

## Рекомендательные системы (RecSys)

$$\text{Уточ} = \text{Округл} (0.3 \cdot R_z + 0.25 \cdot T + 0.15 \cdot P + 0.3 \cdot J)$$

$R_z \times 2$        $P \times 1$   
 $T \times 15$        $J \times 1$

$$\text{Наклон} = \text{Округл} (0.3 \cdot R_z + 0.25 \cdot T + 0.15 \cdot P) / 0.7$$

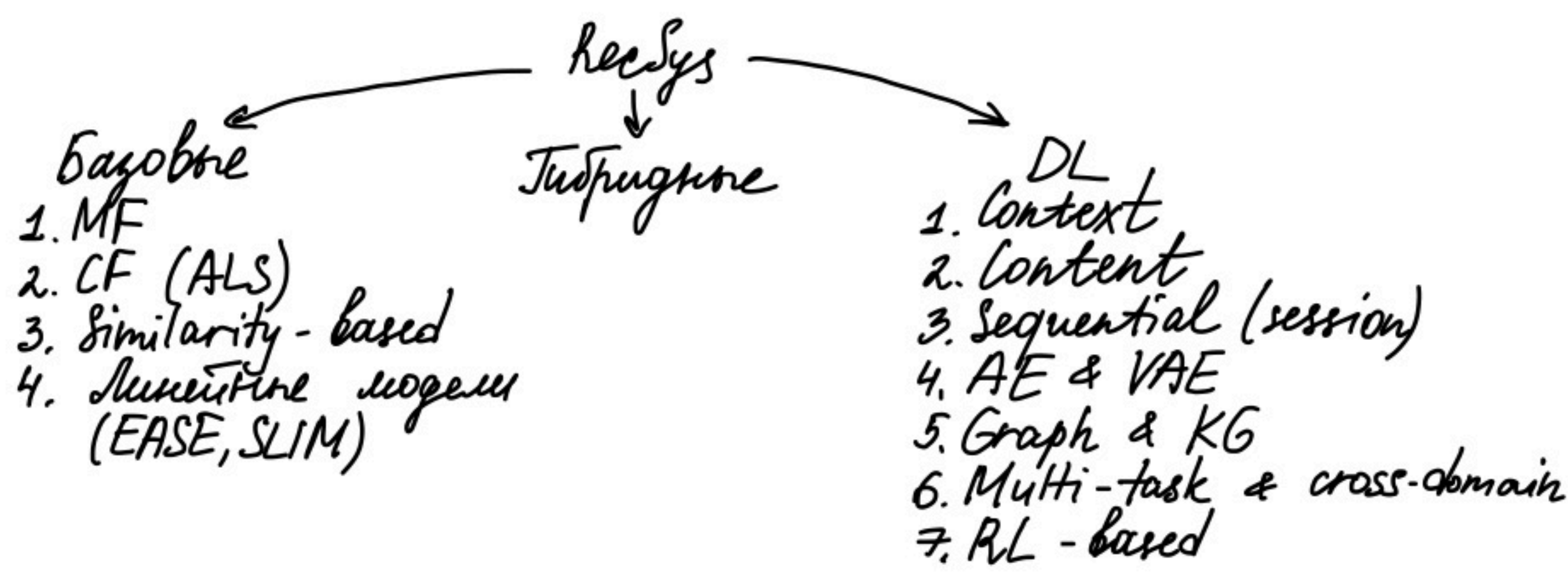
Если  $\text{Наклон} > m$

$$T > n \quad (m = n = 8) - \text{предварительно}$$

16 лекций, 16 семинаров

План:

- Формализация задачи, метрики качества + функции ранжирования
- Занятия 2-12 Подходы в RecSys



• Extra:

- Интерпретируемость
- LLM в RecSys
- A/B тесты online & offline
- API сервис

## Ранжирование как ML задача

В общей постановке:

$X$  - мн-во айтелов (= объектов рекомендаций)

$X^L = \{x_1, x_2, \dots, x_L\}$  - обучающая выборка

$i < j$  на парах  $(i, j) \in \{1, \dots, L\}^2$

$a: X \rightarrow \mathbb{R}$ , что  $i < j \Leftrightarrow a(x_i) < a(x_j)$

$I$  - мн-во айтелов,  $i \in I$

1) totality (полнота)

$$\forall i, j \in I: i \neq j \Rightarrow i \succsim j \vee j \succsim i$$

2) антисимметричность

$$\forall i, j \in I: i \succsim j \wedge j \succsim i \Rightarrow i = j$$

3) транзитивность

$$\forall i, j \in I: i \succsim j \wedge j \succsim k \Rightarrow i \succsim k$$

свойства  
отношения  
линейного  
порядка  
 $\succsim \subset I^2$

# I Поиск (ранжирование выдачи)

query  $\rightarrow d_1, d_2, \dots, d_n \xrightarrow{\quad} d_2 \quad 0.98$   
 $\mathcal{D} = \{(q, d, y)\}$   $y$  - оценки релевантности  $d_1 \quad 0.96$   
 $y \in [0, 1]$  (и.б. и  $y \in \mathbb{R}$ )  $\vdots$   
 $d_n \quad 0.1$   
 $(q, d) \rightarrow p(y=1)$

$\mathcal{D}$  - мн-во док-в

$Q$  - мн-во запросов

$\mathcal{D}_q \subseteq \mathcal{D}$  - мн-во док-в к запросу  $q$

$X = Q \times \mathcal{D}$  - пар (q, d)

$x \equiv (q, d), q \in Q, d \in \mathcal{D}_q$

$Y$  - упоряд. мн-во рейтингов (оценки релевант-сти)

Е.г: CTR/клики, ассессорная разметка

$y: X \rightarrow Y$

$(q, d_1) < (q, d_2) \Leftrightarrow y(q, d_1) < y(q, d_2)$

# II Рекомендации пользователям

$\mathcal{U}$  - мн-во пользователей,  $u \in \mathcal{U}$

$I$  - мн-во айтелов,  $i \in I$

$X = \mathcal{U} \times I$  - пар (пользователь, айтел)

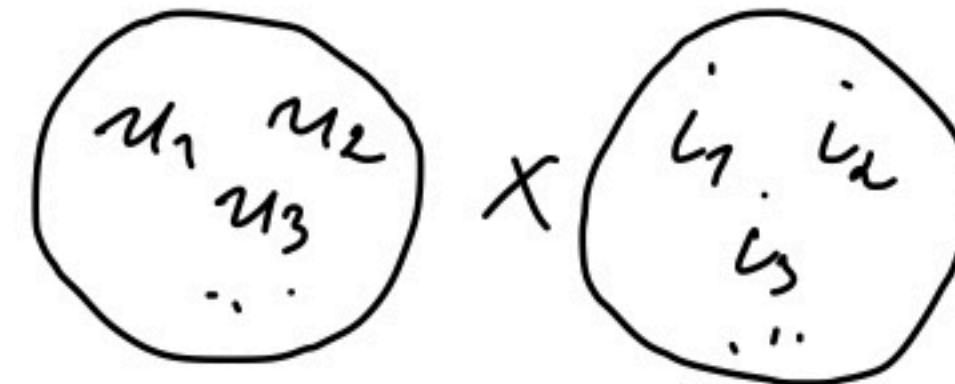
$(u, i_1) < (u, i_2) \Leftrightarrow y(u, i_1) < y(u, i_2)$

$\forall u \in \mathcal{U}: \{y(u, i_1), y(u, i_2), \dots, y(u, i_k)\}$

$|\mathcal{U}| = m$

где  $k \leq n$

$|I| = n$



каждому  
пользователю  
выдам топ-k  
рекомендаций

Данные  $\rightarrow$  explicit feedback  
 $\rightarrow$  implicit feedback

Матрица взаимодействий users x items

	$i_1$	$i_2$	$i_3$
$u_1$	3	0	4
$u_2$	0	1	0
$u_3$	5	0	0

explicit

implicit

е.г.  
(оценки  
пользов)

е.г.  
(число  
просмотров)

	$i_1$	$i_2$	$i_3$
$u_1$	1	0	1
$u_2$	0	1	0
$u_3$	1	0	0

explicit

implicit

е.г.  
(лайк/  
дизлайк)

е.г.  
(клик/  
не клик)

$r(u, i)$  - оценка релевант-сти  
/ результат взаимодействия  
польза  $u$  с айтелом  $i$

проблема - разреженность матрицы

тип итогового списка топ-n рек-ий:

- 1) новинки
- 2) только новые
- 3) микс



# Функции ранжирования

1. point-wise
2. pair-wise
3. list-wise

Примеры 2, 3: BPR, WARP, YotiRank

Метрики оценки качества ранжирования

$y_i$  - истинная релевантность  
 $i$ -го айтема для пользователя,  $y_i \in \mathcal{Y}$   
 $\hat{y}_i$  - предсказание релевантности

Алгоритм:  
 1.  $\forall (u, i)$  получить  $\hat{y}_i, i \in \{1, 2, \dots, |I|\}$   
 2. Отсортировать все  $\hat{y}_i$   
 (по убыванию)  
 3. Отобрать топ  $k$  в рекомендации  
 \* есть способ получать сразу только  $k$  предсказаний (например, через ANN)

Пусть  $y_i \in \{0, 1\}, \hat{y}_i \in \{0, 1\}$  (либо м.б.  $y_i \in \mathbb{R}$  и/или  $\hat{y}_i \in \mathbb{R}$ )

1) Hitrate =  $\frac{\#hits}{\#hits + \#misses}$

1.1. Hitrate@k =  $\frac{\sum_{i=1}^k [y_i = 1]}{|Y_k|}$  - для 1 юзера  
 ~ accuracy (доля правильных ответов)

1.2. На юзере hr@k=1, если хотя бы 1 правильный, k-длина списка рек-и  
 Hitrate@k =  $\frac{1}{|u|} \sum_{u \in U} [\sum_{i=1}^k [y_{ui} = 1] > 0]$  - доля юзеров, у кого хотя бы 1 рек-и правильная

2) Precision@k =  $\frac{\sum_{i=1}^k [y_i = 1]}{k}$

3) Recall@k =  $\frac{\sum_{i=1}^k [y_i = 1]}{|Y|}$

	1	2	3	4	5	6	7	
$u_1$								- нерелевантный
$u_2$								- релевантный
$u_3$								

$|I| = 7$

$u_1$

$p@1 = 0$

$p@3 = \frac{1}{3}$

$p@5 = \frac{2}{5}$

$r@1 = 0$

$r@3 = \frac{1}{3}$

$r@5 = \frac{2}{5}$

для выборки

$hr@3 = 1$

$p@3 = \frac{1}{|u|} \sum_{u \in U} p@k_u = \frac{\frac{1}{3} + \frac{2}{3} + \frac{2}{3}}{3} = \frac{5}{9}$

$r@3 = \frac{1}{|u|} \sum_{u \in U} r@k_u = \frac{\frac{1}{3} + \frac{2}{3} + \frac{2}{3}}{3} = \frac{5}{9}$

Плюсы:

- + интерпретируемость
- + простота
- + применимо и к бинар. классификации
- + работает и на юзере, и на всей выборке

Минусы:

- чувствительность к порогу  $k$
- не учитывают порядок
- бинаризуем оценки релевантности

#### 4) Average Precision (AP@k)

$$AP@k = \sum_{i=1}^k \frac{[y_i=1]}{|Y_k|} \cdot p@i$$

#### 5) Mean Average Precision (MAP@k)

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

1	2	3	4	5

$$AP@5 = 1$$

p@i	1	1	$\frac{1}{3}$	$\frac{2}{4}$	$\frac{2}{5}$
-----	---	---	---------------	---------------	---------------

y <sub>i</sub>	1	1	0	0	0
----------------	---	---	---	---	---

1	2	3	4	5

p@i	0	0	$\frac{1}{3}$	$\frac{2}{4}$	$\frac{3}{5}$
-----	---	---	---------------	---------------	---------------

y <sub>i</sub>	0	0	1	1	1
----------------	---	---	---	---	---

$$AP@5 = \frac{1 \cdot \frac{1}{3} + \frac{1}{2} + \frac{3}{5}}{3} \approx 0.44$$

Плюс: + учитывает порядок

Минус:  $|Y_u| \leq k \Rightarrow$  метрика занижается при большой k

$$MNAP@k = \frac{1}{|U|} \sum_{u \in U} \frac{1}{\min(n_u, k)} \cdot \sum_{i=1}^k y_i \cdot p@i_u$$

	1	2	3	4	5	6	7
u <sub>1</sub>							
u <sub>2</sub>							
u <sub>3</sub>							

где  $n_u$  - число айтемов, с которыми юзер взаимодействовал (вопрос с семинара:)

$$MAP@3 = \frac{1}{3} \cdot \left( \frac{1}{2} + \frac{5}{6} + \frac{7}{12} \right) = \frac{13}{36} \approx 0.64$$

$$MNAP@3 = \frac{1}{3} \cdot \left( \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{5}{6} + \frac{1}{2} \cdot \frac{7}{6} \right) = \frac{5}{9} \approx 0.56$$

#### 6) Discounted cumulative gain (DCG@k)

$$DCG@k = \sum_{i=1}^k g(y_i) \cdot d(i) = \sum_{i=1}^k \frac{2^{y_i} - 1}{\log(i+1)}$$

↑ функция  
гиперосна  
↑ коэффициент  
discount

$$NDCG@k = \frac{DCG@k}{IDCG@k} - \text{это идеальное ранжирование (перестановка, где все релевантные айтемы в топ k)}$$