

Hot Water Leak Detection Using Variational Autoencoder Model

At the Southern California Gas Company, we need to distinguish between various types of abnormal gas usage. SoCalGas has built predictive models to detect such instances. For example, we need to identify water leaks from gas leaks and respond appropriately.

SoCalGas has very large datasets collected from almost six million residential smart meters in our region. The data collected for analyses and modeling includes gas consumption, temperature, service operation times, meter location etc. Via customer pattern analyses data, SoCalGas attempts to identify the different patterns of water leaks.

However, to build predictive models, requires very large sets of accurately labeled and cleaned data. There are a practical resource limitations to a manual labeling process, especially with consumption data that can be noisy. Another modeling challenge is that the meters, which are the data sources, are from different families and in different geographical locations.

For these reasons, we propose the use of a Variational Autoencoder (VAE) model as a semi-supervised learning method. With a certain amount of labeled data, we can train a much larger data set a) without bias from partial data and b) reduce the noise in the larger dataset. This approach has allowed us to differentiate and predict abnormal consumption patterns due to water leak.

The VAE is known as a generative model. The VAE model is an upgraded architecture of a regular autoencoder by replacing the usual deterministic function Q with a probabilistic function $q(z|x)$. A VAE model learns soft ellipsoidal regions in latent space by effectively force filling the gaps where labels are missing. The missing labels are able to be filled via this method when we apply the VAE model to the data. Also, the VAE model is encoded to the latent space and decoded to reconstruct the data. In this process, just like autoencoder, any noise is reduced.

In experiments, we compared VAE results with our consumption data analytics pattern recognition results and also with autoencoder results. This model showed efficient prediction of missing labels and prediction of properties of various areas. Depending on different ratio of labeled data and non-labeled data, the performance and accuracy of model were affected.