HybridBERT4Rec: A Hybrid Recommender System Based on BERT

Sequential Content-Based and Collaborative Filtering

Leon Knorr | December 4, 2023

Table of Contents

Recap: Sequential Modelling & HybridBERT4Rec

Traditional CBF VS Sequential CBF

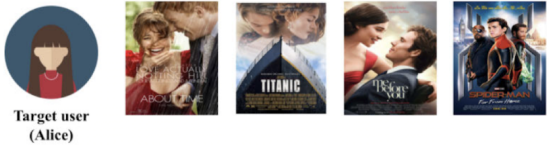

Figure 1: Example history for Alice in traditional CBF [1]

- models **general** user preference
- **BUT:** User preferences **change over time!** [3]

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

- Considers the **order** of historical interactions
- Allows the modelling of “temporary spikes” of interests, as well as the general preferences [3]

HybridBERT4Rec Architecture

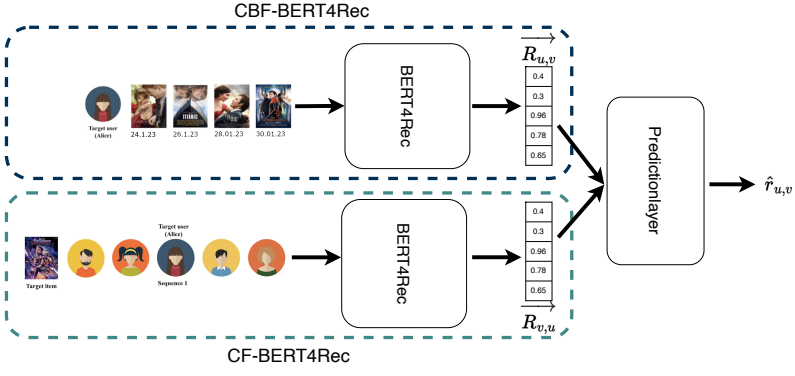

Figure 3: High level overview of HybridBERT4Recs Architecture. [1]

The Setting

The Setting

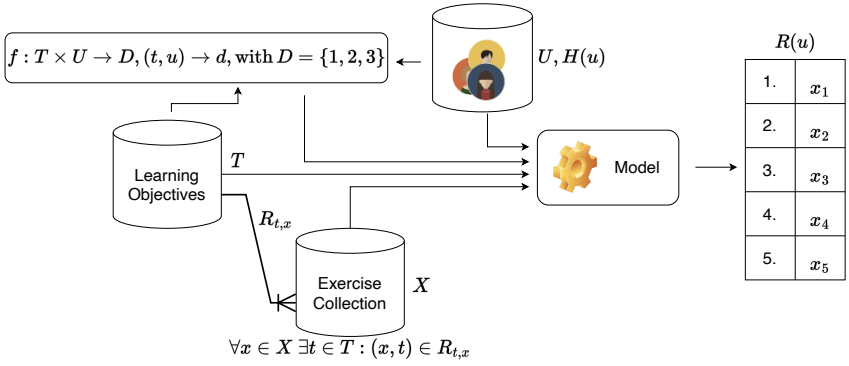

Figure 4: The Setting, consisting of a user collection U and their histories $h(u)$, a collection of learning objectives T and a collection of exercises X, which can be used to predict a ranking $R(u)$ for a given user u.

Model Adaption

CBF-HybridBERT4Rec

- $H(u) := (\{(x_i, t_j, s_k) | (x_i, t_j) \in R_{t,x}\}, \leq)$, with $s_{k-1} \leq s_k$
- $I(u) := (\{x_i | (x_i, t_j, s_k) \in H(u)\}, \leq)$
- $\overrightarrow{R_{u,t}}$: the interaction probability distribution of all items with the user u over the target item

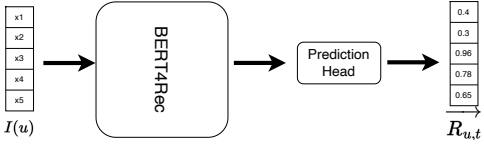

Figure 5: CBF-HybridBERT4Rec architecture and input in the described setting

CF-HybridBERT4Rec

- $u \in N \iff d_{u,t} = d_{u_m,t} \wedge (x,t) \in \{(x,t)|(x,t,s_k) \in H(u)\}$, with $U_m \in U, U \in U, t \in T$, N being the set of neighbors for target (masked) user u_m and learning objective t
- $\overrightarrow{R_{t,u}}$: a user-similarity probability distribution of all users over the target (masked) user

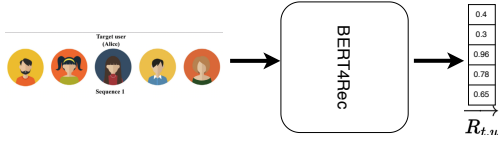

Figure 6: CF-HybridBERT4Rec architecture and input in the described setting

Bringing It All Together

Algorithm 1 HybridBERT4Rec in an E-Learning Setting

for all $u_m \in U$ **do**
    $r_{x,u_m} = \texttt{cbf\_hybridbert4rec}(H(u_m))$
    for all $(x, t) \in R_{t,x}$ **do**
        $r_{u,x} = \texttt{cf\_bert4rec}(u_m, t, x)$
        $\hat{r}_{u,x} = \texttt{prediction\_layer}(r_{x,u}, r_{u,x})$
    end for
end for

- Yields a rating $\hat{r}_{u,x}$ for each exercise and for each user
- Construct an overall rating of exercises by sorting the ratings
- Construct a topic specific rating by filtering for a topic and sorting the ratings

Evaluation

The Problem

- No given Test data with relevance annotations!
- Annotating the whole collection requires $U \times T \times X$ relevance annotations
- **INFEASIBLE** for large U, X and T

Pooling [2]

- For most queries only $N << X$ documents are relevant

$\Rightarrow$ We only annotate the top N results of every query

$\Rightarrow$ $U \times T \times N$ relevance annotations needed

- Compute P@K, R@K, NDCG etc.
- **Shortcoming**: Scores are only **approximations!**

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] Dr. Pedro Ortiz Suarez, Prof. Dr. Goran Glavaš, and Prof. Dr. Simone Paolo Ponzetto. *Information Retrieval & Web Search*. en. Sept. 2022.

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