
Курс: Рекомендательные системы

Занятие 2: Базовые подходы

Подготовил: Лашинин Олег

Дата: 21.01.26

Обо мне

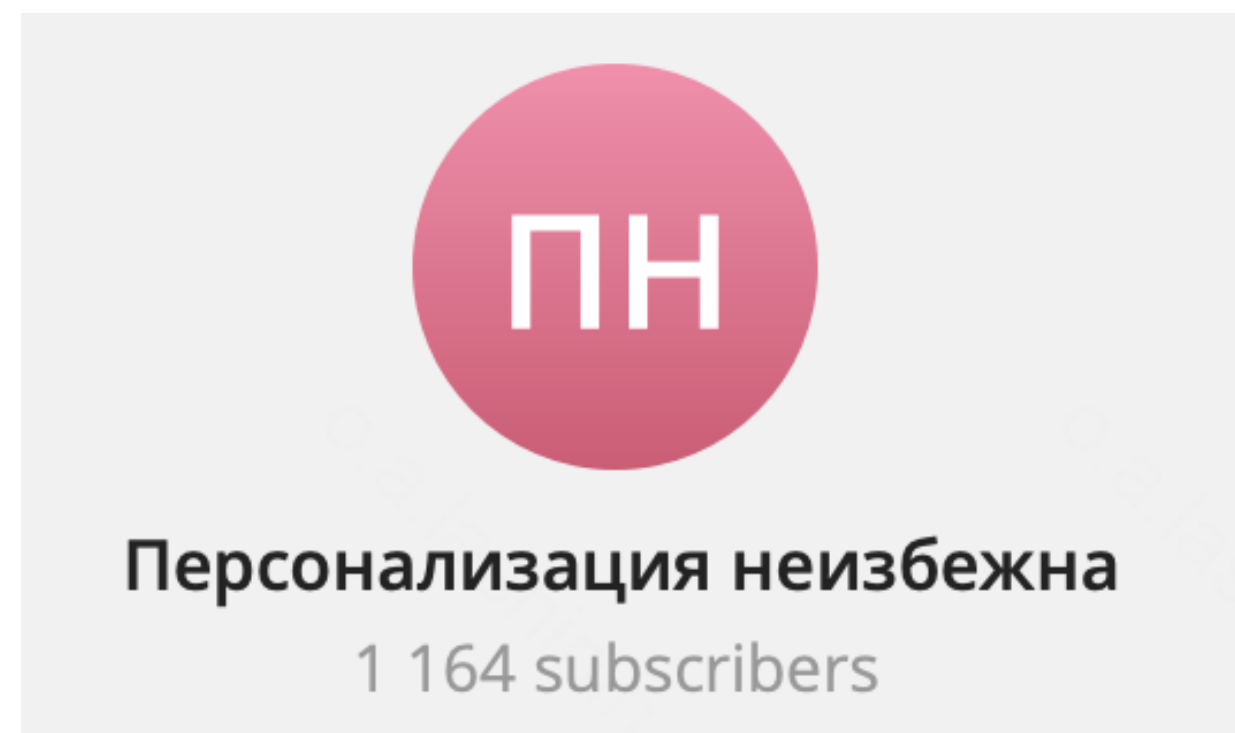
Образование: МФТИ + учусь в аспирантуре

Опыт работы: 6 лет в Т-Банке

Сфера деятельности: R&D + проекты для прода

Публикации: 12 публикаций по RecSys на воркшопах + конференциях

Веду ТГ канал:

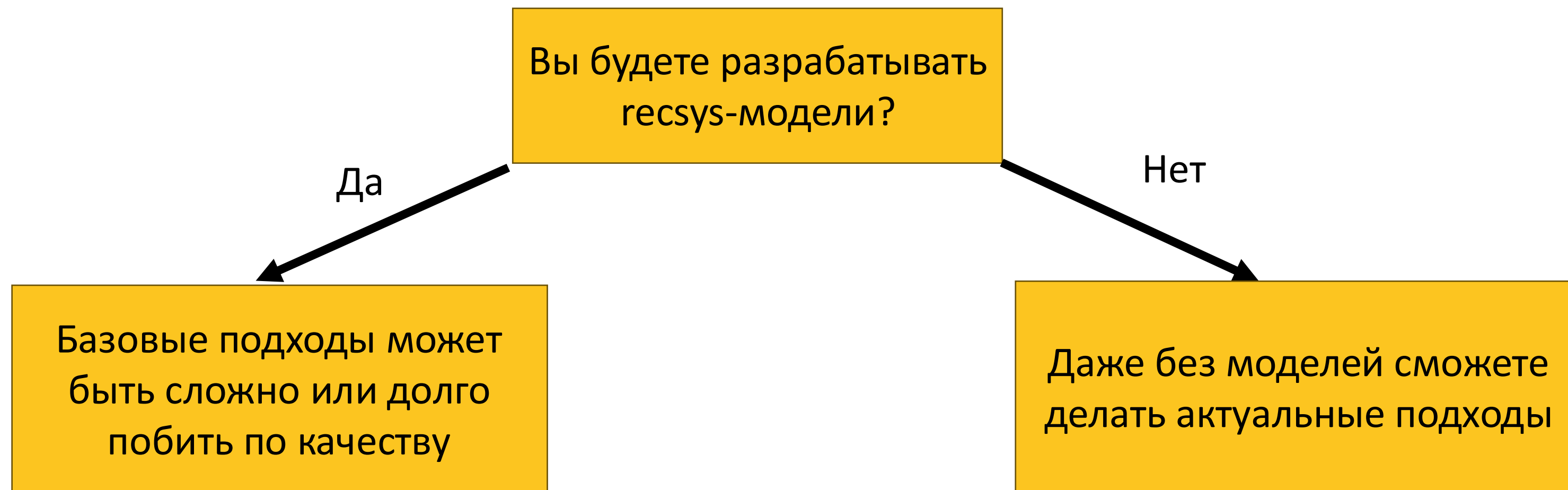


План занятия

- Метрики в RecSys
- Подзадачи рекомендаций
- Статистические персональные и неперсональные подходы
- Разбиение данных для экспериментов
- Что такое проблема холодного старта
- User- и item- расстояния
- Линейные модели (EASE, SLIM)
- Что такое popularity bias

Почему сегодня эта тема?

- Базовые подходы позволяют собрать быстрое и эффективное MVP
- Узнаем общепринятые подходы



Метрики в RecSys

Метрики. Разделение

Два варианта:

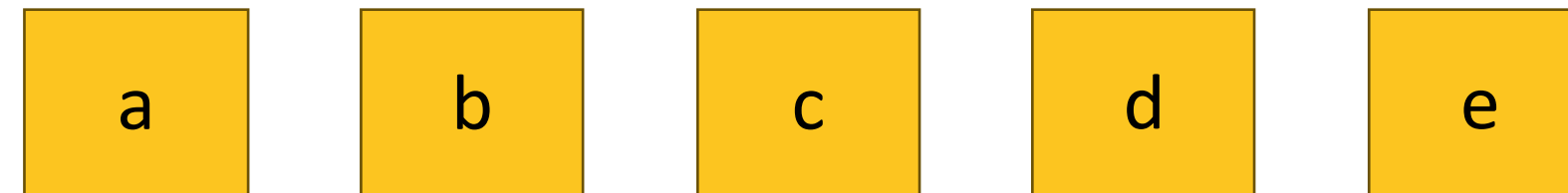
- 1) **Оффлайн** – *не применяем* модель на пользователях. Считаются на исторических данных, угадываем исторические данные, модель не используется/не влияет
- 2) **Онлайн** – *применяем* модель на пользователях, считаем результаты

По сути метрики:

- 1) Accuracy-oriented – сверяем рекомендации с «интересами» пользователя (recall, precision, ndcg...)
- 2) Beyond-accuracy – нам важно что-то кроме точности (разнообразие, новизна...)
- 3) Бизнес метрики – как рекомендации влияют на продукт? (CTR, GMV, Retention...)

Метрики. Как вообще это считается?

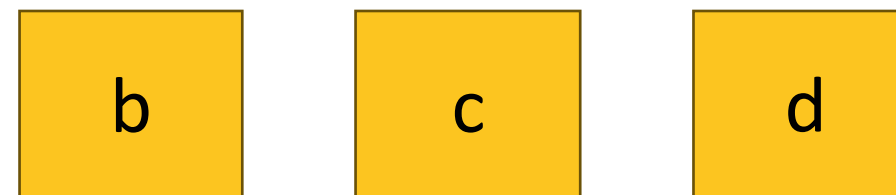
Ваши рекомендации пользователю
на основе обучающих данных:



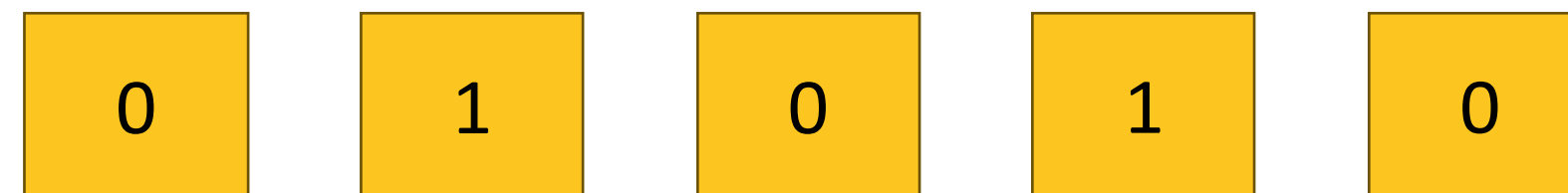
То, что пользователь делал в тесте
(ground-truth разметка):



То, что пользователь видел в тесте
(фильтр для ранжирования):



Релевантность ваших рекомендаций:



Вы либо учитываете то, что пользователь видел, либо нет.

Если не учитываете – ваша задача в том числе угадать, что видел юзер

Соревнования по recsys: надо в том числе угадать текущий recsys на проде

Метрики. Основа accuracy метрик

Пусть есть один запрос (пользователь)

Модель выдала упорядоченный список из k объектов:

$$r_1, r_2, \dots, r_k$$

Истинная релевантность объекта на позиции i :

$$y_i \in \{0, 1\} \quad (1 \text{ — релевантен})$$

(для NDCG допустимо $y_i \geq 0$)

Общее число релевантных объектов:

$$R = \sum_j y_j$$

$$\text{RR@k} = \frac{1}{\min\{i \leq k \mid y_i = 1\}}$$

$$\text{MRR@k} = \mathbb{E}_u[\text{RR@k}(u)]$$

$$\text{Precision@k} = \frac{1}{k} \sum_{i=1}^k y_i$$

$$\text{HitRate@k} = 1 \left(\sum_{i=1}^k y_i > 0 \right)$$

$$\text{Recall@k} = \frac{1}{R} \sum_{i=1}^k y_i$$

$$\text{AP@k} = \frac{1}{k} \sum_{i=1}^k P@i \cdot \text{rel}(i)$$

$$\text{MAP@k} = \frac{1}{|U|} \sum_{u \in U} \text{AP@k}(u)$$

$$\text{DCG@k} = \sum_{i=1}^k \frac{2^{y_i} - 1}{\log_2(i + 1)}$$

$$\text{IDCG@k} = \sum_{i=1}^{\min(k, R)} \frac{1}{\log_2(i + 1)}$$

$$\text{NDCG@k} = \frac{\text{DCG@k}}{\text{IDCG@k}}$$

Метрики. Идея beyond-accuracy

Coverage – доля покрытия каталога вашими рекомендациям (Например, вы рекомендуете 20% от всех айтемов, которые есть в сервисе)

Popularity – какая в среднем популярность порекомендованного контента? 1 – Novelty

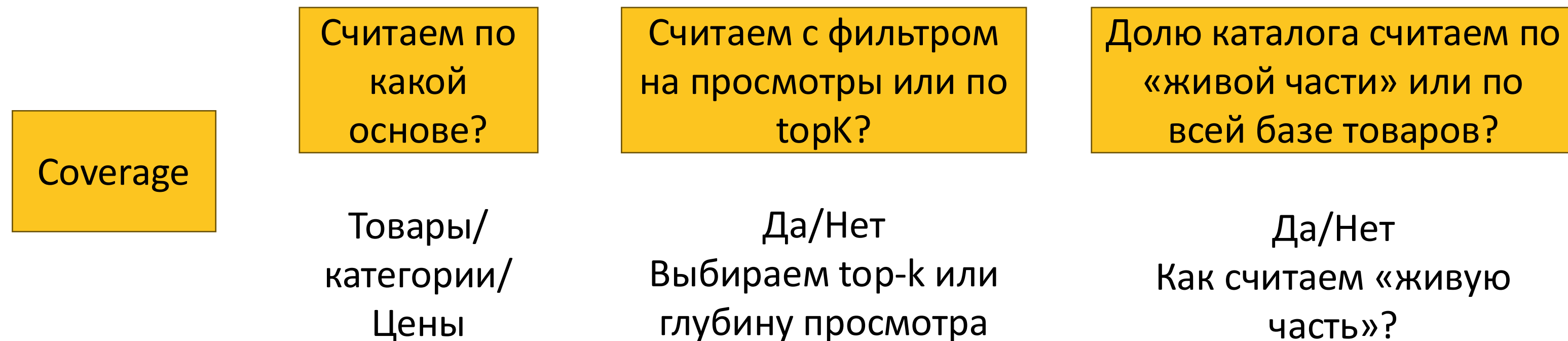
Freshness – насколько свежий в среднем порекомендованный контент (как давно он на платформе?)

Diversity –

1. Внутри одного списка рекомендаций – насколько отличаются айтемы у одного юзера?
2. Между юзерами – насколько у двух случайных юзеров отличны рекомендации?
3. У юзера с течением времени – насколько динамично обновляются рекомендации?

Serendipity – степень удивления юзера релевантными рекомендациями

Метрики. Идея beyond-accuracy. Подробнее



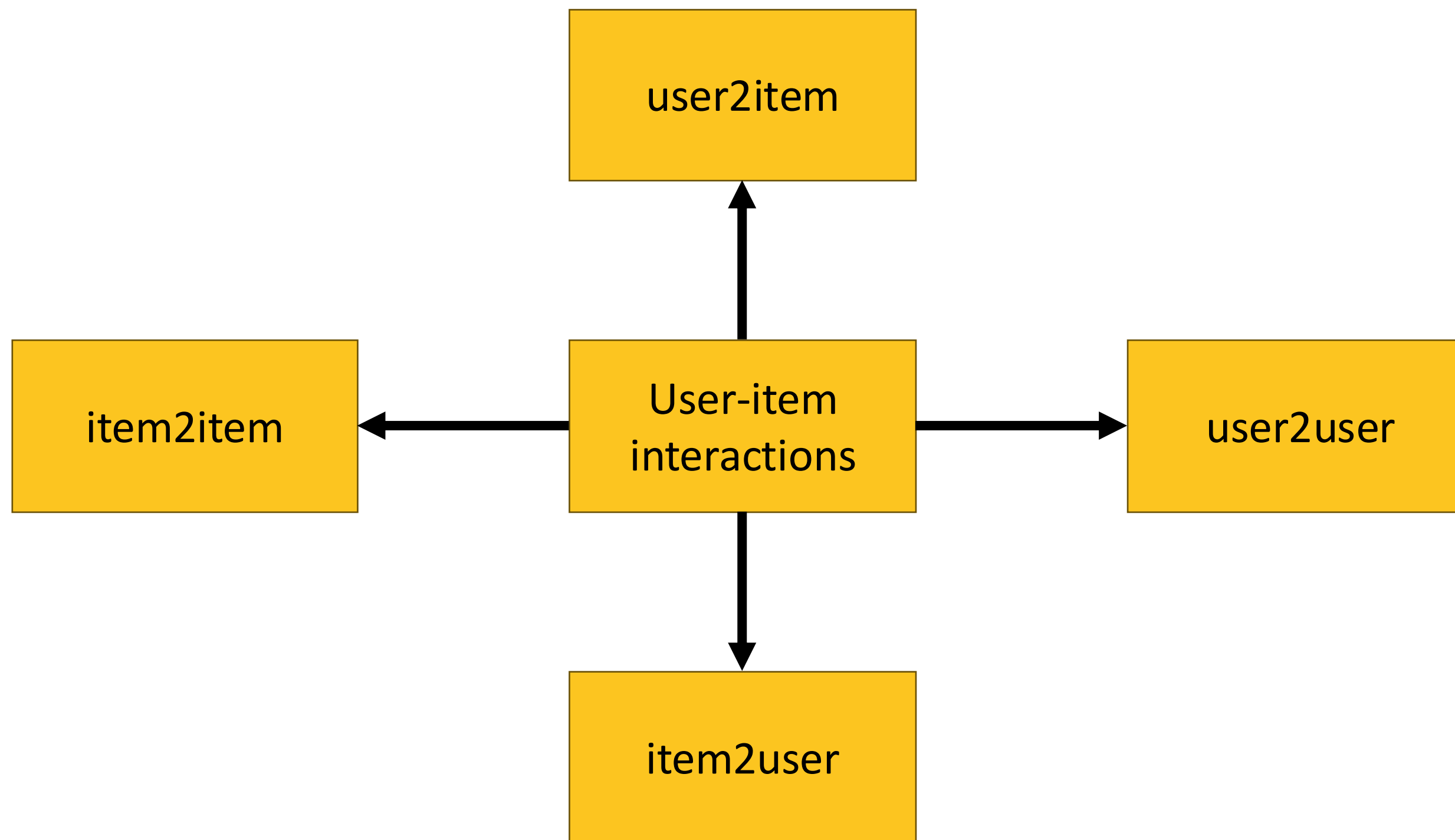
Главное правило – формул великое множество, но вы можете выбрать или придумать то, что подходит под вашу задачу.

Метрики. Сводная таблица

Группа метрик	Считаем в оффлайн?	Считаем в онлайн?	Тип	Нужна ground-truth разметка?
NDCG@K, Recall@K/Precision@K, MAP@K, MRR@K, HitRate@k	Можно	Можно, но осторожно	Accuracy	Да, без нее не посчитать
Coverage, Diversity, Novelty, Serendipity	Можно	Можно	Beyond-accuracy	В основном нет
CTR, GMV, Timespent, Retention, CTA	Почти невозможно	В основном да	Бизнес метрики	Да

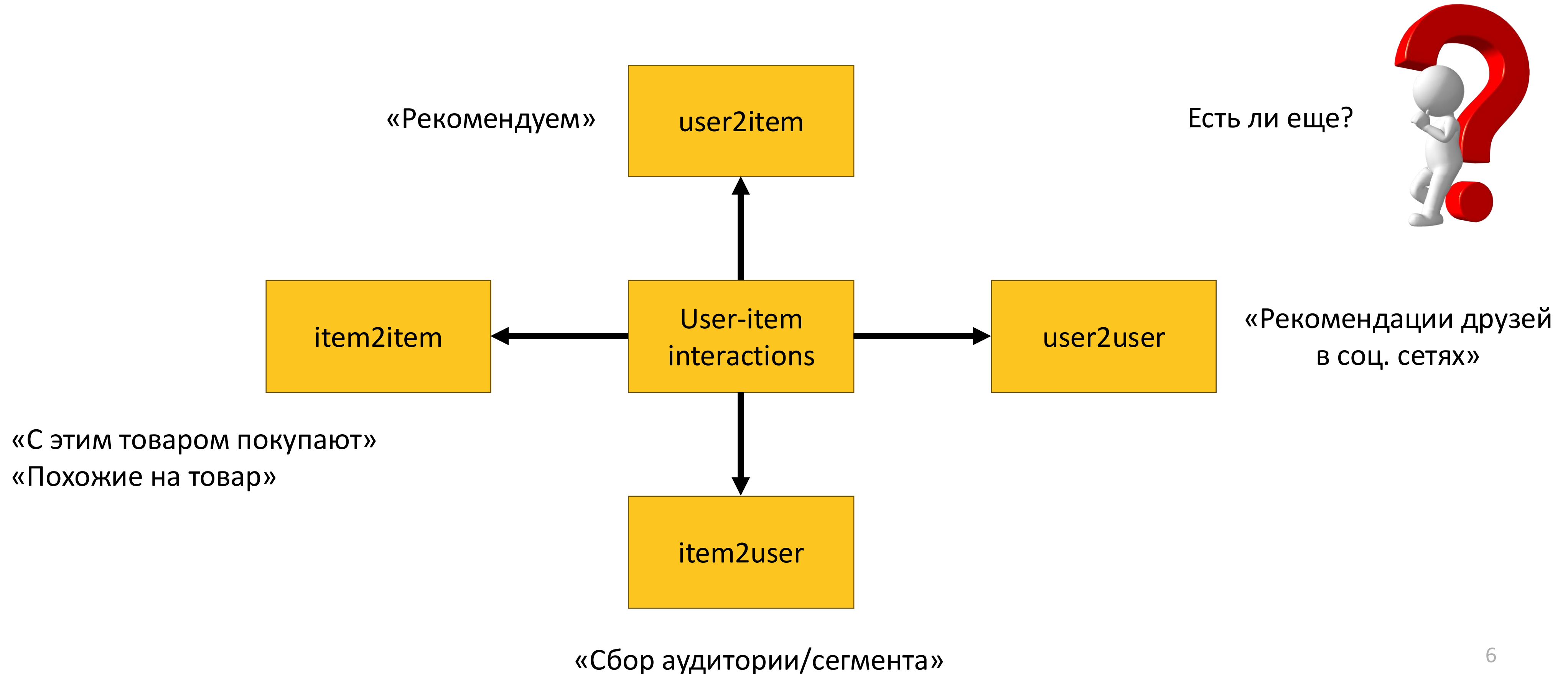
Подзадачи рекомендаций

Какие подзадачи вообще есть?

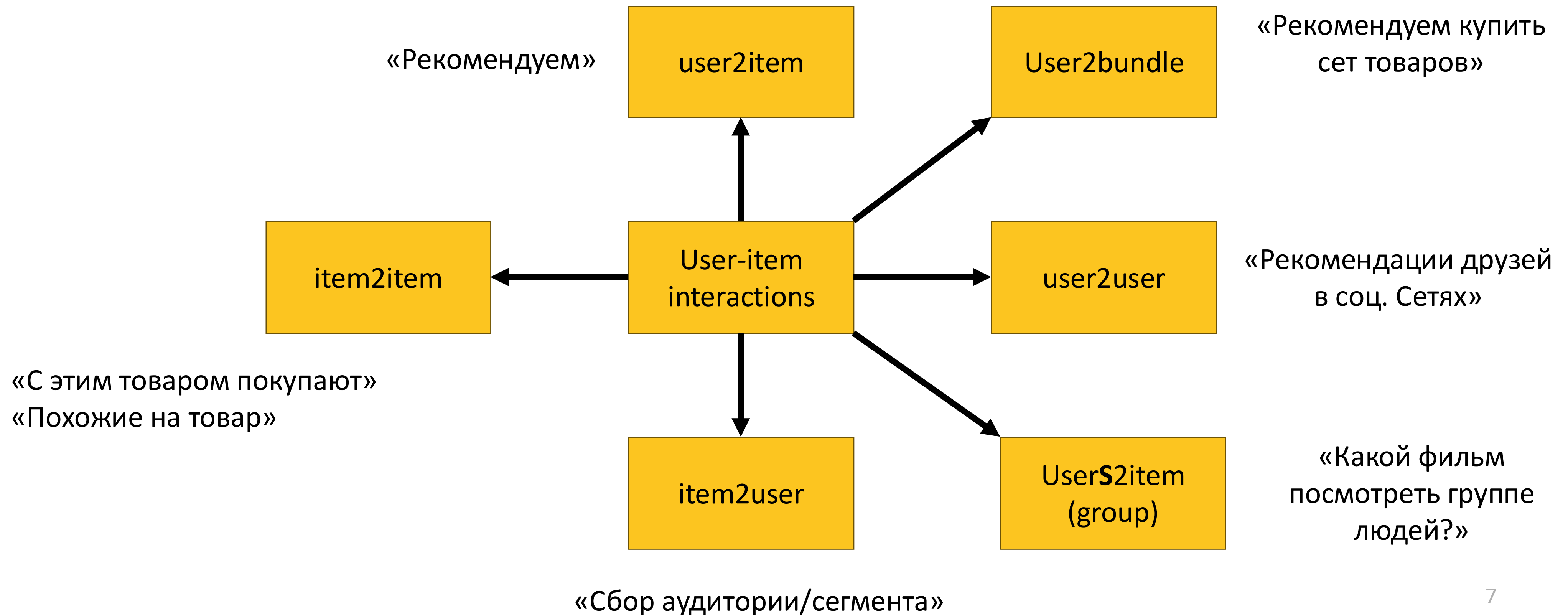


Какие практические кейсы?

Какие подзадачи вообще есть?



Какие подзадачи вообще есть?



Статистические персональные и неперсональные подходы

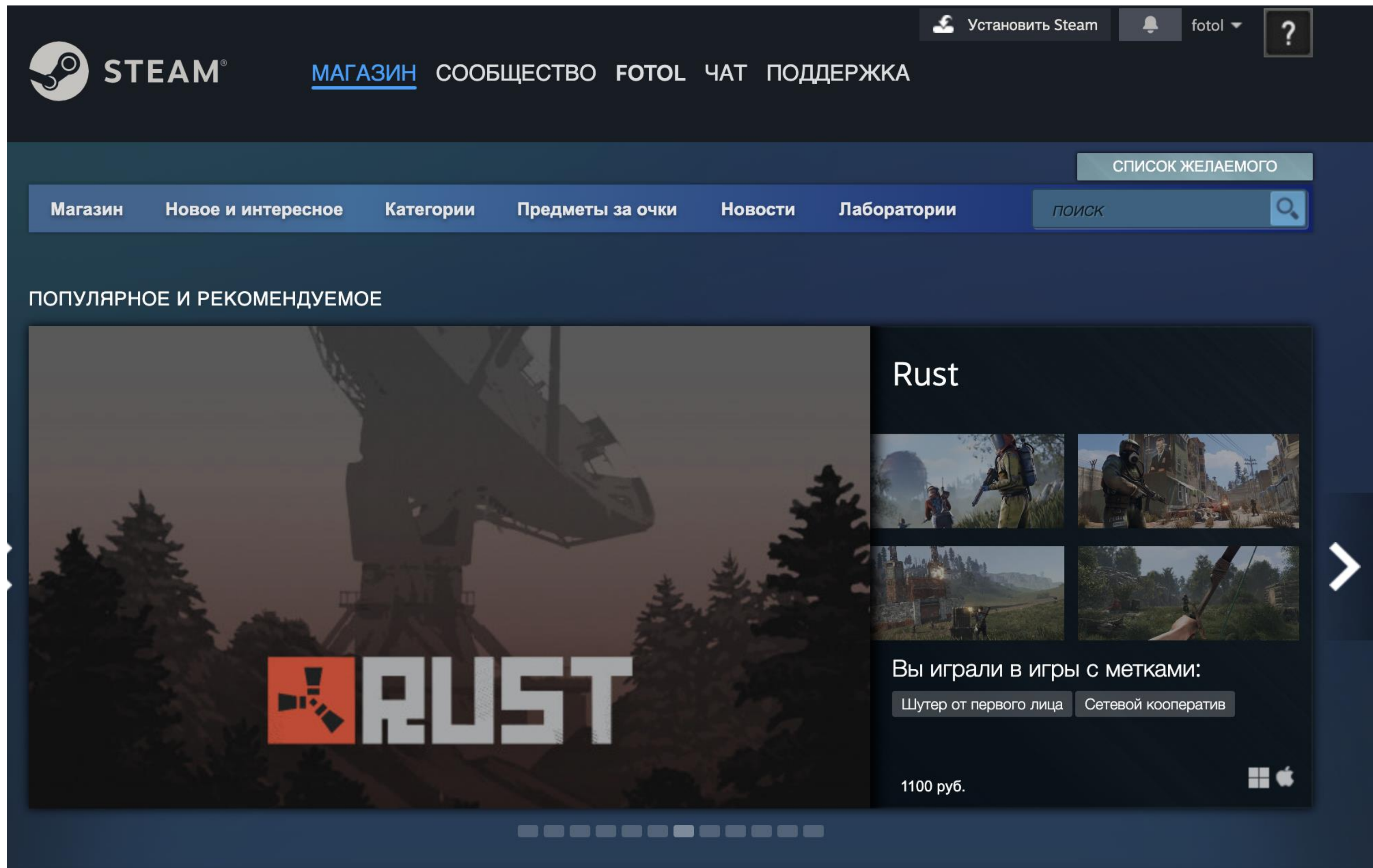
Три важных подхода к user2item

Рекомендуем (персонально)

Популярное

Новое

Три важных подхода к user2item



Три важных подхода к user2item

Мои клипы
Для вас
Подписки
Опубликовать

ЧЕСТНО, СТОК!
28 апр в 23:45
Подписаться

Гении офроуда

63 222
431
21 188
18M

Приложение VK Клипы
Создавайте и смотрите короткие видео в удобном вертикальном формате

Загрузите в App Store
СКАЧАТЬ ИЗ Google Play
Доступно в RuStore

Тренды

#впечатлето2023
1,8M просмотров

#ДачныйСтиль
1,6M просмотров

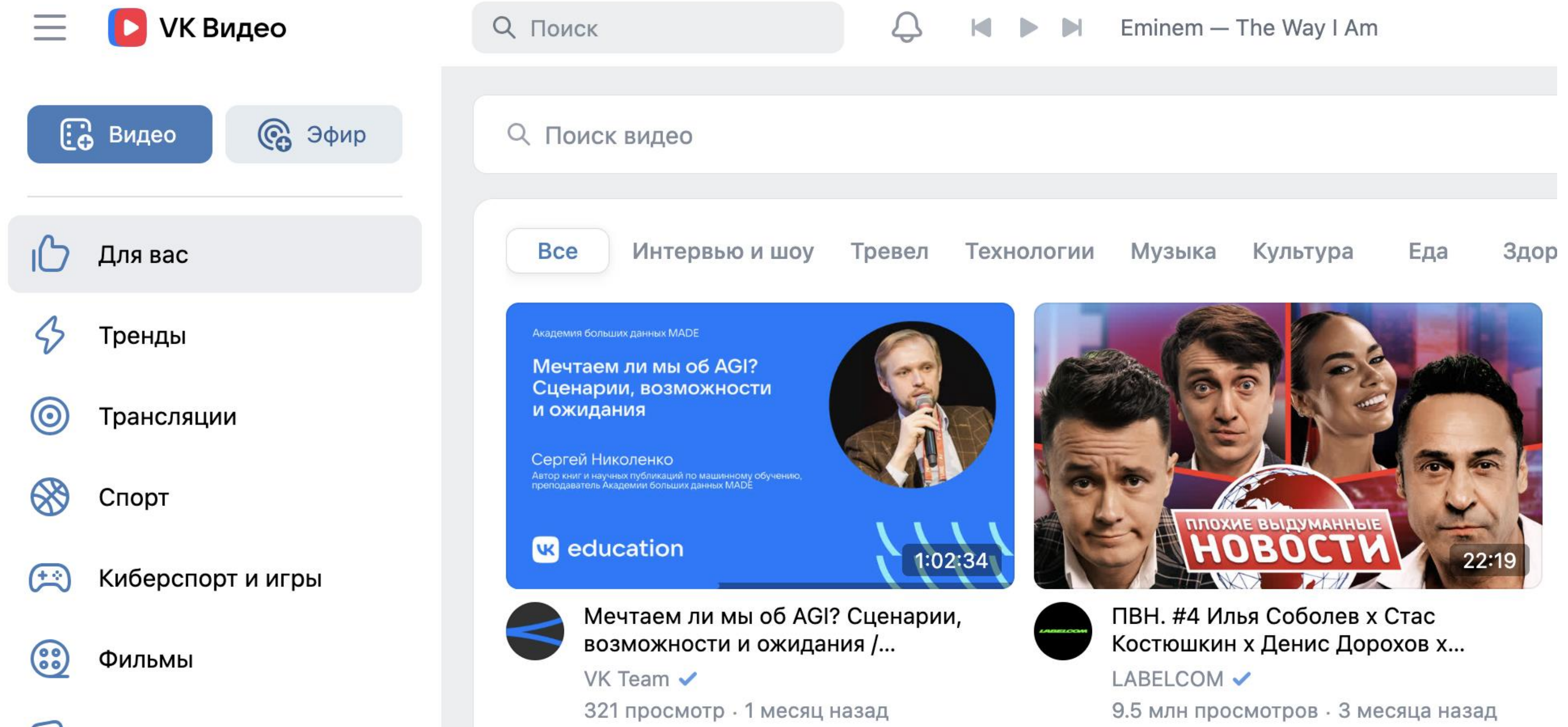
#ДизайнБуквально
1,2M просмотров

#ФотоИзШкафа
3,1M просмотров

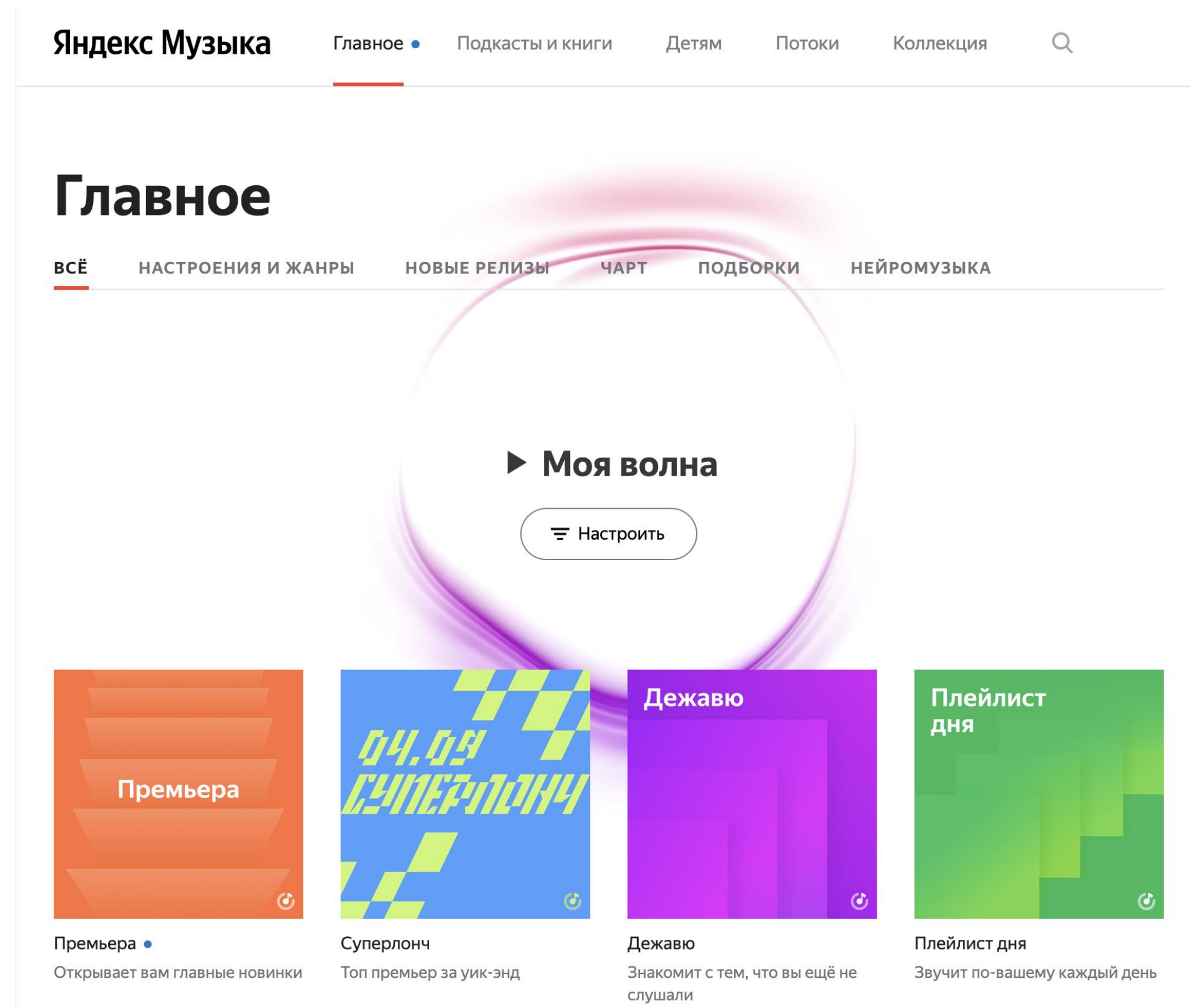
#ЭтоМойГород
2,2M просмотров

#ОтпускаюКамеру
2,4M просмотров

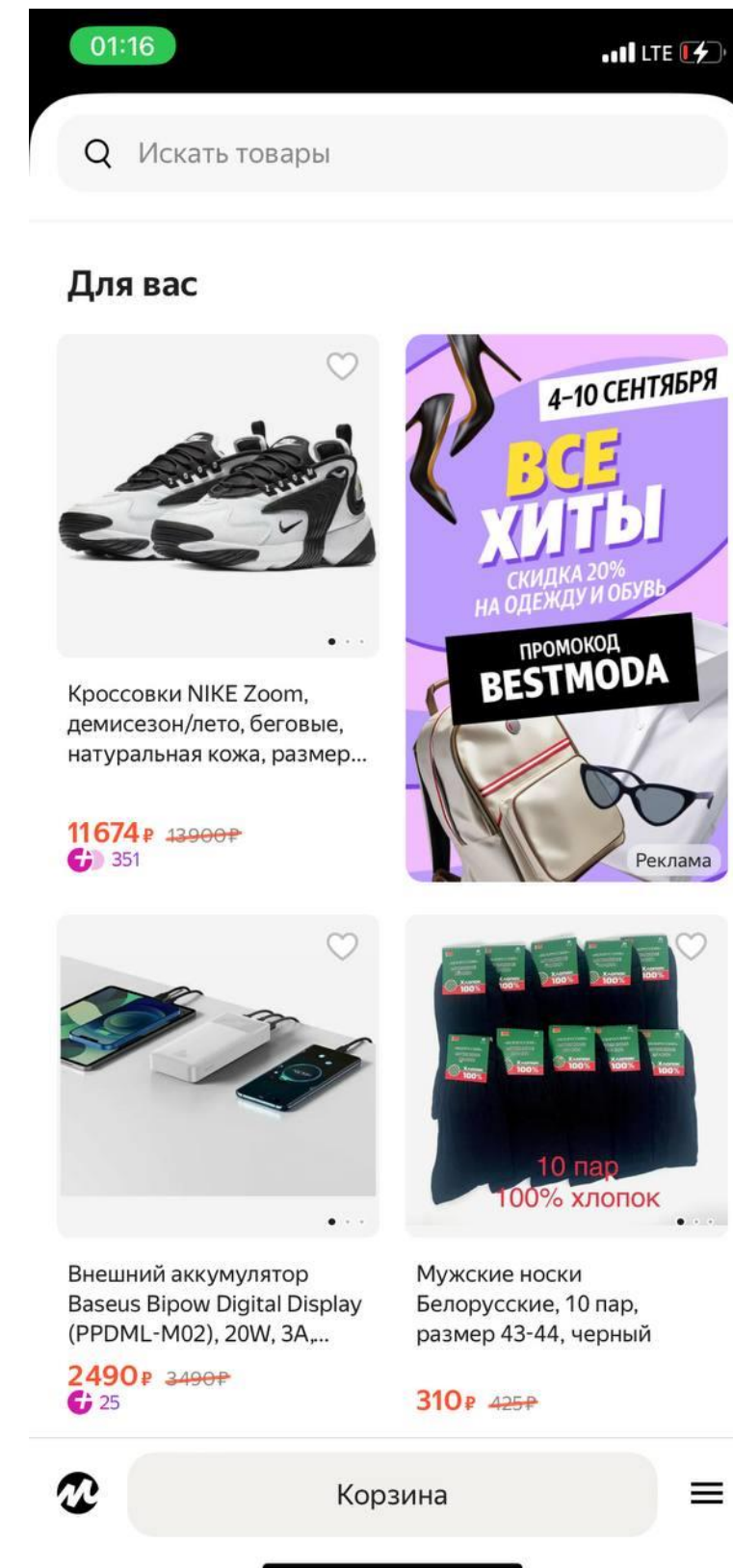
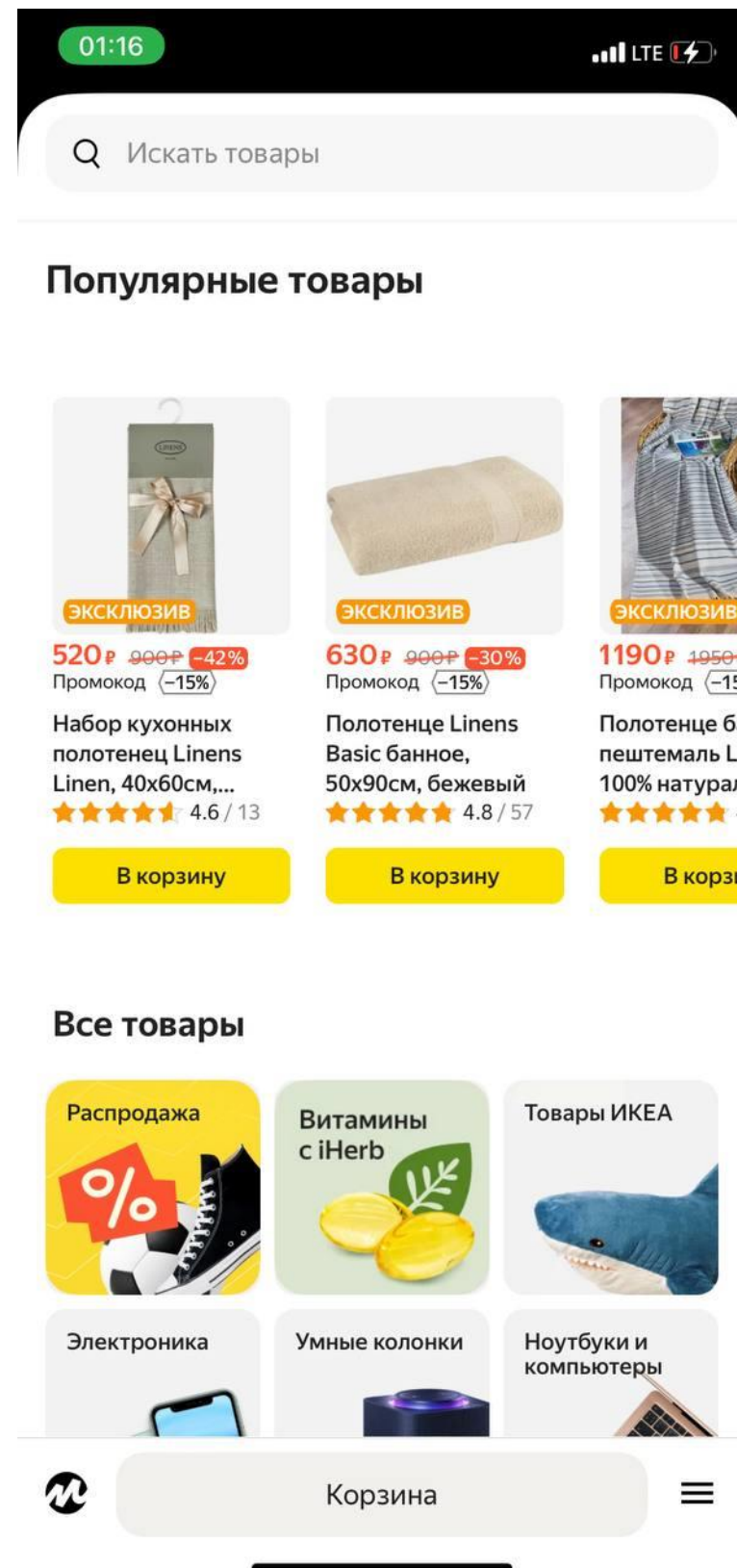
Три важных подхода к user2item



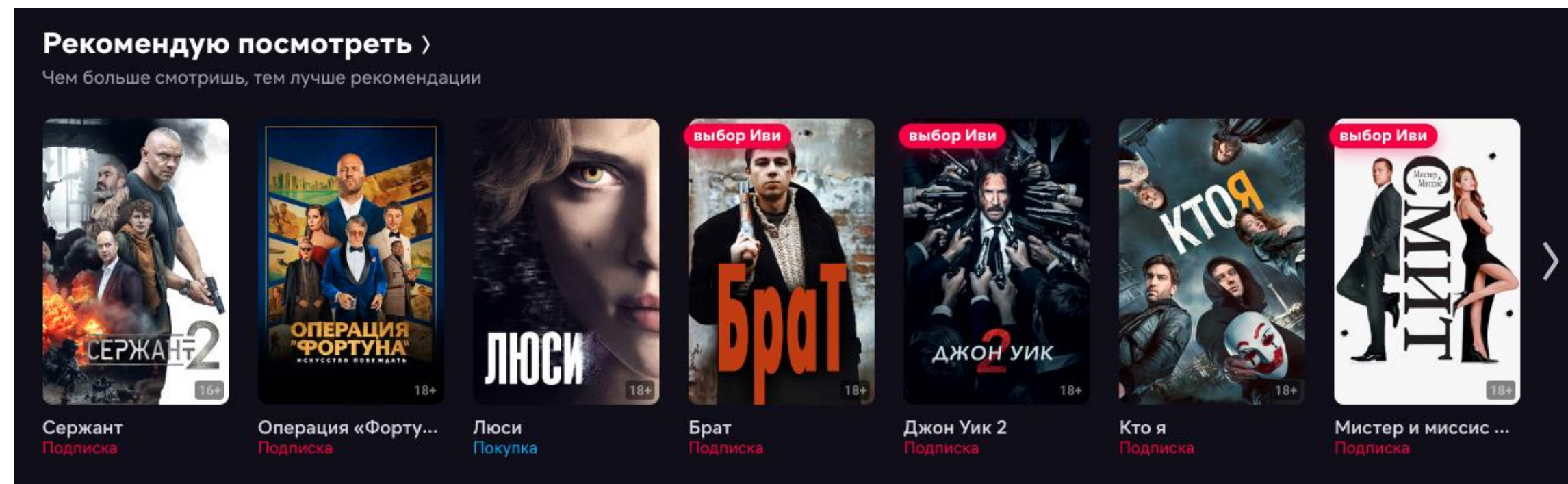
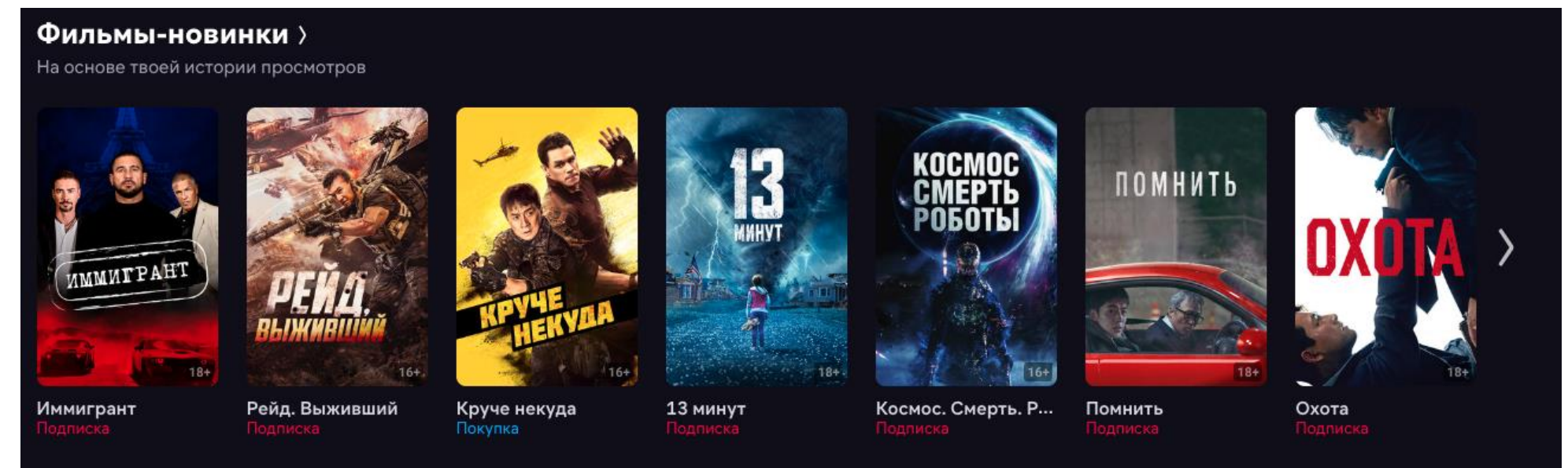
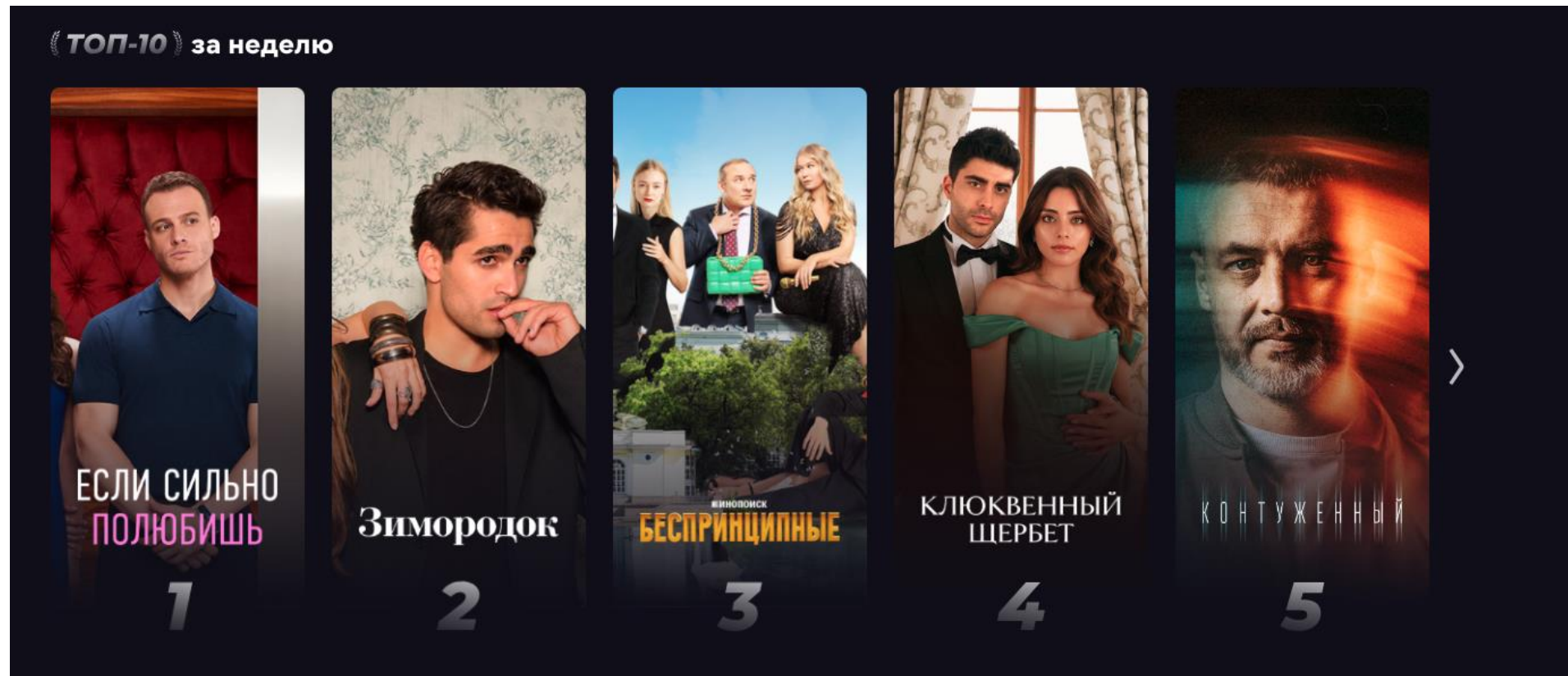
Три важных подхода к user2item



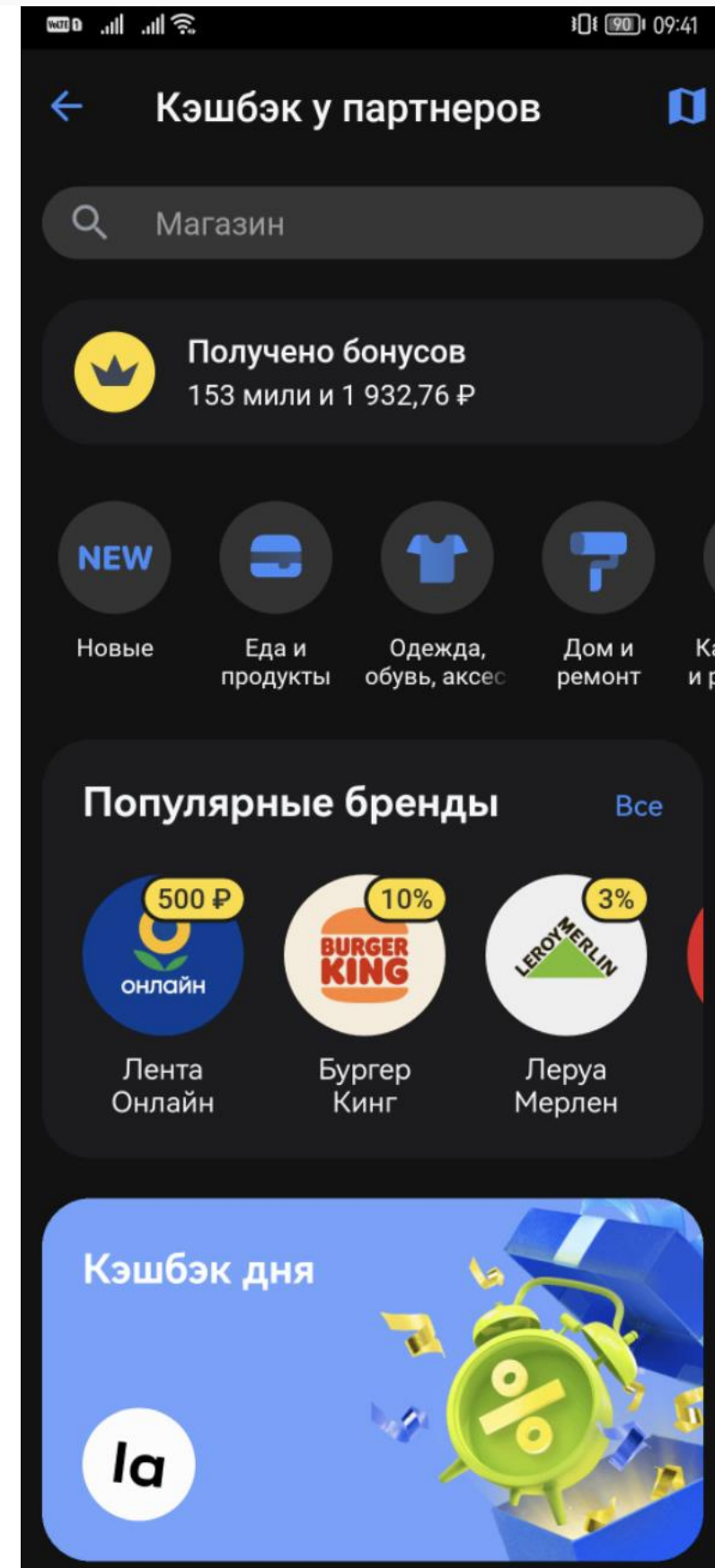
Три важных подхода к user2item



Три важных подхода к user2item



Три важных подхода к user2item



Новое + Популярное = Тренды

Популярное – часто есть в интерфейсах

Рекомендуем

Важно: людям нужны не только персональные рекомендации, но также подборки «новых» и «популярных»

Популярность

User_id	Item_id	Timestamp	label
1	2	124	0
2	3	2512	0
1	1	12	1
3	1	1	1

Что такое популярность айтема?



Популярность



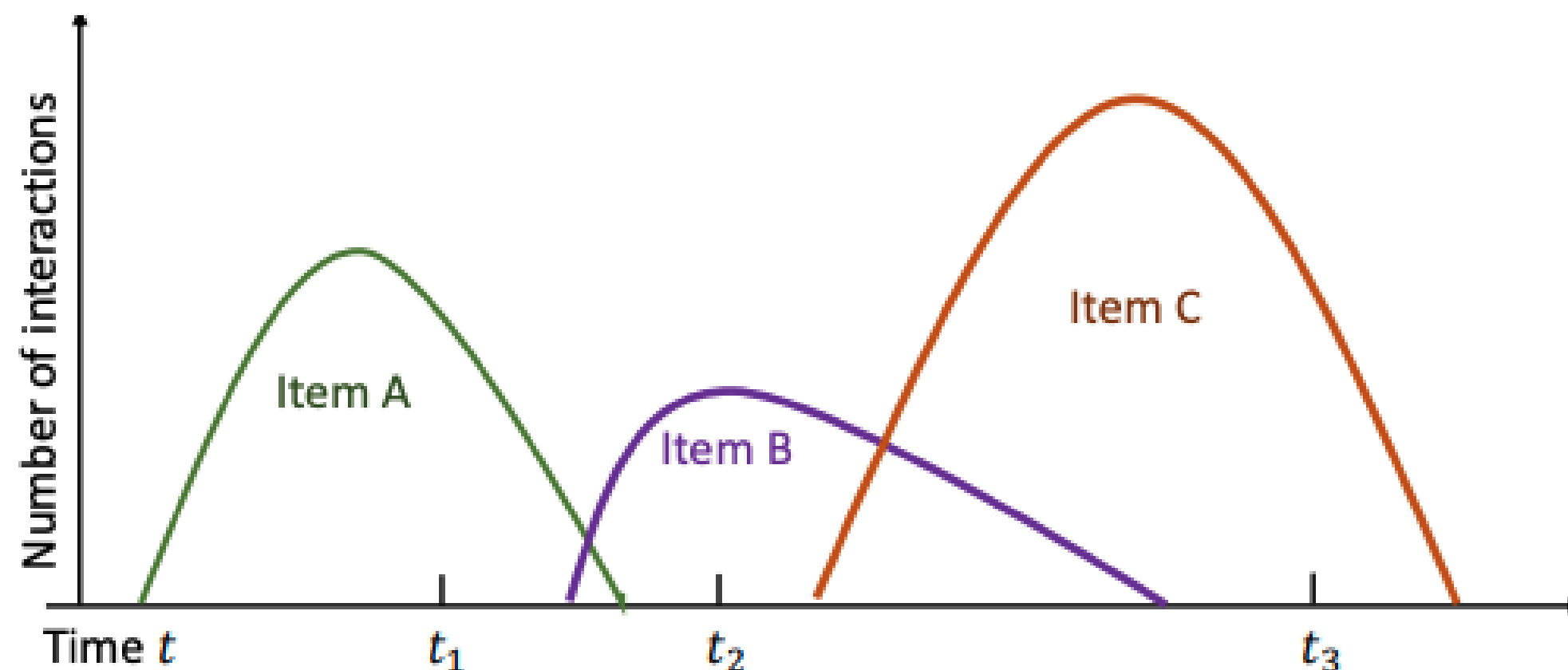
User_id	Item_id	Timestamp	label
1	2	124	0
2	3	2512	0
1	1	12	1
3	1	1	1

Что такое популярность айтема?

$$P(i) = \frac{\text{\# интеракций с айтемом } i}{\text{размер датасета}}$$

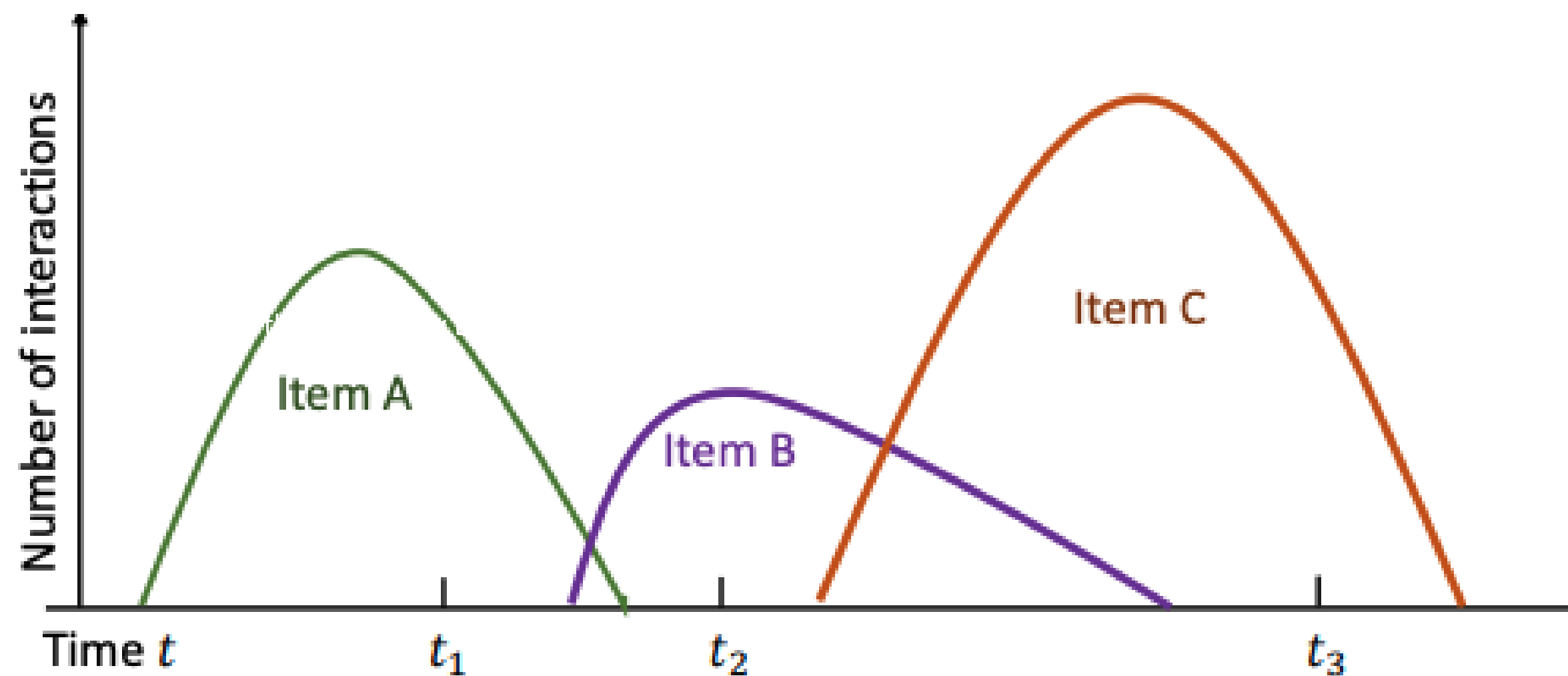
Популярность по времени

$$P(i, t) = \frac{\# \text{ взаимодействий с айтемом } i}{\text{размер датасета}} \mid t > \text{now()} - t_i$$

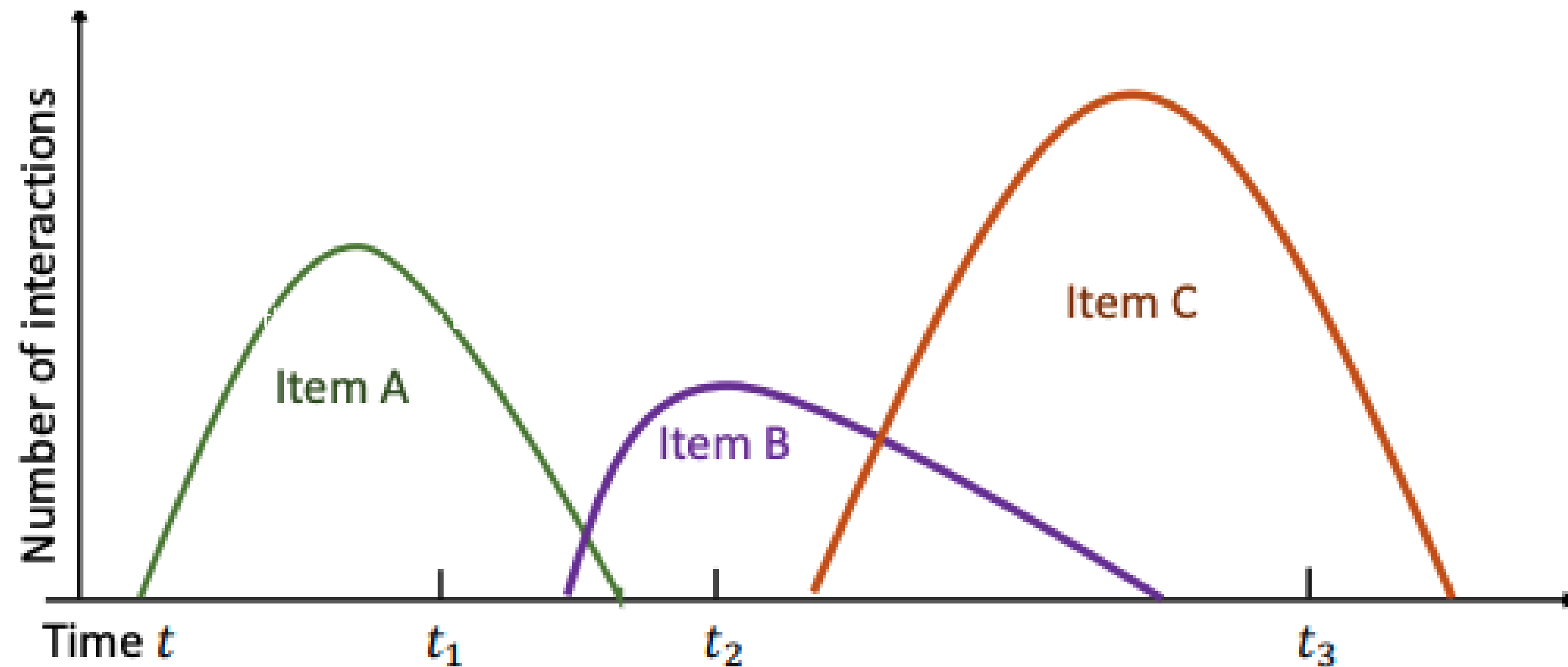


Popularity	HR@5	HR@10	NDCG@5	NDCG@10
MostPop	0.0304	0.0462	0.0198	0.0248
RecentPop	0.0530	0.0845	0.0338	0.0440
DecayPop	0.0532	0.0843	0.0341	0.0441

Что еще важно на этом графике?



Что еще важно на этом графике?

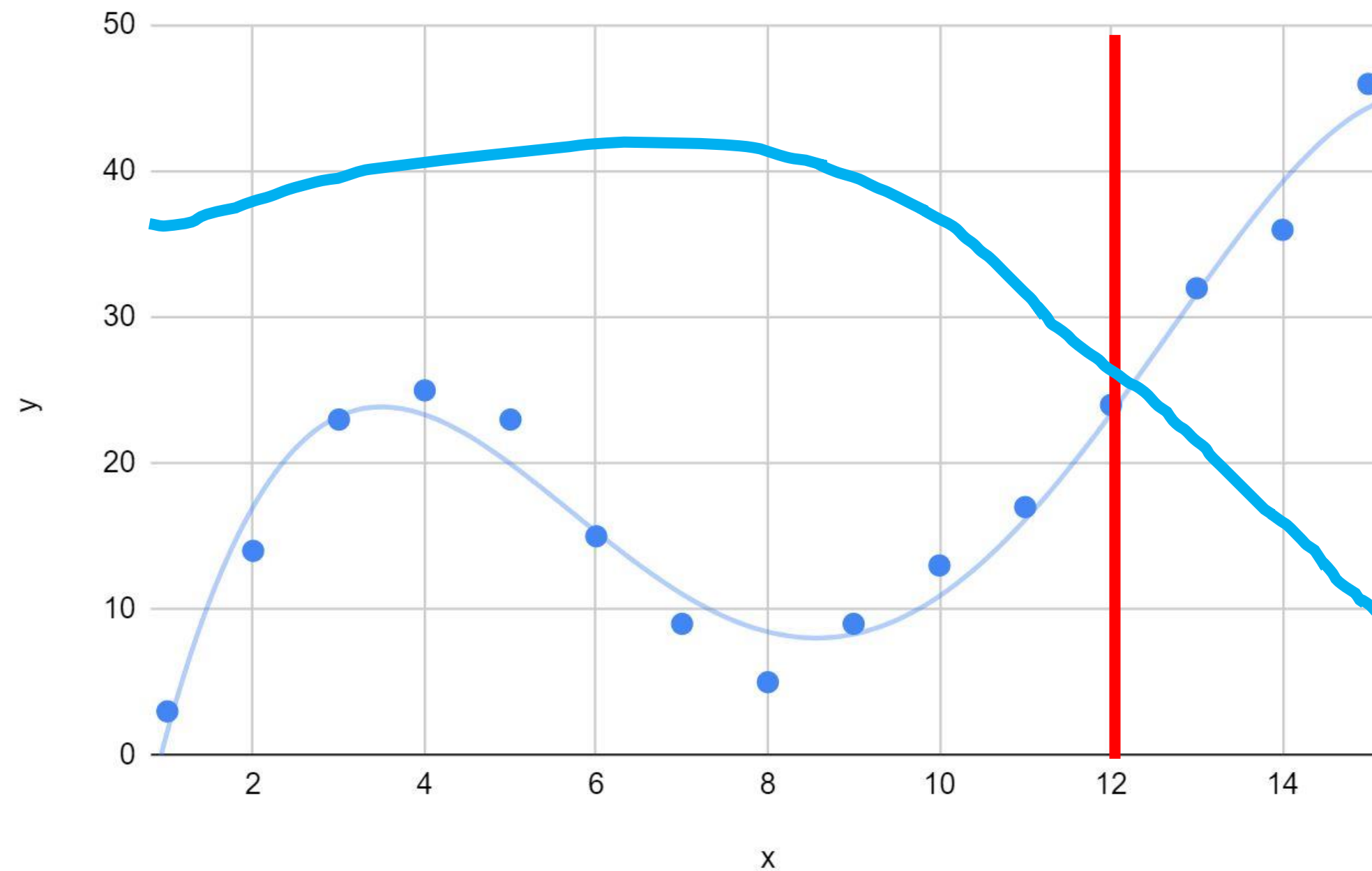


Item A уже не продается -> его нет смысла рекомендовать

Item C в t_2 не известен, модель не сможет его рекомендовать

Популярность как временной ряд

Два айтема и их популярность



История

Как еще можно отсортировать айтемы?



User_id	Item_id	Timestamp	label
1	2	124	0
2	3	2512	0
1	1	12	1
3	1	1	1
4	2	23424	1

Как еще можно отсортировать айтемы?

По Click through rate (CTR)!

User_id	Item_id	Timestamp	label
1	2	124	0
2	3	2512	0
1	1	12	1
3	1	1	1
4	2	23424	1

Item_id	views	clicks	CTR
2	2	1	0.5
3	1	0	0
1	2	2	1



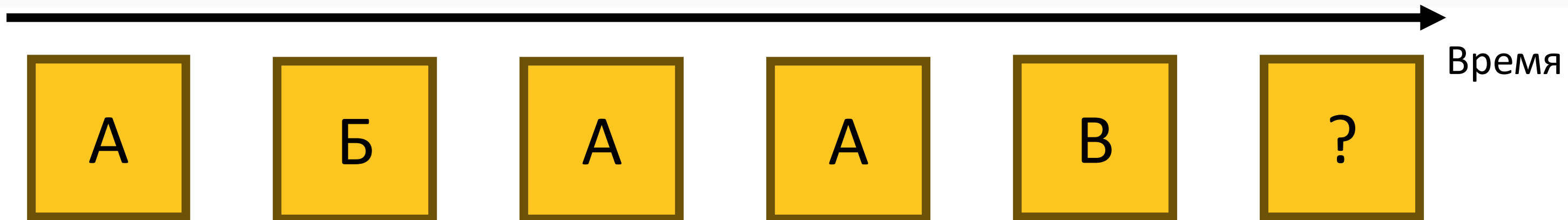
Какие еще проблемы?

Популярность: что надо запомнить

- 1) Можно считать как долю интеракций по каждому айтому
- 2) Меняется со временем, надо учитывать
- 3) У некоторых айтемов падает до 0 со временем
- 4) В литературе: Pop, MostPop, Popular, TopPop, Top
- 5) Рекомендации популярных – часто могут давать «релевантные» рекомендации. Почему?

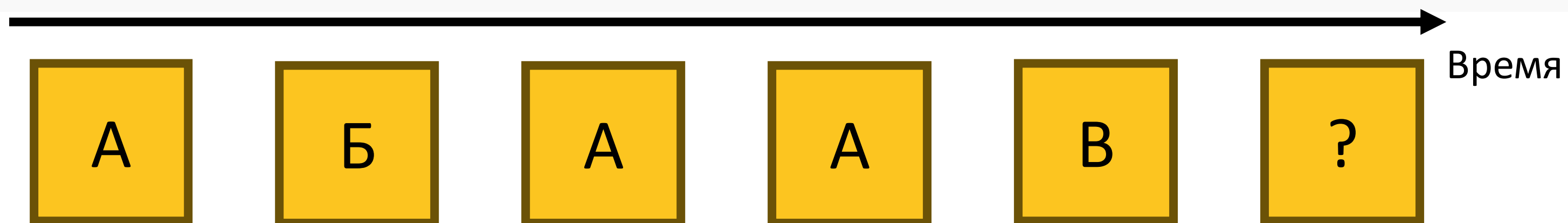


Популярность бывает персональная



$$UP(u, i, t_k) = \frac{\# \text{ взаимодействий } (u, i)}{\text{размер датасета}} \mid t > \text{now()} - t_k$$

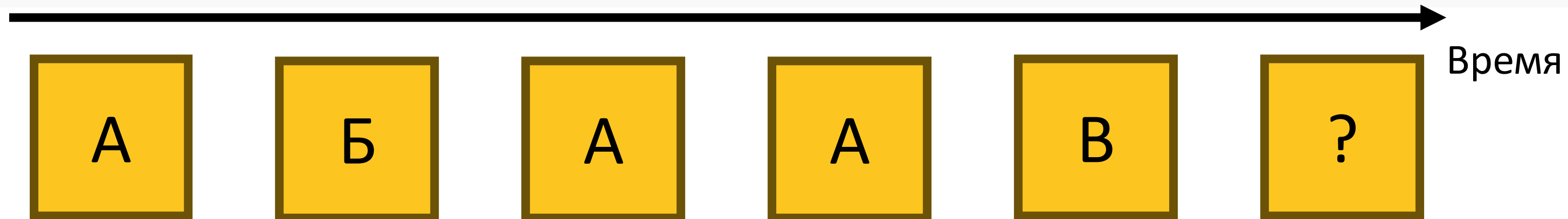
Как объединить TopPersonal + Pop?



$$UP(u, i, t_k) = \frac{\# \text{ взаимодействий } (u, i)}{\text{размер датасета}} \mid t > \text{now()} - t_k$$

$$P(i, t) = \frac{\# \text{ взаимодействий с айтемом } i}{\text{размер датасета}} \mid t > \text{now()} - t_i$$

Как объединить TopPersonal + Pop?



$$UP(u, i, t_k) = \frac{\# \text{ интеракций } (u, i)}{\text{размер датасета}} \mid t > \text{now()} - t_k$$

$$P(i, t) = \frac{\# \text{ интеракций с айтемом } i}{\text{размер датасета}} \mid t > \text{now()} - t_i$$

$$\text{TopPersonal}(u, i, t) = (\#(u, i) + \frac{\# \text{ интеракций с айтемом } i}{\text{размер датасета}}) \mid t > \text{now()} - t_i$$

Разбиение данных для экспериментов

Как разбить данные для экспериментов?



	user_id	item_id	timestamp
0	38	89	92
1	68	32	94
2	9	76	78
3	75	53	47
4	30	48	64
...
95	63	48	13
96	57	50	55
97	47	90	59
98	82	66	93
99	44	92	99

Есть 4 основных способа

Leave One Last	Temporal Split	Random Split	User Split
Item	Basket/Session	User-based	Global

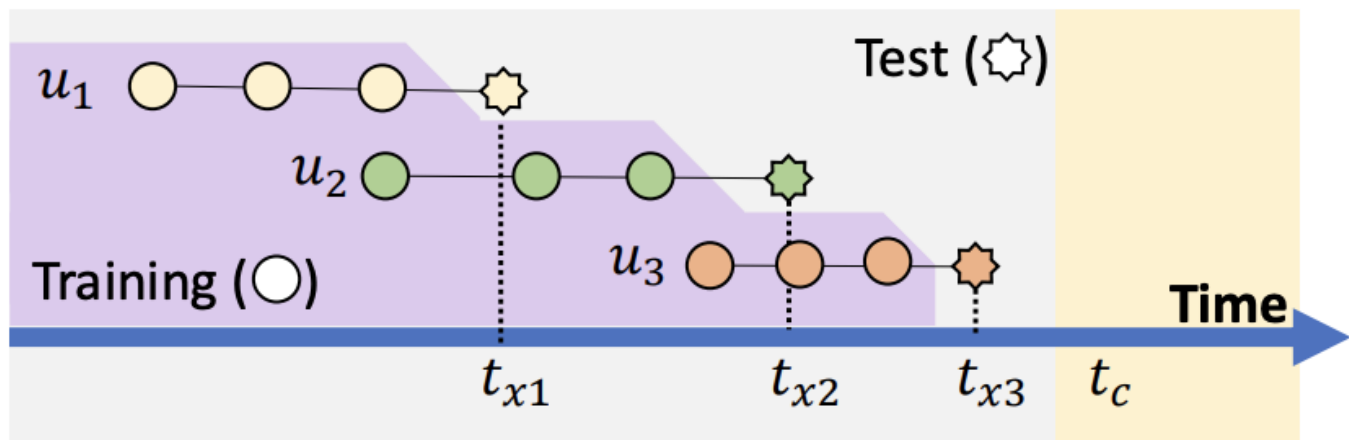


Figure 3: Train/test with leave-one-out split

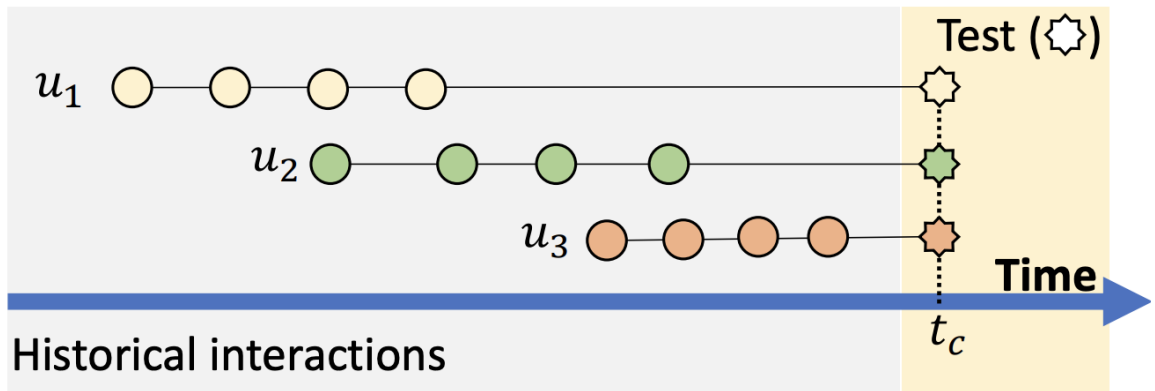


Figure 1: Train/test in practical systems, where t_c indicates the current time point.

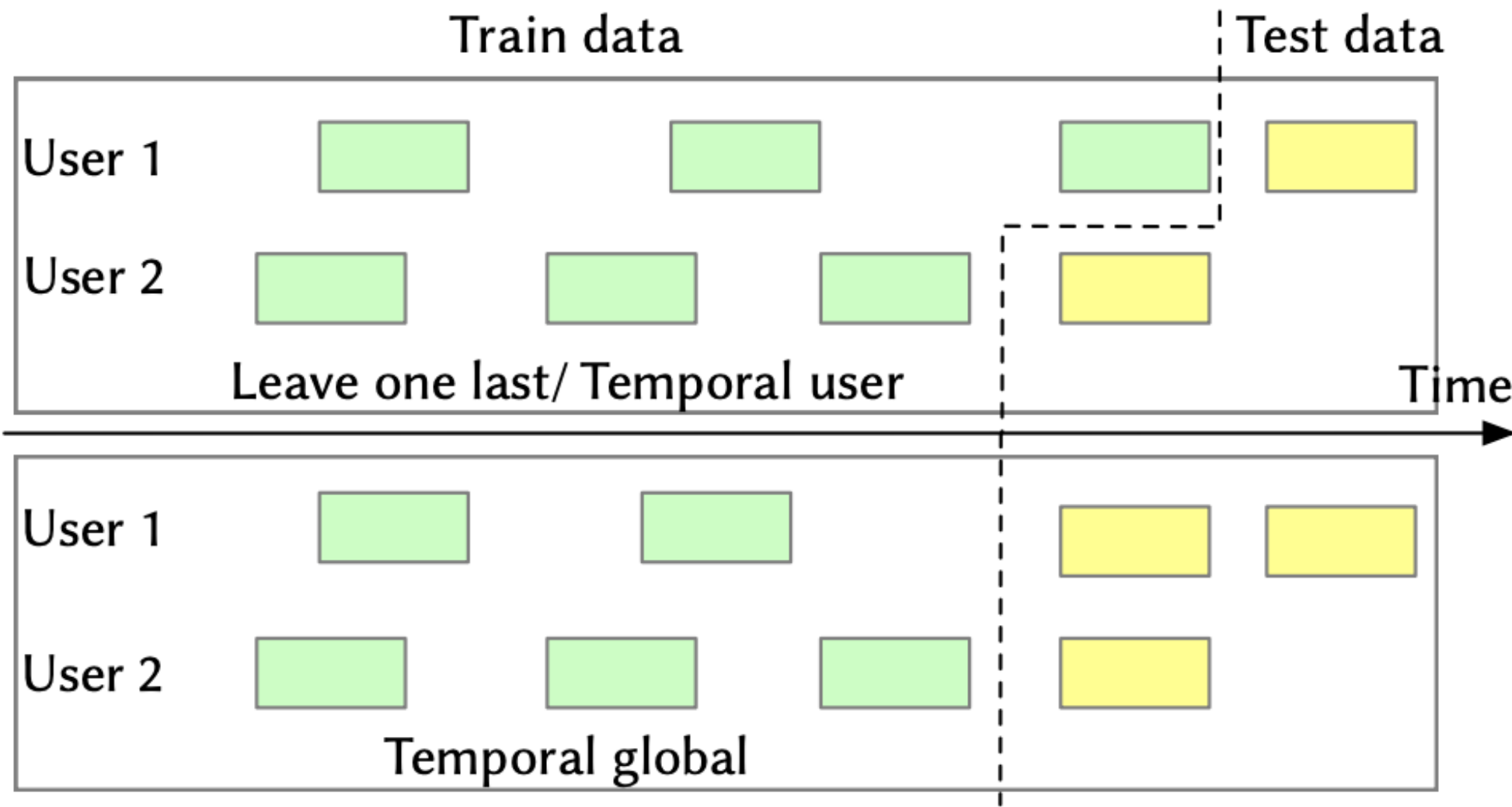


Fig. 1. Temporal global split v.s. Temporal user split/Leave one last

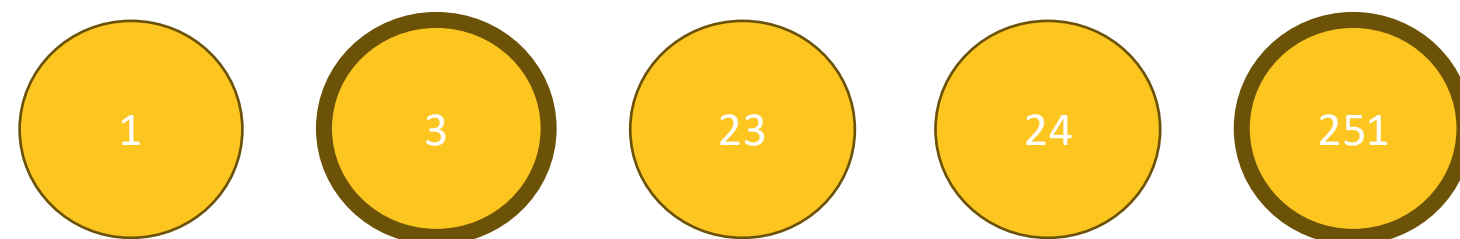
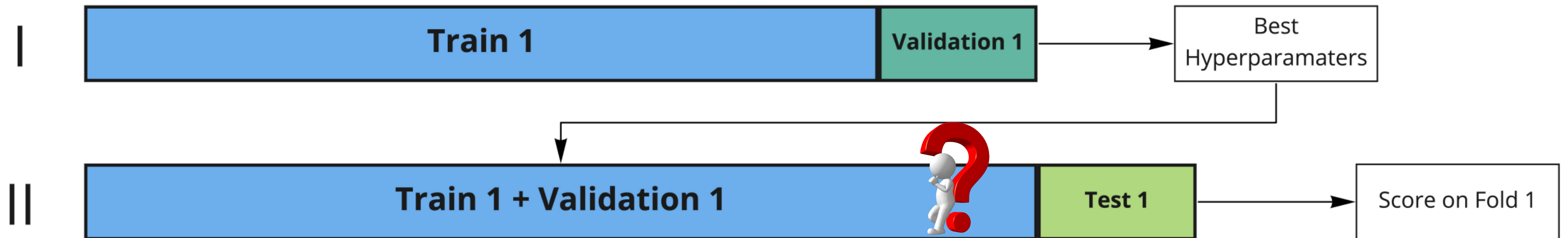
Также есть симуляции + движущееся окно

- Collection: 88 papers in RecSys conferences (2020 – 2022)

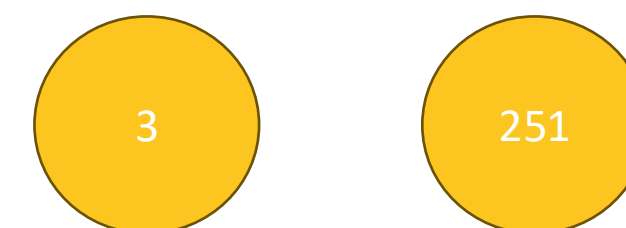
No. papers	Percentage	Train/test split	Global timeline?
30	34%	Random split	No
22	25%	Leave-one-out	No
17	19.5%	Single time point	Partially
15	17%	Simulation-based online	Yes
4	4.5%	Sliding window	Yes

В индустрии почти наверно все используют single time point, но это не так в академии

Что такое оффлайн эксперимент?



Персональные рекомендации



Действия пользователя в тесте

Подробнее про это в следующие
разы

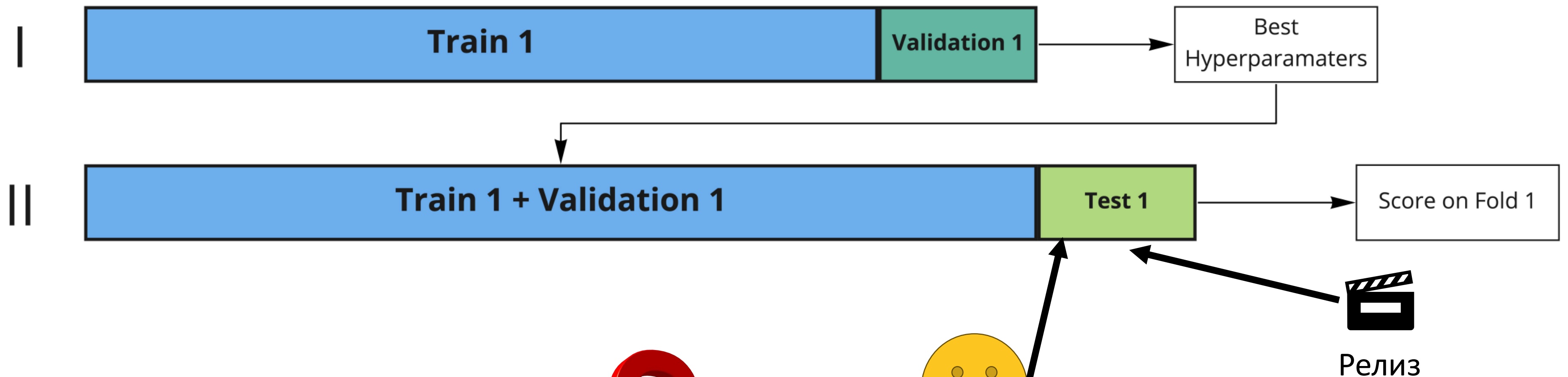
Для оффлайн эксперимента нужно выбрать метрику для оптимизации

RECALL@10_new	0.157	0.180	0.196	0.200	0.199	0.192	0.209	0.198	0.181	0.173
RECALL@25_new	0.255	0.317	0.330	0.339	0.320	0.322	0.345	0.339	0.307	0.302
RECALL@50_new	0.374	0.446	0.446	0.462	0.454	0.447	0.473	0.461	0.420	0.420
PRECISION@10_new	0.041	0.046	0.048	0.049	0.048	0.046	0.049	0.049	0.043	0.041
PRECISION@25_new	0.027	0.032	0.033	0.034	0.032	0.032	0.034	0.033	0.030	0.030
PRECISION@50_new	0.019	0.023	0.023	0.023	0.023	0.023	0.024	0.023	0.021	0.021
NOVELTY@10_new	0.273	0.299	0.352	0.331	0.369	0.345	0.338	0.372	0.334	0.342
NOVELTY@25_new	0.301	0.327	0.380	0.362	0.386	0.375	0.369	0.392	0.361	0.370
NOVELTY@50_new	0.327	0.353	0.406	0.387	0.406	0.401	0.396	0.412	0.385	0.395
MNAP@10_new	0.062	0.075	0.077	0.079	0.082	0.073	0.082	0.085	0.069	0.067
MNAP@25_new	0.070	0.086	0.089	0.090	0.092	0.083	0.093	0.097	0.079	0.078
MNAP@50_new	0.075	0.092	0.094	0.096	0.098	0.089	0.099	0.102	0.084	0.083
MIL_D@10_new	0.535	0.700	0.880	0.797	0.912	0.803	0.819	0.884	0.781	0.778



Как подобрать тут оптимальные гиперпараметры или модель?

Холодный старт по user/item



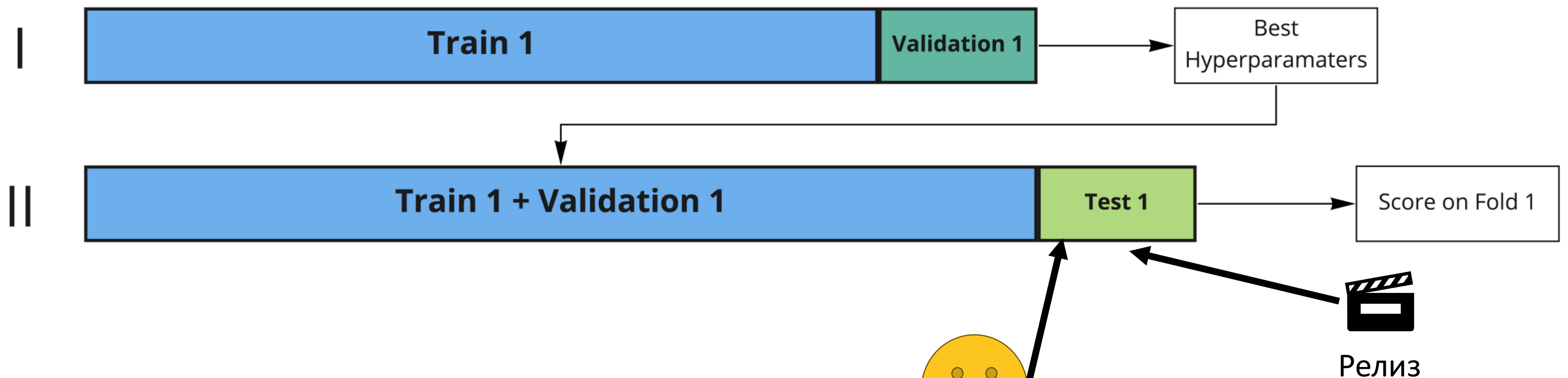
1) Включать ли пользователя в метрики?

2) Выкинуть ли новый айтем из теста?



Регистрация

Холодный старт по user/item

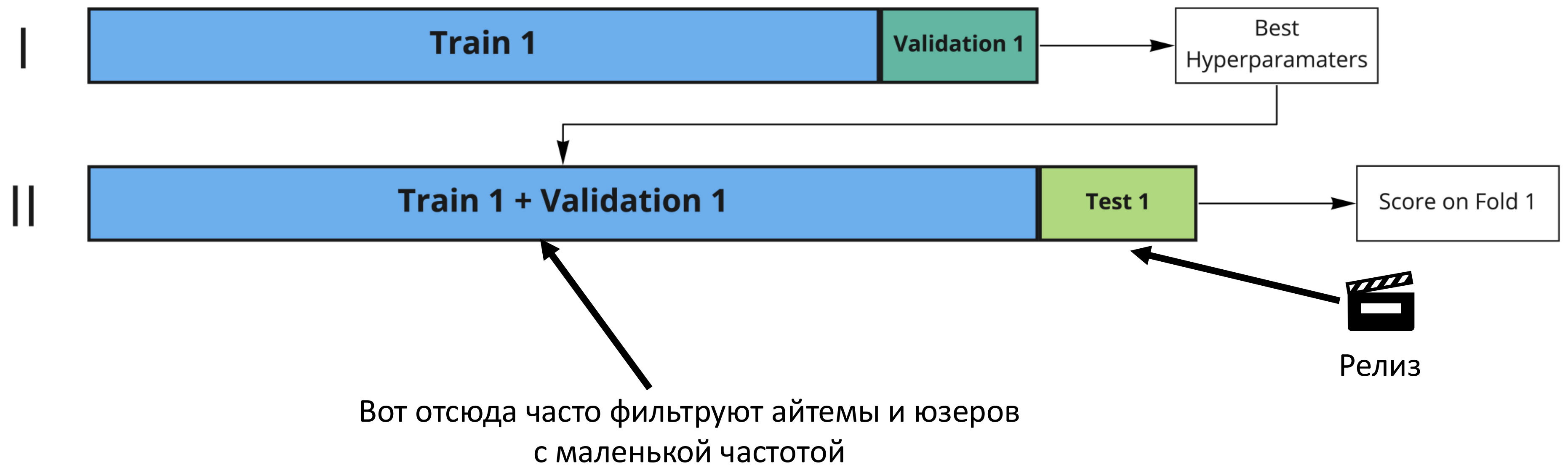


1) Включать ли пользователя в метрики? **Зависит от ситуации**

2) Выкинуть ли новый айтем из теста? **Лучше нет**

Регистрация

Про фильтрацию айтемов



Но когда вы отфильтруете, вы не предскажете новый айтем!

Важно: фильтровать данные лучше только из обучения, но не из теста/валидации!

Холодный старт

«Нужно сделать предсказания, но нет истории взаимодействий»

Способы:

1. Популярные айтемы
2. Построить модель на соцдеме или других данных
3. Собрать данные

UserKNN, ItemKNN

“Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read”

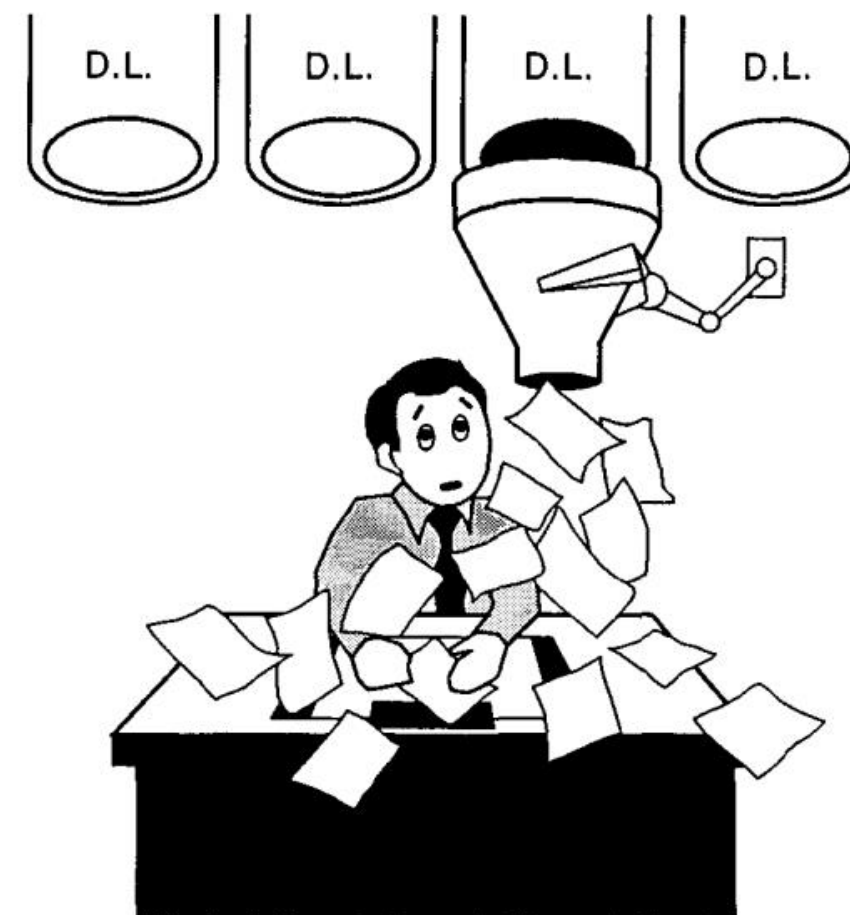
CF – это идея (подход), не конкретный алгоритм!



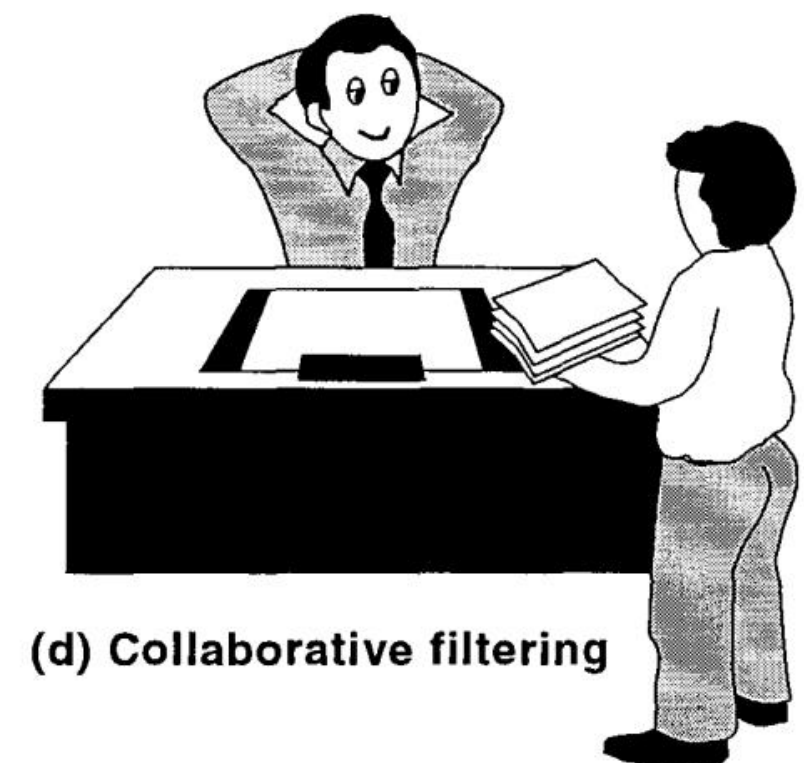
(a) Electronic mail overload



(b) Using distribution lists



(c) Conventional filtering



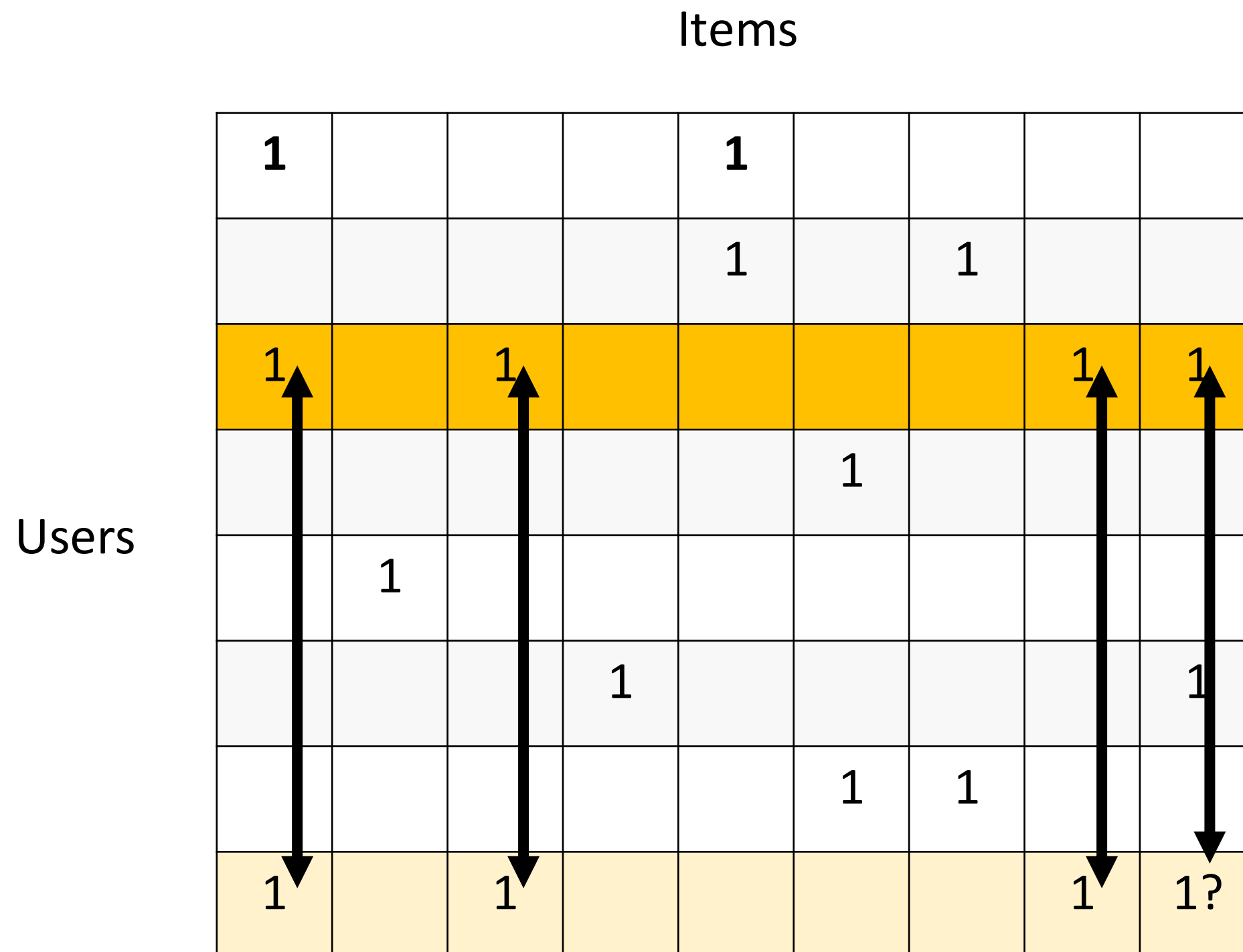
(d) Collaborative filtering

UserKNN, ItemKNN

Items

1				1				
				1		1		
1		1					1	1
					1			
	1							
		1						1
					1	1		
1		1					1	1?

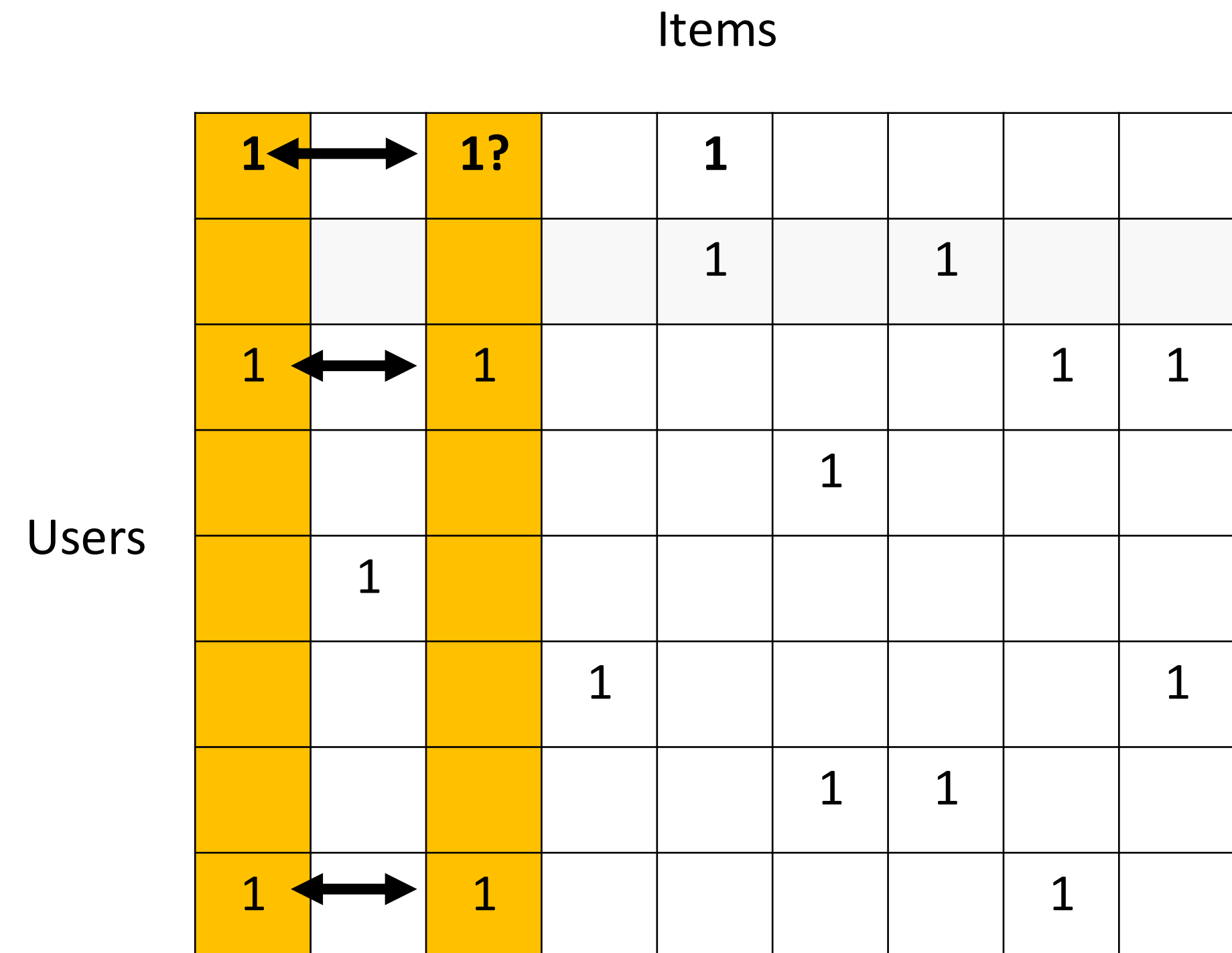
Users



Items

1	1?		1				
			1		1		
1	1					1	1
				1			
	1						
		1					1
					1	1	
1	1					1	

Users



1 – было взаимодействие, отсутствие – нет взаимодействия

UserKNN

Items

Users

1				1				
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	1?

$$\hat{r}_{ui} = \frac{1}{|\mathcal{N}_i(u)|} \sum_{v \in \mathcal{N}_i(u)} r_{vi}$$

1 – было взаимодействие, отсутствие – нет взаимодействия

UserKNN: как определить $N_i(u)$?

	Items								
	1				1				
					1		1		
	1		1					1	1
						1			
Users		1							
				1					1
					1	1			
	1		1					1	1?

$$\hat{r}_{ui} = \frac{1}{|\mathcal{N}_i(u)|} \sum_{v \in \mathcal{N}_i(u)} r_{vi}$$

Table 1 OTUs Expression of Binary Instances i and j

$j \backslash i$	1 (Presence)	0 (Absence)	Sum
1 (Presence)	$a = i \bullet j$	$b = \bar{i} \bullet j$	$a+b$
0 (Absence)	$c = i \bullet \bar{j}$	$d = \bar{i} \bullet \bar{j}$	$c+d$
Sum	$a+c$	$b+d$	$n=a+b+c+d$

$$S_{COSINE} = \frac{a}{\sqrt{(a+b)(a+c)}}^2$$

1 – было взаимодействие, отсутствие – нет взаимодействия

Есть 76+ методов для поиска соседей

A Survey of Binary Similarity and Distance Measures

Seung-Seok Choi, Sung-Hyuk Cha, Charles C. Tappert
Department of Computer Science, Pace University
New York, US

Table 2 Definitions of Measures for binary data

$$S_{JACCARD} = \frac{a}{a+b+c} \quad (1)$$

$$S_{DICE} = \frac{2a}{2a+b+c} \quad (2)$$

$$S_{CZEKANOWSKI} = \frac{2a}{2a+b+c} \quad (3)$$

$$S_{3W-JACCARD} = \frac{3a}{3a+b+c} \quad (4)$$

$$S_{NEI\&LI} = \frac{2a}{(a+b)+(a+c)} \quad (5)$$

$$S_{SOKAL\&SNEATH-I} = \frac{a}{a+2b+2c} \quad (6)$$

$$S_{SOKAL\&MICHENER} = \frac{a+d}{a+b+c+d} \quad (7)$$

$$S_{SOKAL\&SNEATH-II} = \frac{2(a+d)}{2a+b+c+2d} \quad (8)$$

$$S_{ROGER\&TANIMOTO} = \frac{a+d}{a+2(b+c)+d} \quad (9)$$

$$S_{FAITH} = \frac{a+0.5d}{a+b+c+d} \quad (10)$$

$$S_{GOWER\&LEGENDRE} = \frac{a+d}{a+0.5(b+c)+d} \quad (11)$$

$$D_{VARI} = \frac{(b+c)}{4(a+b+c+d)} \quad (23)$$

$$D_{SIZEDIFFERENCE} = \frac{(b+c)^2}{(a+b+c+d)^2} \quad (24)$$

$$D_{SHAPEDIFFERENCE} = \frac{n(b+c)-(b-c)^2}{(a+b+c+d)^2} \quad (25)$$

$$D_{PATTERNDIFFERENCE} = \frac{4bc}{(a+b+c+d)^2} \quad (26)$$

$$D_{LANCE\&WILLIAMS} = \frac{b+c}{(2a+b+c)} \quad (27)$$

$$D_{BRAY\&CURTIS} = \frac{b+c}{(2a+b+c)} \quad (28)$$

$$D_{HELLINGER} = 2\sqrt{1 - \frac{a}{\sqrt{(a+b)(a+c)}}} \quad (29)$$

$$D_{CHORD} = \sqrt{2\left(1 - \frac{a}{\sqrt{(a+b)(a+c)}}\right)} \quad (30)$$

$$S_{COSINE} = \frac{a}{\sqrt{(a+b)(a+c)}} \quad (31)$$

Часто используется Jaccard, Dice, Cosine

$$\hat{r}_{ui} = \frac{1}{|\mathcal{N}_i(u)|} \sum_{v \in \mathcal{N}_i(u)} r_{vi}$$

Table 1 OTUs Expression of Binary Instances i and j

$j \backslash i$	1 (Presence)	0 (Absence)	Sum
1 (Presence)	$a = i \bullet j$	$b = \bar{i} \bullet j$	$a+b$
0 (Absence)	$c = i \bullet \bar{j}$	$d = \bar{i} \bullet \bar{j}$	$c+d$
Sum	$a+c$	$b+d$	$n=a+b+c+d$

Гиперпараметры userKNN

$$s_{ij} = \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\|\mathbf{r}_i\| \|\mathbf{r}_j\| + h}$$

1) Число соседей

2) h – сглаживающий параметр

3) Метрика близости

ItemKNN

		Items							
Users		1 ↔ 1?		1					
				1		1			
		1 ↔ 1					1	1	
					1				
		1							
			1						1
					1	1			
		1 ↔ 1					1		

Пример: почти все к пельменям покупают кетчуп.

Пользователь уже купил Пельмени, возможно ему будет нужен кетчуп

$$\hat{r}_{ui} = \frac{\sum_{j \in \mathcal{N}_u(i)} w_{ij} r_{uj}}{\sum_{j \in \mathcal{N}_u(i)} |w_{ij}|}$$

ItemKNN/ item2item модели

USERS * ITEMS

1				1				
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	

X

ITEMS * ITEMS

.
.	.	.	*
.	.	.	*
.	.	.	*
.
.	.	.	*
.	.	.	*
.	.	.	*
.	.	.	*

USERS * ITEMS

.
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	

$$\hat{r}_{ui} = \frac{\sum_{j \in \mathcal{N}_u(i)} w_{ij} r_{uj}}{\sum_{j \in \mathcal{N}_u(i)} |w_{ij}|}$$

$$r(u, 3) = 1 * \text{sim}(3, 0) + 1 * \text{sim}(3, 4)$$

Матрица схожести айтемов

User/item сравнение

Table 2.2 The space and time complexity of user-based and item-based neighborhood methods, as a function of the maximum number of ratings per user $p = \max_u |\mathcal{J}_u|$, the maximum number of ratings per item $q = \max_i |\mathcal{U}_i|$, and the maximum number of neighbors used in the rating predictions k

	Space	Time	
		Training	Online
User-based	$O(\mathcal{U} ^2)$	$O(\mathcal{U} ^2 p)$	$O(\mathcal{J} k)$
Item-based	$O(\mathcal{J} ^2)$	$O(\mathcal{J} ^2 q)$	$O(\mathcal{J} k)$

Нужно ли это кому-то сейчас?

Top-N Recommendation Algorithms: A Quest for the State-of-the-Art

VITO WALTER ANELLI, Politecnico di Bari, Italy

ALEJANDRO BELLOGÍN, Universidad Autónoma de Madrid, Spain

TOMMASO DI NOIA, Politecnico di Bari, Italy

DIETMAR JANNACH, University of Klagenfurt, Austria

CLAUDIO POMO, Politecnico di Bari, Italy

Rank	Algorithm	Count
1	EASE ^R	185
2	RP ³ β	169
3	SLIM	160
4	UserKNN	154
5	MF2020	115
6	ItemKNN	99
7	MultiVAE	92
8	iALS	90
9	NeuMF	61
10	BPRMF	45
11	MostPop	18
12	Random	0

(a) Overall

Нужно ли это кому-то сейчас?

Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches

Maurizio Ferrari Dacrema
Politecnico di Milano, Italy
maurizio.ferrari@polimi.it

Paolo Cremonesi
Politecnico di Milano, Italy
paolo.cremonesi@polimi.it

Dietmar Jannach
University of Klagenfurt, Austria
dietmar.jannach@aau.at

Table 6: Experimental results for NCF.

	Pinterest			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1663	0.1065	0.2744	0.1412
UserKNN	0.7001	0.5033	0.8610	0.5557
ItemKNN	0.7100	0.5092	0.8744	0.5629
$P^3\alpha$	0.7008	0.5018	0.8667	0.5559
$RP^3\beta$	0.7105	0.5116	0.8740	0.5650
NeuMF	0.7024	0.4983	0.8719	0.5536
	Movielens 1M			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.3043	0.2062	0.4531	0.2542
UserKNN	0.4916	0.3328	0.6705	0.3908
ItemKNN	0.4829	0.3328	0.6596	0.3900
$P^3\alpha$	0.4811	0.3331	0.6464	0.3867
$RP^3\beta$	0.4922	0.3409	0.6715	0.3991
NeuMF	0.5486	0.3840	0.7120	0.4369
SLIM	0.5589	0.3961	0.7161	0.4470

Линейные модели (EASE, SLIM)

Вернемся к таблице с рейтингом моделей

Rank	Algorithm	Count	
1	EASE ^R	185	Разберем сегодня
2	RP ³ β	169	Разберем дальше
3	SLIM	160	Разберем сегодня
4	UserKNN	154	Разобрали
5	MF2020	115	Под вопросом
6	ItemKNN	99	Разобрали
7	MultiVAE	92	Разберем дальше
8	iALS	90	Разберем дальше
9	NeuMF	61	Разберем дальше
10	BPRMF	45	Разберем дальше
11	MostPop	18	Разобрали
12	Random	0	np.random.choice

(a) Overall

SLIM: Sparse Linear Model (2011)

USERS * ITEMS

1				1				
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	

X

ITEMS * ITEMS

0
.	0
.	.	0	
.	.	.	0
.	.	.		0
.	.	.		.	0	.	.	.
.	0	.	.
.	0	.
.	0

=

USERS * ITEMS

1
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	

$$\hat{r}_{ui} = \mathbf{r}_u \mathbf{w}_i^\top$$

minimize_W

subject to

$\frac{1}{2} \|R - RW\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|W\|_1$
 $W \geq 0$
 $\text{diag}(W) = 0.$

SLIM: Sparse Linear Model (2011)

Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches

Maurizio Ferrari Dacrema
Politecnico di Milano, Italy
maurizio.ferrari@polimi.it

Paolo Cremonesi
Politecnico di Milano, Italy
paolo.cremonesi@polimi.it

Dietmar Jannach
University of Klagenfurt, Austria
dietmar.jannach@aau.at

Table 6: Experimental results for NCF.

	Pinterest			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.1663	0.1065	0.2744	0.1412
UserKNN	0.7001	0.5033	0.8610	0.5557
ItemKNN	0.7100	0.5092	0.8744	0.5629
$P^3\alpha$	0.7008	0.5018	0.8667	0.5559
$RP^3\beta$	0.7105	0.5116	0.8740	0.5650
NeuMF	0.7024	0.4983	0.8719	0.5536
	Movielens 1M			
	HR@5	NDCG@5	HR@10	NDCG@10
TopPopular	0.3043	0.2062	0.4531	0.2542
UserKNN	0.4916	0.3328	0.6705	0.3908
ItemKNN	0.4829	0.3328	0.6596	0.3900
$P^3\alpha$	0.4811	0.3331	0.6464	0.3867
$RP^3\beta$	0.4922	0.3409	0.6715	0.3991
NeuMF	0.5486	0.3840	0.7120	0.4369
SLIM	0.5589	0.3961	0.7161	0.4470

Popularity bias на примере SLIM

USERS * ITEMS

1				1				
				1		1		
1		1					1	1
					1			
	1							
			1					1
					1	1		
1		1					1	

X

ITEMS * ITEMS

0
.	0
.	.	0
.	.	.	0
.	.	.	.	0
.	0	.	.	.
.	0	.	.
.	0	.
.	0

=

USERS * ITEMS

1			.	1				
			.	1		1		
1		1	.				1	1
			.		1			
	1		.					
			1					1
			.		1	1		
1		1	?				1	

$$\underset{W}{\text{minimize}} \quad \frac{1}{2} \|R - RW\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|W\|_1$$

$$\text{subject to} \quad W \geq 0$$

$$\text{diag}(W) = 0.$$

Веса выучатся маленькими,
айтем будет редко
рекомендоваться

Popularity bias

Popularity bias – смещение в сторону рекомендаций популярных объектов (их преобладание в выдаче). Несколько фактов:

1. Смещение на популярные изначально есть в данных. Люди сами так ведут себя
2. Модель «рискует», рекомендуя менее популярные айтемы, идет против статистики
3. «Богатые богатеют, бедные беднеют»

Важная проблема с Popular Bias

USERS * ITEMS

1				1				
				1		1		
1		1	1	?			1	1
				1	1			
	1			1				
			1	1				1
				1	1	1		
1		1		1			1	

X

ITEMS * ITEMS

0
.	0
.	.	0
.	.	.	0
.	.	.	.	0
.	0	.	.	.
.	0	.	.
.	0	.
.	0

=

USERS * ITEMS

1				1				
				1		1		
1		1	1				1	1
				1	1			
	1			1				
				1				1
				1	1	1		
1		1		1			1	



Порекомендует ли модель айтем №4?

Важная проблема с Popular Bias

USERS * ITEMS

1				1				
				1		1		
1		1	1	?			1	1
				1	1			
	1			1				
			1	1				1
				1	1	1		
1		1		1			1	

X

ITEMS * ITEMS

0
.	0
.	.	0
.	.	.	0
.	.	.	.	0
.	0	.	.	.
.	0	.	.
.	0	.
.	0

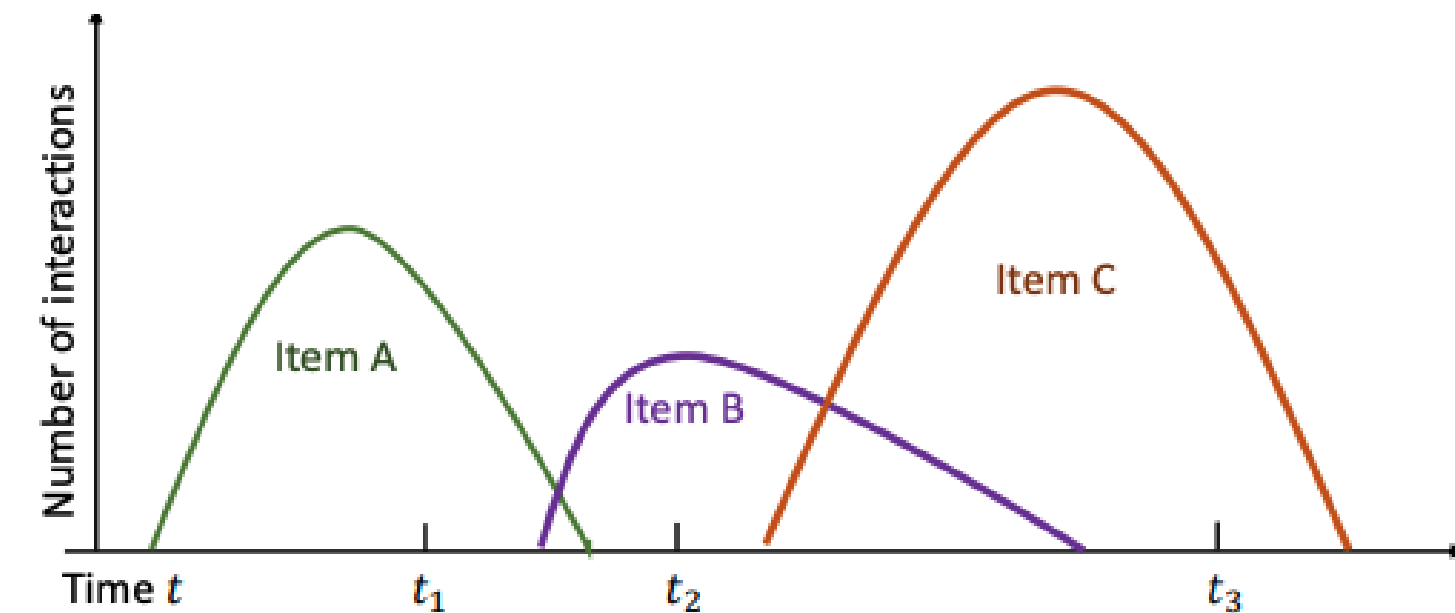
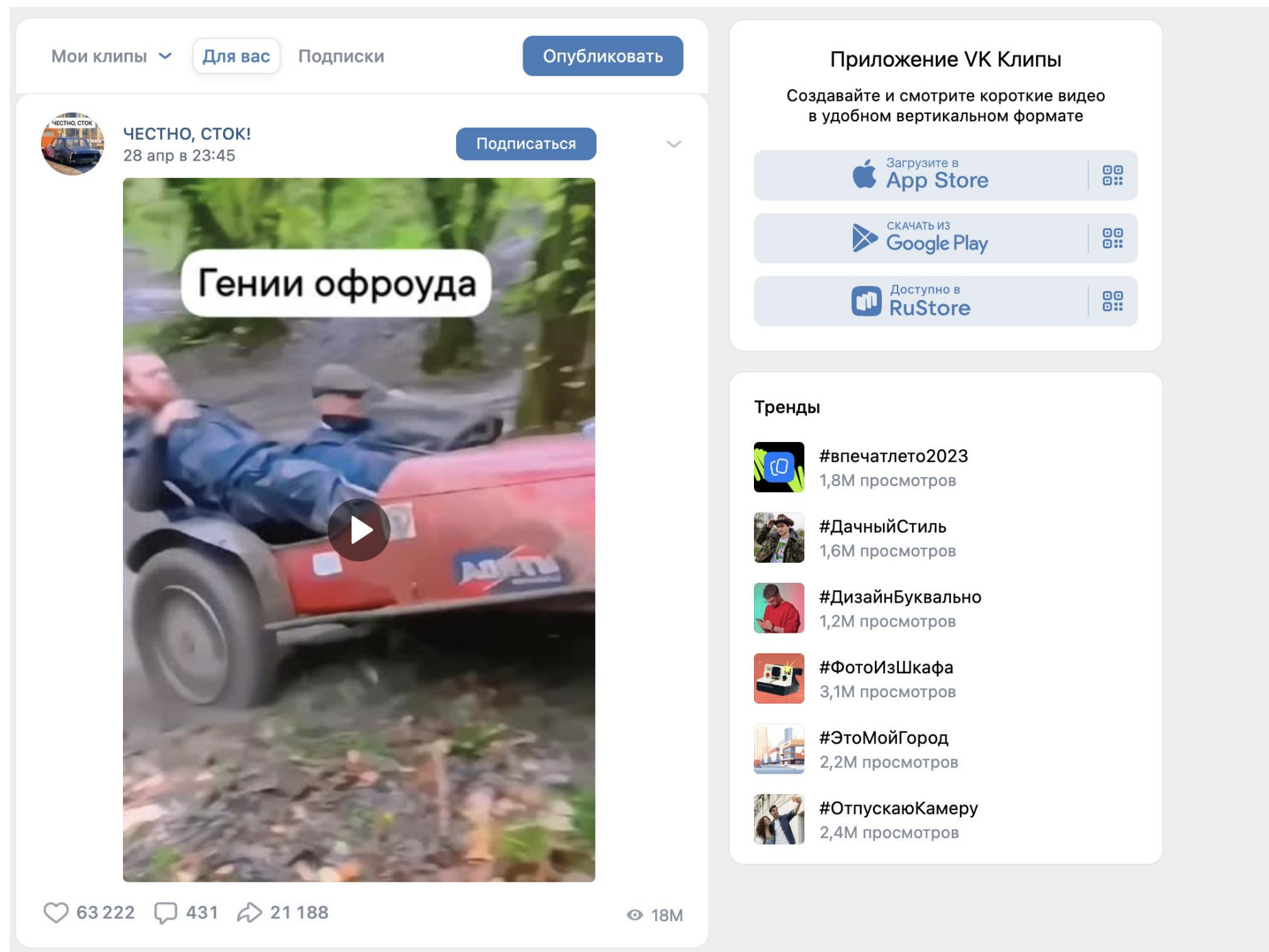
=

USERS * ITEMS

1				1				
				1		1		
1		1	1				1	1
				1	1			
	1			1				
				1				1
				1	1	1		
1		1		1			1	

Но пользователь скорее всего осознанно
не провзаимодействовал с ним

Как залететь в рекомендации или проблемы начинающих авторов



Popularity bias форсирует не только система, но и люди.

Если автор контента стал делать плохой контент, он все равно имеет инерцию/лояльность в сознании подписчиков. Факт привычки

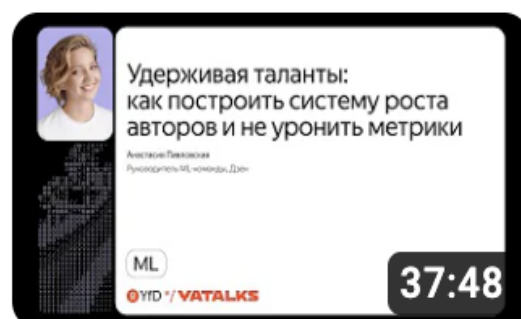
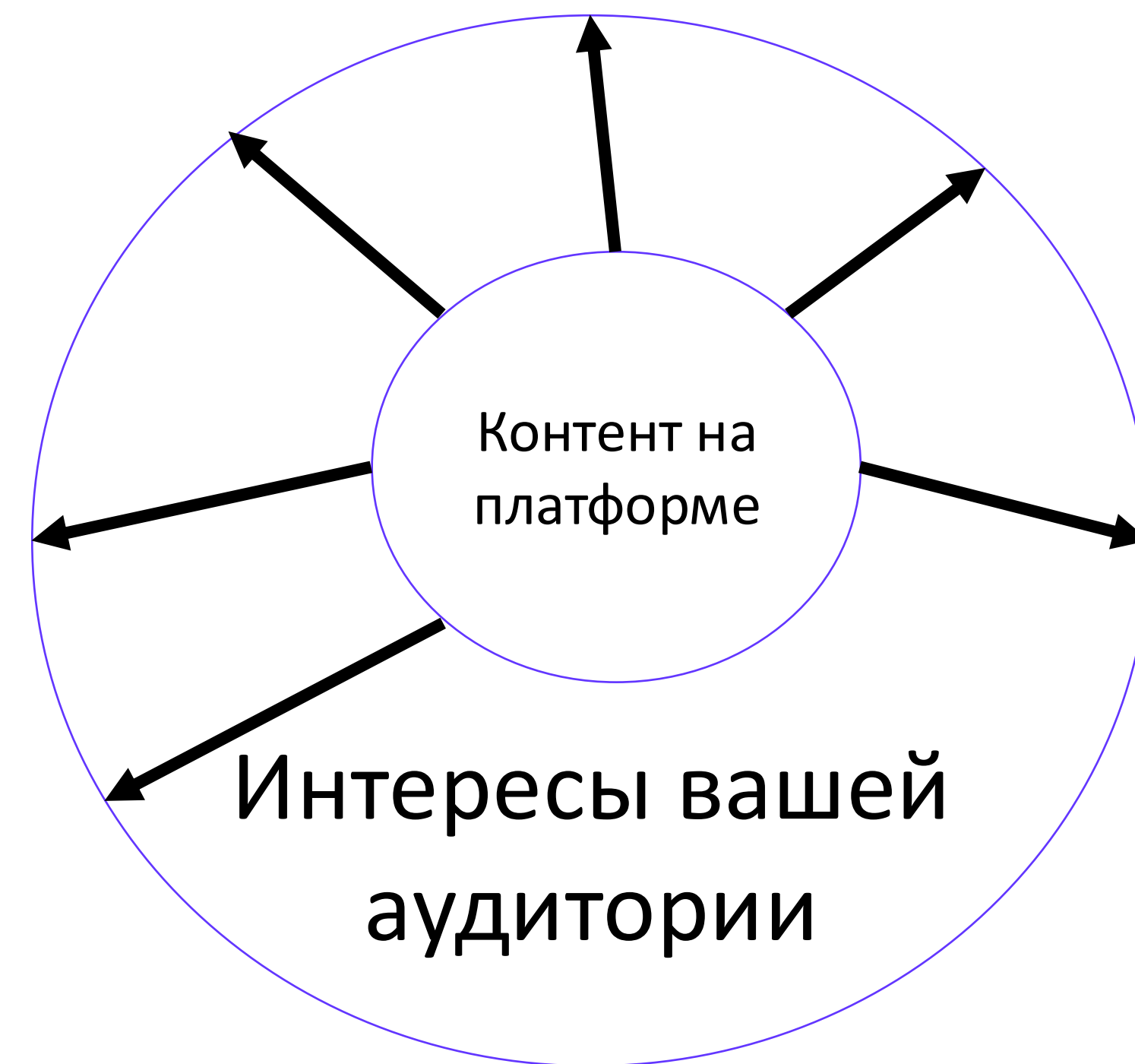
Если новый автор делает качественный контент, система может не поддерживать его, а люди по инерции смотреть популярные каналы

Почему нельзя рекомендовать только популярные?

Если нет трафика на непопулярный контент, его может становиться меньше. Притока новых поставщиков контента может быть меньше.

Но чем меньше такого трафика – тем меньше интересов людей вы покрываете

Позднее вернемся к этой проблеме



Анастасия Павловская – Удерживая таланты: как построить систему роста авторов и не уронить метрики

YfD • 213 views • 1 month ago

EASE (2019): топ-2* модель, которую надо знать

Embarrassingly Shallow Autoencoders for Sparse Data*

Harald Steck
Netflix
Los Gatos, California
hsteck@netflix.com

The predicted score $S_{u,j}$ for an item $j \in \mathcal{I}$ given a user $u \in \mathcal{U}$ is defined by the dot product

$$S_{uj} = X_{u,\cdot} \cdot B_{\cdot,j}, \quad (1)$$

where $X_{u,\cdot}$ refers to row u , and $B_{\cdot,j}$ to column j .

$$\begin{aligned} \min_B \quad & ||X - XB||_F^2 + \lambda \cdot ||B||_F^2 \\ \text{s.t.} \quad & \text{diag}(B) = 0 \end{aligned}$$

EASE

$$\begin{aligned} \underset{W}{\text{minimize}} \quad & \frac{1}{2} \|R - RW\|_F^2 + \frac{\beta}{2} \|W\|_F^2 + \lambda \|W\|_1 \\ \text{subject to} \quad & W \geq 0 \\ & \text{diag}(W) = 0. \end{aligned}$$

SLIM

EASE имеет решение в явном виде

Embarrassingly Shallow Autoencoders for Sparse Data*

Harald Steck
Netflix
Los Gatos, California
hsteck@netflix.com

$$\hat{P} \triangleq (X^T X + \lambda I)^{-1},$$

$$\hat{B}_{i,j} = \begin{cases} 0 & \text{if } i = j \\ -\frac{\hat{P}_{ij}}{\hat{P}_{jj}} & \text{otherwise.} \end{cases}$$

Доказательство
можете почитать в
статье

Преимущества EASE

- 1) Имеет решение в явном виде, не требует обучения по эпохам
- 2) Явное решение оптимально в смысле поставленной задачи
- 3) Можно один раз обучить матрицу B , и использовать ее на разных юзерах
- 4) Имеет всего один гиперпараметр
- 5) Устойчив к варьированию гиперпараметра (в отличие от SLIM). В целом 100 – хорошее значение
- 6) Улучшает качество при добавлении редких айтемов (!)
- 7) Если юзер сделал новую интеракцию – можем обновить рекомендации без дообучения
- 8) Может спокойно работать в онлайн при latency 50 мс под нагрузкой 200 rps с ~10к каталогом
- 9) Можно искать похожие айтемы через его матрицу взаимодействий

Результаты EASE

- 1) Наш коллега из Т-Банка занял ~20/2к мест в соревновании по recsys на Kaggle, где EASE была основной моделью, размер каталога ~90к айтемов
- 2) Я с коллегами выиграл (1/~20 место) хакатон по рекомендациям от Ленты, просто обучив EASE
- 3) Мы с коллегами заняли 2/~20 мест в хакатоне Цифровой Прорыв, обучив EASE

Недостатки EASE

- 1) $O(I^3)$ – сложность вычисления обратной матрицы. Убрать непопулярные айтемы. Используйте scipy!
- 1) Не учитывает последовательность данных
- 2) Не работает с негативными действиям от пользователя (просмотр без клика)

Результаты EASE

Revisiting the Performance of iALS on Item Recommendation Benchmarks

Steffen Rendle
Google
Mountain View, USA
srendle@google.com

Li Zhang
Google
Mountain View, USA
liqzhang@google.com

Walid Krichene
Google
Mountain View, USA
walidk@google.com

Yehuda Koren
Google
Haifa, Israel
yehuda@google.com

Table 2: Quality results on the ML20M and MSD benchmark sorted by Recall@20 scores.

Dataset	Method	Recall@20	Recall@50	NDCG@100	Result from
ML20M	RecVAE [25]	0.414	0.553	0.442	[25]
	H+Vamp (Gated) [14]	0.413	0.551	0.445	[14]
	RaCT [18]	0.403	0.543	0.434	[18]
	Mult-VAE [17]	0.395	0.537	0.426	[17]
	LambdaNet [4]	0.395	0.534	0.427	[25]
	iALS	0.395	0.532	0.425	our result
	EASE [26]	0.391	0.521	0.420	[26]
	CDAE [28]	0.391	0.523	0.418	[17]
	Mult-DAE [17]	0.387	0.524	0.419	[17]
	SLIM [19]	0.370	0.495	0.401	[17]
	iALS	0.363	0.502	0.393	prev. result [13]
	iALS	0.360	0.498	0.386	prev. result [17]
	WARP [27]	0.314	0.466	0.341	[25]
MSD	Popularity	0.162	0.235	0.191	[26]
	EASE [26]	0.333	0.428	0.389	[26]
	iALS	0.309	0.415	0.368	our result
	RecVAE [25]	0.276	0.374	0.326	[25]
	RaCT [18]	0.268	0.364	0.319	[18]
	Mult-VAE [17]	0.266	0.364	0.316	[17]
	Mult-DAE [17]	0.266	0.363	0.313	[17]
	LambdaNet [4]	0.259	0.355	0.308	[25]
	iALS	0.211	0.312	0.257	prev. result [17]
	WARP [27]	0.206	0.302	0.249	[25]
	CDAE [28]	0.188	0.283	0.237	[17]
	Popularity	0.043	0.068	0.058	[26]

У EASE есть развитие

- 1) $O(l^3)$ – сложность вычисления обратной матрицы. Есть модификации SARSA, ELSA, там лучше
- 2) Не учитывает последовательность данных. Есть модификация SLIST
- 3) Не работает с негативными действиям от пользователя (просмотр без клика)

Спасибо за
внимание