



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
НИЖНИЙ НОВГОРОД

Моделирование предпочтений пользователей в динамических системах

Шашкин П.О.

научный руководитель Карпов Н.В.

Магистерская программа «Интеллектуальный анализ данных».

Научно-учебная группа «Анализ мультимедийных данных»

Разработать алгоритм построения персонализированных рекомендаций на основе неявных предпочтений.

1. Построение списка персонализированных рекомендаций на основе неявных предпочтений¹
2. Использование временной динамики данных в рамках факторизационной модели коллаборативной фильтрации²
3. Достижение необходимых показателей быстродействия

¹top-N recommendation with implicit (one-class) feedback

²context-aware matrix factorization for collaborative filtering

Попарные критерии оптимизации для ранжирования¹

$$L_{\text{BPR}} = \sum_{i,j \in D_u} \ln \sigma(f_u(i) - f_u(j))$$

$$L_{\text{WARP}} = \sum_{i,j \in D_u} \Phi(\text{rank}_u(i)) |1 - f_u(i) + f_u(j)|_+$$

$f_u(i)$ модель предпочтений пользователя

D_u множество пар (i, j) , в которых $f_u(i) > f_u(j)$

σ логистическая функция

$\text{rank}_u(i)$ ранк объекта i в выдаче пользователя u

Φ функция взвешивания

¹pairwise ranking loss

- 600-800 тысяч уникальных пользователей в день
- 200-300 новых публикаций в день
- 2 - 3 миллиона просмотров в день
- 86% просмотров в течении суток после публикации материала

Сравнение критериев оптимизации для данных с новостного потрала за одни сутки

	P@10 ¹	R@10 ²
ALS (reco)	0.0283	0.1503
BPR (libMF)	0.0309	0.1972
WARP (lightFM)	0.0332	0.2018

¹точность на первых 10 элементах списка

²полнота на первых 10 элементах списка

$$f_{u,t}(i) = (v_u + v_t) \cdot v_i + b_i^t$$

v_u — вектор пользователя;

v_t — вектор средних предпочтений пользователей в момент времени;

v_i — вектор новости;

b_i^t — скаляр, отражающий популярность новости в момент времени.

Процедура обучения

Input: (user, item, time) triplets

Initialize v_u, v_i, v_t uniformly and set bias term b_i^t to zero

Let Y - maximum number of sampling trials

repeat

 Pick a random known interaction (u, i, t)

 Let $f_{u,t}(i) = (v_u + v_t) \cdot v_i + b_i^t$

 Set $N = 0$

repeat

 Pick a random negative item j that u didn't interacted with

 Let $f_{u,t}(j) = (v_u + v_t) \cdot v_j + b_j^t$

$N = N + 1$.

until $f_{u,t}(j) > f_{u,t}(i)$ or $N \geq Y - 1$;

if $f_{u,t}(j) > f_{u,t}(i) - 1$ **then**

 Make a gradient step (*AdaGrad*) to minimize:

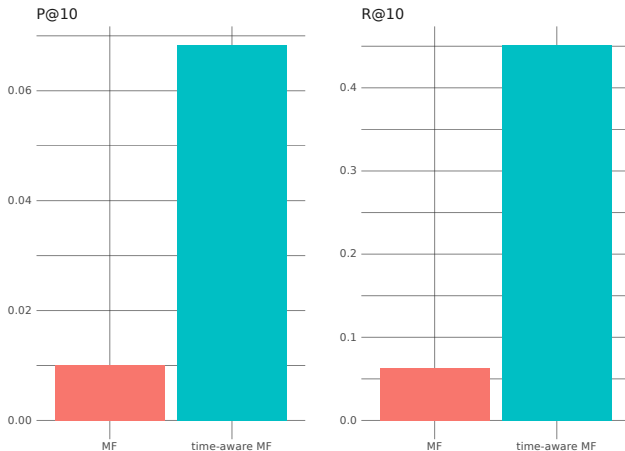
$\ln(\lfloor \frac{Y-1}{N} \rfloor) |1 - f_{u,t}(j) + f_{u,t}(i)|_+$

 Project weights to enforce constraints.

end

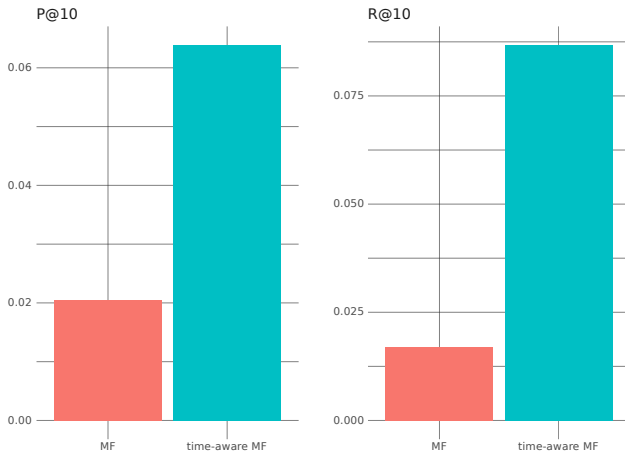
until validation error does not improve;

Сравнение с классической матричной факторизацией на данных с новостного портала за 13 дней¹



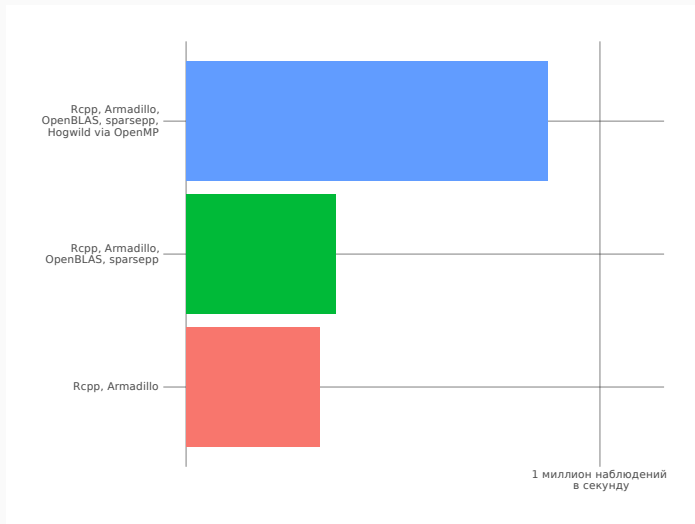
¹Оптимальные значения гиперпараметров: MF - 40 компонент, коэф-т скорости обучения 1.0, 24 эпохи; time-aware MF - 20, 0.5, 26, соответственно.

Сравнение с классической матричной факторизацией на данных о аренде фильмов MovieLens10M¹



¹Оптимальные значения гиперпараметров: MF - 40 компонент, коэф-т скорости обучения 0.25, 14 эпохи; time-aware MF - 20, 0.25, 7, соответственно.

Быстродействие¹



¹Intel® Core™ i7-3520M CPU @ 2.90GHz × 4

Полученное решение:

- ✓ позволяет строить рекомендации на основе просмотров,
- ✓ учитывает временную динамику данных,
- ✓ подходит для большого объёма данных.