

## Variational Explanation Generator: Generating Explanation for Natural Language Inference Using Variational Auto-Encoder

**Authors :** Zhen Cheng, Xinyu Dai, Shujian Huang, Jiajun Chen

**Abstract :** Recently, explanatory natural language inference has attracted much attention for the interpretability of logic relationship prediction, which is also known as explanation generation for Natural Language Inference (NLI). Existing explanation generators based on discriminative Encoder-Decoder architecture have achieved noticeable results. However, we find that these discriminative generators usually generate explanations with correct evidence but incorrect logic semantic. It is due to that logic information is implicitly encoded in the premise-hypothesis pairs and difficult to model. Actually, logic information identically exists between premise-hypothesis pair and explanation. And it is easy to extract logic information that is explicitly contained in the target explanation. Hence we assume that there exists a latent space of logic information while generating explanations. Specifically, we propose a generative model called Variational Explanation Generator (VariationalEG) with a latent variable to model this space. Training with the guide of explicit logic information in target explanations, latent variable in VariationalEG could capture the implicit logic information in premise-hypothesis pairs effectively. Additionally, to tackle the problem of posterior collapse while training VariationalEG, we propose a simple yet effective approach called Logic Supervision on the latent variable to force it to encode logic information. Experiments on explanation generation benchmark—explanation-Stanford Natural Language Inference (e-SNLI) demonstrate that the proposed VariationalEG achieves significant improvement compared to previous studies and yields a state-of-the-art result. Furthermore, we perform the analysis of generated explanations to demonstrate the effect of the latent variable.

**Keywords :** natural language inference, explanation generation, variational auto-encoder, generative model

**Conference Title :** ICDAR 2020 : International Conference on Document Analysis and Recognition

**Conference Location :** Jerusalem, Israel

**Conference Dates :** November 26-27, 2020