утверждаю

Руководитель Департамента

анализа данных, принятия решений

и финансовых технологий

\_\_\_\_\_\_\_\_\_\_\_\_В.И. Соловьев

«15» октября 2019 г.

ПРИЛОЖЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

**РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ**

НАПРАВЛЕНИЕ ПОДГОТОВКИ: **01.04.02 "Прикладная математика и информатика"**

НАПРАВЛЕННОСТЬ ПРОГРАММЫ МАГИСТРАТУРЫ: **"Анализ больших данных и машинное обучение в экономике и финансах"**

ФОРМА ОБУЧЕНИЯ: **Очная**

ГОД ПРИЕМА: **2019, 2020**

Год утверждения программы: 2018

*Одобрено департаментом анализа данных, принятия решений и финансовых технологий*

*Протокол № 3 от «15» октября 2019 г.*

**Содержание Приложения**

|  |  |
| --- | --- |
| **Наименование разделов РПД** | **стр.** |
| Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине | 2 |
| Объем дисциплины (модуль) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся | 3 |
| Содержание дисциплины | 4 |
| Учебно-тематический план | 6 |
| Содержание семинаров, практических занятий | 7 |
| Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы | 9 |
| Перечень вопросов, заданий, тем для подготовки к текущему контролю | 12 |
| Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине | 13 |
| Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем. | 18 |

**2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине**

|  |  |  |  |
| --- | --- | --- | --- |
| **Код компетенции** | **Наименование**  **компетенции** | **Индикаторы достижения компетенции** | **Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции** |
| **ДКН-5** | Способность строить рекомендательные системы и применять их к решению практических задач | 1.Демонстрирует знания основных понятий в области рекомендательных систем и принципов их построения. | **Знать**  Принципы построения неперсонализированных и слабоперсонализированных систем.  Принципы построения контентноориентированных систем  Принципы построения систем коллаборативной фильтрации.  **Уметь**  выделять атрибуты товаров/ пользователей на основе анализа нечисловой информации и строить контентноориентированную систему. |
| 2.Владеет методологией и практическими навыками построения и оценки рекомендательных систем. | **Знать**  Принципы построения продвинутых и гибридных систем.  Специфические метрики оценки качества рекомендательных систем, их преимущества и недостатки.  **Уметь**  Строить систему коллаборативной фильтрации на основе алгоритмов «Пользователь-Пользователь» или «Товар-Товар». Аппроксимировать матрицу рейтингов матрицей меньшего ранга и использовать ее для гибридной рекомендательной системы.  Отбирать, рассчитать и интерпретировать релевантные метрики качества в зависимости от бизнес-целей заказчика системы. |
| **ДКН-6** | Способность создавать ИТ-сервисы, основанные на анализе данных и машинном обучении | 1.Владеет методологией создания ИТ-сервисов, основанных на анализе данных и машинном обучении. | **Знать**  Специализированное ПО и основные инструменты для решения задач машинного обучения.  **Уметь**  Создавать, настраивать и оценивать модели в специализированном ПО |
| 2.Работает с современными программными комплексами для решения задач анализа данных и машинного обучения. | **Знать**  Инструментальные средства визуализации данных.  Инструменты для решения задач машинного обучения.  **Уметь**  Проводить обучение алгоритмов, избегая переобучения. |
| 3.Создает, настраивает и оценивает модели в специализированном программном обеспечении | **Знать**  Инструментальные средства MS Excel, Azure ML, Python/R (если применимо).  **Уметь**  Настраивать обмен данными между специализированным ПО, источником данных и конечным потребителем. |

**4. Объем дисциплины (модуль) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся**

Общая трудоемкость дисциплины составляет 3 зачетные единицы.

Вид промежуточной аттестации - экзамен.

Вид текущего контроля – контрольная работа

*Очная форма обучения - 2019, 2020 год приема*

|  |  |  |
| --- | --- | --- |
| **Вид учебной работы по дисциплине** | **Всего**  **(в з/е и часах)** | **Модуль 6**  **(в часах)** |
| **Общая трудоемкость дисциплины** | **3/108** | **108** |
| ***Контактная работа-***  ***Аудиторные занятия*** | ***30*** | ***30*** |
| *Лекции* | *10* | *10* |
| *Семинары, практические занятия* | *20* | *20* |
| **Самостоятельная работа** | **78** | **78** |
| Вид текущего контроля | Контрольная работа | Контрольная работа |
| Вид промежуточной аттестации | Экзамен | Экзамен |

**5.1. Содержание дисциплины**

**Тема 1. Введение.** Понятие и история развития рекомендательных систем. Основная терминология: пользователи (users), товары (items), рейтинги, предпочтения и рекомендации, атрибуты users и items. Критерии классификации рекомендательных систем: сфера применения, цель, источник рекомендации, внешние факторы, уровень персонализации, надежность и степень конфиденциальности, интерфейс выдачи рекомендации, рекомендательный алгоритм. Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем (Amazon.com, Netflix).

**Тема 2.** **Неперсонализированные рекомендательные системы.** Неперсонализированные и стереотипные рекомендательные системы. Использование описательной статистики при построении рекомендаций, предпочтения большинства. Демографические данные (возраст, пол, национальность, социальный статус, местонахождение) как средство персонализации рекомендаций. Исследование связей между товарами (кросс-продажи). Недостатки неперсонализированных и слабо-персонализированных систем.

**Тема 3.** **Контентно-ориентированные рекомендательные системы.** Построение контентно-ориентированной рекомендательной системы. Фильтрация на основе контента. Персонализаций рекомендаций на основе предпочтения пользователя. Способы получения и обработки информации о предпочтениях пользователя: прямое анкетирование, анализ активности, выделение связи между предпочтениями пользователя и атрибутами товаров. Анализ ключевых слов для формирования атрибутов товаров. TDIDF анализ и его производные для выделения ключевых слов. Математическая модель рекомендательной системы. Построение и актуализация профилей пользователей, взвешивание/нормализация рейтингов. Представление атрибутов пользователей и товаров в векторной форме; расчет рекомендаций на основе оценки расстояния между векторами. Регрессионные/классификационные модели для предсказания рейтинга пользователей. Преимущества и недостатки контентно-ориентированного подхода.

**Тема 4. Коллаборативная фильтрация. Метод ближайших «соседей». Меры близости.** Понятие коллаборативной фильтрации, использование рейтингов пользователей для построения рекомендаций без учета атрибутов. Сравнительный анализ с контентно-ориентированным подходом. Метода ближайших «соседей» и иные методы для нахождения схожих пользователей и товаров. Меры оценки сходства пользователей, преимущества и недостатки. Алгоритм «User-User» для построения рекомендации на основе оценок пользователей со схожими рейтингами. Математическая модель, предпосылки и ограничения. Использование нормализации, снижение «зашумленности» рейтингов. Алгоритм «Item- Item»: рекомендации схожих товаров на основе связей между их рейтингами. Ограничения алгоритма «User-User»: «пробелы» в данных, вычислительная емкость, необходимость актуализации рекомендаций. Двухшаговая реализация алгоритма «Item- Item»: определение сходства товаров на основе их рейтингов, построение предсказания рейтинга на основе рейтинга «соседей». Математическая модель, предпосылки и ограничения.

**Тема 5. Метрики оценки качества рекомендательных систем.** Типы метрик в зависимости от объекта оценки и бизнес-целей. Цели оценки моделей (выбор алгоритма, настройка, сравнение) и ограничения, определение «плохой» и «хорошей» рекомендации. Базовые метрики оценки качества предсказанных рейтингов и рекомендаций. Метрики оценки точности/погрешности: MAE, RMSE, MSE. Метрики поддержки принятия решений: ROC AUC, полнота и точность. Поиск нерелевантных рекомендаций («reversals»). Продвинутые метрики. Метрики, ориентированные на пользователя: охват (доля товаров, для которых имеется рекомендации/персонализированные рекомендации), удержание пользователей, степень принятия рекомендаций пользователем, степень персонализации рекомендации, продвижение непопулярных товаров, степень диверсификации рекомендаций и т.п. Метрики, основанные на рангах (MAP). Общие принципы обучения и валидации моделей: разделение тестовой и обучающей выборки, кросс-валидация, проблема переобучения. Работа с пропущенными данными (холодный старт), устранение эффекта популярного товара и пр. Выбор показателей оценки эффективности (количество просмотров, объем продаж, конверсия и т.п.). Статистический анализ результата.

**Тема 6. Приближение матрицей меньшего ранга.** Задача снижения размерности. Сингулярное разложение матрицы рейтингов (SVD) и приближение матрицей меньшего ранга. Использование градиентного спуска для приближенного разложения матрицы (FunkSVD). Математическая модель FunkSVD. Стохастические алгоритмы для разложения матриц.

**Тема 7. Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем. Продвинутые и гибридные модели рекомендательных систем.** Гибридные модели. Совмещение контентно-ориентированного подхода и коллаборативной фильтрации. Примеры реализаций.

**5.2. Учебно – тематический план**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | **Наименование тем (разделов) дисциплины** | **Трудоемкость в часах** | | | | | | **Формы текущего контроля успеваемости** |
| **Все**  **го** | **Аудиторная работа** | | | | **Самостоятельная работа** |
| Общая,  в т.ч.: | Лекции | Семинары, практические занятия | Занятия в интерактивных формах |
| 1. | Введение | 5 | 3 | 1 | 2 | 2 | 2 | Проверочные вопросы по теории |
| 2. | Неперсонализированные рекомендательные системы | 9 | 3 | 1 | 2 | 2 | 6 | Лабораторные работы  (построение компьютерной системы обработки данных, анализ конкретной ситуации и обсуждение  результатов)  Лабораторные работы  (построение компьютерной системы обработки данных, анализ конкретной ситуации и обсуждение  результатов) |
| 3. | Контентно- ориентированные рекомендательные системы | 10 | 4 | 2 | 2 | 2 | 6 |
| 4. | Коллаборативная фильтрация.  Метод ближайших «соседей». Меры близости | 20 | 6 | 2 | 4 | 4 | 14 |
| 5. | Метрики оценки качества рекомендательных систем | 17 | 3 | 1 | 2 | 2 | 14 |
| 6. | Приближение матрицей меньшего ранга. | 16 | 6 | 2 | 4 | 4 | 10 |
| 7. | Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем.  Продвинутые и гибридные модели рекомендательных систем. | 31 | 5 | 1 | 4 | 4 | 26 | Доклад от коллектива с презентацией. Контрольные теоретически вопросы.  Лабораторные работы.  (построение компьютерной системы обработки данных, анализ конкретной ситуации и обсуждение  результатов) |
|  | В целом по дисциплине | 108 | 30 | 10 | 20 | 20 | 78 | Контрольная работа |
|  | Итого в % |  |  |  |  | 67 % |  |  |

**5.3. Содержание семинаров, практических занятий**

|  |  |  |
| --- | --- | --- |
| **Наименование тем(разделов) дисциплины** | **Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)** | **Формы проведения занятий** |
| ***Тема 1.*** Введение | Понятие и история развития рекомендательных систем.  Критерии классификации рекомендательных систем.  Основная терминология: пользователи, товары (users,items), рейтинги, предпочтения и рекомендаций, атрибуты пользователей и товаров.  *Рекомендуемые источники:*  *а) основная: 1.* | Лекция |
| ***Тема 2.*** Неперсонализированные рекомендательные системы | Построение рейтингов по популярности, средней оценке.  Функции ранжирования.  Ассоциативные правила.  Построение рекомендательной системы на основе кросс-продаж товаров (на примере статистики фильмов).  *Рекомендуемые источники:*  *а) основная: 8.1;*  *б) дополнительная: 8.3;*  *в) Интернет-ресурсы: 9.8,9* | Лабораторная работа №1  Лабораторная работа №2 |
| ***Тема 3.*** Контентно- ориентированные рекомендательные системы | Типы контента.  Обработка текстов.  Практические методы анализа текстов: TF\*IDF, vector model, косинусная мера.  Построение рекомендаций на основе рейтингов, проставленных пользователями для товаров с известными атрибутами.  Построение рекомендаций на основе рейтингов, проставленных пользователями для товаров с неизвестными атрибутами. Извлечение атрибутов на основе программной реализации TFIDF анализа.  *Рекомендуемые источники:*  *а) основная: 8.1,2;*  *б) дополнительная: 8.1, 2;*  *в) Интернет-ресурсы: 9.5-9* | Лабораторная работа № 3.  Лабораторная работа № 4. |
| ***Тема 4.*** Коллаборативная фильтрация.  Метод ближайших «соседей». Меры близости | Метрики, базовые предикторы, регуляризация.  Меры близости: корреляция Пирсона, корреляция Спирмена и др.  Поиск ближайших «соседей»-пользователей/товаров на основе матрицы рейтингов (на примере фильмов).  Построение рекомендаций с помощью алгоритма «User-­user» или «Item­-item» (среда для выполнения задания: Python)  *Рекомендуемые источники:*  *а) основная: 8.1,2;*  *б) дополнительная: 8.1, 2;*  *в) Интернет-ресурсы: 9.5-9* | Лабораторная работа № 5.  Лабораторная работа № 6. |
| ***Тема 5.*** Метрики оценки качества рекомендательных систем | Расчет числовых метрик качества построенных рекомендательных систем и интерпретация результатов.  Бизнес­цели: cross-­selling, вовлеченность, покрытие, разнообразие, неожиданность и др. Способы измерения эффекта от рекомендательных систем: offline эксперименты, user­study, online эксперименты. Метрики: MAE, MSE, RMSE. Precision­recall  Метрики ранжирования: MRR, SRC, DCG, NDCG  *Рекомендуемые источники:*  *а) основная: 8.1,2;*  *б) дополнительная: 8.1, 2, 3*  *в) Интернет-ресурсы: 8.4-9* | Лабораторная работа № 7. |
| ***Тема 6.*** Приближение матрицей меньшего ранга. | Уменьшение размерности матрицы. SVD, Funk SVD, SVD++, ALS.  Разработка и оценка рекомендательной системы с использованием SVD.  *Рекомендуемые источники:*  *а) основная: 8.1,2;*  *б) дополнительная: 8.1, 2;*  *в) Интернет-ресурсы: 9.6-9* | Лабораторная работа № 8. |
| ***Тема 7.*** Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем.  Продвинутые и гибридные модели рекомендательных систем. | Стратегий объединения подходов. Построение гибридной рекомендательной системы.  *Рекомендуемые источники:*  *а) основная: 8.1,2;*  *б) дополнительная: 8.1, 2;*  *в) Интернет-ресурсы: 9.6-9* | Доклад с презентацией.  Лабораторная работа № 9. |

**6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы**

|  |  |  |
| --- | --- | --- |
| **Наименование тем (разделов) дисциплины** | **Перечень вопросов, отводимых на самостоятельное освоение** | **Формы внеаудиторной самостоятельной работы** |
| Введение | Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем (Netflix, Amazon.com, Last.fm, Pandora, LinkedIn, Hulu). | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; |
| Неперсонализированные рекомендательные системы | Стереотипные рекомендательные системы.  Недостатки неперсонализированных и слабо-персонализированных систем. | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы |
| Контентно- ориентированные рекомендательные системы | Способы получения и обработки информации о предпочтениях пользователя.  Расширения и альтернативы TFIDF-анализа | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы |
| Коллаборативная фильтрация. Метод ближайших «соседей». Меры близости | Использование нормализации, снижение «зашумленности» рейтингов.  *k*-ближайших соседей (k-nearest neighbor algorithm, kNN).  Математическое обоснование (доказательство) используемых моделей. | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы |
| Метрики оценки качества рекомендательных систем | Базовые метрики оценки качества предсказанных рейтингов и рекомендаций.  Общие принципы обучения и валидации моделей. | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы * повторение материалы и подготовка к контрольной работе |
| Приближение матрицей меньшего ранга. | Математическая модель FunkSVD.  Стохастические алгоритмы для разложения матриц. | * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы * повторение материала и подготовка к экзамену |
| Примеры успешной реализации бизнес-моделей, основанных на использовании рекомендательных систем.  Продвинутые и гибридные модели рекомендательных систем. | Архитектура гибридных рекомендательных систем. | * подготовка доклада с презентацией о примерах применения рекомендательных систем * работа с текстом лекции, разбор вопросов по теме занятия; * изучение рекомендованных к занятию литературных источников; * подготовка к лабораторным работам, настройка программного обеспечения; * разработка/отладка рекомендательной системы в соответствии с конкретными задачами лабораторной работы * повторение материала и подготовка к экзамену |

**6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю**

***Примерные вопросы к контрольной работе***

* 1. Как работает user-to-user и item-to-item рекомендательная система?
  2. Используя матричные разложения, разработать рекомендации, учитывающие неявные факторы.
  3. Определить метрику для оценки качества рекомендаций, которая лучше всего подходит задаче.
  4. Построить рекомендации фильмов пользователям, по которым еще нет никаких данных.
  5. По заданной матрице предпочтений вычислить похожесть пользователей одним из методов: корреляция Пирсона, косинусное расстояние, расстояние Жаккара, метод ассоциаций.
  6. Классификация методом k ближайших соседей.
  7. Методы оценки качества рекомендательных систем.
  8. Применимость того или иного подхода в зависимости от постановки задачи.

***Примеры заданий контрольной работы***

1. Реализовать и сравнить три варианта рекомендательной системы фильмов на данных, полученных от преподавателя.
2. Реализовать рекомендательную систему по посещениям. Для пользователя u выдать оценку ресурса r; выдать ранжированный список рекомендуемых ресурсов.
3. Реализовать рекомендательную систему на основе рейтингов с учетом требуемой точности прогнозов.
4. Самостоятельно составить матрицу предпочтений. На основе принципа k ближайших соседей рекомендации для пользователя u.
5. Составить оценки точности метода фильтрации путем сравнения прогнозируемых рейтингов непосредственно с фактическим рейтингом пользователей.
6. Дана матрица предпочтений пользователей. Приблизить исходную матрицу, матрицей меньшего ранга.

**7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2 «Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

**Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, знаний и умений**

|  |  |
| --- | --- |
| **Наименование компетенций** | **Примеры заданий для оценки индикаторов достижения компетенций** |
| ДКН-5  Способность строить рекомендательные системы и применять их к решению практических задач | **1.Демонстрирует знания основных понятий в области рекомендательных систем и принципов их построения.**  **Задание 1.**  Составить матрицу предпочтений с известными оценками пользователей, при возможности дополнить эти оценки информацией по клиенту и товару, и заполнить неизвестные значения. Получить численное выражение «похожести» пользователей одним из методов: косинусная мера, коэффициент корреляции Пирсона, Евклидово расстояние.  **2.Владеет методологией и практическими навыками построения и оценки рекомендательных систем.**  **Задание 2.**  Построить систему рекомендация на основе набора данных (данные могут содержать шумы, выбросы, пропущенные значения и требовать дополнительной обработки). Исходный полный набор данных разделить на три набора: набор, предназначенный для обучения модели; публичный тестовый набор данных; закрытый тестовый набор данных. Проверить точность рекомендаций одной из метрик: MAE, RMSE, MSE. |
| ДКН-6  Способность создавать ИТ-сервисы, основанные на анализе данных и машинном обучении | **1.Владеет методологией создания ИТ-сервисов, основанных на анализе данных и машинном обучении.**  **Задание 1.**  Проанализировать матрицу предпочтений с помощью одного или нескольких программных инструментов (Excel/MS Azure/Python/R), сформулировать выводы. Построить рекомендательную систему коллаборативной фильтрации. Сделать выводы о применимости выбранного подхода в зависимости от постановки задачи.  **2.Работает с современными программными комплексами для решения задач анализа данных и машинного обучения.**  **Задание 2.**  Построить рекомендации на основе рейтингов, проставленных пользователями для товаров с неизвестными атрибутами. Извлечь атрибуты на основе программной реализации TFIDF анализа.  **3.Создает, настраивает и оценивает модели в специализированном программном обеспечении**  **Задание 3.**  Настроить сервер базы данных. Установить инструменты для работы с основными библиотеками для анализа данных на Python. После завершения установки, импортировать библиотеки (NumPy, SciPy, Matplotlib, Pandas и д.р.) или их отдельные модули в Jupyter. Обработать собранные данные, чтобы повысить их качество, заменить пропуски в матрице предпочтений используя метод ближайших «соседей» (пользователей/товаров) на основе матрицы предпочтений. |

***Примеры типовых заданий***

Задача 1. С помощью метода ассоциаций найти похожие фильмы по следующим исходным данным.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Задача 2. В таблице приведен набор данных о рейтингах пользователей. Вопросительные знаки относятся к элементам, которые пользователи еще не оценили. Используя корреляцию Пирсона, рассчитать коэффициент схожести между пользователями.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| User 1 | 4 |  | 4 |  | 5 |
| User 2 | 4 | 4 |  |  |  |
| User 3 |  | 3 | 5 | 4 | 5 |
| User 4 | 3 |  | 5 |  |  |
| User 5 | 3 |  |  | 3 | 4 |

Задача 3. В таблице приведен набор данных о действиях пользователей. С помощью асимметричного коэффициента Жаккара рассчитать коэффициент схожести между пользователями.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 |
| User 1 | 1 | 1 | 1 |  |
| User 2 | 1 |  | 1 | 1 |
| User 3 |  | 1 | 1 |  |
| User 4 | 1 | 1 |  | 1 |

Задача 4. В таблице приведен набор данных о рейтингах пользователей. Вопросительные знаки относятся к элементам, которые пользователи еще не оценили. Используя косинус угла между векторами рейтингов, рассчитать коэффициент схожести между пользователями.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 |
| User 1 |  |  | 3 | 4 |
| User 2 |  | 4 |  | 5 |
| User 3 | 4 |  | 4 |  |
| User 4 |  |  | 3 | 4 |

Задача 5. В таблице приведен набор данных о рейтингах пользователей. Вопросительные знаки относятся к элементам, которые пользователи еще не оценили. Спрогнозировать незаполненные ячейки.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| User 1 | 3 | 5 | 4 |  | 5 |
| User 2 |  | 5 |  | 4 |  |
| User 3 |  | 4 | 3 | 4 | 5 |
| User 4 | 3 |  | 4 |  |  |
| User 5 | 3 |  |  | 3 | 4 |

***Теоретические вопросы для подготовки к экзамену***

1. Понятие и история развития рекомендательных систем.
2. Критерии классификации рекомендательных систем.
3. Стереотипные рекомендательные системы.
4. Неперсонализированные рекомендательные системы.
5. Недостатки неперсонализированных и слабо-персонализированных систем.
6. Способы получения и обработки информации о предпочтениях пользователя.
7. Построение контентно-ориентированной рекомендательной системы.
8. Использование нормализации, снижение «зашумленности» рейтингов.
9. Способы получения и обработки информации о предпочтениях пользователя.
10. Базовые метрики оценки качества предсказанных рейтингов и рекомендаций.
11. Анализ ключевых слов для формирования атрибутов товаров.
12. Общие принципы обучения и валидации моделей.
13. Расширения и альтернативы TFIDF-анализа.
14. TDIDF анализ и его производные для выделения ключевых слов.
15. Математическая модель рекомендательной системы.
16. Определение «похожести» интересов пользователей, посчитанное корреляцией Пирсона. Преимущества и недостатки.
17. Построение и актуализация профилей пользователей, взвешивание/нормализация рейтингов.
18. Определение «похожести» интересов пользователей, посчитанное косинусным расстоянием векторов предпочтений. Преимущества и недостатки.
19. Преимущества и недостатки контентно-ориентированного подхода.
20. Определение «похожести» интересов пользователей, посчитанное расстоянием Жаккара. Преимущества и недостатки.
21. Понятие коллаборативной фильтрации.
22. Методики сбора данных.
23. Меры оценки сходства пользователей, преимущества и недостатки.
24. Проблема холодного старта.
25. Алгоритм «User-User» для построения рекомендации на основе оценок пользователей со схожими рейтингами.
26. Актуальность рекомендаций.
27. Использование нормализации, снижение «зашумленности» рейтингов.
28. Подход Slope One.
29. Алгоритм «Item-Item»: рекомендации схожих товаров на основе связей между их рейтингами.
30. Два основных подхода в коллаборативной фильтрации.
31. Ограничения алгоритма «User-User».
32. Переобучение и регуляризация.
33. Двухшаговая реализация алгоритма «Item-Item»: определение сходства товаров на основе их рейтингов.
34. Построение предсказания рейтинга на основе рейтинга «соседей».
35. Гибридные рекомендательные системы.
36. Типы метрик оценки качества рекомендательных систем.
37. Рекомендательные системы, основанные на знаниях.
38. Метрики оценки точности: MAE, RMSE, MSE.
39. Метрики поддержки принятия решений: ROC AUC, полнота и точность.
40. Поиск нерелевантных рекомендаций («reversals»).
41. Задача снижения размерности.
42. Сингулярное разложение матрицы рейтингов (SVD).
43. Математическая модель FunkSVD.

***Пример экзаменационного билета***

Вопрос 1. Критерии классификации рекомендательных систем.

Вопрос 2. Недостатки неперсонализированных и слабо-персонализированных систем.

Вопрос 3. Расширения и альтернативы TFIDF-анализа.

Вопрос 4. Базовые метрики оценки качества предсказанных рейтингов и рекомендаций.

Вопрос 5. Предсказать, какие из неизвестных элементов матрицы будут максимальными в своей строке.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| User 1 | 1 | 2 | 5 |  | 1 |
| User 2 | 5 |  | 1 |  |  |
| User 3 |  |  | 4 | 3 | 4 |
| User 4 | 1 | 1 | 5 | 4 |  |
| User 5 | 2 | 2 | 3 | 4 | 4 |

Вопрос 6. Приблизить матрицу рейтингов матрицей малого ранга посредством SVD.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| User 1 | 4 |  | 4 |  | 5 |
| User 2 | 4 | 4 |  |  |  |
| User 3 |  | 3 | 5 | 4 | 5 |
| User 4 | 3 |  | 5 |  |  |
| User 5 | 3 |  |  | 3 | 4 |

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.**

11.1. Комплект лицензионного программного обеспечения:

1. Windows, Microsoft Office.

2. Антивирус ESET Endpoint Security

11.2. Современные профессиональные базы данных и информационные справочные системы:

1. Информационно-правовая система «Гарант»

2. Информационно-правовая система «Консультант Плюс»

3. Электронная энциклопедия: <http://ru.wikipedia.org/wiki/Wiki>

4.Система комплексного раскрытия информации «СКРИН» -http://www.skrin.ru

11.3. Сертифицированные программные и аппаратные средства защиты информации – не предусмотрены

* 1. Microsoft Excel.
  2. Microsoft Azure
  3. Дистрибутив языка Python 3 (или более поздней версии)