HybridBERT4Rec: A Hybrid Recommender System Based on BERT

Sequential Content-Based and Collaborative Filtering

Leon Knorr | November 14, 2023

Table of Contents

Traditional CBF VS Sequential CBF

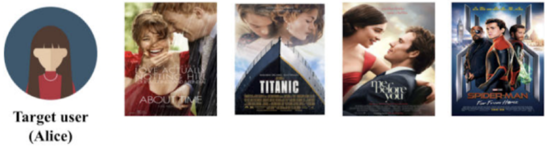

Figure 1: Example history for Alice in traditional CBF [1]

- models **general** user preference

Traditional CBF VS Sequential CBF

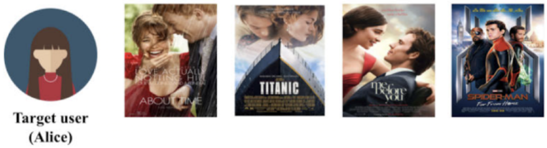

Figure 1: Example history for Alice in traditional CBF [1]

- models **general** user preference
- **BUT:** User preferences **change over time!** [5]

Traditional CBF VS Sequential CBF

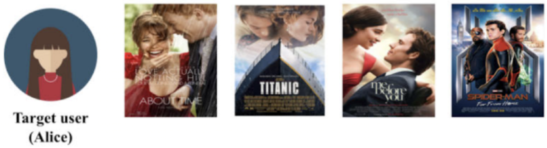

Figure 1: Example history for Alice in traditional CBF [1]

Figure 2: Example history for Alice in sequential CBF [1]

- models **general** user preference
- **BUT:** User preferences **change over time!** [5]

- Considers the **order** of historical interactions
- Allows the modelling of “temporary spikes” of interests, as well as the general preferences [5]

A Common Approach to Sequential modelling

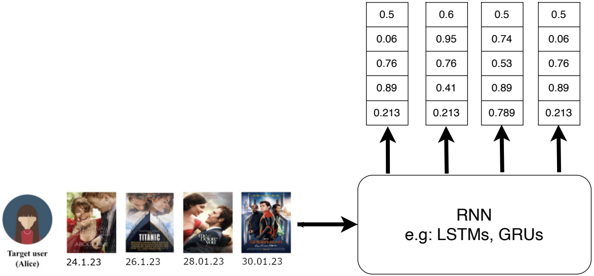

Figure 3: Sequential Content Recommendation with RNNs. [6]

- Suffers from common RNN problems! Especially: **Catastrophic forgetting**, uni-directionality [1]

HybridBERT4Rec

High Level Overview

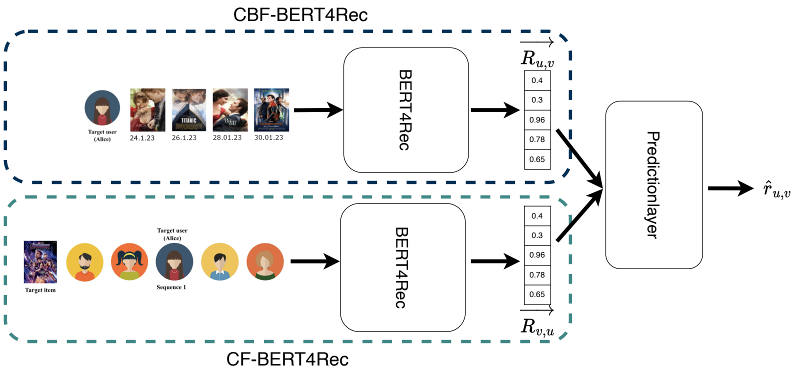

Figure 4: High level overview of HybridBERT4Recs Architecture. [1]

BERT4Rec [4]

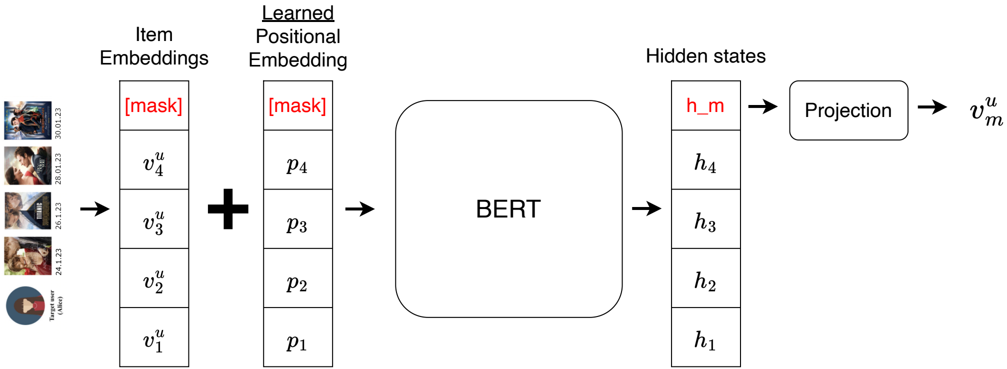

Figure 5: BERT4Rec Architecture, taking item embeddings v_t^u from user u history as input and predicts the next item v_m^u, u is likely to interact with. [4]

CF-HybridBERT4Rec

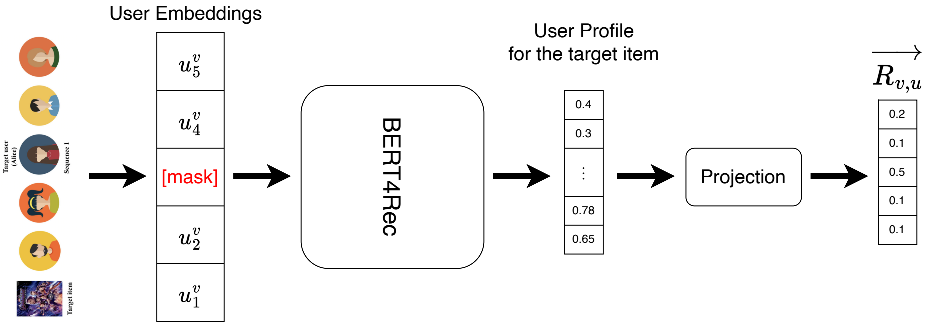

Figure 6: CF-HybridBERT4Rec Architecture, taking user embeddings from all users that have rated the target item v as input and predicts the “target item profile $\overrightarrow{R_{v,u}}$”. [1]

CBF-HybridBERT4Rec

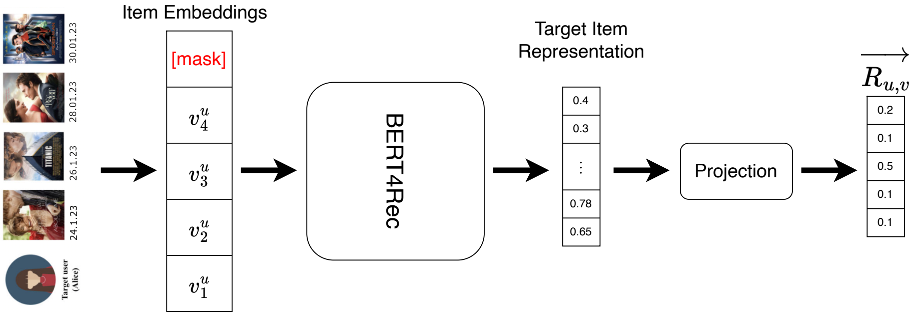

Figure 7: CBF-HybridBERT4Rec Architecture, taking item embeddings from a user *u* history as input and predicts a "target user profile $\overrightarrow{R_{u,v}}$". [1]

Prediction Layer: Combining CF & CBF

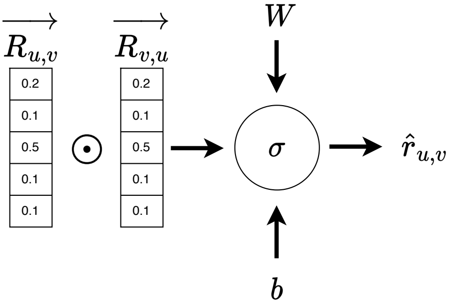

Figure 8: Schematic of HybridBERT4Recs Prediction layer, which uses a generalization of Matrix Factorization based on Neural Networks with Sigmoid activations to predict the rating $\hat{r}_{u,v}$ user u would assign to item v. [1]

Strengths & Weaknesses

Strengths:

- Easy to parallelize
- Bi-directional
- Sequential
- CBF-, CF-model & Prediction layer can be executed independently
- Intermediate results can be cached

Weaknesses:

- Sequence length is limited
- Needs lots of processing power and memory
- Only uses rating information

Model Performance & Experiments

Results

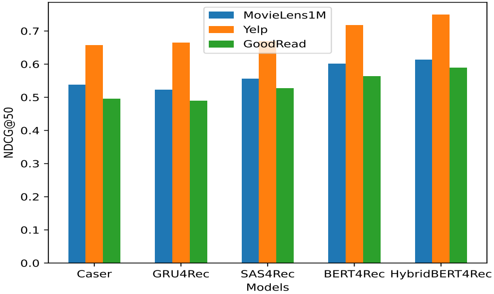

Figure 9: Performance comparison of different recommender models on three datasets as published by the authors of HybridBERT4Rec. [1]

Critique:

Results

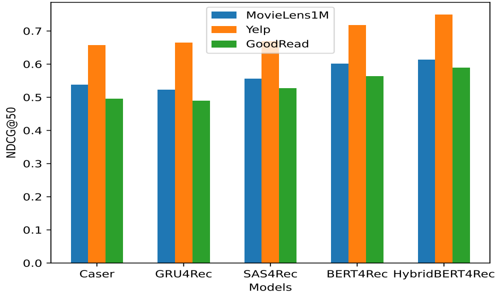

Figure 9: Performance comparison of different recommender models on three datasets as published by the authors of HybridBERT4Rec. [1]

Critique:

- No Hybrid model was evaluated!
- No information about data partitioning
- Generalization performance & real world Applicability is unknown

Applicability to E-Learning

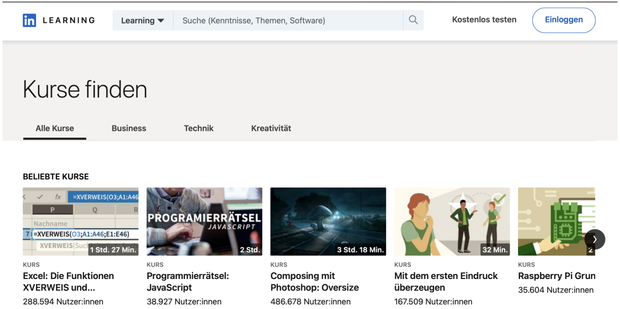

Figure 10: Linked-In Learning landing-page. [3]

Figure 11: Linked-In Learning course overview. [3]

References

[1] Chanapa Channarong et al. "HybridBERT4Rec: A Hybrid (Content-Based Filtering and Collaborative Filtering) Recommender System Based on BERT". In: *IEEE Access* 10 (2022), pp. 56193–56206. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3177610. (Visited on 11/02/2023).

[2] Xiangnan He et al. *Neural Collaborative Filtering*. Aug. 2017. arXiv: 1708.05031 [cs]. (Visited on 11/02/2023).

[3] *LinkedIn Learning mit Lynda: Onlinekurse aus dem Business-, Technik- und Kreativbereich*. https://de.linkedin.com/learning/. (Visited on 11/03/2023).

[4] Fei Sun et al. *BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer*. Aug. 2019. DOI: 10.48550/arXiv.1904.06690. arXiv: 1904.06690 [cs]. (Visited on 11/02/2023).

[5] Shoujin Wang et al. "Sequential Recommender Systems: Challenges, Progress and Prospects". In: (2019), pp. 6332–6338. (Visited on 11/02/2023).

[6] Feng Yu et al. "A Dynamic Recurrent Model for Next Basket Recommendation". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '16. New York, NY, USA: Association for Computing Machinery, July 2016, pp. 729–732. ISBN: 978-1-4503-4069-4. DOI: 10.1145/2911451.2914683. (Visited on 11/02/2023).

Matrix Factorization

Inputs:

- User latent representation p_u
- Item latent representation q_i
- Dimension of latent space K

$$\hat{y_{u,i}} = p_u^T q_i = \sum_{k=1}^{K} p_{uk} q_{ik}$$

source: [2]

BERT

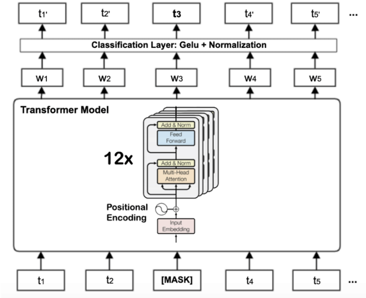

Figure 12: Schematic of BERTs Architecture [1]

[1] source: https://www.researchgate.net/figure/The-Transformer-based-BERT-base-architecture-with-twelve-encoder-blocks_fig2_349546860

Normalized Discounted Cumulative Gain (NDCG)

- Measures the rating lists quality by accessing the items rank and rating

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

$$IDCG_k = \sum_{i=1}^{|REL_k|} \frac{rel_i}{\log_2(i+1)}$$

$$NDCG_k = \frac{DCG_k}{IDCG_k}$$

source: https://en.wikipedia.org/wiki/Discounted_cumulative_gain; [1]