Context Aware Sequence Model

Kyungwoo Song
KAIST, South Korea
gtshs2@kaist.ac.kr

Abstract
Context modeling helps understand the data, such as sentence or user behavior. Contextual information captures the important underlying feature, and it enhances the relationship between data instances or hidden representations. As the importance of the sequential model grows, so does the importance of the sequential contextual modeling. Under the sequential data, we need to consider the context change over time. In this paper, we present our research works on context modeling and its dynamics modeling over time. Furthermore, we extend our research to handle the multi-granularity of sequential context modeling to consider rich context representations.

1 Introduction
The context denotes the generally related thought of the event, and it can be defined on the sentence or user’s behavior. The context modeling helps us to discover the clear meaning of the sentence or understand the user’s behavior. The recent statistical model has enabled us to capture the diverse context. We can model the context of sentence or user’s behavior with the probabilistic graphical model (PGM), and represent the context on the latent space. For example, Latent Dirichlet Allocation (LDA) model the generative process of documents with the context, topic proportion of each document [Blei et al., 2003]. Besides, [Gerrish and Blei, 2012] introduces a generative process of lawmaker’s voting with the context of user behavior.

Recently, the importance of sequential modeling has increased, and it is necessary to handle the context of sequential data. For sentence modeling, there are many kinds of research to incorporate the inductive bias of sequential order such as RNN, and Transformer. For sequential user behavior, movie and music streaming services have interested in recommending the next item, which users will click at the right time. Context modeling is helpful to predict the next words or the user’s next behavior in the sequence. However, static context modeling is not enough to reflect the context changes in the sentence or user’s sequential behavior. We need to handle the dynamics of context to understand the sequential data deeply. Besides, hierarchical sequential context modeling is necessary to capture a more diverse context and represent the relationship between data or hidden representations with sequential information. Figure 1 visualize our propose three models in this thesis, context modeling which consider the relationship between user and sentence (NIPEN) [Song et al., 2018], sequential context modeling which handle the correlations between words (Bivariate Beta-LSTM) [Song et al., 2020], and hierarchical sequential context modeling (HCRNN) [Song et al., 2019].

2 Contributions
We have focused on context modeling and its application to the sequential data to understand the sentence and user’s behavior deeply with hierarchical context modeling.

First, we have focused on static context modeling of sentence and user behavior. We investigate the static context on the legislative roll-call data because legislative processes have both contents of bill (sentence) and quantitative record of legislator’s voting (user behavior). Under the legislative processes, it is challenging to consider the context of the bill (contents) and the contents of the legislator (ideal point) simultaneously. To solve the issue, we assume that contents and ideal points are composed of several topics and the probability of voting YEA increases proportionally to the conformity of the topic of bill and legislator’s ideal point for each
Sequential context modeling is still challenging to understand user behavior and sentence deeply. For sequential modeling of user behavior and sentence, Transformer based on self-attention and positional encoding shows the remarkable performance. However, most of the related works do not incorporate the context explicitly into the attention and positional encoding.

First, we extend the self-attention modeling into the PGM to incorporate the context of the sentence. Traditional multi-head self-attention handles each multi-head independently. However, each head needs to attend different parts to handle the diverse context of sentences and similar parts in the opposite case. Under the PGM, we propose generalized attention, which incorporates the relationship and context between multi-heads. Under the flexible attention module, we can understand the given sentence in detail.

Second, we incorporate the property of user behavior into the positional encoding. Traditional positional doesn’t consider the context of the user’s behavior pattern, and we can improve it by incorporating the stationary pattern for the entire time or locally fixed time. We can understand the user’s sequential behavior if we incorporate the user’s stationary and locally stationary property. We propose a novel time encoding with the kernel method [Williams and Rasmussen, 2006], which reflects the locally-stationary property of the user’s behavior.

4 Conclusions and Future Work
Context modeling helps us to understand user behavior and sentence deeply. We have proposed context modeling and its application to sequential modeling with a hierarchical structure to capture the diverse context of sentence and user behavior. As our future works, we will propose a multi-granularity attention mechanism that improves the traditional attention, which only compares the pair-wise component to compute the attention. We can generalize the pair-wise based traditional attention with multi-granularity attention, which computes attention with n-wise (n=2,3,...). Multi-granularity attention is the generalized version of traditional attention, and it can reflect a more rich context accurately.

References