. Как итог данного преобразования, происходит уменьшение пространственного объема входных данных.

На Рис. 5 показан пример работы слоя субдискретизации с использованием функции максимума. Это означает, что из группы 2x2 мы выбираем максимальное значение, преобразуем заданную группу в данное значение.

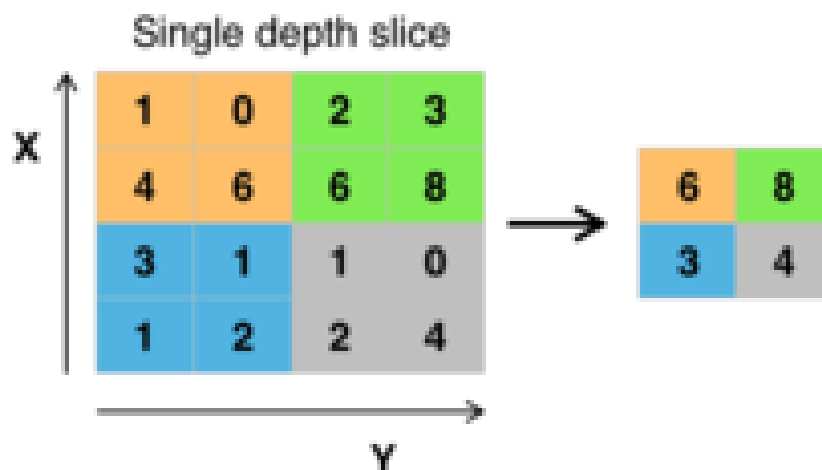


Рис. 5 Субдискретизация

По завершению прохождения данных через сверточные и уплотняющие слои, на выходе получается абстрактная карта признаков. Данная карта признаков имеет небольшую размерность и хранит обобщенные особенности входных данных. Эти выходные данные объединяются и передаются в плотную сеть прямого распространения, например, для того, чтобы произвести обучение для разрешения задачи классификации.

Стоит также упомянуть, о функции активации, которая применяется к скалярному результату каждой свертки. Традиционно в CNN использовался гиперболический тангенс или сигмовидная функция. Однако исследования в 2000х годах показали эффективность использования RELU. Это значительно ускоряет процесс обучения и упрощает вычисления.

2.2.3. LSTM (Long short-term memory)

LSTM [6] – представляет собой разновидность архитектур рекуррентных нейронных сетей [7], отличительной особенностью которой является

соединения с обратной связью и возможность обработки последовательности данных. В частности, LSTM хорошо справляется с обработкой последовательности временных рядов. Архитектура памяти данной сети представлена на Рис.6.

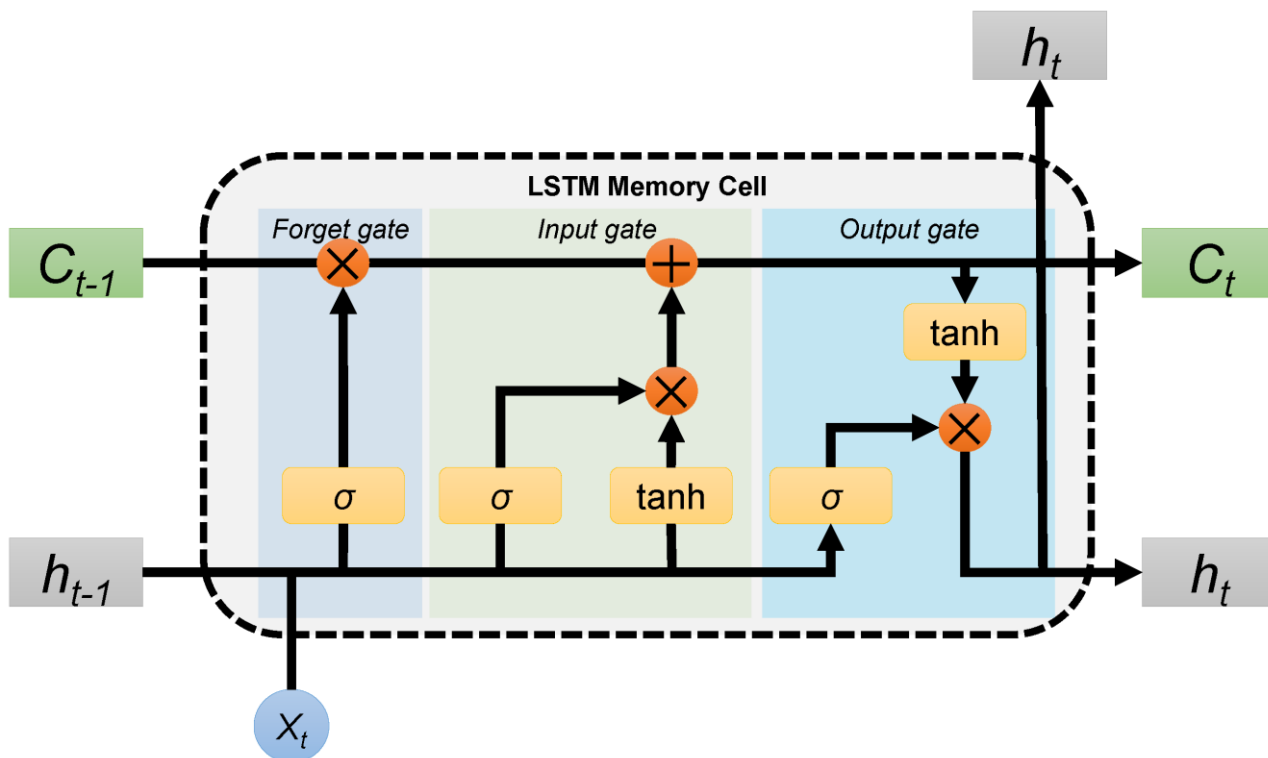


Рис. 6 LSTM-ячейка

Важной составляющей LSTM является LSTM-модуль. Он представляет собой рекуррентный модуль сети, который имеет возможность запоминать значения как на короткий, так и на длительный период. Данная возможность достигается тем, что внутри рекуррентных компонентов LSTM-модуля не используется функцию активации. Это позволяет значению, которое хранится в скрытом состоянии не размывается во времени, а градиент не исчезает при обратном распространении.

На практике LSTM-модули группируются в “блоки”, которые содержат различные LSTM-модули. Сами же LSTM-блоки состоят из 3 или 4 «вентилей», Они выполняют функцию просеивания информации на входах и выходах. Для того, чтобы задать направления оперирования вентилей, используются матрицы весов (W и U). Данные матрицы весов задаются для входных данных и

выходных данных с предыдущего шага соответственно. LSTM-блок регулирует свою память как функцию значений матриц W и U .

Рассмотрим подробнее операции внутри LSTM-блока:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

- x_t - входной вектор данных
- h_t — выходной вектор
- c_t – вектор состояний
- W, U, b – матрицы параметров и смещение
- f_t - вектор вентиля забывания, вес запоминания старой информации
- i_t - вектор входного вентиля, вес получения новой информации
- o_t - вектор выходного вентиля, кандидат на выход
- σ_g - сигмовидная
- σ_c – гиперболический тангенс
- σ_h – гиперболический тангенс

Стоит также отметить преимущества LSTM при тренировке по сравнению с традиционными рекуррентными нейронными сетями. Преимущество заключается в методе обратного распространения ошибки, развёрнутого во времени. Особенностью данного метода является то, что при обратном распространении ошибки в направлении от выходного слоя, ошибка записывается внутри памяти блока. Происходит непрерывное «скармливание»

ошибки обратно каждому из вентиляй, до тех пор, пока в процессе обучения они не начнут отбрасывать значение.

Экспериментальный результат

Так как при обучении используются одномерные (1D) сверточные слои, а входные данные являются распределенными во времени данными с числом измерений 1440 единиц, то при построении сверточных слоев необходимо преобразовать форму входных данных в (3D) согласно Keras API для TimeDistributed layer. Таким образом, было решено использовать преобразования формы входных данных из (1440) в (24, 6, 10).

Опишем конфигурацию запуска:

1. Количество слоев CNN: 2
2. Количество фильтров: 64
3. Размерность ядер в свертках: 3x3
4. Размерность матрицы субдескрипции: 2x2
5. Dropout: 0.5
6. Число нейронов в LSTM слое: 100
7. Число нейронов в плотном слое сети прямого распространения: 100
8. Алгоритм оптимизации: Adam[8]
9. Скорость обучения: 10^{-3}
10. Число учебных эпох: 20

Метрика точности на тестовых данных составила: 44,3%. Если из обучающего и тестового набора убрать спокойные дни, тогда точность возрастает до 71,5%. Таким образом, мы понимаем, что при обучении сеть плохо различает спокойные дни от дней с умеренной степенью штормовых возмущений.

3. ПЛАН РАБОТЫ НА ВЕСЕННИЙ СЕМЕСТР

Реализовать архитектуру автокодировщика (Autoencoder) для того, чтобы подавать в описанные выше архитектуры классификации вместо исходного сигнала – сигнал, преобразованный кодировщиком (Encoder). Использование нейронной сети Автокодировщик дает возможность извлечения зависимостей в данных (за счет минимизации ошибки восстановления) и подавления шума. Предполагается, что это позволит сетям классификации лучше отличать спокойный сигнал от небольших штормовых возмущений. Таким, образом, на выходных данных кодировщика будет обучаться сеть классификации. Итогом работы будет приложение, в котором можно по суточным данным предсказать наличие магнитной бури или определить сутки как спокойные. Данные будут вводиться файлом, выходные данные будут отображать класс магнитных возмущений.

ЗАКЛЮЧЕНИЕ

В ходе работы НИР в осеннем семестре были реализованы 2 архитектуры сетей классификации исходных данных. Первая – на основе стека ограниченных машин Больцмана и сети прямого распространения для последующей процедуры классификации. Вторая – на основе сверточных слоев, сжимающих входные данные в карту признаков и передающие ее затем в LSTM слой и, наконец, сеть прямого распространения. Полученная метрика точности показала, что сети классификации плохо отличают спокойный сигнал от небольшого возмущения. Это объясняется тем, что исходные данные содержат шум и не имеют чистого образа. Именно поэтому была выдвинута гипотеза о обучении сети автокодировщика для того, чтобы повторно обучить сети классификации на выходных данных кодировщика с расчётом на то, что это позволит сетям лучше тренироваться на данных с содержанием шума.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Asifullah Khan. Deep Belief Networks, 2016, 22-23 С.
2. Aleshin M. S., Sharypov R. E.Restricted Boltzmann Machine Training, 2021, 190-200 С.
3. Geoffrey E. Hinton. On Contrastive Divergence Learning, 2021, 2-4 С.
4. Sebastian Ruder. An Overview Of Gradient Descent Optimization Algorithms, 2017, С. 4
5. С.А. Вакуленко, А.А.Жихарева. Практический Курс По Нейронным Сетям. 2018, С. 57
6. Ф.М. Гафаров, А.Ф. Галимянов. Искусственные Нейронные Сети И Их Приложения. 2018, С. 78-92
7. Ф.М. Гафаров, А.Ф. Галимянов. Искусственные Нейронные Сети И Их Приложения. 2018, С. 74-77
8. Sebastian Ruder. An Overview Of Gradient Descent Optimization Algorithms, 2017, С. 7-8