

Рекомендательная система для фильмов

с использованием Больших Языковых Моделей

Команда:

Матренок Семён Сергеевич

Павар Юрий Александрович

Малюшитский Кирилл Дмитриевич

Куратор:

Милоградский Александр Николаевич



Описание задачи

Основной фокус данной работы — использование Больших Языковых Моделей для построения рекомендательной системы для фильмов

- "content based" подход

Система опирается в большей степени на специфику запроса пользователя, а не на множество понравившихся фильмов. Акцент делается на детальное понимание, что требуется конкретному пользователю.

- использование техники Retrieval Augmented Generation (RAG)

Использование RAG позволяет комбинировать достоверную информацию, полученную из базы знаний, с простым и интуитивным интерфейсом взаимодействия пользователя с LLM

- использование текстовых эмбеддингов кратких описаний фильмов

База знаний строится на основе векторной базы данных с использованием метрики близости текстовых эмбеддингов кратких описаний фильмов, что позволяет получать контекстную информацию о фильмах, релевантную для запроса пользователя

Описание данных

Данные состоят из 2 частей:

1. Основной датасет

- Краткое описание сюжетов фильмов, полученных из англоязычной Википедии
- Метаданные 81 741 фильмов, собранные из Freebase
- Метаданные 450 669 персонажей фильмов

2. Пользовательские рейтинги фильмов

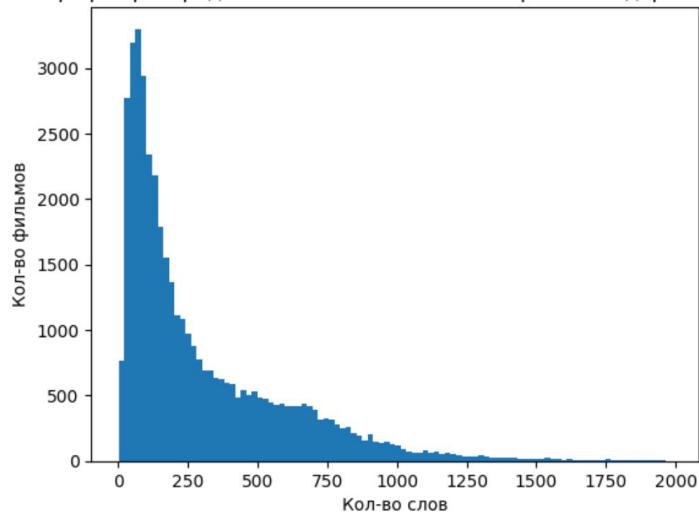
- Метаданные 45 000 фильмов
- Связка фильмов с рейтингами
- 26 млн оценок от 270 000 пользователей для 45 000 фильмов

После объединения этих 2 частей удалось получить матрицу оценок от 269 773 пользователей по 18 730 фильмам

Подготовка данных

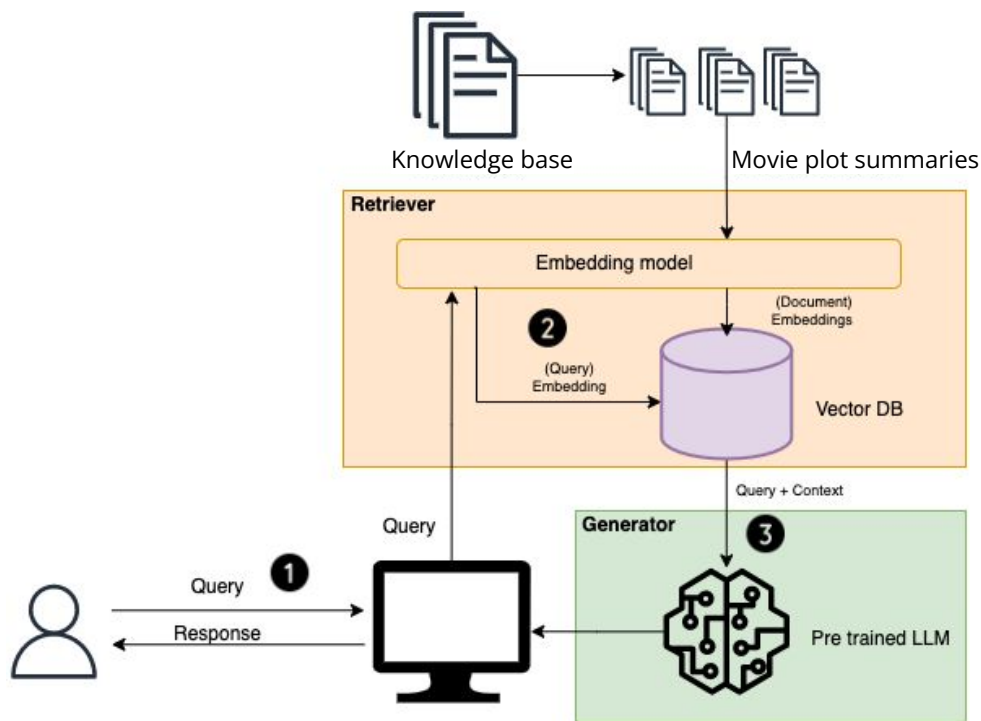
1. Очистка датасета от выбросов (возраст актеров, год выпуска, длительность фильма)
2. Проверка основного используемого признака (кратких описаний фильмов) на возможность использования со стандартными Языковыми Моделями с контекстным окном в 4к токенов

График распределения количества слов в кратких содержаниях



Длина текстовых описаний в основном не превосходит 1500 слов, что примерно соответствует 2000 токенов для лингвистической модели (соответствует длине контекста большинства LLMs)

Архитектура рекомендательной системы



Описание экспериментов

Выбор оптимальной комбинации **LLM + Текстовые эмбединги**

- Сборка RAG пайплайна и инференс необходимых моделей
- Тестирование различных **LLM** (выбор базовой модели, выбор файнтюна, выбор квантизации для инференса) и настройка гиперпараметров (см. подробное описание далее)
- Тестирование различных **текстовых эмбедингов** (выбор модели и настройка гиперпараметров) (см. подробное описание далее)
- Сборка пайплайна с выбранными LLM и текстовыми эмбедингами и финальная настройка гиперпараметров и усовершенствование промпта

Описание экспериментов

Схема оценки качества работы LLM в RAG пайплайне:

- Инференс выбранной LLM
- Промпт инжиниринг: поиск оптимального формата представления мэтчей из векторной базы, различные техники для улучшения релевантности и надежности работы системы (Few-shot, CoT, Zero-CoT, rawinst, sysinst, ...)
- Подбор гиперпараметров (температура, top-p, top-k, typical-p, repetition penalty)
- Визуальная оценка качества работы модели и сравнение с другими моделями*

Исследованные модели:

- Базовые модели: Mistral-7b, Llama-2-7b/13b, Flan-t5-xl
- Инструкт файнтюны: Mistral-7B-Instruct-v0.1/v0.2, Llama-2-7b/13b-chat, zephyr-7b-alpha, neural-chat-7b-v3

* Альтернативой визуальной оценке качества является использование gpt-4 для оценки, однако официально мы не можем использовать API OpenAI из России, и, к тому же, это не выглядит необходимостью

Описание экспериментов

Схема оценки качества работы текстовых эмбеддингов в RAG пайплайне:

- Инференс выбранной LLM для получения текстовых эмбеддингов
- Промпт инжиниринг (для эмбеддинг-моделей, имеющих такой функционал)
- Подбор гиперпараметров (similarity_top_k, min_similarity)
- Визуальная оценка качества работы модели*

Исследованные модели:

- Модели текстовых эмбеддингов: intfloat/multilingual-e5-large-instruct, BAAI/bge-large-en-v1.5, BAAI/bge-base-en-v1.5, llmrails/ember-v1

* Альтернативой визуальной оценке качества является использование gpt-4 для оценки, однако официально мы не можем использовать API OpenAI из России, и к тому же это не выглядит необходимостью

Описание итоговой модели

RAG система на основе LLM и векторной базы данных

- **LLM:** Zephyr 7B Alpha – файнтюн версия модели Mistral 7B, дообученная с использованием DPO на UltraFeedback датасете
- Модель развернута в 4-битной квантизации
- **Текстовые эмбединги:** BAAI/bge-large-en-v1.5 – BERT-like модель, предобученная с использованием RetroMAE дообученная с помощью contrastive learning
- Векторная база данных, встроенная в пакет LlamaIndex

Демонстрация работы системы

M

MovieRec
bot

🔍 ...

/llm 17:08 ✓

Введите запрос: 17:08

I've enjoyed the film King Kong (1933), could you recommend me few similar films? 17:08 ✓

Certainly! Based on the similarities between King Kong (1933) and other films, here are a few recommendations:

1. King Kong (2005) - This is a remake of the 1933 classic, directed by Peter Jackson. It follows a similar plot, with a team of filmmakers traveling to a mysterious island to capture footage of a legendary creature.
2. Godzilla (1954) - This Japanese classic is a monster movie that features a giant radioactive creature wreaking havoc on a city. It has been remade several times, but the original is a must-see for fans of the genre.
3. The Beast from 20,000 Fathoms (1953) - This film, also directed by Jack Arnold, is about a giant prehistoric creature that emerges from the Arctic and begins destroying cities.
4. Them! (1954) - This film is about giant ants that are terrorizing the southwestern United States. It's a classic example of the "giant monster" genre.
5. Cloverfield (2008) - This found-footage film is about a giant monster that attacks New York City. It's a modern take on the genre, with a unique perspective.
6. Pacific Rim (2013) - This film is about giant robots fighting giant monsters, known as Kaiju. It's a fun and exciting action movie that pays homage to the classic monster movies.

M

MovieRec
bot

🔍 ...

/llm 17:08 ✓

Введите запрос: 17:08

I like a dystopian film (forgot its name) where teenagers between the ages of 12 and 18 from twelve districts are selected by lottery to participate in a televised battle (games) to the death, set in a future, post-apocalyptic society. What is the name of the film that i described? What 3 similar films without that i mentioned can you recommend? 17:08 ✓

The film you described is "The Hunger Games" (2012). Here are three similar films that you may enjoy:

1. "The Maze Runner" (2014) - In a post-apocalyptic world, teenagers are selected by lottery to enter a maze filled with deadly creatures and must fight to survive.
2. "Divergent" (2014) - In a dystopian society, teenagers are divided into factions based on their personalities and must fight to the death in an annual competition.
3. "The Giver" (2014) - In a society where emotions are suppressed, a teenage boy is chosen to receive memories of the past and must decide whether to stay in the present or fight for a better future.

17:08

/rec 17:09 ✓

Введите запрос: 17:09

recommend me films like Harry Potter but not films from Harry Potter series 17:09 ✓

Спасибо. Теперь, пожалуйста, выберите количество

M

MovieRec
bot

🔍 ...

/llm 17:07 ✓

Введите запрос: 17:07

Recommend me some happy new year films 17:07 ✓

Based on the context information provided, here are some happy new year films that you might enjoy:

1. Happy New Year (1987) - This film follows two jewel thieves who plan to rob a Florida jewelry store on New Year's Eve. However, their plans are disrupted when one of them falls in love with a woman who owns an antique shop next to the store.
2. Rudolph's Shiny New Year (nan) - In this animated television special, Rudolph and his friends go on a quest to find the next Baby New Year before midnight on New Year's Eve. Along the way, they encounter various obstacles and villains, but ultimately save the day and ensure a happy new year.
3. Operation "Happy New Year!" (2006) - This Russian comedy follows a group of people who meet in an emergency room on New Year's Eve and learn a few things about life and happiness.
4. New Year's Eve (2011) - This romantic comedy follows several interconnected stories that take place on New Year's Eve in New York City. The film stars an all-star cast, including Halle Berry, Jessica Biel, and Robert De Niro.
5. The Holiday (2006) - While not specifically a "happy new year" film, this romantic comedy does take place during the holiday season and features a happy ending. The film follows

Демонстрация работы системы

Без использования LLM
(запрос к векторной базе)

MovieRec bot

Введите запрос: 16:45

recommend me films like Harry Potter but not films from Harry Potter series 16:45 ✓

Спасибо. Теперь, пожалуйста, выберите количество рекомендаций: 16:45

4 16:45 ✓

For your query 'recommend me films like harry potter but not films from harry potter series' we can recommend next movies:

Movie name: Harry Potter and the Prisoner of Azkaban
Movie release year: 2004
Movie runtime: 141
Movie languages: English Language
Movie genres: Fantasy Adventure, Adventure, Children's/Family, Mystery, Fantasy, Children's Fantasy, Family Film
Synopsis: {{further2}} Now 13-years-old, Harry Potter has been spending the summer at Privet Drive, absorbing most of his time studying new spells. When Uncle Vernon's sister, Marge, comes for a visit and infuriates Harry by insulting his parents, he accidentally causes her to inflate and fly away. Harry loses his temper and threatens to curse Vernon but flees, fed up with his life at Privet Drive. The Knight Bus appears and delivers Harry to the Leaky Cauldron, where Minister of Magic Cornelius Fudge tel

Movie name: Harry Potter and the Half-Blood Prince
Movie release year: 2009
Movie runtime: 153

С использованием LLM

MovieRec bot

Введите запрос: 16:45

recommend me films like Harry Potter but not films from Harry Potter series 16:45 ✓

Here are some film recommendations that are similar to the Harry Potter series but not part of the series itself:

1. The Chronicles of Narnia: The Lion, the Witch and the Wardrobe (2005)
2. The Lord of the Rings trilogy (2001-2003)
3. The Dark is Rising (2009)
4. The Golden Compass (2007)
5. The Hunger Games trilogy (2012-2015)
6. The Maze Runner trilogy (2014-2018)
7. The Mortal Instruments: City of Bones (2013)
8. The Perks of Being a Wallflower (2012)
9. The Secret Life of Bees (2008)
10. The Voyage of the Dawn Treader (2010)

These films all have elements of fantasy, adventure, and coming-of-age stories, which are similar to the Harry Potter series. 16:46

Введите запрос: 16:46

recommend me films like Harry Potter 16:46 ✓

Based on the similarities between the Harry Potter series and other popular fantasy films, here are some recommendations:

С использованием LLM (модель сама понимает, что интересны фильмы **не** из той же серии)

MovieRec bot

Введите запрос: 16:46

recommend me films like Harry Potter 16:46 ✓

Based on the similarities between the Harry Potter series and other popular fantasy films, here are some recommendations:

1. The Lord of the Rings trilogy (2001-2003) - This epic trilogy, based on J.R.R. Tolkien's novel, follows the journey of hobbit Frodo Baggins as he sets out to destroy the One Ring, which has fallen into the hands of the evil Sauron. Like Harry Potter, this series features a group of young heroes on a quest to save the world from an evil force.
2. The Chronicles of Narnia series (2005-2010) - This series, based on C.S. Lewis' novels, follows the adventures of four siblings who are transported to the magical land of Narnia. Like Harry Potter, this series features a group of young heroes on a quest to save the world from an evil force.
3. The Hunger Games series (2012-2015) - This series, based on Suzanne Collins' novels, is set in a dystopian future where teenagers are forced to compete in a deadly game. Like Harry Potter, this series features a group of young heroes on a quest to save the world from an evil force.
4. The Mortal Instruments series (2013) - This series, based on Cassandra Clare's novels, follows the adventures of a young woman named Clary Fray as she discovers that she is a Shadowhunter, a member of a secret society that protects the world from demons. Like Harry Potter, this series features a group of young heroes on a quest to save the world from an evil force.

Описание сервиса и бота

Рекомендательная система

- Выдача N похожих фильмов по введенному описанию
- Выдача N похожих фильмов по набору описаний
- Получение рекомендаций в формате чат-бота с использованием LLM

Дополнительный функционал

- N самых кассовых фильмов
- Информация о фильме по введенному названию
- Информация об актере по ФИО

Результаты и планы

Уже сделано:

- Построена промежуточная система на основе Retrieval модуля
- Реализована итоговая архитектура на основе RAG пайплайна с Retrieval модулем и LLM
- Проведены тесты языковых моделей и среди них выбраны оптимальные
- Проведена работа по prompt engineering
- Проведено ручное тестирование корректности работы системы

Дальнейшие планы:

- Построение бейзлайн решения на основе ML моделей
- Сравнение промежуточной системы на основе Retrieval модуля с бейзлайн решением при помощи выбранных метрик (NDCG, precision, recall)
- Интеграция ML модели в content-based RAG пайплайн для совмещения преимуществ двух подходов (классического на основе ML модели и content-based на основе LLM)
- Эксперименты с различными гиперпараметрами и эвристиками для улучшения качества и релевантности контекста, получаемого из Retrieval модуля