

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:
a39bdb15d680d5b0adb1ced0af5e1efb14747dc0

СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ (Сколтех)

Рабочая программа
дисциплины

Введение в рекомендательные системы

Преподаватель

Фролов Евгений

Аннотация

Этот курс дает широкий обзор темы рекомендательных систем и охватывает как теоретические, так и практические вопросы их разработки в рамках ключевых направлений. Цель курса - познакомить студентов с различными типами рекомендательных систем, подходами к их созданию и применению, и помочь им усвоить основные принципы построения практичных решений. В рамках данного курса студенты научатся перекладывать задачи, возникающие в реальных сервисах персонализации, на язык математических формулировок и строить соответствующие решения на основе подходящих методик, алгоритмов и программных инструментов.

1. Основная информация

| | |
|-----------------------------|--------------|
| Академический уровень курса | Магистратура |
| Количество кредитов | 3 |

Предварительные требования к курсу / рекомендации

Курс численной линейной алгебры. Навыки программирования на Python как минимум на среднем уровне, умение работать с пакетами Numpy, Scipy, Pandas, Matplotlib.

Тип оценки - дифференцированная

Отображение оценок в процентах

| | |
|----|----|
| A: | 86 |
| B: | 76 |
| C: | 66 |
| D: | 56 |
| E: | 46 |
| F: | 0 |

| Вид задания | Вес активности, % |
|---------------------|-------------------|
| Домашнее задание | 40 |
| Компьютерные классы | 30 |
| Финальный проект | 30 |

2. Содержание курса

| Тема | Краткое содержание | Лекции (час) | Семинары (час) | Лабораторные занятия (час) |
|--|---|--------------|----------------|----------------------------|
| Подробный обзор области рекомендательных систем | Общие концепции, основные проблемы, последние тенденции. Социально-экономическое влияние услуг персонализации. Примеры использования и междисциплинарные приложения. | 2 | 1 | |
| Общая формулировка проблемы и основные методы рекомендации | Общие задачи по рекомендациям. Предварительная обработка данных и манипулирование ими. Неперсонализированные системы рекомендаций. Базовые предсказатели. Системы рекомендаций на основе контента и разработка функций. | 1 | 1 | 1 |
| Оценка рекомендательных систем | Общий ход эксперимента. Методы разделения данных и выборка несогласованных результатов. Подходы к перекрестной проверке. Оценка в автономном режиме и онлайн. Общие показатели оценки. | 1 | 2 | |
| Коллаборативная фильтрация: подход, основанный на запоминании. | Частый поиск шаблонов и ассоциативных правил. Пример использования рекомендаций в "длинном хвосте". Алгоритм Amazon "от элемента к элементу". Методы, основанные на сходстве. Методы, основанные на KNN. Сложность, масштабируемость, применимость методов, основанных на памяти. | 1 | 1 | 1 |
| Коллаборативная фильтрация: подход, основанный на моделях | Модели скрытых факторов. Матричные доработки в сравнении с рекомендательными задачами top-n. Методы аппроксимации низкого ранга и методы, основанные на SVD. Слабое и сильное обобщение. Метод предварительного запуска и свертывания. Устранение искажений модели с помощью нормализации данных. | 1 | 2 | |
| Взвешенная матричная факторизация | Сложность взвешенной оптимизации. Распространенные алгоритмы оптимизации и их реализации. Методы оценки достоверности и отрицательной выборки. Комплексное и онлайн-обучение. Сложность и масштабируемость. Задачи распределенного обучения. | 1 | | 2 |
| Специальные методы коллаборативной фильтрации | Различные способы включения понятия подобия в модели. Асимметричные подходы. Расширенные задачи оптимизации. Обучение ранжированию: попарные и списочные задачи, сложность и плавные аппроксимации. | 1 | 2 | |

| | | | | |
|--|--|---|---|---|
| Системы холодного пуска и гибридные рекомендательные системы | Типы режимов холодного пуска. Распространенные способы устранения проблем с холодным пуском. Активное обучение. Таксономия гибридных рекомендательных систем. Простые методы включения дополнительной информации. Модели скрытых факторов и гибридные алгоритмы. | 1 | 1 | 1 |
| Контекстно-зависимые рекомендательные системы | Понимание контекстуальных данных. Различные типы подходов с учетом контекста. Методы предварительной и последующей фильтрации. Методы матричной и тензорной факторизации с учетом контекста. Отзывы пользователей в контексте: влияние на обучение и оценку. | 1 | 1 | 1 |
| Рекомендации, учитывающие последовательность действий | Последовательный, основанный на сеансах и учитывающий сеансы подходы. Правила последовательной ассоциации. Принципы цепей Маркова. Последовательный KNN. Модели скрытых факторов для последовательных данных. | 1 | 1 | 1 |
| Искусственные нейронные сети в рекомендательных системах | Нейронная коллаборативная фильтрация. Общая формулировка и сравнение со стандартными подходами. Автоэнкодеры в рекомендательных системах. Методы последовательного прогнозирования, модели на основе GRU, модели с самостоятельным вниманием к последовательностям. Связь с обработкой естественного языка. Включение дополнительной информации. | 1 | 3 | 2 |

| Вид задания | Краткое содержание задания |
|---------------------|--|
| Домашнее задание | Реализуйте модель коллаборативной фильтрации в наборе данных по вашему выбору. Объясните, как эту модель можно использовать в режиме "холодного запуска" и в условиях онлайн-обучения. Проведите анализ сложности. Выполните настройку гиперпараметров и сообщите об окончательных результатах оценки для стандартной задачи с рекомендациями top-n с оптимальной конфигурацией. Пример модели: матричная факторизация на основе достоверности |
| Компьютерные классы | Учащиеся будут участвовать в нескольких онлайн-конкурсах с таблицей лидеров, которые будут проводиться в течение ограниченного времени. Задача состоит в том, чтобы предоставить решение для задания "Лучшие рекомендации" на основе предоставленного набора данных. Результаты конкурсных работ должны превышать предварительно установленные базовые показатели. |
| Финальный проект | Проектная работа в группах. У студентов будет возможность выбрать одну из предложенных тем или выбрать свою собственную. Темы могут включать разработку и демонстрацию нового приложения рекомендательной системы (в качестве доказательства концепции) или внедрение нового метода из новейшей литературы. Это также может быть более масштабируемая или более точная модификация существующего метода. Результатом будет отчетный документ, сопровождаемый исходным кодом, в котором представлены доказательства тщательного экспериментирования и обсуждаются |

| | |
|--|---|
| | полученные результаты с точки зрения новизны/практичности/ потенциального воздействия. |
|--|---|

3. Результаты обучения

| Знания |
|---|
| <p>Перевод описания реальных проблем с рекомендациями в соответствующие математические формулировки и генерация соответствующей задачи машинного обучения.</p> <p>Понимание различных методов и алгоритмов создания рекомендательных систем, их преимуществ, недостатков и применимости для решения практических задач.</p> <p>Понимание общей методологии проведения тщательного анализа и оценки реализованных решений.</p> <p>Знание современных методов и техник в различных областях, связанных с рекомендательными системами.</p> |

| Навыки |
|--|
| <p>Создание прототипов алгоритмов рекомендательных систем, анализ их сложности, практичности и применимости в различных рекомендательных сценариях.</p> <p>Проведение тщательного сравнительного анализа набора конкурирующих алгоритмов с целью выбора, наиболее подходящего для заранее определенной рекомендательной задачи.</p> <p>Делать надежные выводы из тщательно проработанных результатов оценки и отстаивать выбор конкретного подхода или метода для решения проблемы с практическими рекомендациями.</p> |

| Опыт |
|--|
| <p>Работа с реалистичными наборами данных. Извлечение, обработка и использование разнородных источников данных для построения точных рекомендательных систем.</p> <p>Создание алгоритмов практических рекомендаций, которые могут быть перенесены в производственные системы.</p> <p>Внедрение полного цикла машинного обучения - от первоначальной идеи и обработки набора данных до работающего алгоритма рекомендаций, который удовлетворяет требованиям реального приложения рекомендательной системы.</p> |

4. Критерии оценки

| | |
|---------------------|---|
| Домашнее задание | <p>Реализуйте матричную факторизационную модель, основанную на достоверности, в наборе данных по вашему выбору. Разработайте такую функцию взвешивания достоверности, которая будет исправлять ошибки в прогнозировании наиболее популярных продуктов, которые не были использованы пользователем.</p> <p>Проведите анализ сложности подхода. Выполните настройку гиперпараметров и сообщите об окончательных результатах оценки для стандартной задачи top-n recommendations с оптимальной конфигурацией. Объясните, в чем преимущество/ потеря производительности вашего метода по сравнению со стандартным неявным методом ALS.</p> <p>Будут оцениваться следующие критерии:</p> <ol style="list-style-type: none"> 1. Эффективность и корректность реализации метода. 2. Строгость настройки гиперпараметров. 3. Корректность процедуры оценки и сравнение с исходными данными. 4. Согласованность анализа эффективности решения. |
| Компьютерные классы | <p>Ваша задача - отправить решение на веб-сайт онлайн-конкурса. Решение должно быть в виде рекомендаций для группы тестируемых пользователей по заданному набору данных.</p> <p>На основе предварительного анализа данных выберите несколько наиболее перспективных подходов для обучения</p> |

| | |
|------------------|---|
| | <p>рекомендательной системы на основе предоставленных данных. Предложите лучшее решение, созданное с помощью вашей модели <code>topperforming</code>.</p> <p>Заявка будет оцениваться в зависимости от позиции в таблице лидеров. Решение должно быть лучше, чем предварительно определенный (и распространяемый) базовый уровень.</p> |
| Финальный проект | <p>Пример проекта: создайте рекомендательную систему, которая не только рекомендует пользователю новые фильмы, но и объясняет, почему этому пользователю может понравиться рекомендованный фильм. Пояснение — это краткое текстовое описание сюжета фильма, адаптированное к интересам пользователя (например, "Вам понравится этот фильм, потому что это драма с сильной сюжетной линией и неожиданным поворотом сюжета").</p> <p>Будут оцениваться следующие критерии:</p> <ol style="list-style-type: none"> 1. Эффективность и корректность применяемых методов или общая полезность выбранного подхода к определенной задаче. 2. Соответствие выбранного подхода к оценке определенной рекомендательной задаче. 3. Строгость сравнительного исследования или оценочного эксперимента. 4. Согласованность анализа эффективности решения. <p>Оценка основывается на индивидуальном вкладе каждого члена команды.</p> |

5. Оборудование

| |
|-------------------------|
| Программное обеспечение |
| Python |
| Оборудование |
| Персональные компьютеры |

6. Учебники и интернет-ресурсы

| Необходимые учебники | ISBN-13 (or ISBN-10) |
|---|----------------------|
| Personalized Machine Learning, Julian McAuley, Cambridge University Press, 2022. (https://cseweb.ucsd.edu/~jmcauley/pml/pml_book.pdf) | в прессе |
| Рекомендуемые учебники | |
| Recommender Systems: The Textbook. Charu C. Aggarwal, 2016 | 978-3319296579 |
| Collaborative Recommendations: Algorithms, Practical Challenges and Applications. Shlomo Berkovsky, Ivan Cantador, Domonkos Tikk, 2019 | 978-9813275348 |

| Веб-ресурсы (ссылки) | Описание |
|---|---|
| https://recsyswiki.com/wiki/Main_Page | Один из самых долгоживущих ресурсов, посвященных теме рекомендательных систем |
| https://recommender-systems.com | Активно поддерживаемый ресурс, на котором собраны самые свежие новости, связанные с рекомендательными системами |

