ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414915

TOWARDS EFFICIENTLY DIVERSIFYING DIALOGUE GENERATION VIA EMBEDDING AUGMENTATION

Yu Cao[†], *Liang Ding*[†*], *Zhiliang Tian*[§], *and Meng Fang*[‡]

[†] School of Computer Science, The University of Sydney, Australia, [‡] Tencent Robotics X, China
[§] Department of CSE, The Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

Dialogue generation models face the challenge of producing generic and repetitive responses. Unlike previous augmentation methods that mostly focus on token manipulation and ignore the essential variety within a single sample using hard labels, we propose to promote the generation diversity of the neural dialogue models via soft embedding augmentation along with soft labels in this paper. Particularly, we select some key input tokens and fuse their embeddings together with embeddings from their semantic-neighbor tokens. The new embeddings serve as the input of the model to replace the original one. Besides, soft labels are used in loss calculation, resulting in multi-target supervision for a given input. Our experimental results on two datasets illustrate that our proposed method is capable of generating more diverse responses than raw models while remains a similar n-gram accuracy that ensures the quality of generated responses.

***Index Terms*—** Dialogue generation, diversity, data augmentation, embedding, natural language processing

1. INTRODUCTION

Dialogue generation is important in AI applications, in which a model can generate responses for user-issued dialogue history. Former dialogue datasets make the end-to-end training of deep neural models possible, such as DailyDialog [1], PersonaChat [2], etc., and deep neural networks including Seq2seq with attention [3], Transformers [4] have already shown their capability in generating conversation replies. Dialogue generation can be regarded as a **many-to-many** (combination of one-to-many and many-to-one) problem as one specific response may be reasonable for multiple histories and vice versa. However, current models tend to produce generic sentences such as “I don’t know”, caused by the current inherent deterministic training objective as well as insufficient diversity and limited quality of current datasets [5, 6].

Generating more various dialogue responses remains a hard task. In order to tackle it, some researchers try to refine the training objective with extra constraint items [5] or modify the criterion that encourages models to decode more diverse words [7, 8, 9, 10]. However, such models are more likely to generate ungrammatical or uncorrelated responses and the superiority of constructed objectives to cross-entropy loss remains unknown. Another kind of approach enhances the training data directly. Data filtering is commonly used in machine learning and extended to dialogue generation by removing samples with generic responses based on entropy [11] or the predictions of a Seq2seq model [12]. Obtaining more training samples is also a possible solution. Word replacement extends the data scales by randomly replacing original tokens with others based on vocabulary distributions [13, 14]. In addition, new augmented sentences can also be obtained via a learnable model that even keeps interaction between the generation model and both of them will be trained jointly [15]. But these methods merely include more training data and still bridge a one-to-one mapping between input and response within a single sample. And they also add great extra computational load. So it is necessary to propose a simple and effective method regarding the essence of this problem.

In this paper, we propose an embedding augmentation method for dialogue models, where not only the training objective encourages various output, but also soft embeddings enhance the diversity within a single sample. It is inspired by the recent success of the mixup approach that combines training pairs of samples and labels convexly into a single one [16]. Particularly, the original embeddings of tokens in the training samples will be randomly replaced by augmented ones, which are mixtures of the raw one and several semantically similar embeddings conditioned on their distribution. And these selected positions will use soft labels instead of hard one-hot vectors. Such a mechanism ensures that the model can learn a *soft word distribution* rather than a fixed word in both input and output, which is consistent with the multi-source and multi-target purpose, benefiting the generating diversity without using any additional sample. Compared to previous similar work [17, 18], we use both soft embedding and soft labels for a more flexible training process. A compact in-domain similar token prediction model is utilized instead of a deep language model, realizing a more efficient

[*]This work was done when Yu Cao and Liang Ding were interning at Tencent AI Lab. Code is available: `https://github.com/caoyu-noob/embedding_augmentation`.