

The generative approach assumes, it is possible to capture a latent state z and use this state to generate suitable counterfactual candidates. We condition the generation procedure on the factual instance to generate counterfactuals that show sparse differences to the original sequence. The core idea is to sample randomly $e^* \sim p(z|e)$ to generate counterfactual candidates. We can sort each candidate by their *viability* and choose top-K contenders as viable counterfactuals. There are a multitude of approaches to generate the counterfactuals. However, we will limit our exploration to sequential Variational Autoencoders (VAEs) and sequential Generative Adversarial Models (GANs). Both techniques allow us to sample from a smooth latent space conditioned on the factual sequence. VAEs approximate $p(z|e)$ by trying to reconstruct the input using Monte-Carlo methods. GANs require a generator model and a discriminator model. The generator model attempts to fool the discriminator model by generating results that closely resemble true process instances. In contrast, the discriminator tries to distinguish generated instances from real instances.

Model Architecture

TBD