

# BALD-VAE: Generative Active Learning based on the uncertainties of Both Labeled and Unlabeled Data

Deep learning has shown outstanding performances on real world problems, but acquiring sufficient labeled data to train a model is still an on-going issue.

Specifically, manually labeling data is time-consuming and costly. One approach to tackle this issue is active learning. Recently, pool-based methods and generative methods are widely studied among various approaches of active learning. Especially in the uncertainty pool-based methods, a small labeled data set and a large unlabeled data set are given. A model is trained on the labeled data set and then observes the unlabeled data set. The trained model ranks the unlabeled data in order of uncertainty to select the data which has the highest uncertainty. In the generative methods, a generative model is used to generate informative samples. In the previous studies of the uncertainty pool-based active learning, however, the uncertainty of labeled data was not considered. Thus, we propose a new Bayesian active learning by disagreement with variational autoencoder (BALD-VAE), which considers the uncertainty of labeled data when generating informative samples. Basically following the uncertainty pool-based active learning with BALD, the proposed algorithm also utilizes the concept of generative active learning to generate informative data using VAE. Then, the generated data complement the highly uncertain labeled data. To demonstrate the effectiveness, the proposed method is tested on MNIST and CIFAR10 data sets and shown to outperform the previous algorithms.