

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Суворов Антон Дмитриевич
Должность: Ректор
Дата подписания: 13.06.2025 20:30:47
Уникальный программный ключ:
a39bdb15d680d5b0adb1ced0af5e1efb14747dcd

СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ (Сколтех)

Рабочая программа дисциплины	Модели последовательных данных
------------------------------	--------------------------------

Преподаватель	Зайцев Алексей Алексеевич, старший преподаватель, д.ф-м.н.
---------------	--

Аннотация

Описание курса

В этом курсе мы обсудим основные направления современных исследований в машинном обучении на основе последовательных данных. Курс охватывает как базовые алгоритмы обработки последовательностей, так и современные подходы на основе глубинного обучения, которые обычно не обсуждаются на курсах машинного обучения и глубинного обучения.

Кроме того, мы сделаем акцент на том, как с помощью таких подходов может решаться широкий набор прикладных задач. Мы исследуем, как эксперты используют эти концепции для прогнозирования временных рядов, обнаружения отказов механизмов, оценки сходства нефтяных скважин и многое другое. Чтобы предоставить полную картину последовательных моделей данных, мы приводим методы, основы и их использование в конкретных приложениях.

Мы предполагаем, что участники курса знают основы машинного и глубинного обучения, но попытаемся сделать курс как можно более доступным для широкой аудитории магистров и аспирантов Сколтеха.

1. Основная информация

Академический уровень курса	Магистратура Аспирантура
Количество кредитов	6

Предварительные требования к курсу / рекомендации

- Знакомство с Python и машинным обучением
- Опыт работы с платформами глубокого обучения, такими как Pytorch
- Математическая компетентность.

Тип оценки - дифференцированная

Отображение оценок в процентах

A:	86
----	----

B:	76
C:	66
D:	56
E:	46
F:	0

2. Содержание курса

Тема	Краткое содержание	Лекции (час)	Семинары (час)	Лабораторные занятия (час)	Самостоятельная работа (час)
Введение	В недавнем прошлом исследования приложений глубокого обучения для обработки данных временных рядов были ограничены. Мы расскажем о том, как учащиеся могут начать обрабатывать временные ряды, и о перспективных областях исследований.	2	2		10
Однозначность и стационарность. Разложение временных рядов	Классические подходы могут помочь нам понять внутренние зависимости данных. Одним из наиболее известных традиционных подходов к анализу временных рядов является модель ARIMA, которую мы подробно рассмотрим с ее модификациями.	3	3		3
Повторяющиеся нервные сети	Курс представляет собой введение в Рекуррентные нейронные сети (RNNS) и их две основные архитектуры: с долговременной Кратковременной памятью (LSTM) и со стробируемым рекуррентным модулем (GRU). Мы обсуждаем их специфические характеристики, реализации и различные приложения.	1,5	1,5		10
Одномерный CNNs	Мы рассматриваем адаптацию сверточных нейронных сетей (CNN) для обработки временных рядов для решения общих проблем с RNN.	1,5	1,5		10
Внимание. Трансформаторы	Механизм attention зарекомендовал себя как прорыв в моделировании последовательности действий. Мы обсудим, как он работает, углубимся в архитектуру Transformer и рассмотрим его преимущества и недостатки.	1,5	1,5		10
Моделирование длинных последовательностей	Трансформаторы требуют больших вычислительных затрат и затрат на память, если мы стремимся к длительной концентрации внимания. Мы рассмотрим методы, позволяющие обрабатывать длинные последовательности, и то, как они развивались за последние несколько лет.	1,5	1,5		10

<p>Модели глубинного пространства состояний. Существующие контрольные показатели</p>	<p>Длинные последовательности могут быть непомерно дорогими для преобразователей. Здесь мы обсуждаем альтернативу, основанную на представлении последовательности в виде ODE, модели пространства состояний.</p>	1,5	1,5		10
<p>Самостоятельное обучение для последовательных данных</p>	<p>Мы рассмотрим концепцию и детали реализации самоконтроля, обсудив два его основных типа: генеративный и контрастивный. Мы также рассмотрим их преимущества и недостатки, а также современные решения и ограничения.</p>	1,5	1,5		10
<p>Изучение показателей для последовательных данных</p>	<p>Многие подходы к машинному обучению требуют измерения расстояния между точками данных. Метрическое обучение направлено на автоматическое построение метрик расстояния, специфичных для конкретной задачи, из (слабо выраженных) контролируемые данные с помощью машинного обучения. Это особенно интересно для данных временных рядов.</p>	1,5	1,5		10
<p>Обнаружение точки изменения</p>	<p>Задача определения точки изменения (CPD) заключается в определении того, существенно ли меняется поведение ряда данных и когда это происходит. Это важная задача, поскольку она может помочь нам понять причины изменений в данных, например, выявить аварийные ситуации. Мы обсудим теоретическую постановку задачи CPD, а также основные и продвинутые методы ее решения.</p>	1,5	1,5		10
<p>Обработка звука и речи</p>	<p>Аудиоданные представляют собой один из самых больших классов типов данных. Мы выделяем такие данные в отдельные методы, поскольку прямое применение стандартных алгоритмов последовательной обработки данных уступает подходам, основанным на специфике модальности.</p>	3	3		5
<p>Обработка последовательности событий. Нейронные Процессы в временных точках</p>	<p>Временной точечный процесс – это набор математических точек, случайным образом расположенных в математическом пространстве на временной шкале. Мы рассмотрим его определение и моделирование с математической точки зрения. Мы также обсудим классические подходы и реализации глубокого обучения.</p>	1,5	1,5		10
<p>Пространственно-временная</p>	<p>Пространственно-временные данные - один из самых больших классов типов данных. Мы обсудим</p>	1,5	1,5		5

обработка данных. ConvLSTM	специальные инструменты для их анализа и визуализации, а также модели для их обработки.				
Оценка неопределенности для последовательных данных. Начальная загрузка	В некоторых задачах (например, в медицине) точечной оценки недостаточно. Следовательно, для прогнозирования нам необходимо оценить неопределенность. Здесь мы обсудим, как модель может включать оценку неопределенности в случае зависимых последовательных данных	1,5	1,5		5

3. Результаты обучения

Результаты обучения в Сколтехе указаны в соответствии со структурой результатов обучения в Сколтехе

1. ФУНДАМЕНТАЛЬНЫЕ ЗНАНИЯ

1.1. Знание математики и естественных наук

1.4. Междисциплинарное мышление, структура знаний и интеграция

2.1. ПОЗНАНИЕ И СПОСОБЫ РАССУЖДЕНИЯ

2.1.2. Системное мышление

2.1.3. Творческое мышление

2.1.4. Принятие решений (неоднозначных, срочных и т.д.)

2.1.5. Критическое мышление и метапознание

2.2. ОТНОШЕНИЕ И ПРОЦЕСС ОБУЧЕНИЯ

2.2.1. Инициатива и готовность идти на соответствующий риск

2.2.2. Готовность принимать решения в условиях неопределенности

2.2.3. Ответственность, интенсивность, настойчивость, безотлагательность и воля к достижению поставленных целей

2.2.4. Находчивость, гибкость и способность адаптироваться

2.2.5. Самосознание и стремление к самосовершенствованию, обучению на протяжении всей жизни и воспитанию

2.3. ЭТИКА, СПРАВЕДЛИВОСТЬ И ДРУГИЕ ОБЯЗАННОСТИ

2.3.6. Приверженность социальному и профессиональному поведению

3.1. КОММУНИКАЦИЯ В МЕЖДУНАРОДНОЙ СРЕДЕ

3.1.3. Устная презентация и обсуждение

3.1.4. Вопросы, слушание и диалог

3.1.5. Общение на английском языке в научной, деловой и общественной среде

3.2. КОМАНДНАЯ РАБОТА И ЛИДЕРСТВО

3.2.2. Командная работа и управление проектами

3.2.3. Командная координация, принятие решений и лидерство

3.2.4. Рост и эволюция команды

3.2.5. Техническое и междисциплинарное сотрудничество

3.3. СОТРУДНИЧЕСТВО И ИЗМЕНЕНИЯ

3.3.1. Установление разнообразных связей и сетевого взаимодействия

3.3.2. Понимание различных ролей, перспектив и интересов

- 3.3.3. Переговоры и разрешение конфликтов
- 3.3.4. Вовлечение заинтересованных сторон
- 3.3.5. Осуществление намеренных изменений
- 4.1. ПОНИМАНИЕ ГЛОБАЛЬНОГО СОЦИАЛЬНОГО, ЭКОЛОГИЧЕСКОГО И ДЕЛОВОГО КОНТЕКСТА
- 4.1.3. Понимание технических продуктов, систем и инфраструктуры отрасли
- 4.2. ДАЛЬНОВИДНОСТЬ — ИЗОБРЕТЕНИЕ НОВЫХ ТЕХНОЛОГИЙ ПОСРЕДСТВОМ ИССЛЕДОВАНИЙ
- 4.2.2. Фундаментальные исследования, ведущие к новым научным открытиям
- 4.2.3. Исследования, направленные на разработку новых технологий
- 4.2.4. Представление о пользе новой науки и техники
- 4.2.5. Разработка концепций и внедрение их в практику
- 4.3. ВИДЕНИЕ — РАЗРАБОТКА КОНЦЕПЦИИ И ПРОЕКТИРОВАНИЕ УСТОЙЧИВЫХ СИСТЕМ
- 4.3.2. Определение и формулирование целей и задач
- 4.3.6. Управление процессом проектирования, включая планирование, оценку проекта и эффективное принятие решений

4. Задания и выставление оценок

Требование к физической посещаемости (% от числа занятий)	80
---	----

К требованиям о личном посещении:

студентам разрешается пропускать занятия только по уважительной причине, имеющей обоснование, например, медицинское заключение

Тип назначения	Краткое содержание задания	% от итоговой оценки за курс
Test/Quiz	В конце почти всех четных занятий учащиеся проходят контрольную работу в классе или тест на дому. В тестах есть несколько вариантов ответов, и они решаются онлайн на Canvas. Цель этих тестов - закрепить знания, полученные на занятиях. Общее количество тестов равно 6.	20
Отчет	Предварительный отчет. В начале семестра преподаватель представляет студентам набор проектов. Студенты формируют группы из трех-четырёх человек и работают над выбранными проектами. Проекты могут включать в себя повторные исследования, новые работы или аналитические задания. Студенты должны начать работу над проектами с подготовки нескольких разделов, которые будут включены	5

	в отчет. Предварительный отчет должен включать, по крайней мере, следующие разделы: Введение, Связанные с проектом Работы, Методы (включая данные Описание и методология с описанием моделей) и Эксперименты (включая, по крайней мере, описание постановки экспериментов).	
Отчет	Студенты должны написать отчет на 8 страницах в две колонки, описывающий их работу и результаты, включая графики и библиографию. Этот отчет должен быть высокого качества и сопоставим с публикацией в небольшом издании, посвященном исследованиям в области машинного обучения. Цель состоит в том, чтобы представить доклады для публикации на конференциях по машинному обучению.	15
Презентация	Одно из заключительных занятий будет посвящено устным презентациям проектов. Студенты должны подготовить 10-минутную презентацию, описывающую идею своего проекта и полученные результаты. Затем у них будет 10-минутная сессия вопросов и ответов.	10
Отчет	Рецензирование. Учащиеся должны провести рецензирование и оценить работу своих сверстников.	5
Домашние задания	Наш курс будет состоять из двух оценочных заданий, соответствующих классическим моделям (например, ARIMA) и трансформерам. Код должен соответствовать PEP-8 стандарт, быть воспроизводимым (requirements.txt является обязательным) и должен быть подробно описан.	25
Промежуточный экзамен	Промежуточный экзамен	20

5. Критерии оценки

<u>Задание 1 Типа</u>	Командный проект
------------------------------	------------------

Пример задания 1

Являются ли архитектуры трансформаторов RNNs?

- да
- нет

Каковы некоторые недостатки RNN?

- Они не подходят для длинных последовательностей
- Они плохо поддаются обобщению
- Их можно использовать только для задач прогнозирования
- У них нет недостатков

Критерии оценки для задания 1

В конце некоторых занятий участники должны пройти тест в Canvas. Тест будет коротким, с несколькими вариантами ответов и оценками. Каждый из тестов может быть пройден только один раз. Экзамен проводится индивидуально. Недопустимы любые формы академической недобросовестности.

Продолжительность большинства экзаменов составляет около 45 минут. Анкета содержит вопросы, связанные с содержанием, пройденным в тот же день и на предыдущем занятии.

<u>Задание 2 Типа</u>	Отчет
------------------------------	-------

Пример задания 2

Предварительный отчет. Данный результат представляет собой отчет, подтверждающий глубокое понимание цели выбранного проекта. Отчет должен соответствовать шаблону, предоставленному организаторами курса, и содержать, по крайней мере, следующие разделы: Введение, Сопутствующая работа, методы (включая описание данных и методологию с описанием моделей) и эксперименты (включая, по крайней мере, описание постановки экспериментов). Отчет должен быть представлен в формате Canvas. Он не должен содержать прямых формулировок из статей, а все теоретические основы должны быть основаны на существующих статьях с правильным оформлением ссылок. Отчет должен содержать, по крайней мере, следующие разделы: Введение, Сопутствующая работа, методы (включая описание данных и методологию с описанием моделей) и эксперименты (включая, по крайней мере, описание постановки экспериментов).

<u>Задание 3 Типа</u>	Отчет
------------------------------	-------

Пример задания 3

Предоставьте отчет о проекте. Этот результат должен быть 8-страничным отчетом в две колонки с достоверным, логичным и высококачественным текстом, описывающим проект и полученные результаты.

Критерии оценки для задания 3

Отчет должен быть представлен в формате Canvas. Качество работы должно соответствовать качеству публикации, которая может быть опубликована в небольшом издательстве.

<u>Задание 4 Типа</u>	Презентация
------------------------------	-------------

Пример задания 4

Этот результат состоит из представленного файла в формате pdf/pptx со слайдами и 10-минутной устной презентации. Каждый член команды должен принять участие в презентации проекта.

Критерии оценки для задания 4

Предоставьте файл pdf/pptx со слайдами в формате Canvas. Участие в семинаре обязательно. Оценка за презентацию будет выставлена на основе ваших навыков презентации, качества изложения, общей экспозиции и представленных результатов.

<u>Задание 5 Типа</u>	Отчет
------------------------------	-------

Пример задания 5

Экспертная оценка. Экспертная оценка отчета другой команды.

Критерии оценки для задания 5

Рецензирование доклада другой команды должно быть представлено в формате Canvas и включать объективную оценку работы команды с конструктивными комментариями и предложениями. Этот вид деятельности аналогичен работе рецензентов, принимающих решения о публикации статей.

<u>Задание 6 Типа</u>	Домашние задания
------------------------------	------------------

Пример задания 6

Прогнозирование временных рядов с помощью модели ARIMA.

Критерии оценки для задания 6

Наш курс будет состоять из двух контрольных заданий, посвященных классическим моделям (например, ARIMA) и Трансформаторам. Код должен

соответствовать стандарту PEP-8, быть воспроизводимым (requirements.txt обязательно) и быть подробно описан.

<u>Задание 7 Типа</u>	Промежуточный экзамен
------------------------------	-----------------------

Пример задания 7

Промежуточный экзамен

Критерии оценки для задания 7

Вопросы будут связаны с темами курса.

<u>Задание 8 Типа</u>	Участие в занятиях
------------------------------	--------------------

Обязательное

6. Учебники и интернет-ресурсы

Необходимые учебники	ISBN-13 (or ISBN-10)
Ian Goodfellow, Yoshua Bengio and Aaron Courville. Deep Learning, 2015.	9780262035613
Goldberg, Yoav. "Neural network methods for natural language processing." Synthesis Lectures on Human Language Technologies 10.1 (2017): 1-309.	9781627052986
Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, Danilo P. Mandic and Jonathon A. Chambers. 2001.	9780471495178
Документы	DOI or URL
Dey, R. and Salemt, F.M., 2017, August. Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international Midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.	
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. And Polosukhin, I., 2017. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).	

Веб-ресурсы (ссылки)	Описание
Dey, R. and Salemt, F.M., 2017, August. Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.	
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).	

7. Оборудование

Программное обеспечение
Python 3.6+ Jupyter Библиотеки python с открытым исходным кодом, включая библиотеки для машинного обучения: PyTorch, pandas, Matplotlib, seaborn, NumPy

Оборудование
Ноутбук

8. Дополнительные примечания

Комментарии

Мы будем следовать этим правилам при подготовке окончательных работ:

- (1) Если вы не представите ПРОЕКТ, ОТЧЕТ или ПРЕЗЕНТАЦИЮ вовремя, ваша итоговая оценка за все из них будет равна нулю. Три результата являются результатом многонедельной работы. Таким образом, это должна быть совместная заявка. Мы не принимаем заявки с опозданием. Планируйте свою рабочую нагрузку.
- (2) Предварительная оценка ПРОЕКТУ и ОТЧЕТУ будет выставлена на основании схемы оценки, утвержденной в ваших проектных предложениях. Эта оценка может быть снижена в зависимости от их качества, такого как структура, полнота, логика и качество кода.
- (3) Пожалуйста, имейте в виду, что даже если ваша оценка составляет 100% в соответствии со схемой выставления оценок за ПРОЕКТ и ОТЧЕТ, в конечном итоге вы можете получить ноль, если ваши эксперименты будут неполными, или результаты не смогут быть воспроизведены, или в них не будет надлежащих объяснений и кода.
- (4) Оценки за ПРОЕКТ, ОТЧЕТ и ПРЕЗЕНТАЦИЮ в большинстве случаев будут одинаковыми для всех членов команды. По этой причине рабочая нагрузка должна быть поровну распределена между членами команды.
- (5) В некоторых исключительных случаях мы можем изменять оценки отдельных участников пропорционально вкладу каждого члена команды.