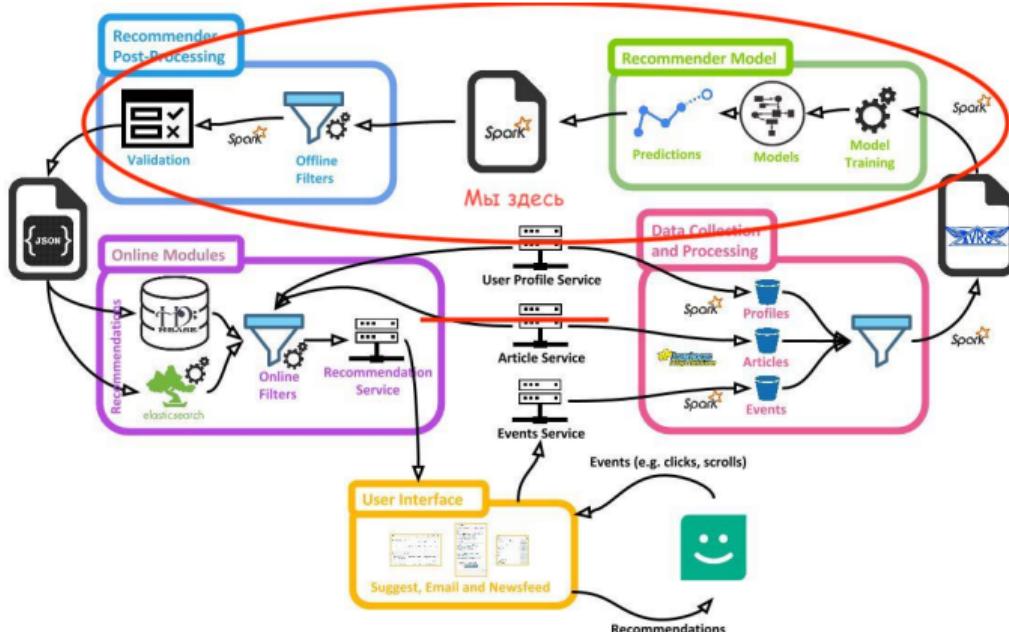


Нерешенные проблемы и новые направления

Николай Анохин

22 апреля 2025 г.

Контекст



Что мы уже умеем

$$\hat{r}_{ui} = f_{\theta}(x_u, x_i, x_c)$$



Проблемы

1. Оцениваем айтемы по-отдельности, а показываем по несколько (лентой)
2. Смещение между распределениями на обучении и применении
3. Модель не объясняет, почему именно эти айтемы подходят пользователю
4. Не учитывается долгострочный эффект рекомендаций

Разнообразие в рекомендательных системах
●○

Разнообразие для exploration
oooooooo

Разнообразие для utility
ooooooo

Итоги
oooo

Разнообразие в рекомендательных системах



Разнообразие / Diversity



Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
●oooooooo

Разнообразие для utility
oooooooo

Итоги
oooo

Разнообразие для exploration



Разнообразие для exploration

Помогает

- пользователям находить новые айтемы
- рекомендеру собирать данные



Набираем айтемы с разными аспектами

f - аспект (признак) айтема, $p(f|i)$ – вероятность найти аспект у айтема i

Распределение аспекта у пользователя

$$p(f|u) = \frac{\sum_{i \in I_u} p(f|i)}{|I_u|}$$

Распределение аспекта в рекомендациях

$$q(f|u) = \frac{\sum_{i \in RL} p(f|i)}{|RL|}$$

Формируем список так, чтобы $q(f|u)$ совпало с $p(f|u)$



Жадное переранжирование

Добавляем в список рекомендаций айтем с максимальным значением

$$(1 - \lambda) \cdot s(u, i) + \lambda \cdot gain(i, RL),$$

пока не получим список нужной длины.

- $s(u, i)$ – релевантность айтема i для пользователя u
- $gain(i, RL) = div(RL \cup \{i\}) - div(RL)$ – улучшение разнообразия при добавлении айтема
- λ – гиперпараметр

Примеры жадного переранжирования

Maximal Marginal Relevance [CG98]

$$MMR = \arg \max_{D_i \in R \setminus S} [\lambda \cdot sim(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} sim(D_i, D_j)] .$$

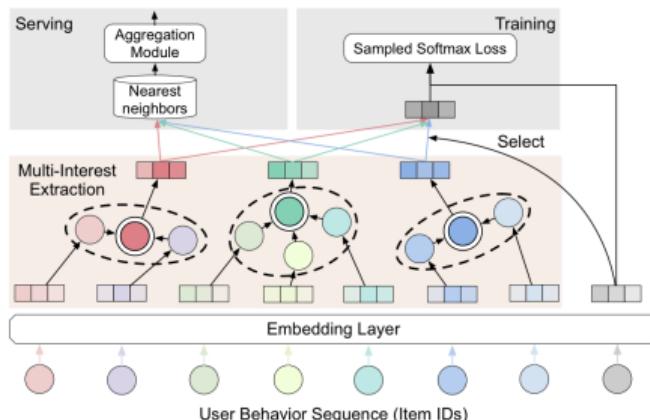
Mean Listing Relevance [AHR⁺20]

$$MLR = \sum_{i=1}^N \left[(1 - \lambda) \cdot c(i) P(I_i) + \lambda \sum_{j < i} \frac{d(I_i, I_j)}{i} \right] ,$$

где $c(i)$ – штраф за позицию в выдаче.

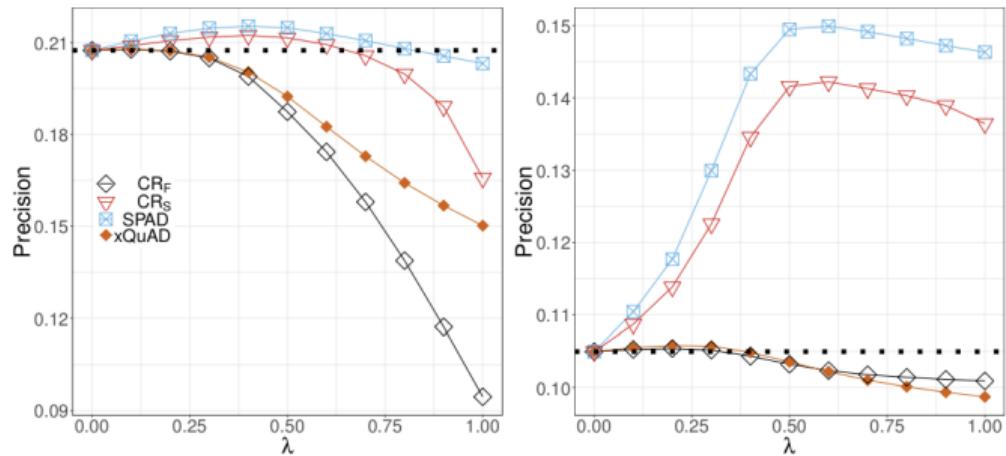
Учим разнообразие вместе с моделью

Controllable Multi-Interest Framework for Recommendation [CZZ⁺20]¹



¹Еще один подход к выученному разнообразию [SVM⁺24]

A Comparison of Calibrated and Intent-Aware Recommendations [KB19]



Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
ooooooo●

Разнообразие для utility
ooooooo

Итоги
oooo

Увеличивая разнообразие, мы готовы жертвовать (сиюминутной) релевантностью. Верим, что это принесет долгосрочный выигрыш.



Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
oooooooo

Разнообразие для utility
●oooooooo

Итоги
oooo

Разнообразие для utility



Разнообразие для utility

Помогает

- убирать избыточные айтемы.



При этом подходе компромисс relevance-diversity не существует!

Determinantal Point Process (DPP) [WRB⁺18]

Дано множество айтемов $S = \{1, 2, \dots, N\}$.

Point Process на S – распределение, определяющее вероятность $P(s)$ любого подмножества $s \subset S$.

DPP (в рекомендациях) – это PP, в котором $P(s)$ тюнится так, чтобы большая вероятность была у подмножеств с релевантными и разнообразными айтемами.



Determinantal Point Process (DPP) contd.

Пусть в рекомендательной выдаче N айтемов, проиндексированных $\{1, \dots, N\}$.

Y – множество индексов айтемов, с которыми было положительное взаимодействие (пример: $Y = \{2, 6, 11\}$).

$$P(Y) = \frac{\det(L_Y)}{\sum_{Y'} \det(L_{Y'})}$$

Determinantal Point Process (DPP) contd.

L – неотрицательно определенная матрица $N \times N$, в которой:

- диагональные элементы отвечают за релевантность
- недиагональные – за попарное сходство.

$$L_{ij} = f(q_i)^T g(\phi_i)^T g(\phi_i) f(q_i)$$

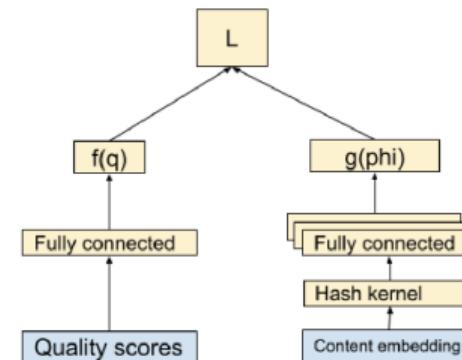


Figure 4: Architecture example for a deep DPP kernel.

Determinantal Point Process (DPP) end.

Алгоритм

На каждом шаге добавляем к Y

$$\max_{v \notin Y} \det(L_Y \cup v),$$

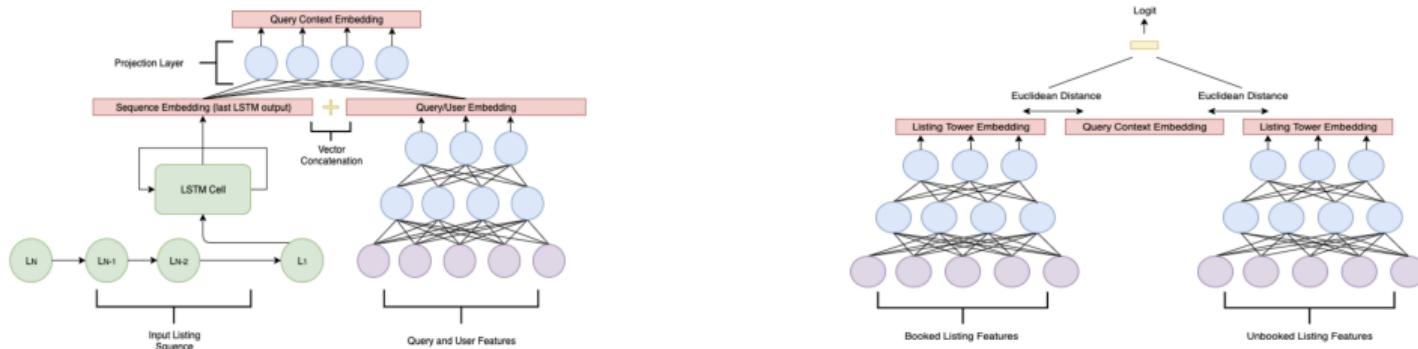
учитывая только последние K айтемов.

| Strategy | Satisfied homepage watchers |
|----------------------|-----------------------------|
| Fuzzy deduping | -0.05% |
| Sliding window | -0.26% |
| Smooth score penalty | -0.41% |
| DPPs | +0.63% |
| Deep DPPs | +1.72% |



Managing Diversity in Airbnb Search [AHR⁺20]

Идея: включить информацию, необходимую для разнообразия в модель, чтобы модель сразу выучила разнообразные рекомендации.²



| Method | MLR | NDCG |
|--------------------------------|--------------|--------------|
| Greedy Ranker | 2.45% | -0.96% |
| Location Diversity Ranker | 1.04% | -0.13% |
| Contextual Features | 0.56% | 0.21% |
| Combined Loss Function | 1.89% | 0.03% |
| Query Context Embedding | 1.97% | 1.26% |

²В [НАН⁺23] эта модель сделана менее гибкой с помощью дополнительных предположений

Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
oooooooo

Разнообразие для utility
ooooooo●

Итоги
oooo

Разнообразие можно рассматривать, как избавление от избыточности. Но для этого понадобятся модели построения списков.



Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
oooooooo

Разнообразие для utility
ooooooo

Итоги
●ooo

Итоги



Итоги

Из-за поточечного предсказания релевантности, приходится дополнительно разнообразить списки рекомендаций

Можно смотреть на разнообразие как на exploration, но при этом быть готовым пожертвовать релевантностью.

Или можно смотреть на разнообразие как на избавление от избыточности, но это добавляет сложность при построении списков.

В любом случае, необходимость разнообразия можно обосновать A/B экспериментом.



Разнообразие в рекомендательных системах
oo

Разнообразие для exploration
oooooooo

Разнообразие для utility
ooooooo

Итоги
ooo



<https://t.me/mlvok>



Литература I

-  Mustafa Abdool, Malay Haldar, Prashant Ramanathan, Tyler Sax, Lanbo Zhang, Aamir Manaswala, Lynn Yang, Bradley Turnbull, Qing Zhang, and Thomas Legrand, *Managing diversity in airbnb search*, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA), KDD '20, Association for Computing Machinery, 2020, p. 2952–2960.
-  Jaime Carbonell and Jade Goldstein, *The use of mmr, diversity-based reranking for reordering documents and producing summaries*, Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (New York, NY, USA), SIGIR '98, Association for Computing Machinery, 1998, p. 335–336.



Литература II

-  Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang, *Controllable multi-interest framework for recommendation*, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2020).
-  Malay Haldar, Mustafa Abdoor, Liwei He, Dillon Davis, Huiji Gao, and Sanjeev Katariya, *Learning to rank diversely at airbnb*, 2023.
-  Mesut Kaya and Derek Bridge, *A comparison of calibrated and intent-aware recommendations*, Proceedings of the 13th ACM Conference on Recommender Systems (New York, NY, USA), RecSys '19, Association for Computing Machinery, 2019, p. 151–159.



Литература III

-  Anima Singh, Trung Vu, Nikhil Mehta, Raghunandan Keshavan, Maheswaran Sathiamoorthy, Yilin Zheng, Lichan Hong, Lukasz Heldt, Li Wei, Devansh Tandon, et al., *Better generalization with semantic ids: A case study in ranking for recommendations*, Proceedings of the 18th ACM Conference on Recommender Systems, 2024, pp. 1039–1044.
-  Mark Wilhelm, Ajith Ramanathan, Alexander Bonomo, Sagar Jain, Ed H. Chi, and Jennifer Gillenwater, *Practical diversified recommendations on youtube with determinantal point processes*, Proceedings of the 27th ACM International Conference on Information and Knowledge Management (New York, NY, USA), CIKM '18, Association for Computing Machinery, 2018, p. 2165–2173.

