

ATRank: An Attention-Based User Behavior Modeling Framework for Recommendation

March 12, 2020

Abstract

This paper proposes an attention based user behavior modeling framework called ATRank. Heterogeneous user behaviors are considered in the proposed model by projecting all types of behaviors into multiple latent semantic spaces. Downstream applications then can use the user behavior vectors via vanilla attention.

1. Introduction

Both the basic RNN and CNN encoders suffer from the problem that the fixed-size encoding vector may not support both short and long sequences well. The attention mechanism is then introduced to provide the ability to reference specific records dynamically in the decoder.

2. Related Work

- 1) Context Aware Recommendation : RNN based methods are studied for recommendation in recent years, though it suffers from several difficulties. RNN has high response time due to its hardships in parallelization and RNN embedding of user behaviors is a fix-sized, aggregated status.
- 2) Attention and Self-Attention : Instead of preserving only one single vector representation for the whole object in the encoder, attention keeps the vectors for each element as well, so that the decoder can reference these vectors at any decoding step. Self-attention considers inner-relations of the data at the encoder side as well.
- 3) Heterogeneous Behavior Modeling : Lots of works have been proposed to learn heterogeneous entity and relation representation by minimizing the distances of the linear projected entities in the relation-type semantic subspace.

3. Proposed Method

- 1) Raw Feature Spaces : Partition the user behavior tuples into different behavior groups according to the target object types. Use group-specific neural nets to build up the behavior embedding.
- 2) Behavior Embedding Spaces : Embed the raw features of any behavior tuple using the same embedding building block.
- 3) Latent Semantic Spaces : Linear projection to put embedding spaces into the same semantic space, where connections can be made.
- 4) Self-Attention Layer : Capture the inner-relationships among each semantic space.

- 5) Downstream Application Network : A point-wise or a pair-wise fully connected neural network. Vanilla attention is performed to produce the final context vector.

4. Experiment

- 1) Results on Single-type Behavior Dataset : ATRank performs better than the competitors especially when the user behavior is dense.
- 2) Results on Multi-type Behavior Dataset

5. Conclusion

This paper proposes an attention-based behavior modeling framework called ATRank. ATRank can model with heterogeneous user behaviors using only the attention model.