

# BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer

Björn Bebensee (2019–21343)  
Topics in Artificial Intelligence

October 29, 2019

In the sequential recommendation setting we want to give a recommendation based on a sequence of previous user behaviours. Previous approaches typically encoded the user’s historical actions or interactions as a series of hidden states from which the model would make the recommendation. Sun et al. [1] argue that this is not optimal as such a left-to-right unidirectional model fails to capture sequences which are not strictly ordered and that the hidden representations fail to adequately capture the historical information. They conclude that models in this setting could benefit from a bidirectional approach. To achieve such bidirectionality the authors propose *BERT4Rec* which is inspired by and borrows from BERT and uses bidirectional self-attention.

In order to train such a model the authors use a Cloze task which masks a target item in a sequence and they train the model to predict the masked item given the surrounding context (the entire sequence). However, as this approach is not strictly consistent with the sequential recommendation setting, they append the [mask] token to the end of the sequence so the model knows what to predict. As usual, the transformer layer consists of a multi-head self-attention sub-layer and a feed-forward sub-layer. To capture item-item interactions,  $L$  transformer layers are stacked where each layer of  $t$  transformers takes the output of all transformers of the previous layer for a sequence of items  $v_1, \dots, v_t$  with  $v_t = [\text{mask}]$ . The inputs for the first transformer layer are the positional embeddings of the  $t$  input items. In the final layer they predict an item for the masked item  $v_t$  based on the hidden representation  $h_t^L$  of the  $t$ -th transformer in the  $L$ -th layer using a two-layer feed-forward neural network with GELU activations and a softmax layer.

Sun et al. evaluate their model on four real-world datasets from various domains (Amazon Beauty, Steam, MovieLens-1M and MovieLens-20M). They find that BERT4Rec consistently outperforms the state-of-the-art on all four datasets in terms of all evaluation metrics. Thus, they have demonstrated that models in the sequential recommendation setting can greatly benefit from bidirectional hidden representations.

## References

- [1] Sun, Fei, et al. "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer." *arXiv preprint arXiv:1904.06690* (2019).