

Рекомендательные системы

Сергей Казенюк

СПбАУ/МИТ

Рекомендательные системы

	Product 1	Product 2	Product 3	Product 4
User 1	3	1	?	5
User 2	?	?	3	5

Предсказываем:

- Оценку продукта 3 пользователем 1
- Оценку продукта 1 и 2 пользователем 2

Рекомендуем пользователю то, что ему
(предположительно) нравится.

Постановка задачи

- Рекомендательная система на основе методов коллаборативной фильтрации (оффлайн-модели):
 - GroupLens
 - k-NN GroupLens (k nearest neighbors)
 - SVD (Singular Value Decomposition)
- Ранжированный список рекомендуемых пользователям продуктов

GroupLens

Рекомендации, основанные на предпочтениях «похожих» пользователей:

$$\hat{r}_{i,a} = \bar{r}_a + \frac{\sum_j (r_{j,a} - \bar{r}_j) w_{i,j}}{\sum_j |w_{i,j}|} \quad (1)$$

Метрика похожести пользователей $w_{i,j}$:

- Коэффициент корреляции Пирсона

$$w_{i,j} = \frac{\sum_a (r_{i,a} - \bar{r}_a)(r_{j,a} - \bar{r}_a)}{\sqrt{\sum_a (r_{i,a} - \bar{r}_a)^2} \sqrt{\sum_a (r_{j,a} - \bar{r}_a)^2}} \quad (2)$$

Использование только нескольких наиболее похожих пользователей

- k-d Tree

SVD (Singular Value Decomposition)

Разложение матрицы рейтингов:

$$R^{m \times n} = Q^{m \times f} \times P^{f \times n} \quad (3)$$

Базовые предикторы:

$$b_{i,a} = \mu + b_i + b_a \quad (4)$$

Предсказания:

$$\hat{r}_{i,a} = \mu + b_i + b_a + q_a^T p_i \quad (5)$$

$$\begin{aligned} b_*, q_*, p_* = \arg \min & \sum_{(i,a)} (r_{i,a} - \mu - b_i - b_a - q_a^T p_i)^2 \\ & + \lambda \underbrace{\left(\sum_i b_i^2 + \sum_a b_a^2 + \|q_a\|^2 + \|p_i\|^2 \right)}_{\text{Регуляризатор}} \end{aligned} \quad (6)$$

Результаты

Рекомендательная система:

- ➊ Методы коллаборативной фильтрации:
 - GroupLens
 - k-NN GroupLens
 - SVD, SVD++ (GraphLab)
- ➋ k-d Tree (kd-tree++)
- ➌ Ранжированный список рекомендуемых пользователям продуктов

Вопросы?

Спасибо за внимание!