|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра вычислительной техники (ВТ)**

**ОТЧЁТ ПО ПРАКТИКЕ**

**Научно-исследовательская работа**

приказ Университета о направлении на практику от «01» сентября 2021 г.

№ 4087-С

|  |  |  |
| --- | --- | --- |
| Отчет представлен к  рассмотрению:  Студент группы ИКБО-04-18 | «\_\_» декабря 2021 | (подпись и расшифровка подписи) |
| Отчет утвержден.  Допущен к защите: |  |  |
| Руководитель практики  от кафедры | «\_\_» декабря 2021 | (подпись и расшифровка подписи) |
|  |  |  |

Москва 2021 г.

|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра вычислительной техники (ВТ)**

**ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ НА ПРАКТИКУ**

**Научно-исследовательская работа**

**Студенту 4 курса учебной группы ИКБО-04-18**

**Зыкову Александру Алексеевичу**

**Место и время практики:** Кафедра вычислительной техники института информационных технологий Московского технологического университета - МИРЭА с 1 сентября 2021г. по 21 декабря 2021г.

**Должность на практике:**  студент

**1. ЦЕЛЕВАЯ УСТАНОВКА:** Исследование выбранной предметной области на предмет осуществления её анализа и актуальности, постановки задачи, реализации под неё технического задания и проектирования на основе структурных и объектных методологий.

**2. СОДЕРЖАНИЕ ПРАКТИКИ:**

2.1. Изучить: научную и техническую литературу, электронные информационно-образовательные ресурсы, применяемые для профессиональной деятельности по теме научно-исследовательской работы.

2.2. Практически выполнить: реализацию технического задания и проектной части с использованием диаграмм структурных и объектных методологий, описать процессы и этапы жизненного цикла выбранного программного обеспечения, согласно условиям стандартизации, охарактеризовать актуальность выбранной предметной области, представить подобные программные разработки и провести анализ предметной области.

2.3. Ознакомиться: с литературой по выбранной предметной области и технической литературой по искусственному интеллекту, а также стандартами и ГОСТами реализации программного обеспечения.

**3. ДОПОЛНИТЕЛЬНОЕ ЗАДАНИЕ:** оформить презентационный материал

**4. ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКИЕ УКАЗАНИЯ:** в процессе практики рекомендуется использовать периодические издания и отраслевую литературу годом издания не старше 5 лет от даты начала прохождения практики

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Руководитель практики от кафедры  «06» сентября 2021 г. | | | *Подпись* | | (Сорокин А. Б.) |
| Задание получил | |  |  | |  |
| «06» сентября 2021 г. | |  | *Подпись* | | (Зыков А.А.) |
|  | |  |  | |  |
|  | |  |  | |  |
| **СОГЛАСОВАНО:** | |  |  | |  |
| Заведующий кафедрой: | | | | | |
| «06» сентября 2021 г. | | | *Подпись* | | (Платонова О.В.) |
| **Проведенные инструктажи:** |  | | |  | |
| Охрана труда: |  | | | «06» сентября 2021 г. | |
| Инструктирующий | *Подпись* | | | Сорокин А. Б. доцент кафедры ВТ | |
| Инструктируемый | *Подпись* | | | Зыков А.А. | |
|  |  | | |  | |
| Техника безопасности: |  | | | «06» сентября 2021 г. | |
| Инструктирующий | *Подпись* | | | Сорокин А. Б. доцент кафедры ВТ | |
| Инструктируемый | *Подпись* | | | Зыков А.А. | |
|  |  | | |  | |
| Пожарная безопасность: |  | | | «06» сентября 2021 г. | |
| Инструктирующий | *Подпись* | | | Сорокин А. Б. доцент кафедры ВТ | |
| Инструктируемый | *Подпись* | | | Зыков А.А. | |
|  | | | |  | |
| С правилами внутреннего распорядка ознакомлен: | | | | «06» сентября 2021 г. | |
|  | *Подпись* | | | Зыков А.А. | |

|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**СОВМЕСТНЫЙ РАБОЧИЙ ГРАФИК ПРОВЕДЕНИЯ  
ПРАКТИКИ**

студента Зыкова А.А. 4 курса группы ИКБО-04-18 очной формы обучения, обучающегося по направлению подготовки 09.03.04 Программная инженерия, профиль «Системы поддержки принятия решений».

|  |  |  |  |
| --- | --- | --- | --- |
| **Неделя** | **Сроки**  **выполнения** | **Этап** | **Отметка о выполнении** |
| 1 | 06.09.2021 | Подготовительный этап, включающий в себя организационное собрание (Вводная лекция о порядке организации и прохождения научно-исследовательской работы, инструктаж по технике безопасности, получение задания на практику) |  |
| 1-6 | 06.09.2021-11.10.2021 | Исследовательский этап (Формирование текстового описания, актуальности темы исследования и анализа предметной области) |  |
| 7 | 18.10.2021 | Представление руководителю структурированного материала об объекте исследования согласно содержанию выше указанного этапа |  |
| 7-10 | 18.10.2021-08.11.2021 | Исследовательский этап (Формирование текстового описания, постановки задачи и технического задания |  |
| 11 | 15.11.2021 | Представление руководителю структурированного материала об объекте исследования согласно содержанию выше указанного этапа |  |
| 11-14 | 22.11.2021-29.11.2021 | Представление руководителю текстового описания процессов проектирования на основе объектных диаграмм и реализации их в соответствующих инструментальных средах |  |
| 15 | 07.12.2021 | Представление руководителю структурированного материала об объекте исследования согласно содержанию выше указанного этапа |  |
| 16 | 14.12.2021 | Представление руководителю предварительной версии отчета с обеспечением согласованности материала по всем его частям, полученных на предыдущих этапах |  |
| 17 | 21.12.2021 | Подготовка окончательной версии отчета и презентационного материала по практике (Оформление материалов отчета и презентации в полном соответствии с требованиями на оформление письменных учебных работ студентов) |  |

Содержание практики и планируемые результаты согласованы с руководителем практики от профильной организации

Руководитель практики от  
кафедры \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/Сорокин А.Б., к.т.н., доцент/

Обучающийся \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/Зыков А.А./

**Согласовано:**

Заведующий кафедрой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/Платонова О.В., к.т.н., доцент/

Содержание

[Содержание 6](#_Toc91506778)

[1 текстовое описание 8](#_Toc91506779)

[1.1 Описание видов рекомендательных систем 9](#_Toc91506780)

[1.1.1 Рекомендации, основанные на контенте 9](#_Toc91506781)

[1.1.2 Рекомендательные системы, основанные на коллаборативной фильтрации 12](#_Toc91506782)

[1.1.3 Рекомендательная система, основанная на знаниях 15](#_Toc91506783)

[1.1.4 Гибридные рекомендательные системы 16](#_Toc91506784)

[2 Актуальность темы исследования 18](#_Toc91506785)

[3 Постановка задачи и техническое задание 21](#_Toc91506786)

[3.1 Основные задачи 21](#_Toc91506787)

[3.2 Техническое задание 21](#_Toc91506788)

[3.2.1 Введение 21](#_Toc91506789)

[3.2.2 Назначение разработки 21](#_Toc91506790)

[3.2.3 Требования к программе или программному изделию 22](#_Toc91506791)

[3.2.4 Требования к надежности 22](#_Toc91506792)

[3.2.5 Время восстановления после отказа 23](#_Toc91506793)

[3.2.6 Отказы из-за некорректных действий оператора 23](#_Toc91506794)

[3.2.7 Условия эксплуатации 24](#_Toc91506795)

[3.2.8 Требования к численности и квалификации персонала 24](#_Toc91506796)

[3.2.9 Требования к составу и параметрам технических средств 24](#_Toc91506797)

[3.2.10 Требования к информационной и программной совместимости 25](#_Toc91506798)

[3.2.11 Требования к исходным кодам и языкам программирования 25](#_Toc91506799)

[3.2.12 Требования к программным средствам, используемым программой 25](#_Toc91506800)

[3.2.13 Требования к защите информации и программ 26](#_Toc91506801)

[3.2.14 Специальные требования 26](#_Toc91506802)

[3.2.15 Требования к программной документации 26](#_Toc91506803)

[3.2.16 Технико-экономические показатели 26](#_Toc91506804)

[3.2.17 Стадии и этапы разработки 27](#_Toc91506805)

[3.2.18 Содержание работ по этапам 28](#_Toc91506806)

[4 Этап проектирования 30](#_Toc91506807)

[4.1 Реализация общей структуры 32](#_Toc91506808)

[4.2 Реализация рекомендательной системы 33](#_Toc91506809)

[Заключение 35](#_Toc91506810)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 36](#_Toc91506811)

1. текстовое описание

Тема работы: построение различных видов рекомендательных систем фильмов для реализации наилучшего решения.

Нейронные сети являются одной из областей искусственного интеллекта. Основной их целью является прогнозирование и, как следствие, моделирование различных ситуаций в зависимости от исходных данных. Задачи, решаемые типичной нейронной сетью:

* это классификация;
* прогнозирование;
* распознавание.

Нейронные сети могут учиться и развиваться самостоятельно, накапливая опыт на основе допущенных ими ошибок.

Как основные можно выделить следующие типы нейронных сетей:

1. Нейронные сети прямого распространения (feed forward neural networks, FF или FFNN)
2. Сети радиально-базисных функций (radial basis function, RBF)
3. Нейронная сеть Хопфилда (Hopfield network, HN)
4. Цепи Маркова (Markov chains, MC или discrete time Markov Chains, DTMC)
5. Машина Больцмана (Boltzmann machine, BM)
6. Ограниченная машина Больцмана (restricted Boltzmann machine, RBM)
7. Автокодировщик (autoencoder, AE)
8. Разреженный автокодировщик (sparse autoencoder, SAE)
9. Вариационные автокодировщики (variational autoencoder, VAE)
10. Шумоподавляющие автокодировщики (denoising autoencoder, DAE)
11. Сеть типа «deep belief» (deep belief networks, DBN)
12. Свёрточные нейронные сети (convolutional neural networks, CNN) и глубинные свёрточные нейронные сети (deep convolutional neural networks, DCNN)
13. Развёртывающие нейронные сети (deconvolutional networks, DN)

-- тут должен быть глобальный рисунок сетей -

Рисунок 1 – основные виды нейронных сетей

* 1. Описание видов рекомендательных систем
     1. Нейронные сети прямого распространения

Это направленный ациклический граф, что означает, что в сети нет обратных связей или петель. У него есть входной слой, выходной слой и скрытый слой. Как правило, может быть несколько скрытых слоев. Каждый узел в слое — нейрон, который можно рассматривать как основной процессор нейронной сети.

Искусственный нейрон — это основная единица нейронной сети (рисунок 1.2).

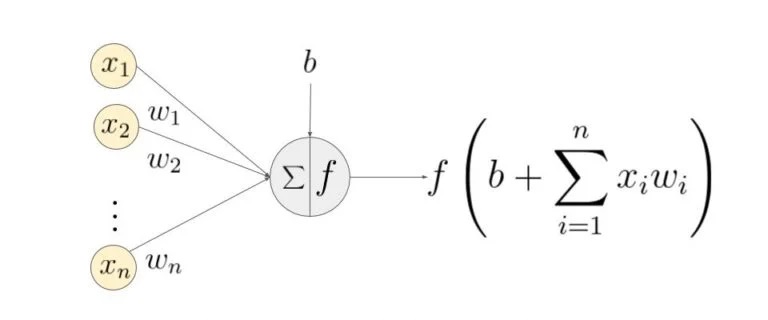


Рисунок 1.2 – схема работы нейрона

Как видно выше, он работает в два этапа: вычисляет взвешенную сумму своих входных данных, а затем применяет функцию активации для нормализации суммы. Функции активации могут быть линейными или нелинейными. Также, есть веса, связанные с каждым входом нейрона. Это параметры, которые сеть должна приобрести на этапе обучения.

где:

P(dj) – профиль документа, представленный вектором взвешенных термов (слов) из значений термов;

dj – текстовый документ;

wkj – вес терма, рассчитанный с помощью метрики TF-IDF.

Рекомендация строится на основании оценки близости контента профиля пользователя и контента объекта (Формула 2).

где:

r(u, d) – мера близости пользователя и объекта;

P(u) – профиль пользователя, представленный вектором взвешенных термов (слов) на основе просмотренных пользователем объектов;

P(d) – профиль объекта, представленный вектором взвешенных термов (слов) из значений термов.

При таком подходе моделируются вкусы пользователя, основываясь на предыдущих оценках.

Использования ключевых слов – важный шаг для многих рекомендательных систем. Но их выделение так же является проблемой, особенно в случае с видео-контентом. Существует два решения данной проблемы: делегировать разметку данных экспертам или непосредственным пользователям. В качестве примера источника размеченных данных по фильмам можно использовать MovieLens. Каждому фильму присвоены тэги (Рисунок 3), каждые отвечающие за свою категорию, в том числе: жанр (Рисунок 2), особенности сюжета, место действия, актерский состав. Совокупность этих тэгов хранит в себе достаточно информации, для создания точного представления о фильме.

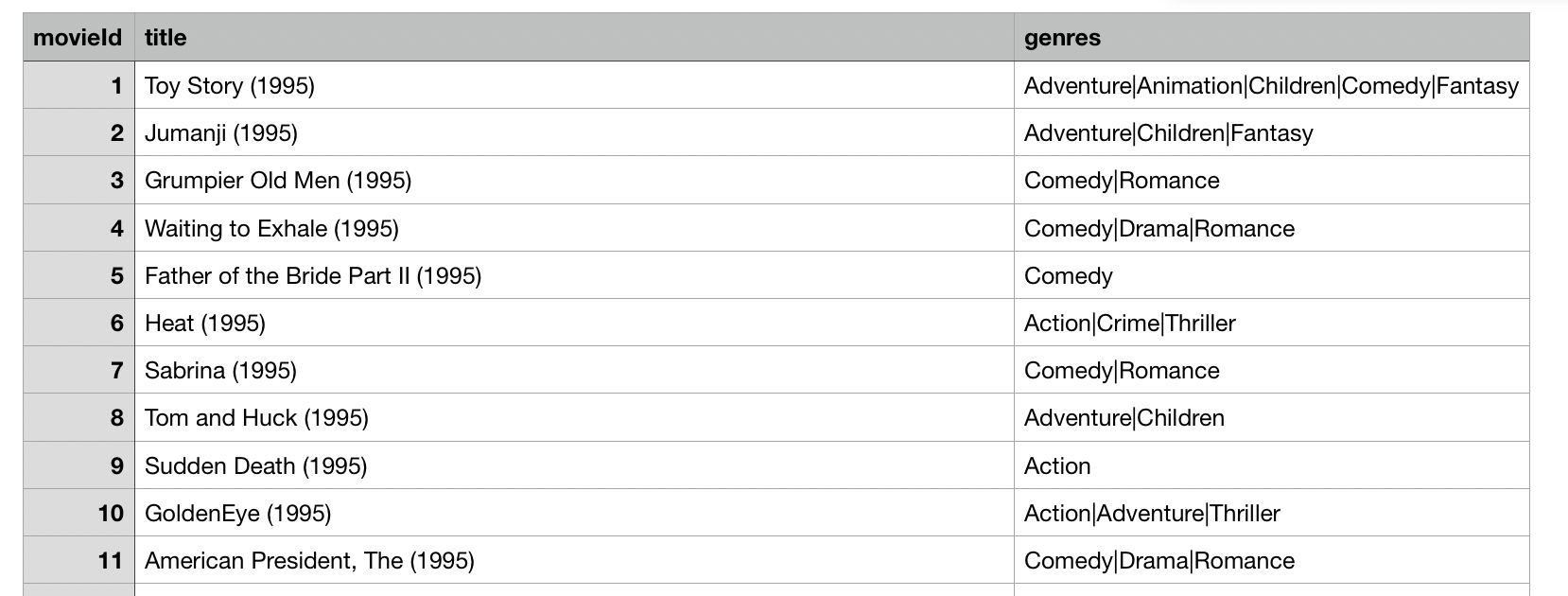


Рисунок 2 - Пример данных о фильмах (жанры, название, id)

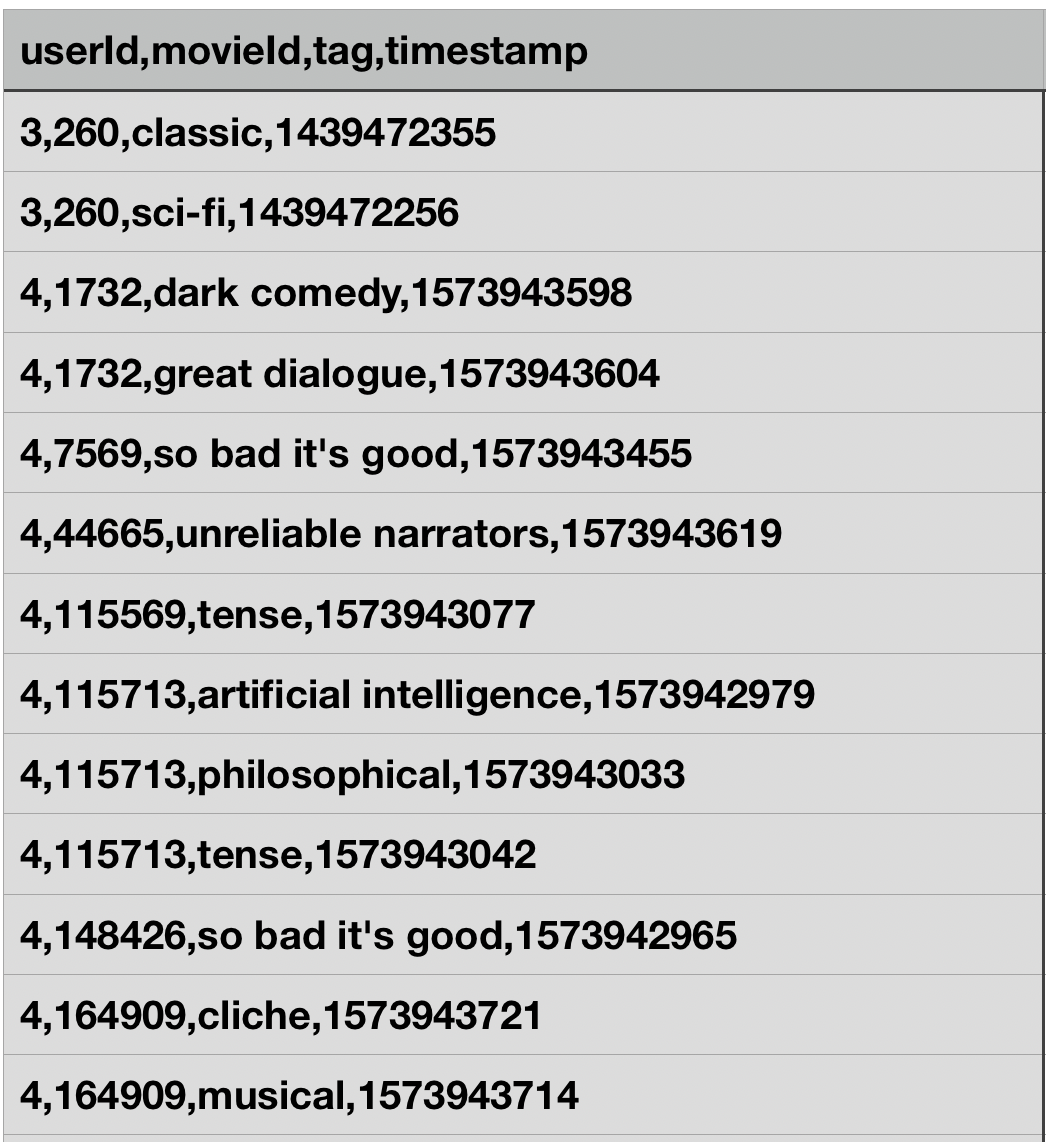


Рисунок 3 - Пример пользовательских тегов для фильмов

Системы, использующие разметку пользователями, используются шире, чем использующие разметку экспертов. Показательными ресурсами являются Delicious, MovieLens и Flickr. Преимуществом таких систем являются более разнообразные теги, но качество тегов страдает, часть из которых оказывается нерелевантной.

При подходе, с использованием пользовательской разметки, необходимо решить две проблемы: рекомендация тегов и способ рекомендации фильмов, с проставленными тегами. Первую проблему необходимо решить для повышения качества тегов. Самым простым способом решения второй проблемы является использование тегов в качестве ключевых слов и последующее их использование в рекомендательных системах, основанных на контенте.

* + 1. Сети радиально-базисных функций

Рекомендательные системы, основанные на коллаборативной фильтрации наиболее распространены. При данном подходе алгоритмом формируются пользовательские предпочтения, основанные на истории поведения.

Различают два вида рекомендательных систем, основанных на коллаборативной фильтрации: основанные на пользователях и основанные на продуктах.

В случае с фильтрацией, основанной на пользователях, предполагается, что пользователю понравятся вещи, которые нравятся пользователям с похожими вкусами. В первую очередь при таком подходе необходимо установить, что похожий вкус – это показатель, что пользователям понравились одни и те же альтернативы (Формула 3).

где:

u, v – пользователи;

N(u), N(v) – набор альтернатив, понравившихся пользователям u и v соответственно;

Suv – схожесть пользователей u и v.

Существует множество алгоритмов схожести, формула 4 – один из примеров такого алгоритма, где рассчитывается привлекательность i для пользователя u (Формула 4).

где:

pui – привлекательность альтернативы i для пользователя u;

suv – схожесть пользователей u и v;

pvi – привлекательность i для пользователя v.

Таблица 1 - Схожесть пользователей по понравившимся вещам

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

В таблице 1 из того, что пользователям A и C понравились схожие вещи, делается вывод, что пользователю A понравится вещь D, которая понравилась пользователю C.

Коллаборативная фильтрация, основанная на продуктах, отличается тем, что в качестве рекомендации выбирается альтернатива, похожая на предыдущие, понравившиеся конкретному пользователю. Прежде всего, при данном подходе, необходимо определить эту схожесть между двумя альтернативами (Формула 5).

где:

i, j – альтернативы;

N(i), N(j) – наборы пользователей, которым нравятся i и j соответственно;

sij – схожесть альтернатив i и j.

Таким образом, привлекательность элемента i для пользователя u может быть получена следующим образом (Формула 6).

где:

pui – привлекательность альтернативы i для пользователя u;

sij – схожесть альтернатив i и j;

puj – привлекательность альтернативы j для пользователя u.

Таблица 2 - Схожесть вещей по похожим пользователям

|  |  |  |  |
| --- | --- | --- | --- |
| User/Item | Item 1 | Item 2 | Item 3 |
| User 1 | + |  | + |
| User 2 | + | + | + |
| User 3 | + |  | recommend |

В таблице 2 из того, что вещи Item 1 и Item 3 понравились схожим пользователям User 1 и User2, делается вывод, что Item 1 и Item 3 - схожи.

Оба алгоритма, как основанный на пользователях, так и на конкретных вещах – являются алгоритмами поиска соседей. Томас Хоффман в своей работе [1] вывел, что пользователь не интересуется предметом напрямую, он обращает внимание лишь на некоторые свойства конкретного предмета, таким образом, его модель учится выделять эти свойства для определения и прогнозирования поведения пользователя.

Помимо модели, связанной с использованием матриц, в коллаборативной фильтрации широко используется модель графов.

* + 1. Нейронная сеть Хопфилда

Рекомендательная система является основанной на знаниях, когда она опирается не на оценки пользователя, а на его поисковые запросы. Эта система позволяет подсказать пользователю возможный результат, который он пытается найти или подсказать возможное дополнение к его запросу.

В данном случае, подходящим будет являться результат, который не обязательно полностью удовлетворяет поисковому запросу. При данном подходе необходимо правильно знать относительную важность каждой особенности продукта и то, что их объединяет. Например, связь со стоимостью: чем дешевле – тем лучше. В любом случае, необходимо вернуть наиболее точный результат в соответствии с запросом. Но, как и в случае с рекомендациями, основанными на контенте, эти связи могут выставляться как конкретными экспертами, так и пользователями системы.  
Существует несколько способов персонализации таких рекомендательных систем:

1. С использованием обратной связи пользователя
2. С учетом поисковых конструкций, предпочитаемых пользователем, можно предлагать похожие поисковые конструкции
3. Менять приоритетность критериев используя динамику изменения пользовательской истории поиска

Данная рекомендательная система может быть использована эффективнее в комбинации с другими рекомендательными системами, или при «холодном» старте, когда пользователь не успел оставить достаточное количество данных для анализа. Данная работа не предполагает реализации рекомендательной системы, основанной на знаниях, но она достойна упоминания.

* + 1. Цепи Маркова

Гибридные рекомендательные системы являются наиболее популярными на данный момент. Причина – совмещение коллаборативной и основанной на контенте фильтраций (Рисунок 4) может быть наиболее эффективным[2].

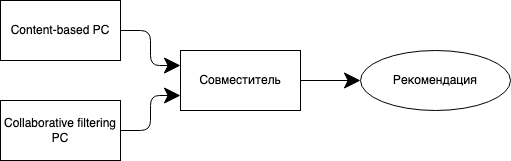


Рисунок 4 - Схема гибридной рекомендательной системы

Можно выделить семь методов гибридизации:

1. С использованием весов: добавление оценок к разным компонентам рекомендации.
2. С использованием переключения: менять способ фильтрации в зависимости от компоненты рекомендации
3. Смешанный: вывод результатов из разных рекомендательных систем
4. Комбинация особенностей: извлечение особенностей из разных источников и их совмещение
5. Выделение особенностей: вычисление особенностей одним рекомендательным алгоритмом и дальнейшая обработка этих данных
6. Каскад: генерация «грубого» результата с использованием рекомендательной техники и помещение его поверх предыдущего результата
7. Вложенный: использование результата одного алгоритма рекомендации в качестве входных данных для другого

Несмотря на большое количество комбинаций в теории, на практике не все из них оказываются более эффективными и подходящими для решения конкретной задачи. Основной принцип гибридных рекомендательных систем - избежать недостатков отдельного подхода или компенсировать иx.

* + 1. Машина Больцмана
    2. Ограниченная машина Больцмана
    3. Автокодировщик
    4. Разреженный автокодировщик
    5. Вариационные автокодировщики
    6. Шумоподавляющие автокодировщики
    7. Сеть типа «deep belief»

1. Актуальность темы исследования

Проблема поиска подходящего товара или услуги является вечной: ежедневно человек делает выбор в пользу наиболее подходящего, начиная с выбора яблок в продуктовом магазине, заканчивая выбором подходящего врача. С развитием технологий кинематограф не стал исключением и фильмы начали распространять через онлайн-кинотеатры. Например, по итогам 2020 года российская аудитория онлайн-кинотеатров выросла на 52% и составила 8,5 млн человек, согласно исследованию "ТМТ Консалтинг".

В некоторых случаях, крупные хиты могут выходить в обход кинотеатров, сразу в «онлайне», или появляться в интернете через неделю-две после премьеры.

С ростом спроса на просмотр фильмов в интернете растет и предложение, что приводит к появлению новых игроков на этом рынке. На данный момент конкуренция растет, идет борьба за пользователей. Среди мировых «игроков» можно выделить Netflix, Amazon Prime, Disney+ и Apple TV. Более того, спрос оказался на столько высок, что привел к появлению локальных сервисов в России, таких как: Ivi, Кинопоиск, Окко, Амедиатека, Megogo и многие другие.

Таким образом, для огромного количества пользователей сети Интернет, доступ к фильмам и сериалам значительно упростился. Но возникла другая проблема: что посмотреть при огромном выборе контента? Искать подходящий фильм вручную может оказаться непростой задачей, требующей времени.  
По частоте посещений пользователей можно условно разделить на группы:

1. 1-2 раза в месяц
2. каждый день или почти каждый день
3. несколько раз в месяц
4. два-три раза в неделю

В качестве основы для такого разделения пользователей была использована инфографика от Telecom Daily за первое полугодие 2021 года (Рисунок 5).



Рисунок 5 - Частота посещений онлайн-кинотеатров в России по данным TELECOM DAILY

Всех пользователей объединяет одна проблема - трата времени на подбор подходящего фильма для просмотра. Более того, чем больше просматривается контента, тем сложнее становится найти что-то подходящее. Нередко случаются ситуации, когда в качестве рекомендованного к просмотру контента всем предлагаются одни и те же популярные варианты, когда как менее известные фильмы, не смотря на их возможную актуальность для конкретного пользователя, просто теряются среди подобных вариантов.

Релевантные рекомендации сокращают время, необходимое для поиска товаров и услуг, и значительно увеличивают вероятность попадания в поле зрения пользователя других объектов, которые смогут его заинтересовать. В результате повышается лояльность и удовлетворенность пользователей веб-сервисами. Как правило, пользователи также взаимодействуют с большим количеством товаров, что приводит к увеличению потребления контента и росту прибыли. Кроме того, информационные бюллетени, персонализированные рекламные материалы и push-уведомления побуждают пользователей возвращаться, увеличивают частоту посещений постоянными пользователями и уменьшают отток клиентов.

Одна из основных задач рекомендательной систем – увеличить вовлеченность пользователя в использование сервиса. Например, 80% просматриваемого контента на Netflix – это рекомендованные пользователям фильмы или сериалы. Благодаря упрощению пользовательского опыта за счет рекомендательной системы им удалось увеличить процент постоянных пользователей и сэкономить 1 миллиард долларов на привлечении пользователей (на момент 2016 года). Данная рекомендательная система не сможет представлять конкуренции такому гиганту, но при этом не требует больших вложений, улучшая пользовательский опыт, который впоследствии конвертируется в увеличение статистики посещения и просмотров, что напрямую влияет на увеличение прибыли и сокращение трат на привлечение и удержание пользователей.

Сегодня каждой компании просто необходимо наладить процесс сбора данных и уметь грамотно и эффективно использовать их в бизнесе, тем самым оптимизируя и улучшая пользовательский опыт, снижая расходы, увеличивая выручку и средний чек, и повышая рентабельность бизнеса в целом.

1. Постановка задачи и техническое задание
   1. Основные задачи

Основными задачами являются:

1. Анализ различных видов РС
2. Выбор наиболее эффективного подхода к реализации РС
3. Реализация наиболее эффективной РС
4. Реализация веб-интерфейса для взаимодействия с РС
   1. Техническое задание
      1. Введение

В качестве объектов для реализации выступает рекомендательная система фильмов Reccorsys (Recommendation core system) и пользовательский web-интерфейс. Данная рекомендательная система может применяться в приложениях и сервисах, связанных с фильмами (их просмотром или оценкой).

* + 1. Назначение разработки

Итоговая программа используется в качестве помощника конечному пользователю в вопросе выбора фильма для просмотра, основываясь на его предпочтениях, а также предпочтениях других пользователей. Она позволит пользователю упростить проблему выбора фильма для просмотра, тем самым экономя время. Для создания удачного пользовательского опыта реализуется web-интерфейс для взаимодействия с программой.

* + 1. Требования к программе или программному изделию

Выполняемые функции:

* Обработка входных данных
* Получение новых данных в качестве результата работы алгоритма рекомендации
* Вывод данных

Входные данные должны быть представлены в формате csv или json.

Система должна удовлетворять критериям эффективности:

* Максимально допустимое время ответа системы – 10 сек
* Максимальный используемый объем оперативной памяти – 2 гб
  + 1. Требования к надежности

Надежное (устойчивое) функционирование программы должно быть обеспечено рядом организационно-технических мероприятий, а именно:

* Организацией бесперебойного питания технических средств;
* Выполнением рекомендаций Министерства здравоохранения и социального развития РФ, изложенных в приказе от 14 октября 2011 г. № 1175н «Об утверждении межотраслевых типовых норм времени работы по сервисному обслуживанию оборудования телемеханики, сопровождению и доработке программного обеспечения»
* Выполнением требований ГОСТ 51188-98. Защита информации. Испытания программных средств на наличие компьютерных вирусов;
* Необходимым уровнем квалификации сотрудников профильных подразделений.
  + 1. Время восстановления после отказа

Время восстановления после отказа, вызванного сбоем электропитания технических средств (иными внешними факторами), не должно превышать времени, необходимого для урегулирования этих проблем.

Время восстановления после отказа, вызванного неисправностью технических средств, фатальным сбоем (крахом) приложения, не должно превышать времени, требуемого на устранение неисправностей технических средств и переустановки программных средств.

Для сохранности данных производится их резервное копирование каждые 7 дней.

* + 1. Отказы из-за некорректных действий оператора

Предотвратить отказы программы вследствие некорректных действий оператора (пользователя) при взаимодействии с программой. Во избежание возникновения отказов программы по указанной выше причине следует тщательно проверять вводимые данные и ограничить возможности пользователя, во избежание ошибок программы.

* + 1. Условия эксплуатации

Климатические условия эксплуатации, при которых должны обеспечиваться заданные характеристики, должны удовлетворять требованиям, предъявляемым к техническим средствам в части условий их эксплуатации.

* + 1. Требования к численности и квалификации персонала

Минимальное количество персонала, требуемого для работы

программы, должно составлять не менее 2 штатных единиц: системный

программист и конечный пользователь программы – оператор. Системный программист должен иметь техническое образование. В перечень задач, выполняемых системным программистом, должны входить:

* задача поддержания работоспособности технических средств;
* задачи установки (инсталляции) и поддержания работоспособности системных программных средств – операционной системы;
* задача установки (инсталляции) программы.
  + 1. Требования к составу и параметрам технических средств

Для сервера:

В состав технических средств должен входить сервер или персональный компьютер с настроенным окружением для запуска python-файлов и JavaScript-файлов. Минимальные требования к аппаратной части:

* Оперативная память (ОЗУ) — объемом не менее 2 Гб;
* Объём свободного дискового пространства (SSD) — не менее 100 Гб;
* Процессор – Intel Core i5 (или выше);
* Доступ в интернет

Для клиента:

В состав технических средств должен входить персональный компьютер с доступом к интернету и скоростью соединения не менее 10 мб/с.

* + 1. Требования к информационной и программной совместимости

Пользовательский интерфейс должен быть интуитивно понятным. На данных этапах разработки доступа к подстройке системы извне не предусмотрен.

* + 1. Требования к исходным кодам и языкам программирования

Исходные коды рекомендательной системы и серверной части web-приложения должны быть реализованы на языке программирования Python версии 3.0 и выше. Клиентская часть должна быть реализована на языке программирования JavaScript спецификации ES6 (ES2015) и выше. В качестве среды разработки предлагается использовать Visual Studio Code.

* + 1. Требования к программным средствам, используемым программой

Системные программные средства, используемые программой, должны быть представлены операционными системы Windows 7 и выше, MacOS версии 10.9 и выше. Для выполнения математических операций в Python-коде используется библиотека math, для работы с массивами – numpy. Для клиентской части используются библиотеки: react – для расширения возможностей языка и добавления «реактивности», typescript – для типизации, node-sass – для стилизации.

* + 1. Требования к защите информации и программ

В системе должен быть обеспечен надлежащий уровень защиты информации в соответствии с законом о защите персональной информации и программного комплекса в целом от несанкционированного доступа – «Об информации, информационных технологиях и о защите информации» Федеральный закон от 27.07.2006 г. No 149-ФЗ.

* + 1. Специальные требования

Программа должна обеспечивать взаимодействие с пользователем (оператором) посредством графического пользовательского интерфейса.

* + 1. Требования к программной документации

Программная документация для разработчика системы не требуется. Руководство пользователя не требуется.

* + 1. Технико-экономические показатели

Работы по проекту разделены на 5 этапов, суммарная длительность

которых составила 60 дней. Договорная цена разработки составила 300 000

рублей. Плановая прибыль от проекта составляет 20% от текущего уровня.

* + 1. Стадии и этапы разработки

Стадии разработки:

Разработка должна быть проведена в пять стадий:

* разработка технического задания;
* постановка задачи и анализ предметной области;
* проектирование программного обеспечения;
* разработка программного кода;
* тестирование и дальнейшая модификация.

Этапы разработки:

На стадии разработки технического задания должен быть выполнен этап разработки, согласования и утверждения настоящего технического задания.

На стадии постановки задачи и анализа предметной области должны

быть выполнены перечисленные ниже этапы работ:

* определить и вербально описать предметную область как сферу человеческой деятельности;
* предметную область определить, как объект со всем комплексом понятий и знаний о ее функционировании;
* обозначить цель разработки программного обеспечения. В соответствии с целью определить задачи.

На стадии проектирования программного обеспечения должны быть выполнены перечисленные ниже этапы работ:

* определить входные и выходные данные программного обеспечения;
* предоставить проектную реализацию в виде графических диаграмм, основанных на структурном подходе.

На стадии разработки программного кода должен быть выполнен этап написания кода на определенном языке программирования.

На стадии тестирование и дальнейшая модификация должны быть выполнены перечисленные ниже этапы работ:

* оценка разрабатываемого программного обеспечения, чтобы проверить его возможности, способности и соответствие ожидаемым результатам;
* описать дальнейшее видение программной реализации.
  + 1. Содержание работ по этапам

На этапе разработки технического задания должны быть выполнены

перечисленные ниже работы:

* определение и уточнение требований к техническим средствам;
* определение требований к программе;
* определение стадий, этапов и сроков разработки программы и
* документации на неё;
* выбор языков программирования;
* согласование и утверждение технического задания.

На этапе постановка задачи и анализ предметной области должны быть выполнены перечисленные ниже работы:

* описать сценарии использования программного обеспечения;
* определить квалификацию пользователей и условия, в которых они будут работать;
* определить объемы передаваемых данных и модель данных, которая будет использоваться, внесите ясность, в то, какие типы данных будут использоваться;
* описать примеры аналогичных реализаций, уточнить их преимущества и недостатки.

На этапе проектирование программного обеспечения должны быть

выполнены перечисленные ниже работы:

* создать функциональную модели программного обеспечения (IDEF0);
* создать базу данных основанную на диаграмме «сущность-связь» (IDEF1);
* создать модель потоков данных (нотация Гейна-Сэрсона).

На этапе разработки программного кода должна быть выполнена работа по программированию и отладке программы. Должна быть выполнена разработка программных документов в соответствии с требованиями ГОСТ 19.101-77.

На этапе тестирование и дальнейшая модификация должны быть выполнены перечисленные ниже работы:

* разработка, согласование и утверждение программы и методики испытаний;
* корректировка программы и программной документации по результатам испытаний.

1. Этап проектирования

В качестве методологии разработки ПО выбрана каскадная модель с промежуточным контролем (Рисунок 6).



Рисунок 6 - Каскадная модель с промежуточным контролем

Достоинства каскадной модели:

* постоянство системных требований на протяжении всего жизненного цикла разработки;
* каждый этап заканчивается набором проектной документации, которая должна отвечать критерию полноты;
* простота исполнения и применения модели обусловлена понятностью действий;
* легкость планирования работ и сроков реализации программного обеспечения на основании логической связанности последовательности этапов жизненного цикла.

Недостатки каскадной модели:

* большая дистанция жизненного цикла приводит к неадекватности чёткого формирования требований, а иногда даже к невозможности их изменений;
* низкая пластичность в управлении проектом;
* логическая последовательность определяет линейность этапов разработки, возвращение к предыдущим шагам приводит к срыву разработки программного обеспечения;
* промежуточный продукт не может использоваться;
* разработка специальных систем на основе гибкого моделирования невозможна;
* разработка больших информационных систем вызывает сложности интеграции всех программных продуктов;
* модель обусловлена большими рисками при реализации уникальных программных продуктов, в связи с невозможностью реализации гибкого моделирования.
  1. Реализация общей структуры

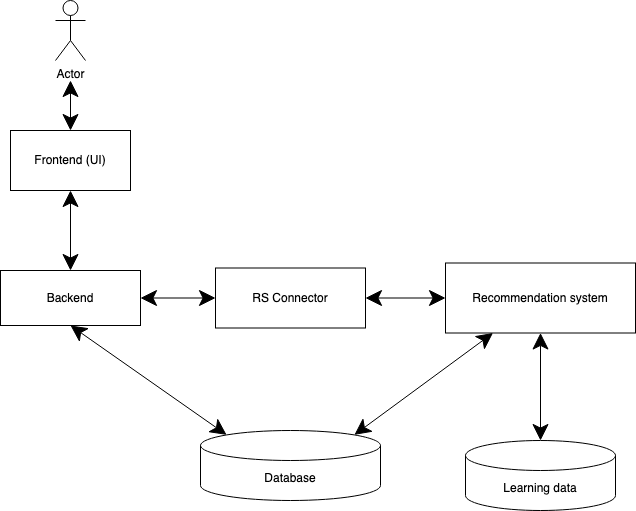


Рисунок 7 - Общая структура приложения

Общая структура приложения включает в себя (Рисунок 7):

* Frontend - web-интерфейс для взаимодействия пользователя с системой, реализуется с использованием языка программирования Javascript (спецификации ES6 и выше), библиотек React и node-sass;
* Backend – серверная часть пользовательского интерфейса, реализуется с использованием языка программирования Python и фреймворка Django;
* RS Connector (Recommendation service connector) – сервис-коннектор между серверной частью пользовательского интерфейса и рекомендательной системой, реализуется с использованием языка программирования Python;
* Recommendation system – рекомендательная система фильмов, реализуется с использованием языка программирования Python;
* Database – база данных, в которой хранятся данные пользователей ресурса, а так же предустановленные данные, необходимые для ускорения работы программы, представлена в формате базы данных SQL;
* Learning data – база данных с обучающей выборкой, необходимая для «холодного старта», представлена в формате набора файлов формата .csv.
  1. Реализация рекомендательной системы



Рисунок 8 - Структура рекомендательной системы

Рекомендательная система состоит из (Рисунок 8):

* RS Interface – интерфейс рекомендательной системы для связи с внешними потребителями (RS Connector);
* RS Controller – контроллер рекомендательной системы, в котором хранится логика для обработки и систематизации данных, полученных в результате работы рекомендательной системы и обработки запросов от интерфейса рекомендательной системы. Также имеет механизмы обращения к базе данных и к обучающей выборке;
* Hybrid RS – реализация гибридной рекомендательной системы, взаимодействует рекомендательными системами, такими как: основанная на контенте (Content-based RS), коллаборативной фильтрации (Collaborative Filtering RS);
* Content-based RS – рекомендательная система, основанная на контенте, принимает запросы от Hybrid RS и возвращает результаты работы рекомендательного алгоритма;
* Collaborative Filtering RS – рекомендательная система, основанная на коллаборативной фильтрации, принимает запросы от Hybrid RS и возвращает результаты работы рекомендательного алгоритма.

Заключение

Были установлены цели и задачи работы, проведено исследование предметной области и собрана пользовательская статистика. Сформировались четкие условия и требования к продукту, проведен архитектурный анализ и построение верхнеуровневой архитектуры. Сформировано техническое задание, определены программные средства реализации ПО.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Thomas Hofmann and Jan Puzicha. Latent class models for collaborative filtering. In IJCAI, volume 99, pages 688–693, 1999
2. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.  
   Knowledge and Data Engineering, IEEE Transactions on, 17(6):734–749, 2005.
3. Akiko Aizawa. An information-theoretic perspective of tf–idf measures // Information Processing & Management, 2003. Vol. 39, Issue 1, pp. 45-65