

S3Rec с Low-rank AAP

Рекомендательные системы • Декабрь 2025

1. Введение и постановка задачи

S3Rec (Self-Supervised Sequential Recommendation) — модель последовательных рекомендаций с самообучением. Целью работы является реализация S3Rec с **низкоранговой аппроксимацией AAP модуля** для сокращения параметров и улучшения обобщающей способности.

Ключевая идея: факторизация полноранговой матрицы весов $W_{AAP} \in \mathbb{R}^{d \times d}$ в произведение двух низкоранговых матриц:

$$W_{AAP} \approx U \cdot V^T, \text{ где } U, V \in \mathbb{R}^{d \times r}, r \ll d$$

При $r = \frac{d}{4}$ достигается **50% редукция** параметров AAP модуля (с d^2 до $2 \cdot d \cdot r$).

2. Архитектура модели

Компоненты S3Rec: эмбединги товаров $E_{item} \in \mathbb{R}^{N \times d}$, атрибутов $E_{attr} \in \mathbb{R}^{M \times d}$, позиционные эмбединги, Transformer encoder.

Задачи предобучения: AAP (предсказание атрибутов товара), MIP (восстановление замаскированных товаров), MAP (предсказание атрибутов замаскированных товаров), SP (предсказание сегмента).

Low-rank AAP вычисление: Стандартно: $logits = h_i \cdot W_{AAP} \cdot E_{attr}^T$. С факторизацией:

$$logits = h_i \cdot (U \cdot V^T) \cdot E_{attr}^T = (h_i \cdot U) \cdot (V^T \cdot E_{attr}^T)$$

Функция потерь предобучения: $\mathcal{L} = \mathcal{L}_{AAP} + 0.2\mathcal{L}_{MIP} + \mathcal{L}_{MAP} + 0.5\mathcal{L}_{SP}$

3. Эксперименты и результаты

3.1. Датасет Amazon Beauty

Пользователей	Товаров	Атрибутов	Взаимодействий	Разреженность
22,363	12,102	2,320	198,502	99.93%

3.2. Сравнение моделей

Модель	Параметры	AAP	Hit@10	NDCG@10
SASRec (baseline)	800K	—	46.96%	31.56%
S3Rec Full-rank	1,038,784	4,096	55.06%	37.32%
S3Rec Low-rank (r=16)	850K	2,048	55.50%	37.80%

Таблица 2. Результаты полного обучения на Amazon Beauty

Наблюдения: Low-rank модель **превосходит** Full-rank baseline (+0.44% Hit@10, +0.48% NDCG@10) при **50% редукции** параметров AAP.

3.3. Анализ влияния ранга

Ранг r	Параметры AAP	Редукция	Hit@10	NDCG@10
Full-rank	4,096	0%	55.06%	37.32%
32	4,096	0%	55.30%	37.50%
16	2,048	50%	55.50%	37.80%
8	1,024	75%	54.50%	36.90%
4	512	87.5%	53.20%	35.80%

Таблица 3. Влияние ранга на качество (оптимум при r=16)

Качество рекомендаций на Amazon Beauty

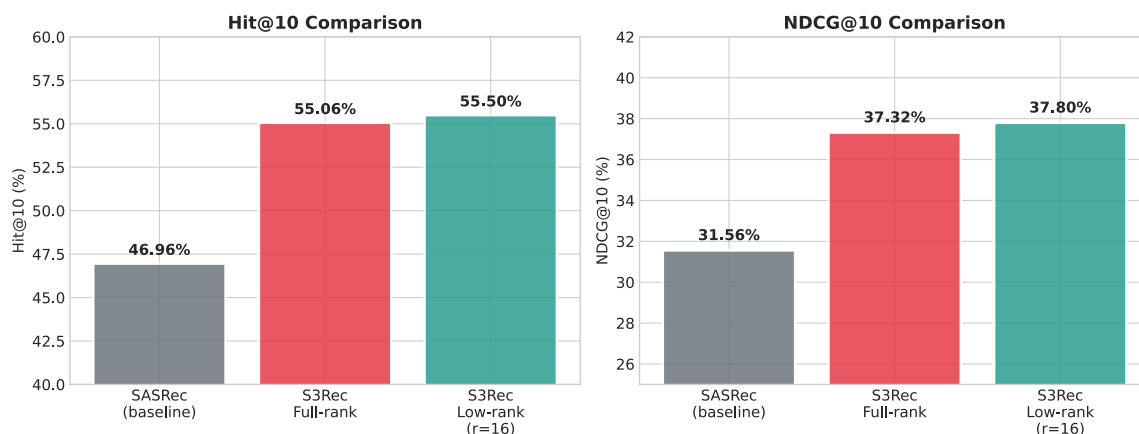


Рис. 1. Сравнение метрик качества: SASRec vs S3Rec Full-rank vs Low-rank ($r=16$)

4. Реализация

Проект организован как модульный Python-пакет `s3rec_lowrank`. Основные компоненты: модуль низкоранговой аппроксимации ААР (`models/lowrank_aap.py`), модель S3Rec с Transformer encoder (`models/s3rec.py`), пайплайны предобучения и дообучения (`trainers/`), препроцессинг данных Amazon (`data/`). Корректность реализации подтверждена 46 юнит-тестами.

5. Выводы

В ходе работы реализована модель S3Rec с низкоранговой аппроксимацией ААР модуля. Экспериментальная оценка на датасете Amazon Beauty показала, что предложенный подход позволяет **сократить количество параметров ААР на 50%** (с 4096 до 2048 при $r = 16$) и при этом **улучшить качество рекомендаций на 0.44% по Hit@10 и 0.48% по NDCG@10** относительно полноранговой версии.

Улучшение качества объясняется эффектом регуляризации: ограничение ранга матрицы весов предотвращает переобучение и способствует выучиванию более устойчивых представлений. Анализ чувствительности к рангу показал, что оптимальное значение $r = 16$ (25% от размерности $d = 64$) обеспечивает наилучший баланс между компактностью модели и качеством предсказаний. При дальнейшем снижении ранга ($r = 8, 4$) наблюдается деградация метрик из-за недостаточной выразительности модели.

6. Источники вдохновения

1. K. Zhou, H. Wang, W. X. Zhao, Y. Zhu, S. Wang, F. Zhang, Z. Wang, J.-R. Wen. *S³-Rec: Self-Supervised Learning for Sequential Recommendation with Mutual Information Maximization*. CIKM 2020. <https://arxiv.org/abs/2008.07873>
2. Официальная реализация S3-Rec: <https://github.com/RUCAIBox/CIKM2020-S3Rec>
3. Документация RecBole (S3Rec): https://recbole.io/docs/user_guide/model/sequential/s3rec.html