

# Автоматическое дополнение плейлистов в рекомендательной системе пользователей

Кислинский Вадим Геннадьевич

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам  
(практика, В. В. Стрижов)/Группа 574, весна 2018

## Цель

Решить задачу top-N рекомендаций

## Метод

Матричная факторизация, учитывающая дополнительные знания о плейлистах и треках

## Обзор классических методов

- 1 Paolo Cremonesi al. *Performance of recommender algorithms on top-n recommendation tasks*. 2010

## Метод совместной факторизации

- 2 Dimitrios Rafailidis al. *Modeling the Dynamics of User Preferences in Coupled Tensor Factorization*. 2013

## Алгоритм LCE

- 3 Martin Saveski al. *Item Cold-Start Recommendations: Learning Local Collective Embeddings*. 2014

## Дано

- 1  $U$  - множество из  $n$  плейлистов,
- 2  $I$  - множество из  $m$  треков,
- 3  $D = \{(u, i) | u \in U, i \in I\}$  - множество транзакций,
- 4  $R$  - матрица  $n \times m$ , где  $R_{ui} = 1$ , если  $(u, i) \in D$ ,
- 5  $X_U$  - матрица  $n \times v$  признакового описания плейлистов,
- 6  $X_I$  - матрица  $m \times w$  признакового описания треков

## Задача

Для плейлиста  $u$  построить вектор  $\mathbf{r}$  из  $m$  элементов, которые означают насколько треки подходят данному плейлисту

## Оптимизационная задача

$$\begin{aligned} \arg \min_{\mathbf{W}, \mathbf{H}_I, \mathbf{H}_U} & (\alpha \|\mathbf{X}_U - \mathbf{W}\mathbf{H}_U\|^2 + (1 - \alpha) \|\mathbf{R} - \mathbf{W}\mathbf{H}_I\|^2 + \\ & + \lambda (\|\mathbf{W}\|_F^2 + \|\mathbf{H}_U\|_F^2 + \|\mathbf{H}_I\|_F^2)) \\ \text{s.t. } & \mathbf{W} \geq 0, \mathbf{H}_U \geq 0, \mathbf{H}_I \geq 0 \end{aligned}$$

## Вычисляем $\mathbf{r}$ для пользователя $u$

Пусть  $\mathbf{x}$  - признаковое описание плейлиста.

Решим систему  $\mathbf{x} = \mathbf{H}_U^T \mathbf{w}$  относительно  $\mathbf{w}$ , и определим

$$\mathbf{r} = \mathbf{H}_I^T \mathbf{w}$$

## Метрики качества

$R$  - список top-N рекомендаций,  $G$  - список настоящих треков плейлиста

$$R@call = \frac{|G \cap R_{1:|G|}|}{|G|}$$

$$Presicion = \frac{|G \cap R_{1:|G|}|}{|R_{1:|G|}|}$$

## Данные

Выборка из 20000 плейлистов, содержащая 265464 различных треков, количество транзакций - 1302790.

## Базовый алгоритм - PureSVD

Неизвестные значения матрицы  $\mathbf{R}$  заполняются нулями и делается SVD разложение, полученной матрицы.

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{\Sigma}\mathbf{Q}^T$$

$$\mathbf{R} = \mathbf{P}\mathbf{Q}^T, \mathbf{P} = \mathbf{U}\mathbf{\Sigma} = \mathbf{R}\mathbf{Q}$$

$$\hat{r}_u = \mathbf{r}_u \mathbf{Q}\mathbf{Q}^T$$

# Результаты эксперимента

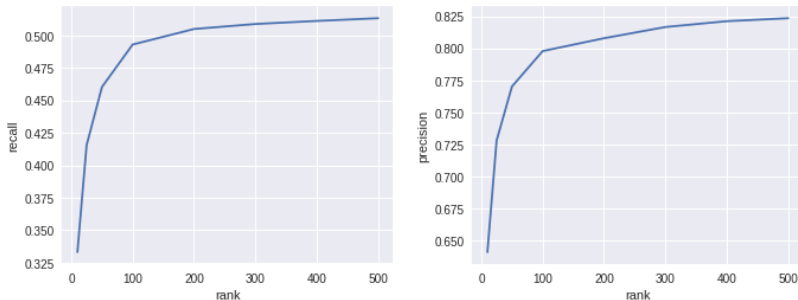


Рис.: Зависимость метрики от ранга разложения.