

# Автоматическое дополнение плейлистов в рекомендательной системе\*

*Кислинский В. Г., Фролов Е., Воронцов К. В.*

kislinskiy.vg@phystech.edu; evgeny.frolov@skolkovotech.ru; vokov@forecsys.ru

Московский физико-технический институт

Работа посвящена исследованию метода совместной матричной факторизации в задаче top-N рекомендаций для автоматического продолжения плейлистов. Предлагается модель матричной факторизации, учитывающая дополнительную информацию о плейлистах и треках. Данный метод будет иметь, не только преимущество алгоритмов коллаборативной фильтрации, которые способны выявить скрытые свойства пользователей и объектов, но также сможет учитывать контекстную информацию. В данном методе будет введена дополнительная регуляризация, основанная на предположении, что если объекты близки в пространстве признаков, то они также близки и в латентном факторном пространстве. Для анализа качества представленного алгоритма проводятся эксперименты на выборке из миллиона плейлистов MPD.

**Ключевые слова:** задача top-N рекомендаций, совместная матричная факторизация, алгоритм LSE, латентное факторное пространство, коллаборативная фильтрация.

## 1 Введение

Существует два основных подхода к задаче автоматического продолжения плейлистов: музыкальный информационный поиск и рекомендательные системы[1]. В данной работе предлагается один из методов рекомендательных систем. Рекомендательные системы имеют несколько основных техник, которые можно разделить на две основные группы: коллаборативная фильтрация и content-based подход. Content-based системы работают на основе свойств объектов, оценивают похожесть объектов, создают на основе этого профили пользователей и рекомендуют им наиболее подходящие к профилю объекты. Коллаборативная фильтрация использует историю действия пользователей в системе для получения новых рекомендаций. Такие методы обладают хорошей устойчивостью и способны выявлять скрытые свойства объектов, что улучшает релевантность рекомендаций. Основным недостатком этого подхода является проблема холодного старта - неспособность строить рекомендации для новых пользователей и рекомендовать новые объекты.

В данной работе предполагается, что использование свойств объектов и пользователей улучшит качество рекомендаций, получаемых методами коллаборативной фильтрации и решит проблему холодного старта. Будет исследована модель совместной матричной факторизации, подобный подход предлагается в статье[2], где матрица признаков объектов факторизуются вместе с матрицей "рейтингов". Существуют различные подходы учета дополнительной информации в модели матричной факторизации, введение дополнительной регуляризации[3], совместная факторизация с матрицами похожести объектов и пользователей[4].

## 2 Постановка задачи

Задано множество треков  $\mathcal{T} = \{t_i\}_{i=1}^m$  и множество плейлистов  $\mathcal{P} = \{p_i\}_{i=1}^n$ , где  $p_i \in \mathcal{T}$ ,  $|p_i| < |T|$ . Каждый трек описывается исполнителем и альбомом, обозначим  $\mathcal{S} = \{s_i\}_{i=1}^k$

Работа выполнена при финансовой поддержке РФФИ, проект № 00-00-00000. Научный руководитель: Воронцов К. В. Консультант: Фролов Е.

- множество исполнителей,  $\mathcal{L} = \{l_i\}_{i=1}^d$  - множество альбомов, а каждый плейлист имеет название,  $\mathcal{A} = \{a_i\}_{i=1}^r$  - множество названий. Определим матрицу  $\mathbf{R} \in \mathbb{R}^{n \times m}$ , следующим образом  $\mathbf{R}_{ij} = 1$ , если  $t_j \in p_i$ , иначе ноль, также зададим бинарную матрицу  $\mathbf{X} \in \mathbb{R}^{n \times (k+d+r)}$ , которая описывает треки, каких авторов содержатся в плейлисте -  $k$  первых столбцов, каких альбомов  $d$  следующих столбцов, имя плейлиста -  $r$  последних столбцов. Надо для нового плейлиста  $p$  найти  $N$  наиболее подходящих треков, для этого будет строиться вектор  $\mathbf{r} \in \mathbb{R}^{m \times 1}$ ,  $i$  - ый элемент которого означает насколько трек  $t_i$  подходит плейлисту  $p$ .

### 3 Описание метода

Задача матричной факторизации заключается в нахождение двух матриц меньшей размерности, произведение которых приближает исходную. В задаче рекомендаций ищется приближение матрицы рейтингов, в наших терминах  $\mathbf{R} \approx \mathbf{UV}$ , где  $\mathbf{U}$  - матрица профилей плейлистов,  $\mathbf{V}$  - матрица профилей треков. Предполагая, что профили плейлистов зависят от того какие исполнители, альбомы входят в плейлист, какие названия у плейлистов, можно записать  $\mathbf{X} \approx \mathbf{UH}$ , где  $\mathbf{H}$  - матрица профилей авторов, альбомов, названий. Таким образом приходим к следующей задаче оптимизации:

$$U, V, H = \arg \min_{\mathbf{U}, \mathbf{V}, \mathbf{H}} \alpha \|\mathbf{R} - \mathbf{UV}\|_F^2 + (1 - \alpha) \|\mathbf{X} - \mathbf{UH}\|_F^2 + \lambda (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 + \|\mathbf{H}\|_F^2),$$

$$s.t. \mathbf{U} \geq 0, \mathbf{V} \geq 0, \mathbf{H} \geq 0$$

где  $\lambda$  положительный коэффициент регуляризации, а  $\alpha \in [0, 1]$ , если  $\alpha > 0.5$ , то матрица  $U$  больше зависит от распределения треков по плейлистам, если  $\alpha < 0.5$ , то наоборот, предполагается, что  $U$  зависит от распределения авторов, альбомов по плейлистам и названий плейлистов.

#### 3.1 Введение дополнительной регуляризации на основе близости плейлистов в пространстве признаков

Близость плейлистов в пространстве признаков можно оценивать евклидовым расстоянием. Составим граф близости плейлистов, в котором вершинами будут плейлисты, каждый плейлист соединим ребром с  $v$  ближайшими плейлистами. Матрица  $\mathbf{A}$  - матрица смежности графа близости плейлистов. Предполагая, что если плейлисты близки в реальном пространстве признаков, то они также близки в пространстве профилей, можно записать сумму, которую надо минимизировать:

$$S = \frac{1}{2} \sum_{i,j=1}^n \|u_i - u_j\|^2 A_{ij} = \sum_{i=1}^n (u_i^T u_i) D_{ii} - \sum_{i,j=1}^n (u_i^T u_j) A_{ij} = Tr(\mathbf{U}^T \mathbf{D} \mathbf{U}) - Tr(\mathbf{U}^T \mathbf{A} \mathbf{U}) = Tr(\mathbf{U}^T \mathbf{L} \mathbf{U})$$

где  $\mathbf{D}$  диагональная матрица, где на  $ii$  месте стоит сумма  $i$  строки матрицы  $\mathbf{A}$ ,  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ . Перепишем (1), учитывая  $S$ :

$$U, V, H = \arg \min_{\mathbf{U}, \mathbf{V}, \mathbf{H}} \alpha \|\mathbf{R} - \mathbf{UV}\|_F^2 + (1 - \alpha) \|\mathbf{X} - \mathbf{UH}\|_F^2 + \beta Tr(\mathbf{U}^T \mathbf{L} \mathbf{U}) + \lambda (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 + \|\mathbf{H}\|_F^2)$$

$$s.t. \mathbf{U} \geq 0, \mathbf{V} \geq 0, \mathbf{H} \geq 0$$

## **3.2 Оптимизационная проблема**

## **3.3 Получение рекомендаций**

## **4**

## **5 Базовый вычислительный эксперимент**

### **Литература**

- [1] Geoffray Bonnin, Dietmar Jannach. Automated Generation of Music Playlists: Survey and Experiments. ACM Computing Surveys (CSUR). 2014
- [2] Martin Saveski, Amin Mantrach. Item Cold-Start Recommendations: Learning Local Collective Embeddings. Proceeding RecSys '14 Proceedings of the 8th ACM Conference on Recommender systems Pages 89-96. 2014
- [3] Yifan Chen, Xiang Zhao. Leveraging High-Dimensional Side Information for Top-N Recommendation. CoRR. 2017
- [4] Cold-Start Item and User Recommendation with Decoupled Completion and Transduction Iman Barjasteh, Rana Forsati, Farzan Masrour, Abdol-Hossein Esfahanian, Hayder Radha. Cold-Start Item and User Recommendation with Decoupled Completion and Transduction. 2015