

# Ранжирование

Доклад на НИС Машинное обучение и приложения

---

Даша Ничвидюк

НИУ ВШЭ ФКН БПМИ162

1. Модели ранжирования
2. Применения в реальной жизни
3. Оценивание ранжирования

# Модели ранжирования

---

- RankNet – первая идея pairwise-подхода.
- Пусть у нас есть кое-какие прямые данные для обучения (т.е. про некоторые подмножества документов эксперт сказал, какие более релевантны, какие менее).
- Подход к решению: давайте обучать функцию, которая по данному вектору атрибутов  $x \in \mathbb{R}^n$  выдаёт  $f(x)$  и ранжирует документы по значению  $f(x)$ .

RankNet изначально был разработан с использованием нейронных сетей, но базовая модель может отличаться и не ограничивается только нейронными сетями. Функция стоимости для RankNet направлена на минимизацию количества инверсий в ранжировании. Здесь инверсия означает неправильный порядок среди пары результатов, то есть когда мы оцениваем результат с более низким рангом над результатом с более высоким рангом в ранжированном списке. RankNet оптимизирует функцию стоимости с помощью SGD.

# LambdaRank

- Проблема с RankNet в том, что оптимизируется число попарных ошибок, а это не всегда то, что нужно.
- Градиенты RankNet – это не то же самое, что градиенты NDCG:



- Как оптимизировать, скажем, NDCG?

Как правило, RankSVM включает в себя три этапа в период обучения:

- Отобразить сходство между запросами и страницами, по которым щелкнули, в определенное пространство функций.
- Расчитать расстояния между любыми двумя векторами, полученными на шаге 1.
- Сф задачу оптимизации, которая похожа на стандартную классификацию SVM, и решает эту проблему с помощью обычного решателя SVM.

## Применения в реальной жизни

---





**Ранжирование** — применительно к работе поисковой системы — **это** процесс выстраивания найденных по запросу пользователя страниц в порядке наибольшего соответствия искомому запросу

[Ответы@Mail.Ru: Что такое ранжирование?](#)

[otvet.mail.ru > 42557472](#)

Нашлось 99 млн результатов

684 показа в месяц

**W Ранжирование — Википедия**

[ru.wikipedia.org](https://ru.wikipedia.org) > Ранжирование ▾

**Ранжирование** — сортировка сайтов в поисковой выдаче, применяемая в поисковых системах. Существует множество факторов для **ранжирования**... [Читать ещё >](#)

**FB Ранжирование - что это значит?**

[fb.ru > article/147437/ranjirovanie---chto-eto...](https://fb.ru/article/147437/ranjirovanie---chto-eto...) ▾

Тогда ответим **простыми словами** на вопрос о том, **ранжирование — что это**. Под данным термином подразумевается соответствие содержания веб-ресурса запросу конкретного пользователя и размещение результатов поиска на более высоких позициях. Именно поэтому веб-мастеру так важно проводить... [Читать ещё >](#)

NETFLIX

Browse ▾

DVD

Search



Joshua ▾

### Top Picks for Joshua



### Trending Now



### Because you watched Narcos



### New Releases



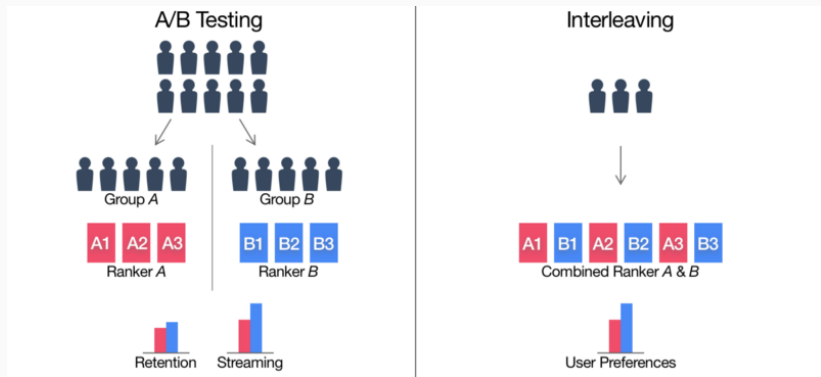
# Оценивание ранжирования

---

Онлайн-метрики доступны в работающей системе, можно использовать в эксперименте, например, среднее время просмотра сайта пользователем.

- Онлайн-модели отличаются от оффлайн-моделей тем, что их главная цель – как можно быстрее «поймать» изменения популярности тех или иных продуктов.
- Данных тут недостаточно, чтобы такие изменения можно было поймать методами коллаборативной фильтрации.
- Поэтому онлайн-методы обычно меньше персонализированы, индивидуальных данных не наберётся просто.

# Interleaving



# А/В-тестирование против чередования

В традиционном А/В-тестировании популяция делится на две группы, одна из которых подвергается алгоритму ранжирования А, а другая - В. Измеряются и сравниваются основные показатели оценки, такие как сохранение и потоковая передача, между двумя группами. В отличие от этого, чередование подвергает одну группу участников смешанному ранжированию ранжеров А и В. Предпочтение пользователя для алгоритма ранжирования определяется путем сравнения доли часов просмотра видео, рекомендованных ранжерами А или В.

- <https://medium.com/@nikhilbd/intuitive-explanation-of>
- <https://medium.com/netflix-techblog/interleaving-in>
- [https://en.wikipedia.org/wiki/Learning\\_to\\_rank](https://en.wikipedia.org/wiki/Learning_to_rank)
- <https://github.com/esokolov/ml-course-hse/blob/master>