How to Make Context More Useful? An Empirical Study on Context-Aware Neural Conversational Models

Zhiliang Tian,[1] **Rui Yan,**[2*] **Lili Mou,**[3] **Yiping Song,**[4] **Yansong Feng,**[2] **Dongyan Zhao**[2]
[1]Baidu Inc., China tianzhiliang@baidu.com
[2]Institute of Computer Science and Technology, Peking University, China
[3]Key Laboratory of High Confidence Software Technologies, MoE, China
Institute of Software, Peking University, China
[4]Institute of Network Computing and Information Systems, Peking University, China
{ruiyan,songyiping,yansong,zhaody}@pku.edu.cn doublepower.mou@gmail.com

Abstract

Generative conversational systems are attracting increasing attention in natural language processing (NLP). Recently, researchers have noticed the importance of context information in dialog processing, and built various models to utilize context. However, there is no systematic comparison to analyze how to use context effectively. In this paper, we conduct an empirical study to compare various models and investigate the effect of context information in dialog systems. We also propose a variant that explicitly weights context vectors by context-query relevance, outperforming the other baselines.

1 Introduction

Recently, human-computer conversation is attracting increasing attention due to its promising potentials and alluring commercial values. Researchers have proposed both retrieval methods (Ji et al., 2014; Yan et al., 2016) and generative methods (Ritter et al., 2011; Shang et al., 2015) for automatic conversational systems. With the success of deep learning techniques, neural networks have demonstrated powerful capability of learning human dialog patterns; given a user-issued utterance as an input query q, the network can generate a reply r, which is usually accomplished in a sequence-to-sequence (`Seq2Seq`) manner (Shang et al., 2015).

In the literature, there are two typical research setups for dialog systems: single-turn and multi-turn. Single-turn conversation is, perhaps, the simplest setting where the model only takes q into consideration when generating r (Shang et al., 2015; Mou et al., 2016). However, most real-world dialogs comprise multiple turns. Previous utterances (referred to as *context* in this paper) could also provide useful information about the dialog status and are the key to coherent multi-turn conversation.

Existing studies have realized the importance of context, and proposed several context-aware conversational systems. For example, Yan et al. (2016) directly concatenate context utterances and the current query; others use hierarchical models, first capturing the meaning of individual utterances and then integrating them as discourses (Serban et al., 2016). There could be several ways of combining context and the current query, e.g., pooling or concatenation (Sordoni et al., 2015). Unfortunately, previous literature lacks a systematic comparison of the above methods.

In this paper, we conduct an empirical study on context modeling in `Seq2Seq`-like conversational systems. We emphasize the following research questions:

- **RQ1**. *How can we make better use of context information?* Our study shows that hierarchical models are generally better than non-hierarchical ones. We also propose a variant of context integration that explicitly weights a context vector by its relevance measure, outperforming simple vector pooling or concatenation.
- **RQ2**. *What is the effect of context on neural dialog systems?* We find context information is useful to neural conversational models. It yields longer, more informative and diversified replies.

To sum up, the contributions of this paper are two-fold: (1) We conduct a systematic study on context modeling in neural conversational models. (2) We further propose an explicitly con-

*Corresponding author.

Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 231–236
Vancouver, Canada, July 30 - August 4, 2017.
https://doi.org/10.18653/v1/P17-2036