Проблема длинного хвоста в музыкальных рекомендациях

Матеха Ярослав Кириллович

30.10.2026

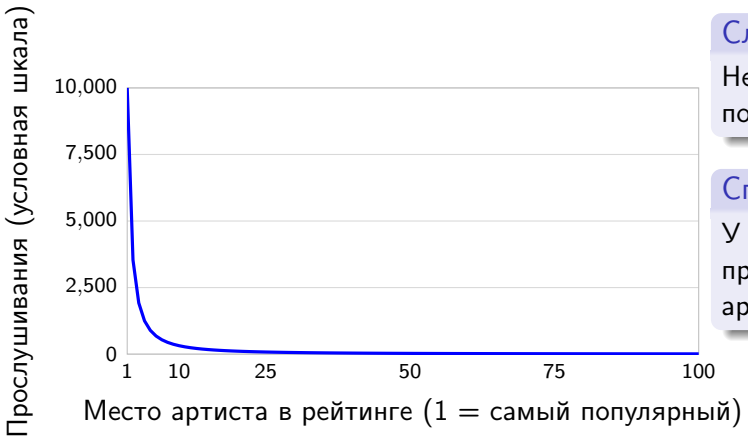
Как мы выбираем музыку?

- Знакомые артисты и плейлисты
- Советы друзей и соцсетей
- Рекомендации сервиса
- Случайное открытие



Задача: рекомендовать релевантное, но не застревать на знакомом

Что такое длинный хвост?



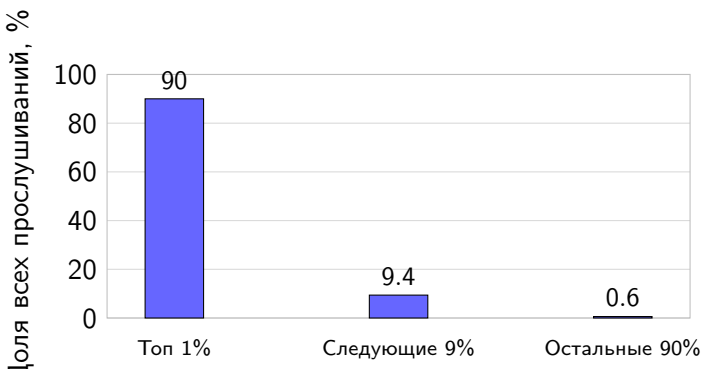
Слева - «голова»

Немного артистов собирают почти все прослушивания.

Справа - «хвост»

У многих артистов меньше прослушиваний, но таких артистов больше.

Неравномерность популярности



- Топ 1% артистов получает 90% всех прослушиваний.
- Следующие 9% - ещё 9.4%.
- Оставшиеся 90% - 0.6%.

Почему хиты получают преимущество?



Популярность может усиливаться самой системой рекомендаций.

Пример: The Beatles и малоизвестный артист

The Beatles

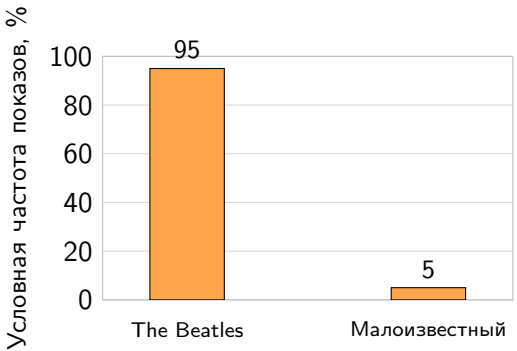
Большая история прослушиваний, много связей с другими артистами и альбомами - система «уверена» в их оценке.

Малоизвестный исполнитель

Меньше данных и связей, но может идеально подходить конкретному слушателю.

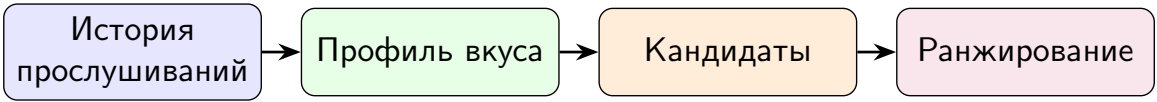
Вывод

Система чаще показывает то, о чём у неё больше данных, - но частота показов **не равна** качеству рекомендации для конкретного человека.



Числа иллюстративные: показывают, как часто система предлагает артиста, а не насколько он подходит слушателю.

Как работает рекомендательная система?



- Коллаборативная фильтрация: похожие слушатели
- Content-based: похожее звучание
- Гибрид: комбинация сигналов

Петля обратной связи



Риск

Система принимает собственное влияние на выбор за независимое предпочтение слушателя.

Exposure bias

Популярный трек

Показан 1000 раз; выбрали 100 раз.

$$CTR = \frac{100}{1000} \times 100\% = 10\%$$

Малоизвестный трек

Показан 20 раз; выбрали 4 раза.

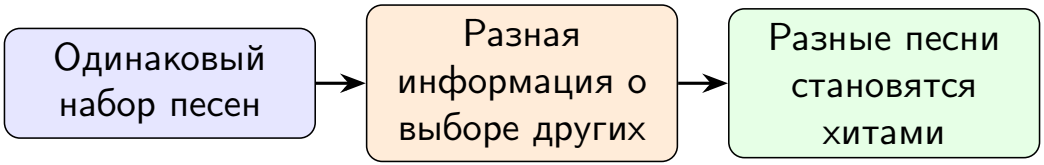
$$CTR = \frac{4}{20} \times 100\% = 20\%$$



CTR (Click-Through Rate, показатель кликабельности) - доля показов, которые закончились выбором трека слушателем.

Социальное влияние: реальный эксперимент

- В 2006 году Salganik, Dodds и Watts провели эксперимент с незнакомыми песнями в онлайн-музыкальной среде.
- Участники могли видеть, какие песни выбирали другие.
- Социальная информация меняла выборы и усиливала неравенство успеха песен.



Почему это важно?

Популярность зависит не только от качества песни, но и от того, что другие уже выбрали.

Salganik, Dodds & Watts (2006), *Science*. DOI: 10.1126/science.1121066.

Проблема или особенность индустрии?

- **Для пользователей:** Иллюзия выбора. Библиотеки огромны, но алгоритмы часто ведут к одним и тем же хитам, ограничивая кругозор.
- **Для артистов:** «Богатые становятся богаче». Попадание в топ-плейлист решает всё, а 100 000+ новых треков ежедневно остаются незамеченными.
- **Для платформ:** Двойственная ситуация.
  - С одной стороны, алгоритмы усиливают хиты, чтобы удержать внимание пользователей.
  - С другой — программы вроде **«Discovery Mode»** позволяют лейблам влиять на алгоритм за снижение роялти, что рискует превратиться в современную «пейолу» и уже вызывает судебные иски.

Вывод

Это реальная проблема, которая балансирует между **эффективностью** (удержание пользователя) и **справедливостью** (шанс для новых артистов).

Почему одной точности недостаточно?

Точность

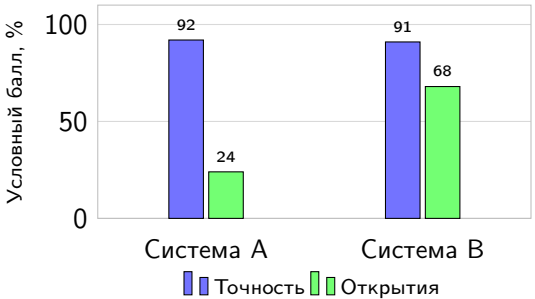
Насколько часто рекомендации действительно подходят человеку?

Чего она не показывает?

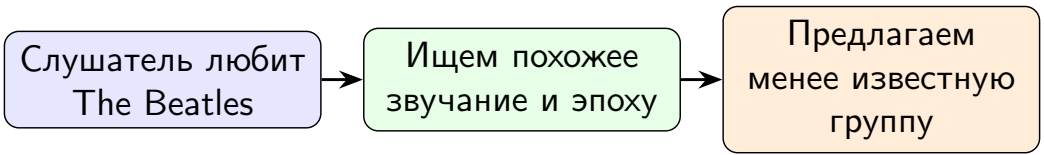
Открывает ли человек новых артистов и не повторяется ли один и тот же стиль.

Вывод

Две системы с почти одинаковой точностью могут давать совершенно разный опыт.



Новизна

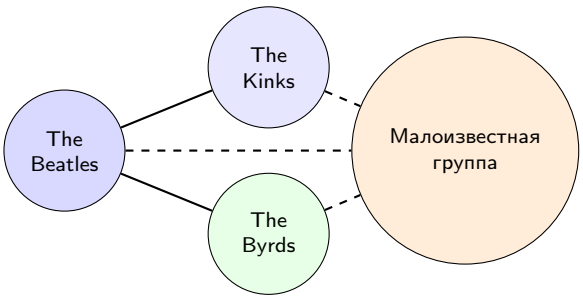


Новизна

насколько рекомендация незнакома конкретному слушателю. Новая песня должна всё ещё подходить его вкусу.

- Не только другие хиты.
- Не случайный трек из любого жанра.
- Поиск рядом с уже любимой музыкой.

Граф музыкального сходства



- Кружок - артист.
- Линия - сходство по музыке или аудитории.
- Пунктир - возможный путь к менее очевидной рекомендации.

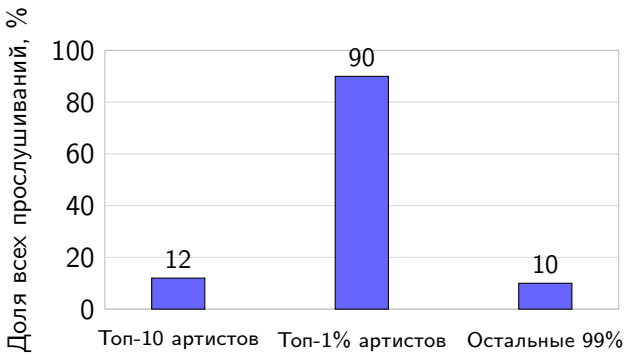
Сравнение метрик

Метрика	Вопрос	Ограничение
Accuracy / Recall@K	Есть ли релевантные треки?	Может поощрять повторы
Novelty	Предлагается ли малоизвестное?	Не гарантирует интерес
Serendipity	Есть ли удачные открытия?	Сложно измерить одним числом
Catalog coverage	Какую часть каталога используем?	Не описывает отдельный список
Intra-list diversity	Насколько список разнообразен?	Разнообразие может быть нерелевантным

Вывод

Нужно оценивать набор метрик, а не одну цифру.

Реальные данные Spotify (2024)

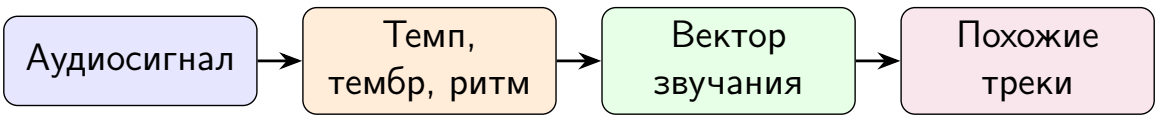


Данные: Spotify (2024), Alpha Data.

Spotify, 2024 год

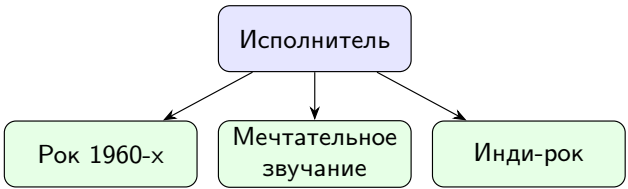
- Всего за год - около **4,8 трлн** прослушиваний.
- Топ-10 артистов собрали более **12%** всех стримов.
- Топ-1% артистов (около 16 тыс.) аккумулировали **90%** прослушиваний.
- На остальные **99%** артистов (более 1,5 млн) приходится лишь **10%**.

Content-based: анализ аудио



- Помогает рекомендовать треки без истории прослушиваний
- Ограничение: сходство звука не равно совпадению вкуса

Теги и социальные связи

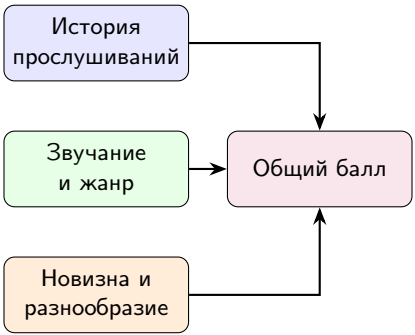


Зачем нужны эти признаки?

- Жанр и эпоха связывают похожую музыку.
- Настроение помогает находить похожую атмосферу.
- Теги дополняют аудио и историю прослушиваний.

Гибридная система

$$S(u,i) = \alpha R(u,i) + \beta N(u,i) + \gamma D(L,i)$$



- $S(u,i)$ - итоговый балл трека i для слушателя u.
- $R(u,i)$ - релевантность: насколько трек подходит вкусу; зависит от истории прослушиваний и похожести.
- $N(u,i)$ - новизна для этого слушателя; выше, если он раньше не слушал артиста или трек.
- $D(L,i)$ - вклад трека в разнообразие списка L; выше, если он не дублирует уже выбранные песни.
- α, β, γ - веса важности трёх целей. Большая α усиливает точность, β - новые открытия, γ - разнообразие.

Reranking

Reranking - система уже выбрала подходящие треки, а затем переставляет их, чтобы список не был однообразным.

До: повторы

A1 A2 A3 A4

Все четыре трека - одного артиста A.

После: разнообразие

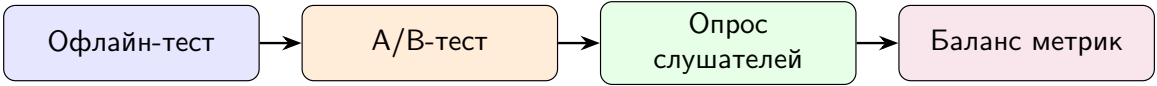
A1 B1 C1 A2

Треки артистов A, B и C чередуются.

Важно

Нужно сохранить соответствие вкусу, но уменьшить повторы и добавить музыкально близкие альтернативы.

Как проверить улучшение?



- **Офлайн-тест** - быстрая проверка на исторических данных.
- **A/B-тест** - проверка на реальных пользователях.
- **Опрос** - проверка субъективного впечатления.
- **Баланс метрик** - улучшение одного показателя не должно ломать другие.

Как проверить улучшение? (детали)

Офлайн-тест

прогон модели на исторических логах без участия пользователей. Быстро и дёшево, но не учитывает exposure bias.

Опрос

прямая оценка удовлетворённости и удачности открытий. Ловит то, чего не видно в логах: «понравилось ли», а не «кликнул ли».

A/B-тест

две группы пользователей видят разные версии системы; сравнивают поведение - прослушивания, пропуски, удержание.

Баланс метрик

улучшение точности не должно ухудшать разнообразие и новизну. Метрики проверяют вместе, а не по одной.

Компромиссы и ограничения

Если оптимизировать только точность

- Больше хитов
- Меньше риска
- Меньше открытий

Если оптимизировать только разнообразие

- Много разных треков
- Может падать релевантность
- Растёт риск случайности

Итог

Что помогает

- **Гибридные модели:** сочетают историю, звук и теги.
- **Reranking:** убирает однообразие, перемешивая список.
- **Новизна:** поиск незнакомого, но похожего.

Главные трудности

- **Exposure bias:** алгоритм не видит то, что не показывал.
- **Петля обратной связи:** популярность усиливает сама себя.
- **Баланс:** точность легко ломает разнообразие.

Литература


Celma, Ò. & Cano, P. (2008). *From Hits to Niches? or How Popular Artists Can Bias Music Recommendation and Discovery*. KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize.


Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311(5762), 854–856. https://doi.org/10.1126/science.1121066


Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems. *IEEE TKDE*, 17(6), 734–749.


Kunaver, M. & Požrl, T. (2017). Diversity in recommender systems – A survey. *Knowledge-Based Systems*, 123, 154–162.