ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9747307

RETRIEVAL BIAS AWARE ENSEMBLE MODEL FOR CONDITIONAL SENTENCE GENERATION

Yiping Song[1], *Zheng Xie*[*1], *Jianping Li*[1], *Luchen Liu*[2], *Ming Zhang*[2], *Zhiliang Tian*[3]

[1] National University of Defense Technology, Changsha, China
[2] The School of Computer Science, Peking University, Beijing, China
[3] The Hong Kong University of Science and Technology, Hong Kong SAR, China
{songyiping,xiezheng81, jianpingli65}@nudt.edu.cn
{liuluchen,mzhang_cs}@pku.edu.cn, ztianac@cse.ust.hk

ABSTRACT

Conditional sentence generation aims to generate proper target sentences with the given condition, and has shown great promise in many text generation applications such as dialogue systems and poetry generation. The ensemble of retrieval and generation-based models retrieve texts according to the input condition to assist the generation-based model. Those approaches obtain great performance tasks as they can absorb both merits to generate informative and coherent sentences. However, the input condition and its retrieved results are usually not highly consistent due to the quality of retrieval. It leads to a retrieval bias between the condition and its retrieved result, and then text generation augmented by such results becomes unreliable. To fix this issue, we propose RBAEM, a Retrieval Bias Aware Ensemble Model. RBAEM employs two CVAEs (Conditional variational Auto-encoder) to represent the retrieved target and the ground truth with latent vectors, and then diminishes the bias by decreasing the distance of two corresponding distributions. The extensive experiments on two tasks show that the proposed methods excel the existing state-of-the-art generation models.

Index Terms— Generation, Ensemble, Retrieval Bias

1. INTRODUCTION

Conditional sentence generation is one of the basic tasks in the field of natural language processing. It aims to generate a proper target sentence under the constraints of the given condition. Many applications such as machine translation [1, 2], dialogue systems [3, 4, 5, 6] and poem generation [7, 8], are of substantial value both in industry and academia. Previous studies on this task can be roughly categorized into three groups: the retrieval-based methods, the generation-based methods, and the ensemble of these two.

Given the condition, the retrieval-based methods [9, 10, 11] retrieve the most relevant condition along with the corresponding target sentence. Since the corpus in the repository is from human beings, the retrieved sentences are usually vivid and informative, where the lack of informativeness is a major issue in dialog systems [12, 13, 14]. The generation-based methods [15, 16] tailor a specific target sentence for the given condition, so the generated sentences are highly consistent with the given condition. Recently, the ensemble of retrieval-based model and generation-based methods have become the mainstream for this task as it can absorb both merits [17, 18].

Some ensemble methods uses a re-ranker [18, 19] to combine the generation and retrieval models. Song et al. [18] use a GBDT-based classifier to score all the candidate sentences from both retrieval and generation modules. Tanaka et al. [19] sort these candidates using several hand-crafted features. The majority of works integrate the results of two methods more deeply [20, 17, 21, 22]. Tian et al. [20] propose to abstract the sentences of different clusters in the repository, and use the abstracted sentences to assist the target generation process. Zhang et al. [17] propose to successively feed the outputs of the condition encoder and the prototype (retrieved sentence) encoder to the decoder of the target sentence generation. These methods directly fuse the retrieved sentence into the generation model and ignore the fact that there remains an inconsistency between the given condition and the retrieved condition, which we call retrieval bias in this paper.

To consider the retrieval bias between the original inputs and their retrieved results, researchers proposed to edit the retrieved sentences by replacing or deleting the tokens. Wu et al. [23] proposed to construct an edit vector by explicitly encoding the lexical differences between the input text and its retrieved text. Cai et al. [24] extract the sentence skeleton according to the gap between the input and the retrieved results, and then revise the retrieved results by removing the tokens that mismatch the skeleton. Those methods focus on the token-level modifications but may be hard to model the non-lexical difference between the input and its retrieved text.

In this paper, we propose to exclude the retrieval bias between the given condition and retrieved condition, so it

*Corresponding author