УДК 004.58:004.738.5

Кренц Олеся Павловна

студент

4 курс, факультет «Информатика и вычислительная техника»

Донской государственный технический университет

Россия, г. Ростов-на-Дону

Газизов Андрей Равильевич, доцент

Заведующий кафедрой «Информационная безопасность в вычислительных системах и сетях»

Донской государственный технический университет

Россия, г. Ростов-на-Дону

ИНТЕГРАЦИЯ ВИЗУАЛЬНЫХ ПРИЗНАКОВ В ГИБРИДНЫЕ РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ ДЛЯ СТРИМИНГОВЫХ ПЛАТФОРМ

Аннотация. *Автоматизация управления пользовательским опытом в вещательных и стриминговых сервисах — это критически важное для сохранения пользователей. Автор рассматривает способ увеличения точности предсказания интересов пользователя путем комбинации методов collaborative filtering с глубокими нейронными сетями. Акцент делается на получение высокоуровневых эмбеддингов изображений из видео и изучении влияния эмбеддингов на распределение рекомендаций в категории "Длинный хвост" каталога. Сравнение простой матричной факторизации с предложенной гибридной мультимодальной сетью по метрике ранжирования AUC показывает улучшение последнего метода. Результаты исследования позволяют сделать вывод о необходимости применения мультимодальных методов борьбы с систематическими ошибками рекомендательных систем.*

Ключевые слова: *мультимодальные сети, рекомендательные системы, глубокое обучение, компьютерное зрение, ранжирование контента, смещение популярности.*

DEVELOPMENT AND ANALYSIS OF A HYBRID MULTIMODAL RECOMMENDER SYSTEM FOR STREAMING PLATFORMS

Abstract. *Automating user experience management in broadcast and streaming services is critical to user retention. The author considers a way to increase the accuracy of predicting user interests by combining collaborative filtering methods with deep neural networks. The emphasis is on obtaining high-level embeddings of images from videos and studying the influence of embeddings on the distribution of recommendations in the "Long tail" category of the catalog. A comparison of simple matrix factorization with the proposed hybrid multimodal network using the AUC ranking metric shows an improvement in the latter method. The results of the study allow us to conclude that it is necessary to use multimodal methods to combat systematic errors in recommendation systems.*

Keywords: *multimodal recommender systems, deep learning, computer vision, content ranking, popularity bias, long tail.*

Введение

Современная индустрия цифрового видеовещания функционирует в условиях жесточайшей борьбы за удержание внимания потребителя. Согласно профильным исследованиям рынка, подавляющая часть пользовательских просмотров (до 80% на ключевых глобальных платформах) генерируется автоматическими рекомендательными блоками. Это превращает алгоритмы ранжирования контента в главный инструмент снижения оттока аудитории.

Тем не менее традиционные подходы часто страдают от эффекта «пузыря фильтров», зацикливая пользователя на узком круге мейнстримных позиций. Настоящее исследование направлено на разработку и валидацию модели, способной эффективно осваивать «длинный хвост» (*Long Tail*) доступного медиакаталога посредством глубокого анализа визуальной и стилистической составляющей видеоматериалов.

Предобработка данных и извлечение признаков

Экспериментальная база исследования построена на основе расширенного набора данных MovieLens 20M, в который были интегрированы дополнительные метаданные (жанровые теги, хронологические маркеры) и мультимедийные компоненты (постеры фильмов и их официальные трейлеры).

При формировании признакового пространства учитывались следующие критические факторы:

- **Лог пользовательских взаимодействий:** Векторизованная история оценок и просмотров (*User-Item Interactions*).
- **Индекс популярности контента:** Частотный показатель просмотров, который, с одной стороны, указывает на общие тренды, но с другой — негативно влияет на метрику новизны (*Novelty*) выдачи.
- **Мультимодальные дескрипторы контента:** Набор внутренних признаков (визуальных и звуковых), характеризующих эстетическую и атмосферную специфику произведения независимо от его известности широкой публике.

Методология Feature Extraction: Для извлечения семантических векторов из видеоряда трейлеров применялась предварительно обученная сверточная нейросеть ResNet-50. Этот подход позволил оцифровать такие

неочевидные для текстовых тегов характеристики, как цветовая палитра, специфика освещения и динамика монтажных склеек.

Анализ системных ошибок базового алгоритма

Первичная оценка качества работы классического алгоритма матричной факторизации (*Matrix Factorization*) выявила устойчивые паттерны неоптимального распределения рекомендаций:

1. Ложноотрицательные срабатывания (False Negatives)**:** Система систематически игнорировала редкие, авторские или нишевые картины, которые идеально соответствовали эстетическим вкусам пользователя, но не имели плотной сетки просмотров в общей матрице.

2. Ложноположительные срабатывания (False Positives): Алгоритм навязывал массовые высокобюджетные блокбастеры пользователям с выраженной склонностью к камерному кинематографу, что обусловлено встроенным искажением популярности (*Popularity Bias*).

Для минимизации данных дефектов были разработаны специализированные синтетические признаки:

- Visual_Style_Vector: Плотный эмбеддинг, кодирующий визуальную тональность и жанровую эстетику (например, нуар, неоновые палитры киберпанка или пастельные тона артхауса). Это позволило находить скрытые связи между фильмами разных категорий.

- Atmosphere_Match_Score: Метрика скалярного сходства между текущим визуальным вектором фильма и совокупным профилем предпочтений пользователя, сформированным на основе истории его просмотров.

Архитектура исследуемых моделей

В рамках экспериментальной проверки были сопоставлены два концептуально разных поколения рекомендательных систем:

- Модель 1 (Matrix Factorization / Neural CF): Базовое решение (*Baseline*), опирающееся на нейросетевую коллаборативную фильтрацию. Данный подход демонстрирует высокую точность при анализе явных сигналов (кликов, лайков), однако полностью изолирован от семантики и содержания самого контента.

- Модель 2 (Hybrid Multimodal Network): Предлагаемая комплексная архитектура, сочетающая в себе совместное обучение коллаборативного блока и контентной ветви глубокого анализа видеоряда.

Валидация систем проводилась по двум ключевым осям: точности ранжирования (метрика *AUC*) и степени диверсификации выдачи (*Diversity*).

Результаты экспериментов и обсуждение

Тестирование моделей на отложенной выборке показало существенное преимущество мультимодального подхода при оценке общей способности алгоритма к корректному ранжированию объектов.

Сравнительный анализ точности ранжирования

В Таблице 1 зафиксированы итоговые показатели площади под ROC-кривой (AUC) для обеих архитектур.

Таблица 1. Сравнение качества работы моделей

Архитектурное решение	**AUC (Area under the curve)**
Matrix Factorization	0.745
Hybrid Multimodal Network	0.812

Существенный прирост метрики AUC до 0.8120 у гибридной сети объясняется тем, что интеграция визуальных признаков компенсирует

недостаток информации в условиях «холодного старта» для новых или малоизвестных фильмов.

При этом статистически значимое улучшение ($\Delta AUC = +0.067$, $p < 0.01$ по t-тесту) свидетельствует о робастности гибридного подхода к разреженным данным пользовательских взаимодействий. Это даёт основания рекомендовать multimodal-архитектуру как приоритетное решение для продакшн-развёртывания, особенно в сценариях с высокой долей новых товаров в каталоге.

Оптимизация порога уверенности (Confidence Threshold)

Для практического внедрения модели в продакшн-среду критически важно найти баланс между точностью рекомендаций и разнообразием каталога, поскольку избыточный спам популярными фильмами вызывает выгорание пользователя. В Таблице 2 представлены результаты симуляции работы системы при различных значениях порога отсечения вероятности.

Таблица 2. Влияние порога уверенности на бизнес-метрики системы

Порог	Точность	Разнообразие	Покрытие Каталога
0.5	0.62	Высокое	85%
0.6	0.68	Среднее	60%
0.7	0.75	Низкое	35%

На основе полученных данных в качестве рабочего был выбран динамический порог 0.60. Данная конфигурация позволяет гарантировать, что 68% предложенных фильмов будут релевантны интересам пользователя, но при этом система сохраняет доступ к 60% всего медиакаталога, успешно выводя из «тени» скрытые нишевые произведения.

Заключение

Применение методов глубокого обучения и компьютерного зрения открывает новые возможности для автоматической персонализации медиаконтента. Проектирование визуальных признаков (*Visual Feature Engineering*) на базе глубоких сверточных сетей позволило преодолеть ключевые ограничения классической коллаборативной фильтрации.

Гибридная мультимодальная сеть продемонстрировала превосходство (AUC 0.8120), подтвердив гипотезу о том, что учет внутренней стилистики и атмосферы видео значительно улучшает обнаруживаемость (*discovery*) сложного контента и повышает долгосрочную вовлеченность пользователей.

Библиографический список

1. Воронцов, К. В. Машинное обучение и интеллектуальный анализ данных : учебное пособие / К. В. Воронцов. – Москва : МЦНМО, 2018. – 256 с. – Текст : непосредственный.
2. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; пер. с англ. А. А. Слинкина. – Москва : ДМК Пресс, 2018. – 652 с. – Текст : непосредственный.
3. Лукашов, К. А. Нейронные сети и глубокое обучение : учебное пособие для вузов / К. А. Лукашов. – Москва : Издательство Юрайт, 2021. – 234 с. – Текст : непосредственный.
4. Дрозд К.В. Актуальные вопросы педагогики и образования. — М. : Юрайт, 2022. — 265 с.
5. Кругликов В.Н., Оленникова М.В. Интерактивные образовательные технологии. — М. : Юрайт, 2021. — 353 с.