

ВНЕДРЕНИЕ В ВЕБ-ПЛАТФОРМУ МОДЕЛИ РЕКОМЕНДАЦИЙ НА БАЗЕ FAISS

Шныра Б.О., Андриевская Н. К.

Донецкий национальный технический университет»

кафедра автоматизированных систем управления

E-mail: mi_shnyra@mail.ru

Аннотация:

Шныра Б.О., Андриевская Н.К. Внедрение в веб-платформу модели рекомендаций на базе FAISS. В статье рассматривается процесс внедрения интеллектуальной рекомендательной системы на основе мультимодальной модели CLIP и библиотеки FAISS в веб-платформу. Целью исследования являлась разработка высокопроизводительного решения, способного генерировать персонализированные рекомендации на основе анализа изображений и текстовых описаний. В работе использована предобученная модель CLIP, адаптированная для задачи формирования рекомендаций. Для обеспечения скорости и масштабируемости применены алгоритмы библиотеки FAISS, включая L2-нормализацию векторов и косинусную меру близости. Результаты подтверждают, что сочетание CLIP и FAISS является перспективным для создания семантически осмысленных рекомендаций в креативных индустриях.

Annotation:

Shnyra B.O., Andrievskaya N.K. Implementation of the recommendation model with FAISS in the web platform. The article discusses the process of implementing an intelligent recommendation system based on the multimodal CLIP model and the FAISS library in a web platform. The aim of the research was to develop a high-performance solution capable of generating personalized recommendations based on the analysis of images and text descriptions. The paper uses a pre-trained CLIP model adapted for the task of recommendations. To ensure speed and scalability, FAISS algorithms are used, including L2-normalization of vectors and cosine proximity measure. The results confirm that the combination of CLIP and FAISS is promising for creating semantically meaningful recommendations in the creative industries.

Ключевые слова: высокоразмерные векторные пространства, рекомендательные системы, библиотека FAISS, embedding-пространства, индексация данных.

Keywords: high-dimensional vector spaces, recommendation system, FAISS library, embedding spaces, data indexing.

Введение. В современном цифровом обществе объем информации стремительно растет, и с каждым днем становится все более актуальной задача эффективного обмена идеями и знаниями между людьми. Интернет предоставляет огромные возможности для коммуникации, однако часто отсутствует структурированная платформа, которая позволяла бы пользователям не только делиться своими мыслями и опытом, но и находить единомышленников, получать рекомендации и вдохновение для новых начинаний.

Возникла потребность в создании веб-ориентированной информационной системы общественной платформы обмена идеями, которая объединит людей с общими интересами, предоставит удобные инструменты для публикации и обсуждения идей, а также будет способствовать генерации новых знаний и воплощению замыслов. Важной компонентой такой системы является рекомендательная модель [1-2], которая позволит пользователям получать персональные рекомендации на основе их интересов и предпочтений, находить схожие объекты и расширять кругозор.

Внедрение модели рекомендаций в веб-систему включает в себя интеграцию предварительно обученной модели в веб-приложение, предвычисление эмбедингов для

эффективного поиска, а также реализацию механизмов генерации и отображения персонализированных рекомендаций пользователю. Использование современных инструментов, таких как CLIP и FAISS, позволяет обеспечить высокую производительность и качество рекомендаций на платформе обмена творческими идеями.

Характеристика разработанной рекомендательной модели Clip. В рамках ранних исследований была разработана инновационная рекомендательная система, основанная на мультимодальной архитектуре CLIP (Contrastive Language-Image Pre-training)[3-4].

Библиотека FAISS от Facebook AI Research позволяет реализовать передовые методы векторного поиска, основанные на глубоких математических принципах машинного обучения.

Ключевой формулой для измерения близости между векторами служит евклидово расстояние:

$$D_{euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (1)$$

где x, y - векторы признаков, n - размерность пространства признаков.

Данная формула позволяет количественно оценить расстояние между точками в многомерном пространстве, что критично для задач семантического поиска.

Процесс формирования embedding-пространства описывается следующей математической моделью:

$$E_{embedding} = \frac{f_{image}(I) + f_{text}(T)}{2}, \quad (2)$$

где f_{image} - функция извлечения признаков изображения, f_{text} - функция извлечения текстовых признаков, I - входное изображение, T - текстовое описание.

Принципиальным этапом является L2-нормализация векторов:

$$U_{normalized} = \frac{v}{||v||} = \frac{v}{\sqrt{\sum_{i=1}^n v_i^2}}, \quad (3)$$

где v - исходный вектор, $||v||$ - евклидова норма вектора, n - число измерений.

Нормализация обеспечивает единую шкалу измерения для различных векторных представлений.

Алгоритмы поиска и рекомендаций, ключевая формула семантической близости:

$$Similarity(q, v_i) = \max \{\cos(q, v_i)\}, \quad (4)$$

где q - вектор пользовательского запроса, v_i - векторные представления в базе, $\cos()$ - косинусная мера близости.

Агрегация пользовательского профиля, усреднение embedding просмотренных объектов:

$$U_{embedding} = \frac{1}{|V|} \sum_{v \in V} v, \quad (5)$$

где $U_{embedding}$ - усредненный embedding пользователя, V - множество просмотренных векторов, $|V|$ - число просмотров.

Вычислительная эффективность сложность основных операций:

– Индексация: $O(n \cdot d)$;

– Поиск: $O(\log n)$,

где n - число векторов, d - размерность пространства признаков.

Модель представляет собой передовое решение для персонализированных рекомендаций, сочетающее в себе семантический анализ мультимодальных данных (изображений и текста). Параметры модели приведены в таблице 1 ниже.

Таблица 1- Параметры разработанной модели Clip

Параметры	CLIP-based Model
Архитектура	Предобученная CLIP с адаптацией
Входные данные	Изображения + текст
Embedding Dim	512
Loss Function	InfoNCE
Предобучение	Использует предобученные веса CLIP
Сложность модели	Высокая
Вычислительные затраты	Умеренные

В рамках комплексной оценки разработанной CLIP-based рекомендательной модели проведен многоуровневый анализ эффективности с использованием передовых метрик машинного обучения [5]. Представленная ниже таблица с метриками отражает результаты тестирования рекомендательной модели.

Таблица 2- Метрики разработанной модели Clip

Метрика	CLIP-based Model
Recall@1	0,3244
Precision@1	0,3244
F1@1	0,3244
Recall@5	0.5913
Precision@5	0.4224
F1@5	0.4928
Recall@10	0.7037
Precision@10	0.4373
F1@10	0.5394
Median Rank	3.00
MRR	0.4472

Представленная CLIP-based модель демонстрирует высокий потенциал для практического применения в рекомендательных системах, обеспечивая достаточно качественный семантический поиск (F-мера для 10 результатов в выдаче составляет 0.54) и персонализацию для данной веб-системы.

Выбор средства разработки программного обеспечения системы. Основным выбором для реализации проекта стал язык Python — универсальный и многоцелевой язык программирования.

Для разработки интерфейса системы был выбран фреймворк Django. Django является высокоуровневым веб-фреймворком на языке программирования Python, который предназначен для быстрого создания веб-сайтов и веб-приложений.

Язык разметки гипертекста HTML послужил основой для создания структурированной и семантически корректной разметки веб-страниц. Каскадные таблицы стилей CSS использовались для визуального оформления и стилизации элементов, созданных с помощью HTML.

Facebook AI Research Similarity Search (FAISS) – разработка команды Facebook AI Research для реализации нейросетевых алгоритмов, в том числе алгоритма быстрого поиска ближайших соседей и кластеризации в векторном пространстве. Высокая скорость поиска позволяет работать с очень большими данными – с векторами объёмом до нескольких миллиардов [6].

PyTorch — современная библиотека глубокого обучения, развивающаяся под крылом Facebook [7].

Для обеспечения надежного хранения, быстрой обработки и безопасного доступа к данным была выбрана система управления базами данных (СУБД) PostgreSQL.

Разработка схемы взаимодействия объектов системы. В архитектуре программного обеспечения и разработке комплексных систем важным аспектом является понимание того, как отдельные компоненты и объекты взаимодействуют друг с другом. Это необходимо для обеспечения гармоничного функционирования всей системы и для достижения запланированных конечных результатов в рамках проектных требований. Данная схема представлена на изображении 1.

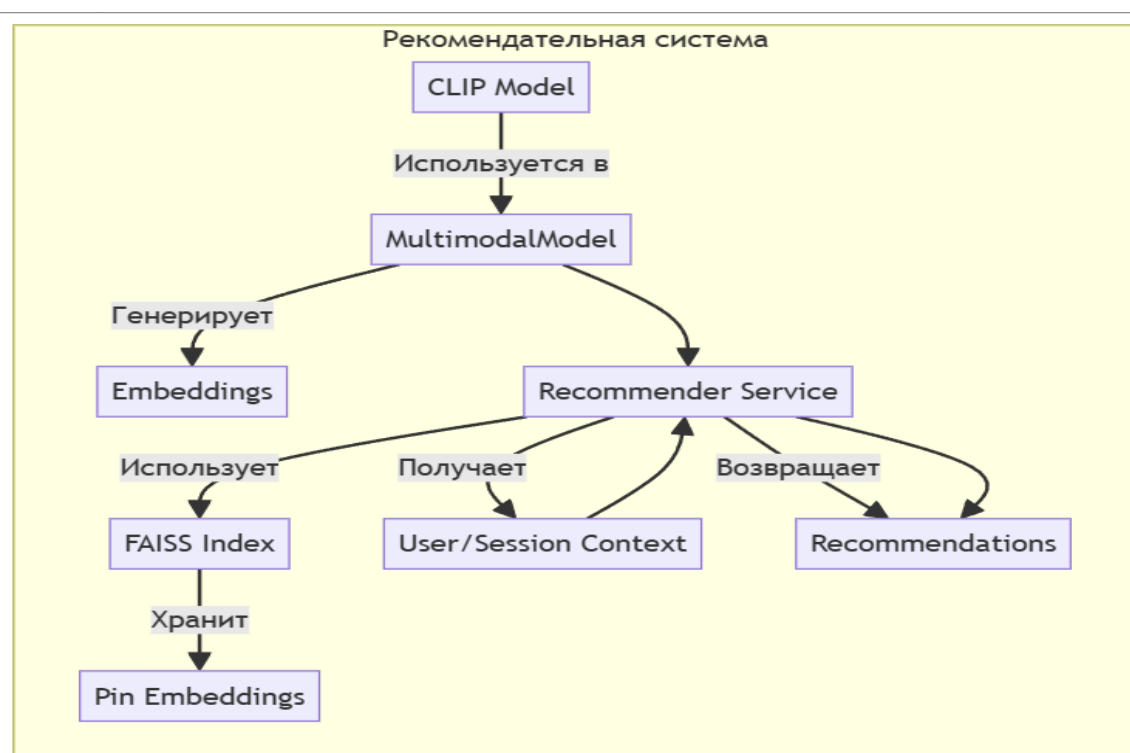


Рисунок 1 – Схема взаимодействия объектов системы

Результаты работы системы. Протестируем как происходит персонализированные рекомендации для пользователей на платформе. Посмотрим изначальный главный экран как идеи там находятся на рисунке 2:

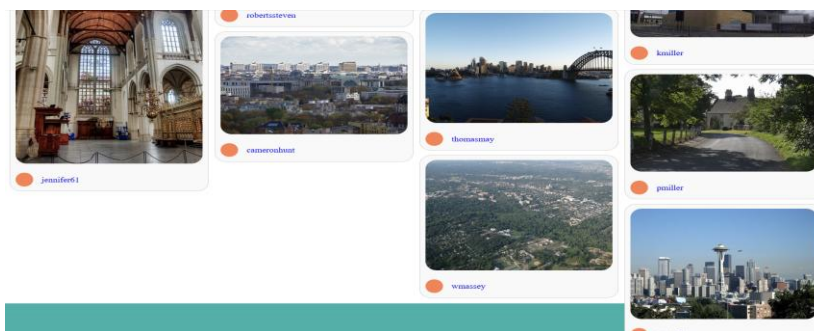


Рисунок 4 – Персонализированные рекомендации пользователя

Как видим, веб-платформа при визуальной оценке неплохо строит предпочтения для пользователя. Модель демонстрирует хороший потенциал, особенно при увеличении числа рекомендаций. В численном выражении для 10 рекомендаций полнота поиска составляет 0.7, точность поиска составляет 0.43, F-мера для рекомендаций составляет 0.54.

Заключение. Результаты работы подтверждают, что сочетание современных методов мультимодального анализа (CLIP) и высокопроизводительных алгоритмов поиска (FAISS) является перспективным направлением для создания интеллектуальных рекомендательных систем в креативных индустриях. Интеграция предобученной модели CLIP позволила эффективно обрабатывать мультимодальные данные (изображения и текст), формируя семантически значимые эмбединги, которые стали основой для персонализированных рекомендаций. Библиотека FAISS обеспечила высокую скорость и масштабируемость системы.

Литература

1. Шныра, Б.О. Использование сверточных нейросетей для анализа визуального контента веб-платформы / Б.О. Шныра, Н.К. Андриевская // Информатика, управляющие системы, математическое и компьютерное моделирование» (ИУСМКМ-2024): сб. материалов XV Междунар. науч.-техн. конф. в рамках X Междунар. Науч. Форума ДНР; Т.2 / Ред. кол.: Аноприенко А.Я. (пред.); Васяева Т.А.; Карабчевский В.В. [и др.]; от. ред. Р.В. Мальчева. – Донецк : ДонНТУ, 2024. – С. 690 – 696.
2. Шныра, Б.О. Использование контентной фильтрации рекомендательной системы для веб-платформы / Б.О. Шныра, Н.К. Андриевская // Бизнес-инжиниринг сложных систем: модели, технологии, инновации: Сборник материалов IX международной научно-практической конференции, Донецк, 20 ноября 2024 года. – Донецк - Екатеринбург: Донецкий национальный технический университет, 2024.
3. ruClip – мультимодальная модель для русского языка [Электронный ресурс] \ Режим доступа: <https://habr.com/ru/companies/sberdevices/articles/564440/>. Дата обращения: 01.05.2025.
4. Мультимодальная языковые модели: как нейросеть учатся видеть и слышать [Электронный ресурс] \ Режим доступа: <https://habr.com/ru/articles/892172/>. Дата обращения: 01.05.2025.
5. Показатели для оценки систем рекомендаций и ранжирования [Электронный ресурс] \ Режим доступа: <https://www.evidentlyai.com/ranking-metrics/evaluating-recommender-systems>. Дата обращения: 08.05.2025.
6. Faiss быстрый поиск лиц и клонов на многомиллионных данных [Электронный ресурс] \ Режим доступа: <https://habr.com/ru/companies/okkamgroup/articles/509204/> Дата обращения: 11.05.2025.
7. PyTorch- фреймворк машинного обучения [Электронный ресурс] \ Режим доступа: <https://habr.com/ru/companies/okkamgroup/articles/509204/> Дата обращения: 11.05.2025.