Лекция 4.
ALS, NCF.

Миша Печатов

План

- Задача матричной факторизации
- explicit/implicit feedback
- ALS
- Ранжирующие функции
- NCF

Матричная факторизация

Матричная факторизация

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

factor

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Матричная факторизация

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Проблемы MF подходов

Проблемы MF подходов

- Холодный старт
- Вычислительная сложность/масштабируемость
- Учитывает только счетчики интеракций(рейтинги)
- Небольшое разнообразие

Фидбэк пользователя
Implicit vs explicit

Фидбэк пользователя
Implicit vs explicit

- Клик
- Лайк
- Досмотр видео до конца
- Рейтинг
- Покупка
- Добавление в корзину
- Просмотр поста в ленте
- Добавление в избранное

Фидбэк пользователя
Implicit vs explicit

- Клик
- Лайк
- Досмотр видео до конца
- Рейтинг
- Покупка
- Добавление в корзину
- Просмотр поста в ленте
- Добавление в избранное

Фидбэк пользователя
Implicit vs explicit

- Клик
- Лайк
- Досмотр видео до конца
- Рейтинг
- Покупка
- Добавление в корзину
- Просмотр поста в ленте
- Добавление в избранное

Фидбэк пользователя
Implicit vs explicit

- Клик
- Лайк
- Досмотр видео до конца
- Рейтинг
- Покупка
- Добавление в корзину
- Просмотр поста в ленте
- Добавление в избранное

Свойства неявного фидбэка:

1. Отсутствуют негативы
2. Довольно шумный
3. Значение показывает уверенность,

   а не предпочтения
4. Нужны специальные метрики для оценки

Alternating Least Squares

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

factor

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.11	0.32	0.4
0.33	-0.5	-0.2
0.11	-0.2	-0.3
0.71	0.52	-0.4
0.57	0.62	0.99
-0.9	-0.7	0.22
0.42	0.21	0.67

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.11	0.32	0.4
0.33	-0.5	-0.2
0.11	-0.2	-0.3
0.71	0.52	-0.4
0.57	0.62	0.99
-0.9	-0.7	0.22
0.42	0.21	0.67

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.11	0.32	0.4
0.33	-0.5	-0.2
0.11	-0.2	-0.3
0.71	0.52	-0.4
0.57	0.62	0.99
-0.9	-0.7	0.22
0.42	0.21	0.67

n x k

x

item

0.12	0.33	0.42	0.58	0.61
0.13	0.24	0.36	0.12	-0.2
0.94	-0.7	0.21	0.42	0.26

factor

k x m

$$\sum (r_{i,j} - x_u^T \cdot y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Implicit ALS

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

Collaborative filtering for implicit feedback datasets
Y Hu, Y Koren, C Volinsky
2008 Eighth IEEE international conference on data mining, 2008 · ieeexplore.ieee.org

A common task of recommender systems is to improve customer experience through personalized recommendations based on prior implicit feedback. These systems passively track different sorts of user behavior, such as purchase history, watching habits and browsing activity, in order to model user preferences. Unlike the much more extensively researched explicit feedback, we do not have any direct input from the users regarding their preferences. In particular, we lack substantial evidence on which products consumer

РАЗВЕРНУТЬ

Сохранить Цитировать Цитируется: 4343 Похожие статьи Все версии статьи (14)

- Используется в проде во всех крупных компаниях
- Де-факто является индустриальным стандартом
- Отлично параллелится и показывает хорошие метрики

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}$$

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases} \qquad c_{ui} = 1 + \alpha r_{ui}$$

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases} \qquad c_{ui} = 1 + \alpha r_{ui}$$

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

factor

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

k x m

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

item

user

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

n x m

=

factor

user

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

n x k

x

item

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

factor

k x m

$$\min_{x_\star, y_\star} \sum_{r_{u,i} \text{ is known}} (r_{ui} - x_u^T y_i)^2 + \lambda(\|x_u\|^2 + \|y_i\|^2)$$

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

Algorithm 1 The ALS algorithm

```
procedure ALS(R, f, λ; X, Y)
    X ← 0, Y ← random initial guess
    repeat
        for row u ← 1, m do
            x_u ← (Y^T Y + λI)^{-1} Y^T r_u by solving the linear system
                (Y^T Y + λI) x_u = Y^T r_u
        end for
        for column i ← 1, n do
            y_i ← (X^T X + λI)^{-1} X^T r_i by solving the linear system
                (X^T X + λI) y_i = X^T r_i
        end for
    until reaching the maximum iterations
end procedure
```

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

Algorithm 1 The ALS algorithm

```
procedure ALS(R, f, λ; X, Y)
    X ← 0, Y ← random initial guess
    repeat
        for row u ← 1, m do
            x_u ← (Y^T Y + λI)^{-1} Y^T r_u by solving the linear system
                (Y^T Y + λI) x_u = Y^T r_u
        end for
        for column i ← 1, n do
            y_i ← (X^T X + λI)^{-1} X^T r_i by solving the linear system
                (X^T X + λI) y_i = X^T r_i
        end for
    until reaching the maximum iterations
end procedure
```

Отлично параллелится по юзерам/айтемам

+ можем объяснять рекомендации!

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$\hat{p}_{ui} = y_i^T x_u$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$\hat{p}_{ui} = y_i^T x_u = y_i^T (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$\hat{p}_{ui} = y_i^T x_u = y_i^T \underbrace{\boxed{(Y^T C^u Y + \lambda I)^{-1}}}_{W^u} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$\hat{p}_{ui} = y_i^T x_u = y_i^T \underbrace{\boxed{(Y^T C^u Y + \lambda I)^{-1}}}_{W^u} Y^T C^u p(u)$$

$$s_{ij}^u = y_i^T W^u y_j$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. Объяснение рекомендаций

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$\hat{p}_{ui} = y_i^T x_u = y_i^T \underbrace{\boxed{(Y^T C^u Y + \lambda I)^{-1}}}_{W^u} Y^T C^u p(u)$$

$$s_{ij}^u = y_i^T W^u y_j$$

$$\hat{p}_{ui} = \sum_{j:r_{uj}>0} s_{ij}^u c_{uj}$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

=

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

X

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	

=

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

X

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	
	1	1		

=

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

X

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	
	1	1		

=

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6

X

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Implicit ALS. One step ALS

1		1	1	
	1	1		1
	1		1	
1		1		
1	1	1		
			1	1
1	1		1	
	1	1		

=

0.1	0.3	0.4
0.3	-0.5	-0.2
0.1	-0.2	-0.3
0.7	0.5	-0.4
0.5	0.6	0.9
-0.9	-0.7	0.2
0.4	0.2	0.6
0.3	-0.4	0.2

X

0.1	0.3	0.4	0.5	0.6
0.1	0.2	0.3	0.1	-0.2
0.9	-0.7	0.2	0.4	0.2

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

Функции ранжирования

Функции ранжирования

Функции ранжирования

- pointwise(поточечные)

Функции ранжирования

- pointwise(поточечные)
  - Оцениваем каждый айтем независимо
  - Любой лосс для задачи классификации или регрессии

Функции ранжирования

- pointwise(поточечные)
- pairwise(попарные)

Функции ранжирования

- pointwise(поточечные)
- pairwise(попарные)
  - Оцениваем пары релевантный vs нерелевантный айтем
  - Штрафуем если у нерелевантного айтема скор больше чем у релевантного
  - BPR, WARP, YettiRank

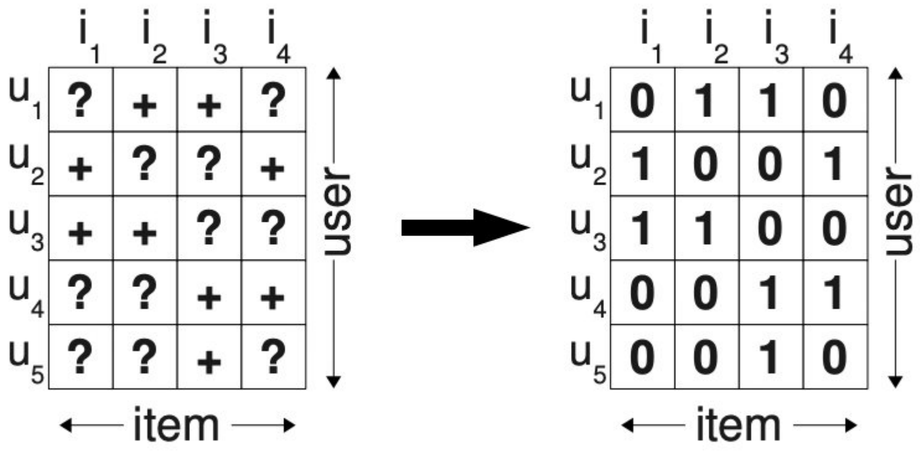
Функции ранжирования

- pointwise(поточечные)
- pairwise(попарные)
- listwise(списочные)

Функции ранжирования

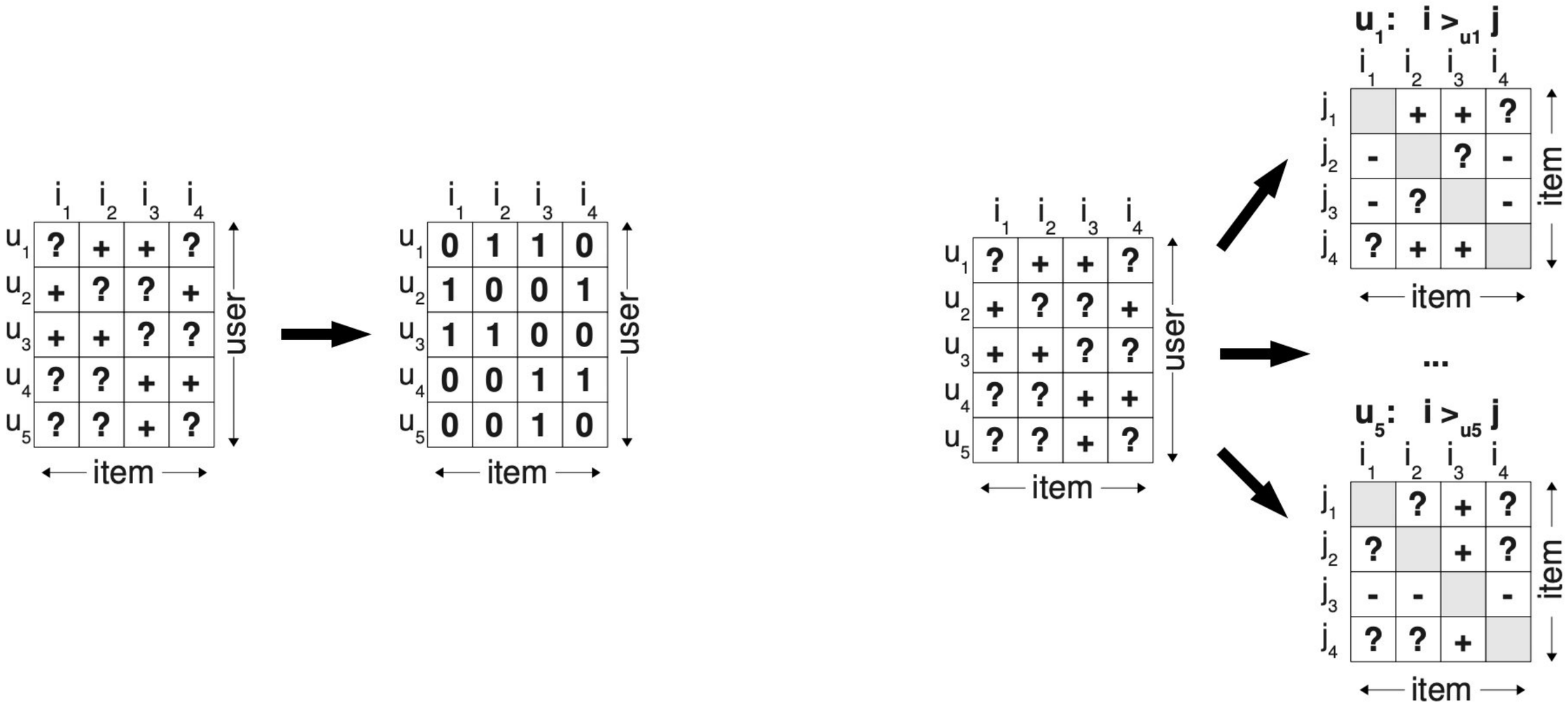
- pointwise(поточечные)
- pairwise(попарные)
- listwise(списочные)
  - Оптимизируем метрику ранжирования
  - Довольно сложные, поэтому не так часто используются на практике
  - LambdaRank, RankNet

BPR: Bayesian Personalized Ranking from Implicit Feedback



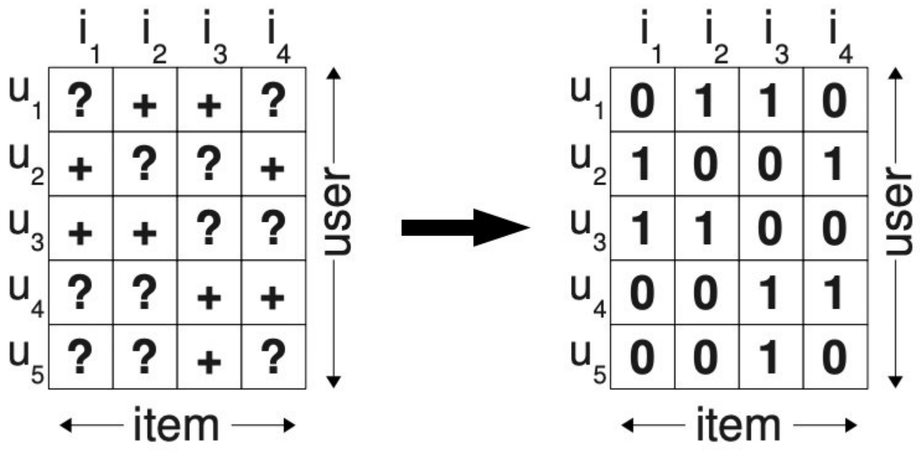
Rendle S. et al. BPR: Bayesian personalized ranking from implicit feedback //arXiv preprint arXiv:1205.2618. – 2012.

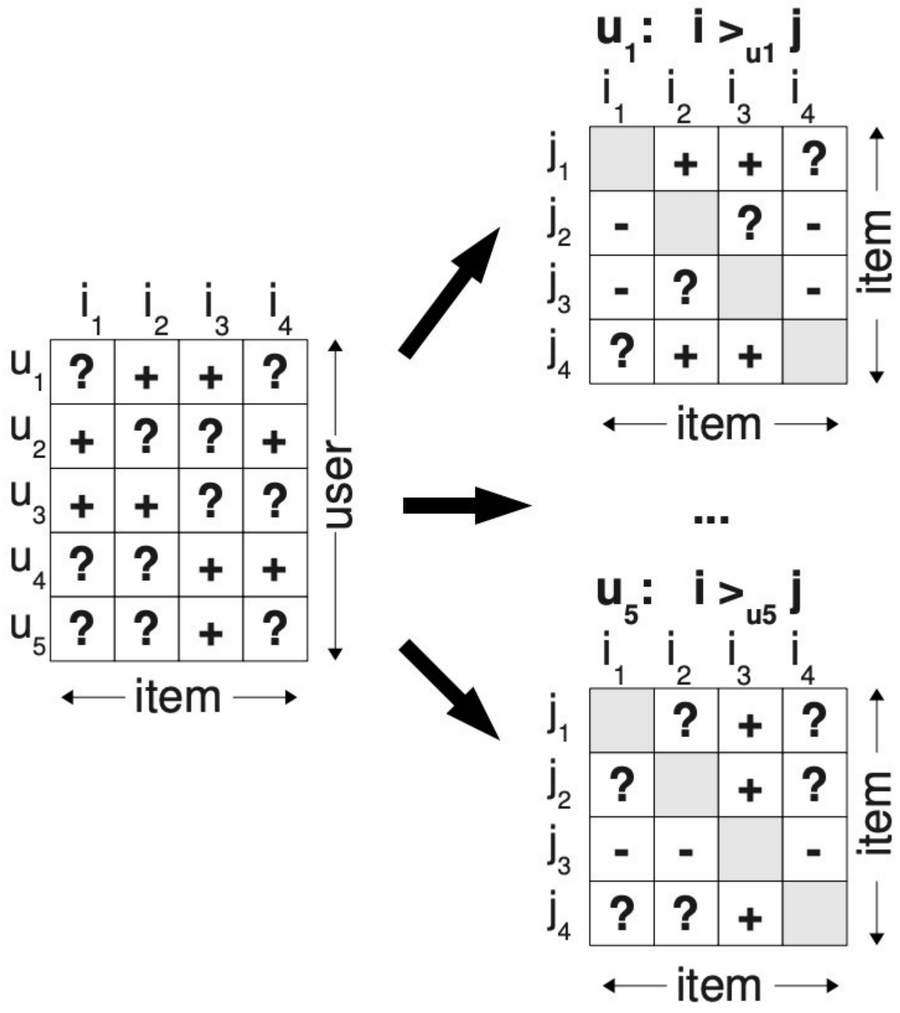
BPR: Bayesian Personalized Ranking from Implicit Feedback



Rendle S. et al. BPR: Bayesian personalized ranking from implicit feedback //arXiv preprint arXiv:1205.2618. – 2012.

BPR: Bayesian Personalized Ranking from Implicit Feedback





$$D_S := \{(u, i, j) | i \in I_u^+ \wedge j \in I \setminus I_u^+\}$$

$$L_{\text{bpr}} = -\frac{1}{N_S} \sum_{j=1}^{N_S} \log \sigma(r_i - r_j) \qquad \sigma(x) := \frac{1}{1 + e^{-x}}$$

Rendle S. et al. BPR: Bayesian personalized ranking from implicit feedback //arXiv preprint arXiv:1205.2618. – 2012.

WARP:Weighted Approximate-Rank Pairwise

Algorithm 1 K-OS algorithm for picking a positive item.

We are given a probability distribution P of drawing the ith position in a list of size K. This defines the choice of loss function.
Pick a user u at random from the training set.
Pick $i = 1, \ldots, K$ positive items $d_i \in \mathcal{D}_u$.
Compute $f_{d_i}(u)$ for each i.
Sort the scores by descending order, let $o(j)$ be the index into d that is in position j in the list.
Pick a position $k \in 1, \ldots, K$ using the distribution P.
Perform a learning step using the positive item $d_{o(k)}$.

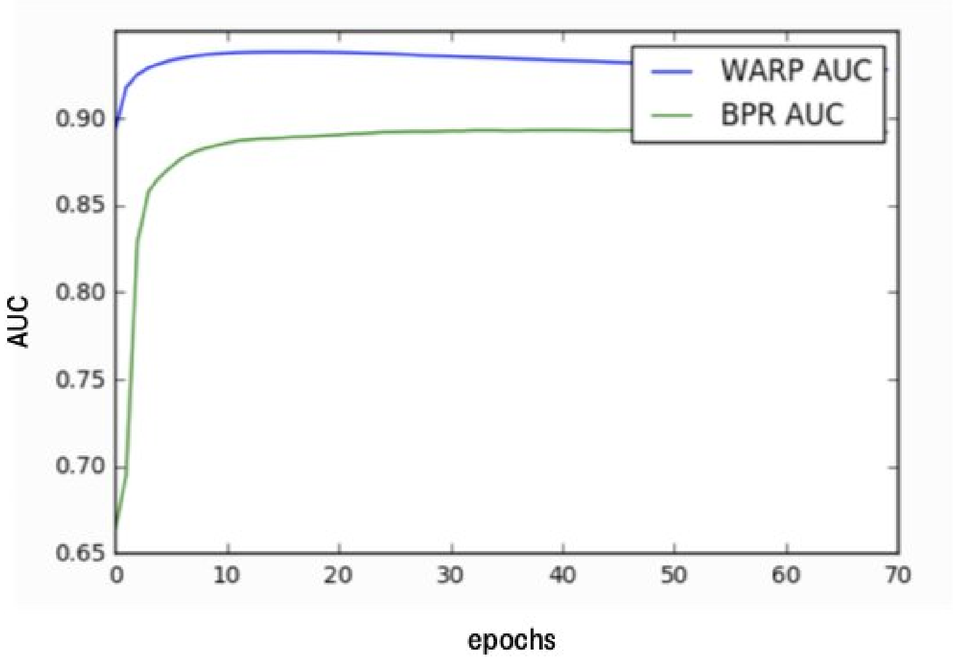
Algorithm 2 K-OS WARP loss

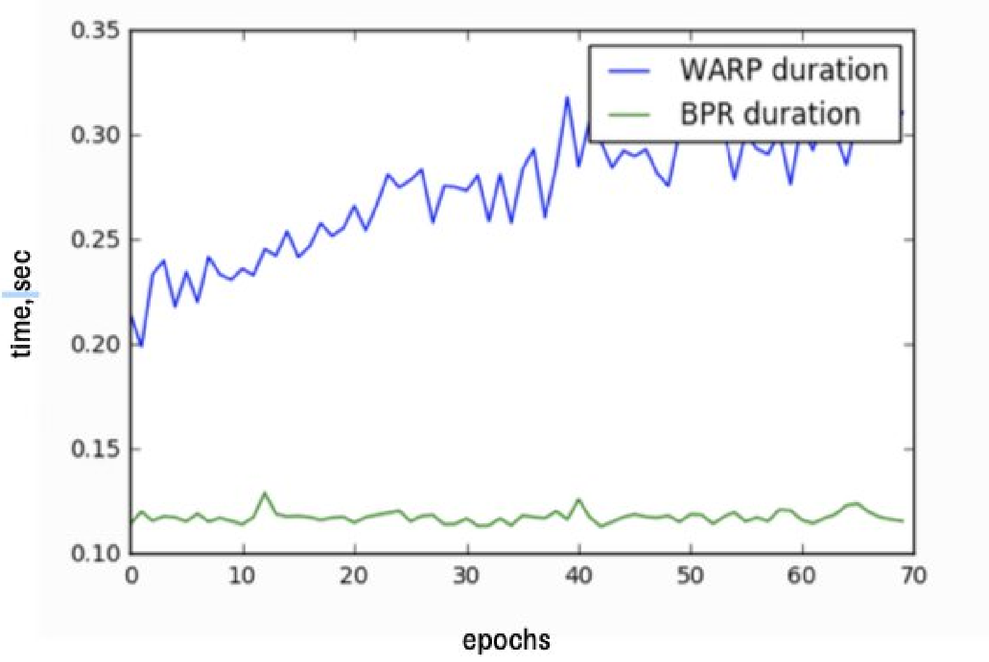
Initialize model parameters (mean 0, std. deviation $\frac{1}{\sqrt{m}}$).
repeat
  Pick a positive item d using Algorithm 1.
  Set $N = 0$.
  repeat
    Pick a random item $\bar{d} \in \mathcal{D} \setminus \mathcal{D}_u$.
    $N = N + 1$.
  until $f_{\bar{d}}(u) > f_d(u) - 1$ or $N \geq |\mathcal{D} \setminus \mathcal{D}_u|$
  if $f_{\bar{d}}(y) > f_d(u) - 1$ **then**
    Make a gradient step to minimize:
    $$\Phi(\frac{|\mathcal{D} \setminus \mathcal{D}_u|}{N}) \max(0, 1 + f_{\bar{d}}(u) - f_d(u)).$$
    Project weights to enforce constraints, e.g. if $||V_i|| > C$ then set $V_i \leftarrow (CV_i)/||V_i||$.
  end if
until validation error does not improve.

- Рассматриваем тройки (x, i, j)
- Если отранжировали неправильно, делаем шаг оптимизации
- Чем больше троек без ошибок, тем меньше шаг

Weston J., Yee H., Weiss R. J. Learning to rank recommendations with the k-order statistic loss //Proceedings of the 7th ACM Conference on Recommender Systems. – 2013. – С. 245-248.

WARP vs BPR



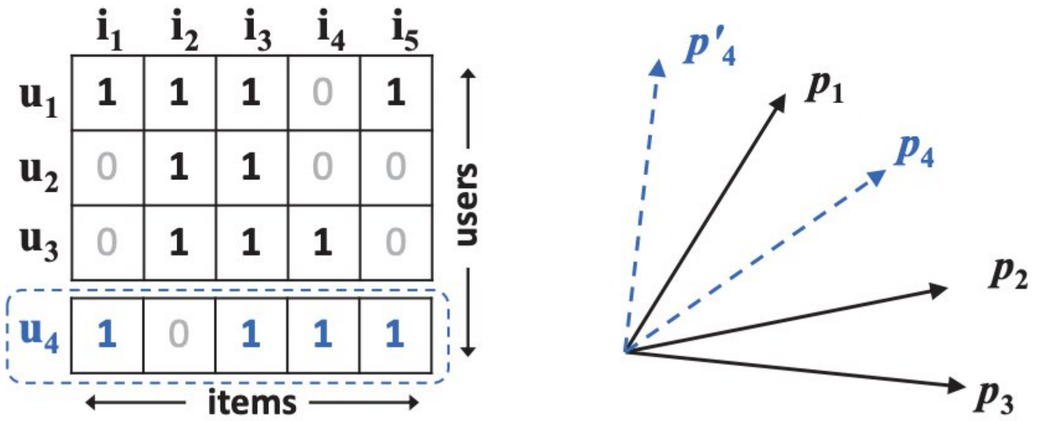


- WARP лучше по качеству
- Нужно меньше эпох чтобы сойтись
- Каждая последующая эпоха занимает больше времени, чем предыдущая

https://making.lyst.com/lightfm/docs/examples/warp_loss.html#learning-to-rank-using-the-warp-loss

Neural Collaborative Filtering

NCF

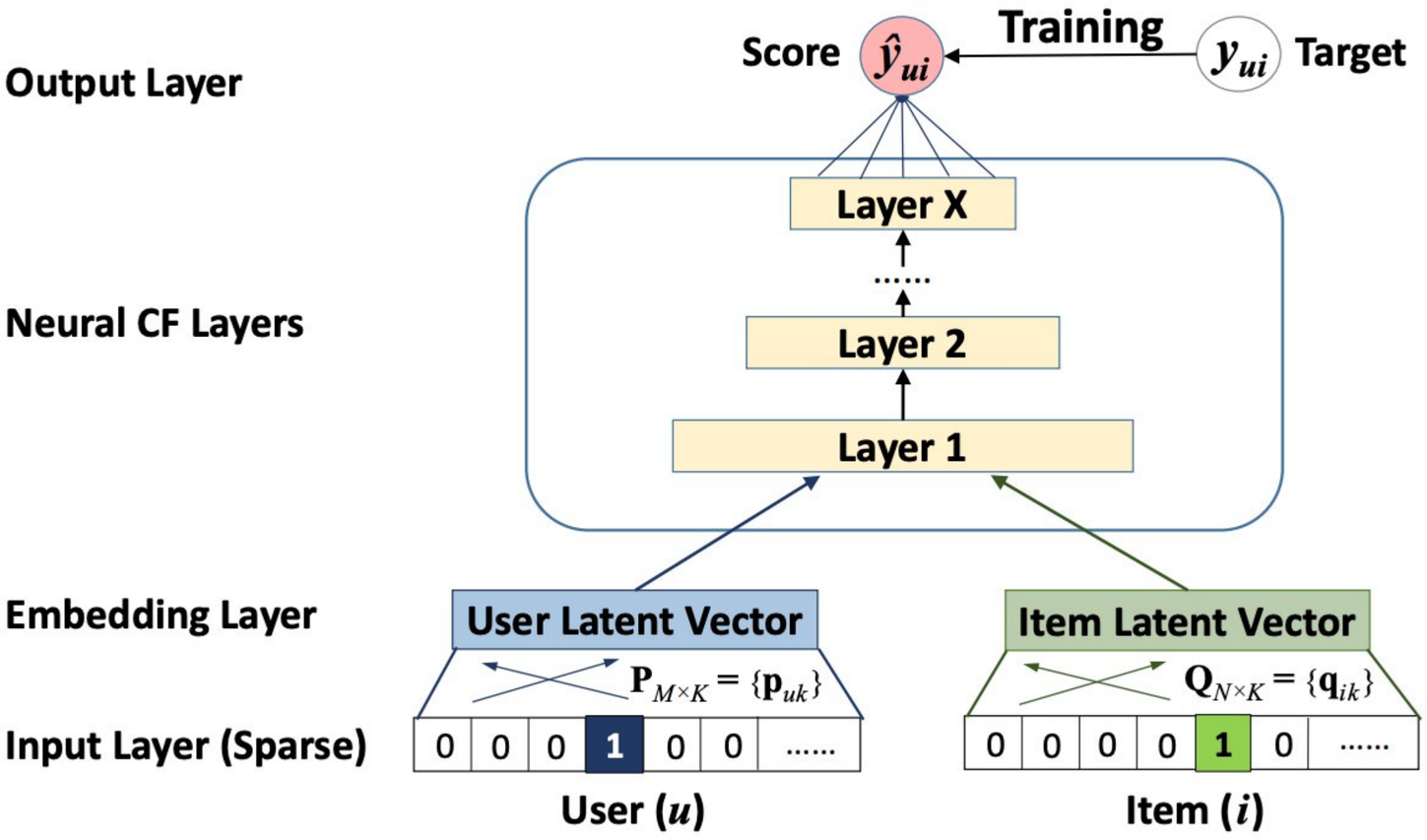


(a) user–item matrix (b) user latent space

Figure 1: An example illustrates MF's limitation. From data matrix (a), u_4 is most similar to u_1, followed by u_3, and lastly u_2. However in the latent space (b), placing $\mathbf{p}_4$ closest to $\mathbf{p}_1$ makes $\mathbf{p}_4$ closer to $\mathbf{p}_2$ than $\mathbf{p}_3$, incurring a large ranking loss.

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

NCF



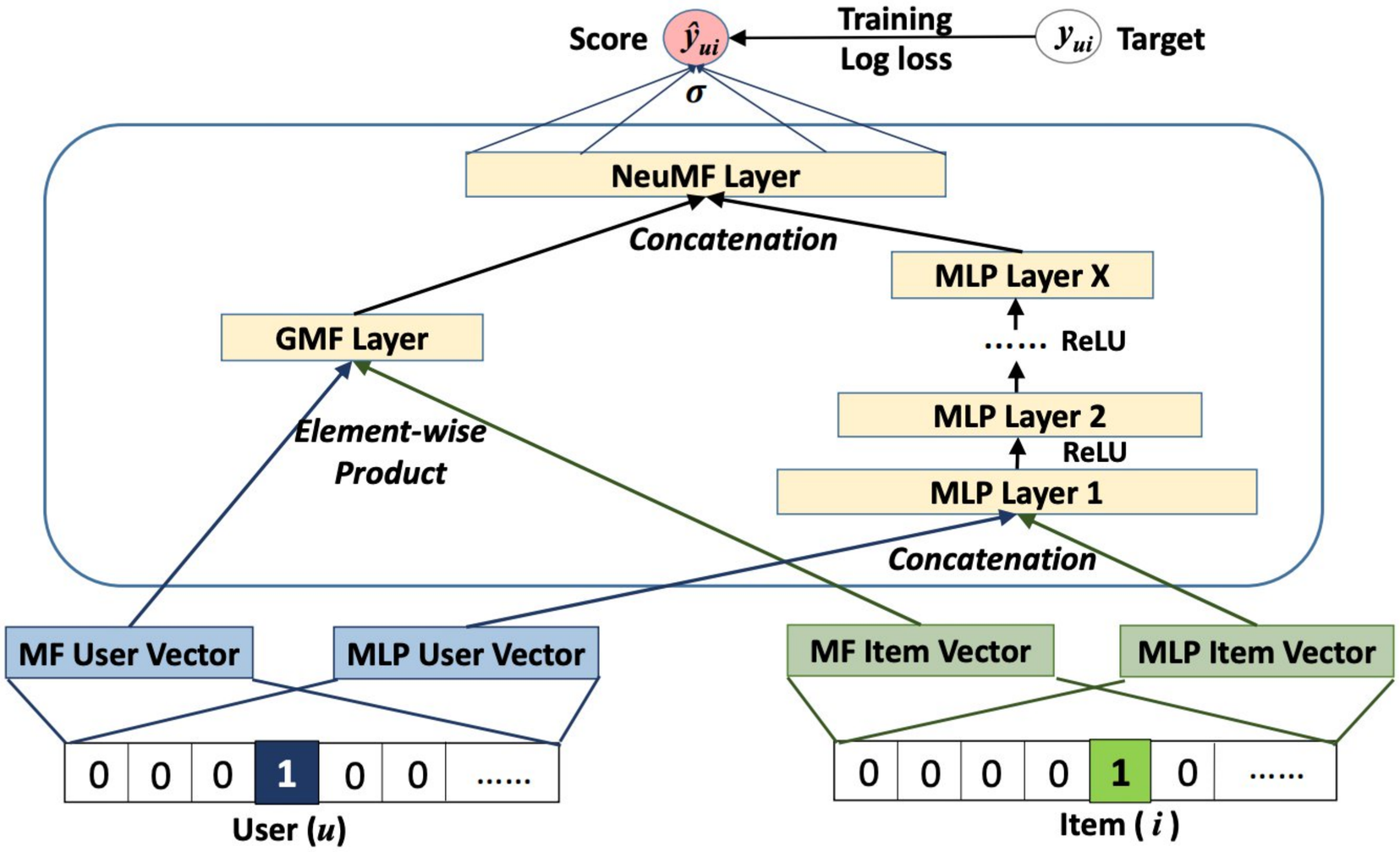
$$L_{sqr} = \sum_{(u,i)\in\mathcal{Y}\cup\mathcal{Y}^-} w_{ui}(y_{ui} - \hat{y}_{ui})^2,$$

$$L = -\sum_{(u,i)\in\mathcal{Y}} \log \hat{y}_{ui} - \sum_{(u,j)\in\mathcal{Y}^-} \log(1 - \hat{y}_{uj})$$

$$= -\sum_{(u,i)\in\mathcal{Y}\cup\mathcal{Y}^-} y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log(1 - \hat{y}_{ui}).$$

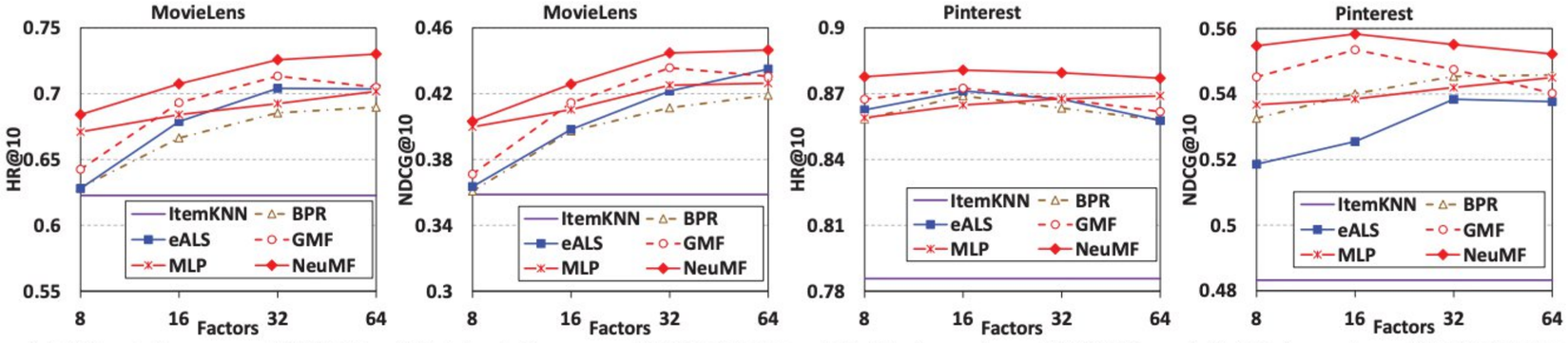
Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

NCF. Neural Matrix Factorization



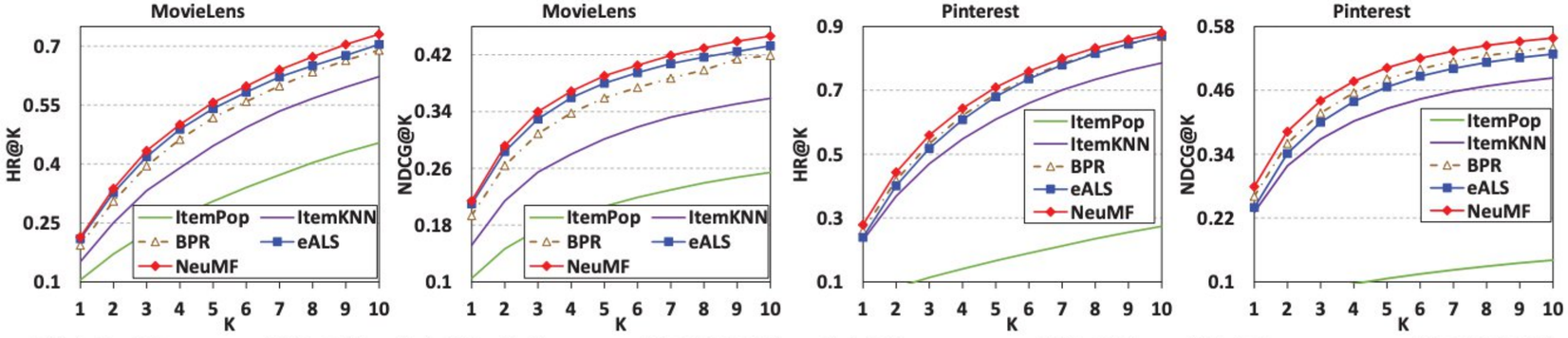
Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

NCF. Neural Matrix Factorization



(a) MovieLens — HR@10 (b) MovieLens — NDCG@10 (c) Pinterest — HR@10 (d) Pinterest — NDCG@10

Figure 4: Performance of HR@10 and NDCG@10 *w.r.t.* the number of predictive factors on the two datasets.



(a) MovieLens — HR@K (b) MovieLens — NDCG@K (c) Pinterest — HR@K (d) Pinterest — NDCG@K

Figure 5: Evaluation of Top-K item recommendation where K ranges from 1 to 10 on the two datasets.

Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets //2008

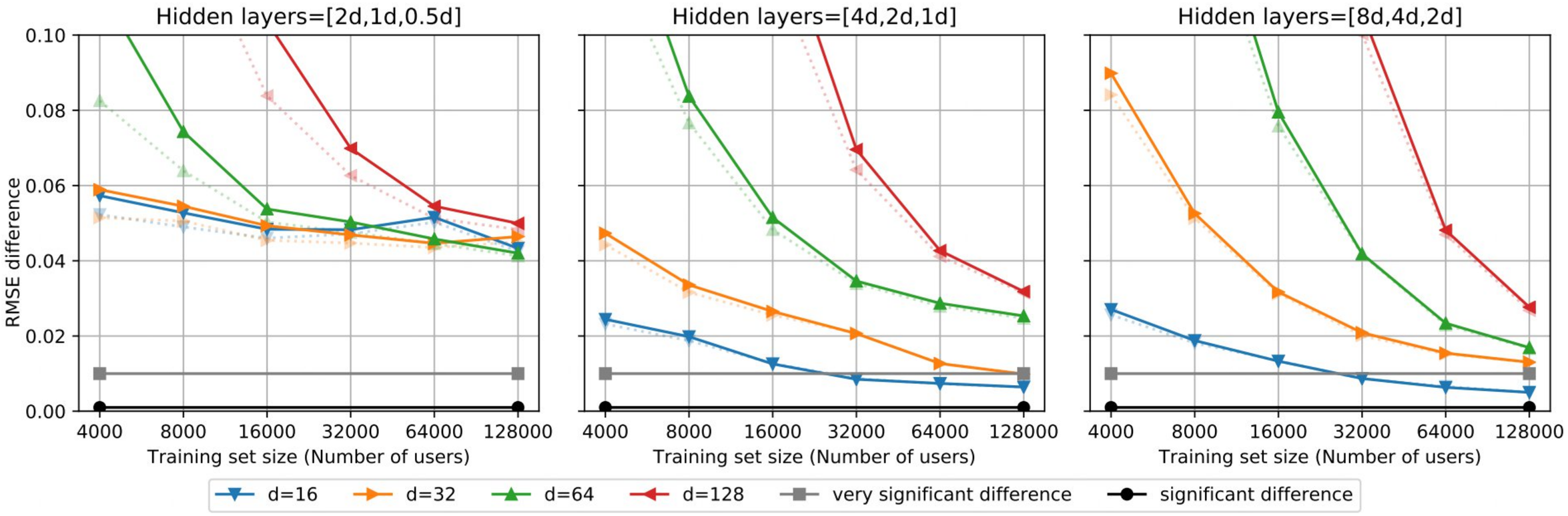
LEARNING A DOT PRODUCT WITH MLP IS HARD

Table 1: Comparison from [8] of NeuMF with various baselines and our results. The best results are highlighted in bold, the second best result is underlined.

Method	Movielens		Pinterest		Result
	HR@10	NDCG@10	HR@10	NDCG@10	from
Popularity	0.4535	0.2543	0.2740	0.1409	[8]
SLIM [25, 30]	<u>0.7162</u>	<u>0.4468</u>	0.8679	<u>0.5601</u>	[8]
iALS [20]	0.7111	0.4383	0.8762	0.5590	[8]
NeuMF (MLP+GMF) [17]	0.7093	0.4349	<u>0.8777</u>	0.5576	[8]
Matrix Factorization	**0.7294**	**0.4523**	**0.8895**	**0.5794**	Fig. 2

Rendle S. et al. Neural collaborative filtering vs. matrix factorization revisited //Proceedings of the 14th ACM Conference on Recommender Systems. – 2020. – С. 240-248.

LEARNING A DOT PRODUCT WITH MLP IS HARD



Rendle S. et al. Neural collaborative filtering vs. matrix factorization revisited //Proceedings of the 14th ACM Conference on Recommender Systems. – 2020. – С. 240-248.

Спасибо за внимание