

Рекомендательные системы

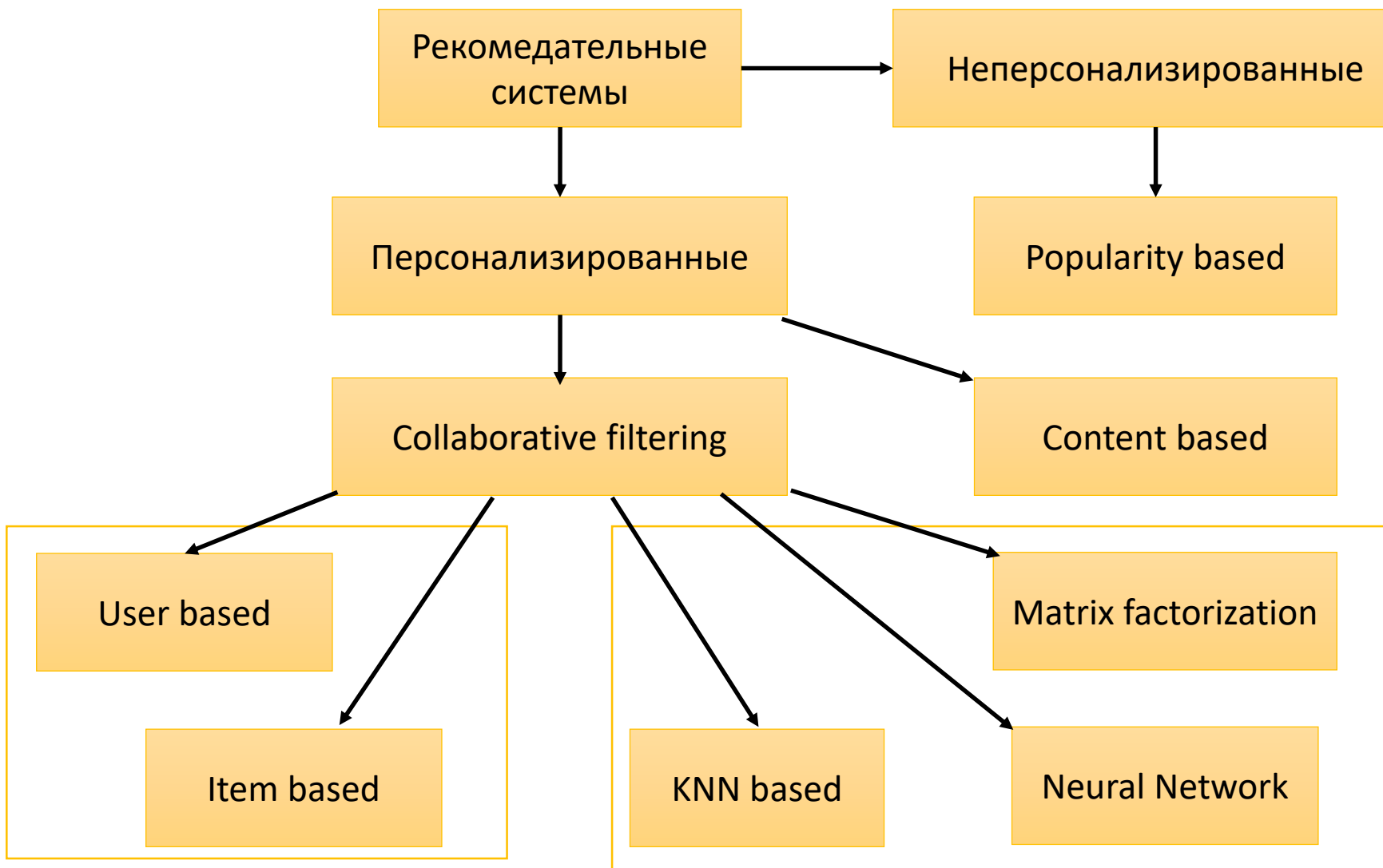
Басалов Юрий
руководитель группы
разработки компании DDPlanet



Рекомендательные системы

- Это подкласс системы фильтрации информации, которая стремится предсказать «рейтинг» или «предпочтение», которое пользователь дал бы какому-то объекту.
- Пользователь непосредственно не взаимодействует с такими системами.
- Метрики, обычно, оценивают точность предсказания «рейтинг» или «предпочтение».
- Основная бизнес-цель - формирование и оптимизация предложений продуктов для пользователей. Другие бизнес-цели лежат за пределами рекомендательных систем.

Основные подходы



Explicit vs Implicit

Explicit (явная) оценка пользователя:

- рейтинг продукта
- сложно получить
- оценка не объективна (в том числе и
рамках одного пользователя)

Implicit (скрытая) оценка пользователя:

- история просмотров, покупок и тд
- легко получить
- много ложно положительно данных
- дискретная



Explicit метрики

- Mean Absolute Error (средняя ошибка)

$$\text{MAE} = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

- Root Mean Squared Error
(среднеквадратичная ошибка)

$$\text{RMSE} = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}$$



Implicit метрики

- Precision (точность)

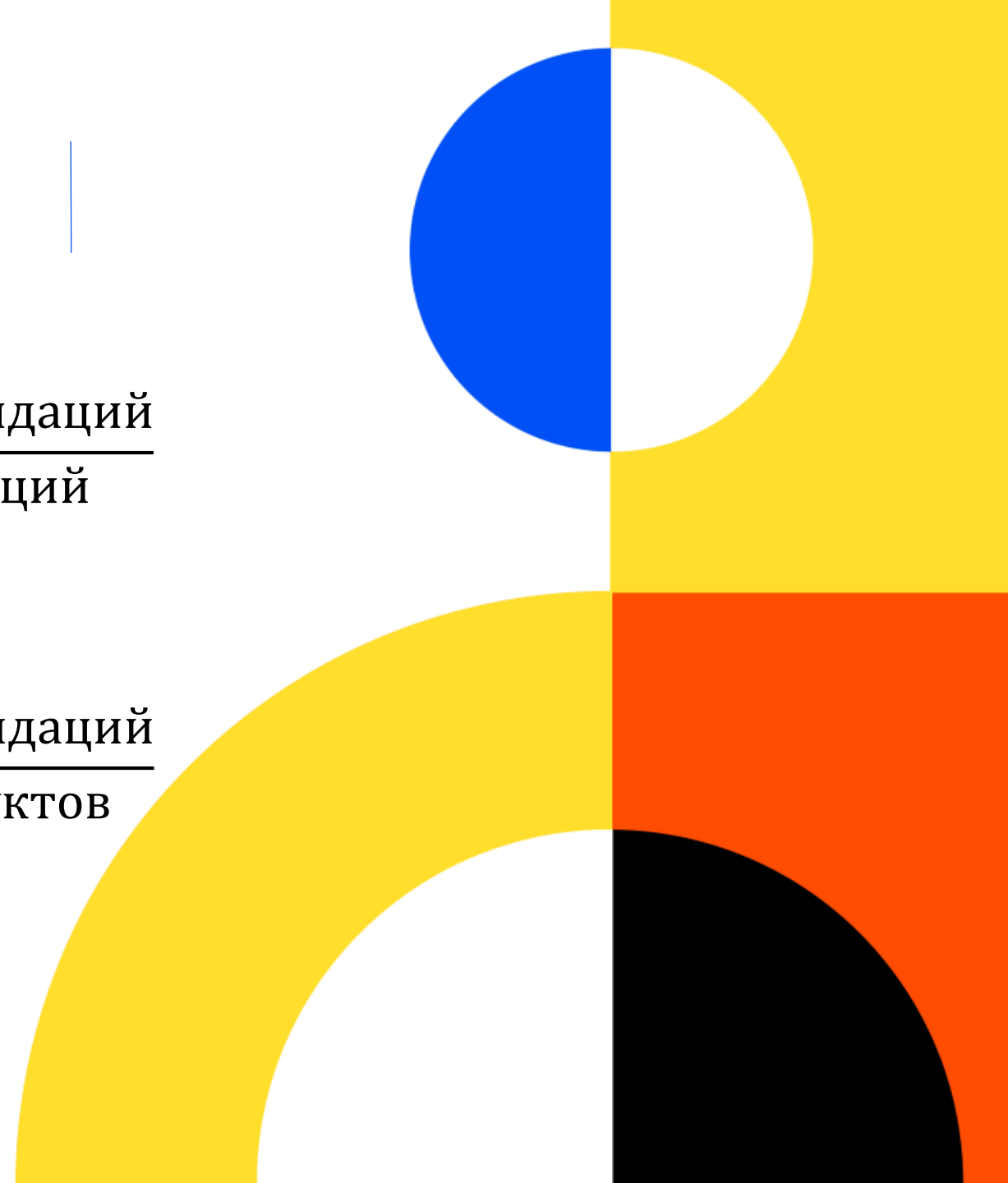
$$P = \frac{\text{кол — во выбранных рекомендаций}}{\text{общее кол — во рекомендаций}}$$

- Recall (точность)

$$R = \frac{\text{кол — во выбранных рекомендаций}}{\text{кол — во выбранных продуктов}}$$

- F-мера

$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$



Implicit метрики

- Average Precision (AP)

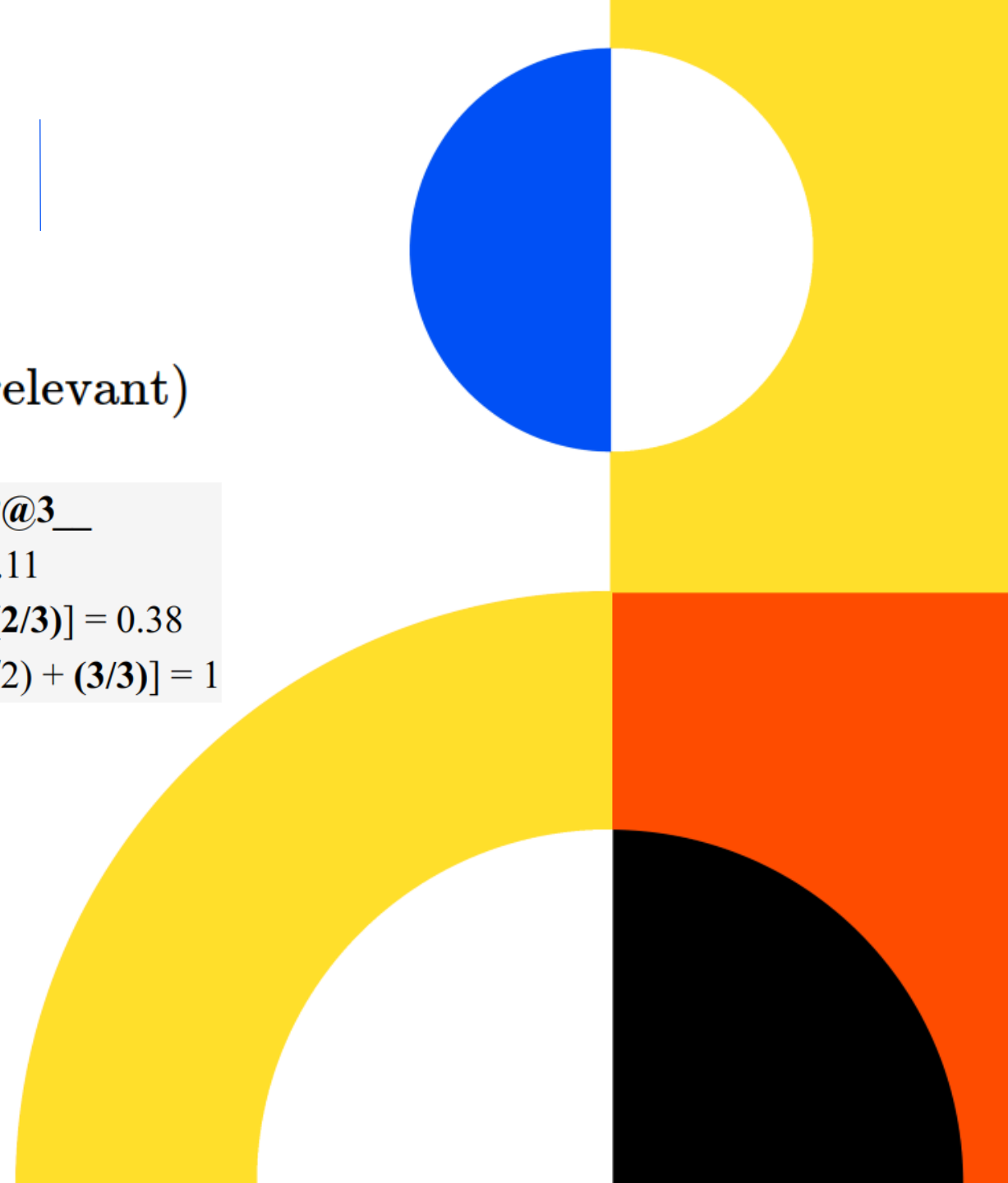
$$AP@N = \frac{1}{m} \sum_{k=1}^N (P(k) \text{ if } k^{th} \text{ item was relevant})$$

Recommendations	Precision @k's	AP@3
[0, 0, 1]	[0, 0, 1/3]	(1/3)(1/3) = 0.11
[0, 1, 1]	[0, 1/2, 2/3]	(1/3)[(1/2) + (2/3)] = 0.38
[1, 1, 1]	[1/1, 2/2, 3/3]	(1/3)[(1) + (2/2) + (3/3)] = 1

- Mean Average Precision (MAP)

$$MAP@N = \frac{1}{|U|} \sum_{u=1}^{|U|} (AP@N)_u$$

- Аналогично вводятся AR и MAR



Общие метрики

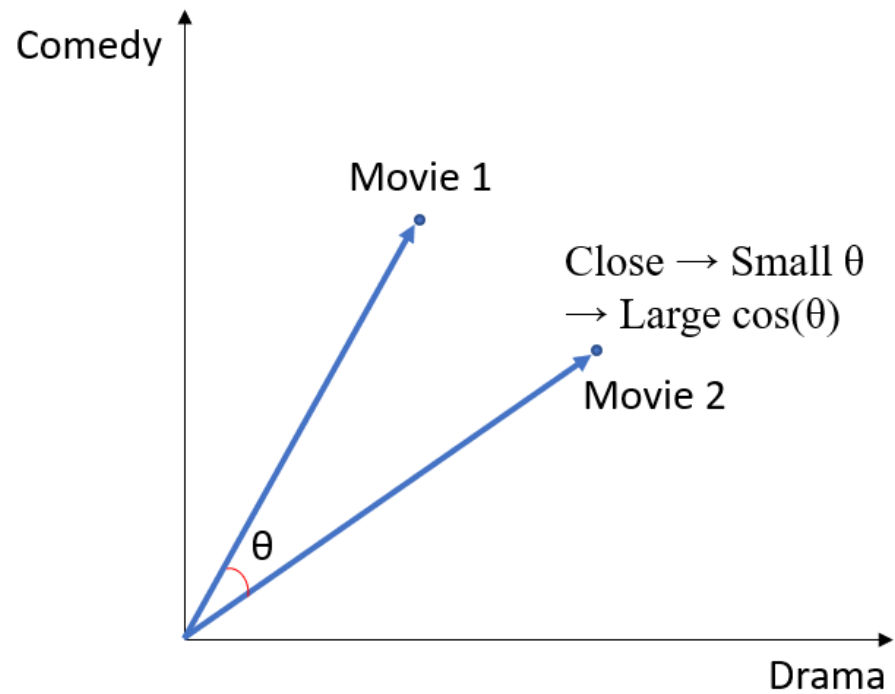
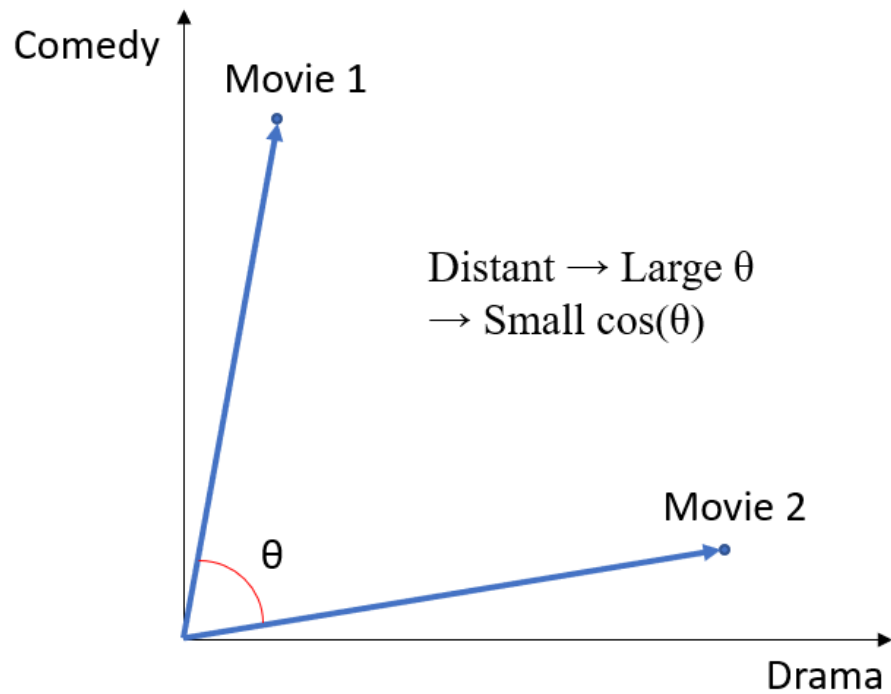
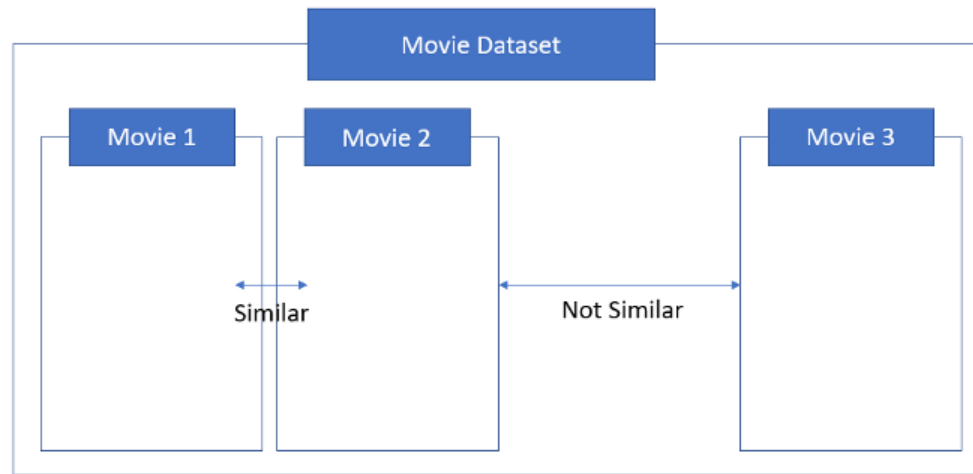
- Coverage (покрытие) - доля продуктов, которые будут рекомендованы
- Novelty (новизна) - степень того, насколько часто пользователям будут рекомендоваться продукты, которые им неизвестны

$$novelty = \frac{1}{|U|} \sum_{\forall u \in U} \sum_{\forall i \in topN} \frac{\log_2(\frac{count(i)}{|U|})}{|N|}$$

- Personalization (персонализация) - насколько рекомендации для разных пользователей различаются

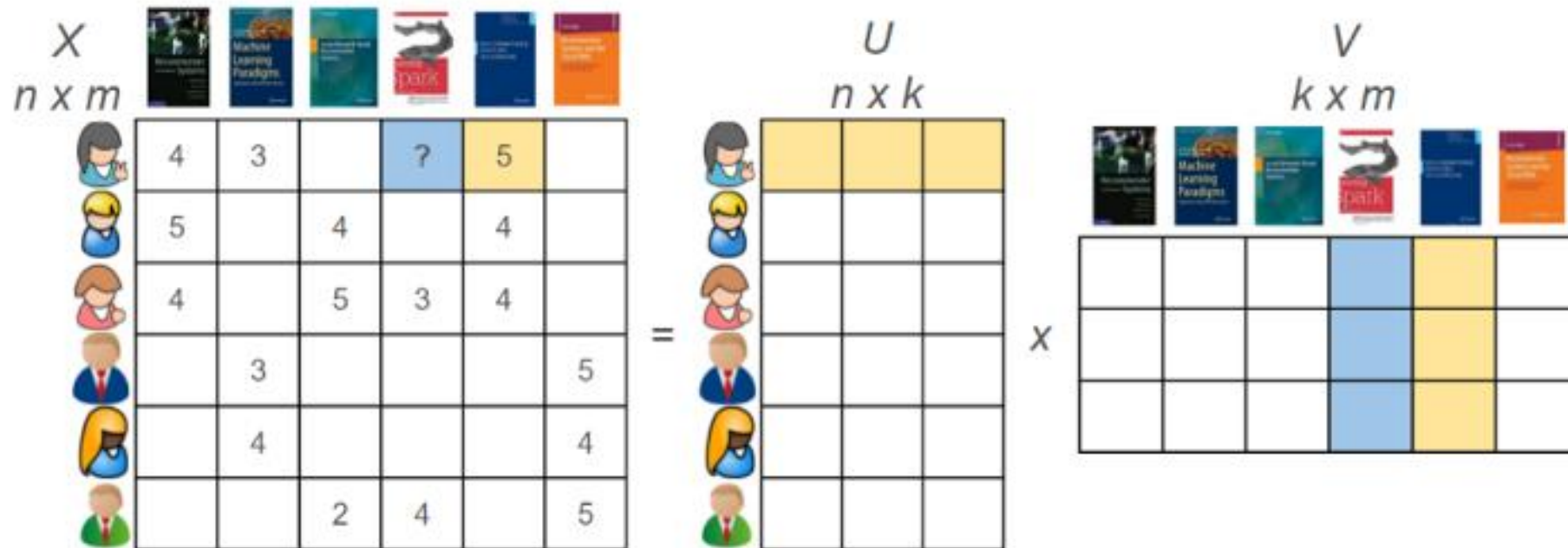
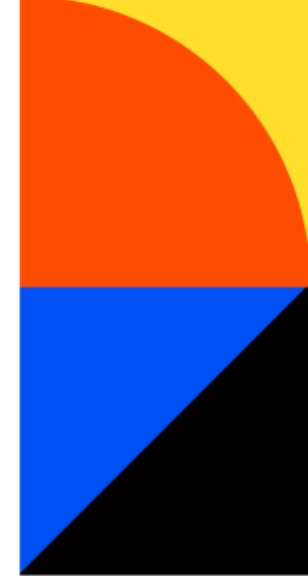


Content based



Collabrative filtering

Matrix
factorization



- LU, SVD разложения
- библиотеки Surprise, Turicreate

Collabrative filtering | KNN

