Hierarchical Context enabled Recurrent Neural Network for Recommendation
Kyungwoo Song*, Mingi Ji*, Sungrae Park, Il-Chul Moon
(* : Equal Contribution)

30-Second Summary

**Question**: Can we detect where the user’s interest changes?

**Our answer**: Yes!

**How?**: Interest drift assumption

- “If the user’s local context (for sub-sequence) and the current item are very different, the user’s temporary interest drift occurs.”

More specific: Hierarchical Context enabled Recurrent Neural Network (HCRNN)

- Incorporate the interest drift assumption
- Design hierarchical contexts (global, local, and temporary)
- Keep local and temporary contexts independently
- Introduce interest drift gate to capture the interest drift

**Motivation**

- A user history is a sequence of user orders or clicks, and the history represents the user’s interest.
- A long user history inevitably reflects the transitions of personal interests over time.
- We can predict next item better if we include modeling on an interest drift of users.

**Model Assumption**

- The user’s interest can be hierarchically ranging from general interest to a temporary (global, local, and temporary)
- Each hierarchical context have different abstract levels of information.
- Interest drift assumption

**Related Works**

- LSTM

**Sequential Recommendation**

- NARM [CIKM-17]: Focus on long-term interest
- STAMP [KDD-18]: Focus on short-term interest
- HCRNN (Ours): Focus on interest drift with long-term and short-term interest modeling

There are no studies which capture user's interest change with hierarchical context modeling

**Model Overview**

- Goal: Predict next item \( y_T \) given \( x_{1:T} \)
- \( x_T \): Current item embedding
- \( \theta \): Global context proportion for \( x_{1:T} \)
- \( M_{global} \): Global context memory
- \( c_t \): Local context (generated by global context, not current item)
- \( h_t \): Temporary context (generated by previous temporary context and current item, not local context)

**Methodology**

- Proportion for sequence (Variational Encoder)
- Attention weight
- Generation of local context with local context gate \( G_l^{(t)} \)
- Generation of temporary context \( h_t \) (separation with local context)
- Interest drift assumption
- Sigmoid function is not sharp
- Introduce the interest drift gate \( G_{id} \) to make \( h_t \) focus on the current input

**Inference**

- Variational inference by optimizing the evidence lower bound (ELBO)

- \( G_{id} \) has a relatively small value
- This small value is caused by the selection of different category items to the previous sub-sequence at \( t=16 \).

**Results**

1) Quantitative Results

- HCRNN-1 > Baselines (NARM, STAMP)
- Necessity of hierarchical context
- HCRNN-3 > HCRNN-2, HCRNN-1
- Interest drift assumption is experimentally justifiable.
- HCRNN-3+Bi > HCRNN-3
- Bi-channel attention with hierarchical contexts may improve the performance experimentally.

2) Context Embedding

- If the genre of the current item is different with previous items, \( r_t \circ G_{id}^{(t)} \) has a smaller value compared to the opposite situation.

3) Gate Analysis

- The bi-channel attentions distinguishes the attentions
- \( a_2^{(c)} \) focuses on the neighbor attention (short-term)
- \( a_2^{(d)} \) reads out through the whole sequence (long-term)

4) Case Study

- Sub-sequence1 (Action) Sub-sequence2 (Musical) Sub-sequence3 (Action/Romance)
- LSTMM, HCRNN, HCRNN-1, HCRNN-2, HCRNN-3, HCRNN-3+Bi

**Class**

- Action
- Musical
- Action/Romance

**Model Overview**

- HCRNN

**HCRNN-I**

- Direct connection between \( c_t \) and \( h_t \)

**HCRNN-2**

- \( r_t = \sigma(c_t \circ h_t \circ a_t \circ h_t \circ h_t) \)

**HCRNN-3**

- \( r_t = \sigma(a_t \circ h_t \circ a_t \circ h_t \circ h_t) \)

**HCRNN-3+Bi**

- \( a_2^{(c)} \) : attention based on the local context (Short-term dependency)
- \( a_2^{(d)} \) : attention based on the temporary context (Long-term dependency)