

# Введение в рекомендательные системы. часть 2

Гиздатуллин Данил

# Холодный старт

- У нового пользователя нет истории просмотра
- У нового объекта нет истории взаимодействия с пользователями

# Холодный старт. Почему это важно?

- Новинки часто интересны пользователям
- Объекты могут иметь короткий срок жизни на сервисе (Новости)
- Нужно сразу показывать всю привлекательность сервиса для нового пользователя

# Проблема новых объектов

Как решить?

- Жанры
- Тэги
- Авторы
- Режиссер, актеры
- Исполнитель
- Страна
- Год производства

и т. д.

# Проблема новых объектов

Обработка сырых данных

- Видео
- Тексты
- Музыка

# Проблема новых пользователей

Как решить?

- Использовать информацию о пользователе (пол, возраст, страна и т. д.)
- Онбординг (предложение выбрать интересный контент из списка при первом заходе)

## Content-based рекомендации

Используем схожесть контента по его фичам, а не по совместному смотрению

$$\hat{r}_{ui} = \bar{r}_i + \frac{\sum_{j \in I_u} \text{sim}(i, j)(r_{uj} - \bar{r}_j)}{\sum_{j \in I_u} \text{sim}(i, j)}, \text{ где } I_u - \text{топ } N$$

айтемов, наиболее похожих на айтем  $i$

# Factorization Machines

Feature vector $\mathbf{x}$															Target $y$							
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
	User				Movie					Other Movies rated						Last Movie rated						



## Factorization Machines

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times k}$$

# Factorization Machines

Преимущества:

- Гибридный подход. Используем сразу коллаборативную часть и фичи объектов и пользователей
- Не надо думать о проблеме холодного старта для айтемов (не совсем верно)

# Onboarding для новых пользователей

- Самый популярный контент внутри каждого жанра
- Просто выбрать любимые жанры
- Разделить контент на непохожие группы и показать пользователю самый популярный из каждой группы

# Onboarding для новых пользователей

Идея:

Разделить контент на непохожие группы и показать пользователю самый популярный из каждой группы

# Что почитать?

- Recommender Systems Handbook, F. Ricci, 2011
- Recommender Systems: The Textbook, Charu C. Aggarwal, 2016
- [Collaborative Filtering for Implicit Feedback Datasets,](#)  
[Yifan Hu](#)
- [Factorization Machines,](#) Steffen Rendle
- [Netflix blog](#)

# Сообщество, конференции, курсы, данные

- [RecSys конференция](#)
- [Recommender-systems specialization Coursera.org](#)
- [Open Data Science](#) канал #recommender\_systems
- [Датасет MovieLens](#)
- [Million Songs Challenge](#)
- [Job Recommendation Challenge](#)

**Спасибо за внимание!**