Hyper parameter of Periormer

October 16, 2023

For all methods, the input length is chosen from 24, 48, 96, 192, 336 for all datasets, the batch size is set to 32, and the learning rate in Adam optimizer is set to 10^{-4}. For all datasets, we split the training, validation, and test set by the ratio of 6:2:2. We set the hyper-parameters for all baselines to the best values achieved in the validation set. We use grid search for all methods in the hyperparameter tuning. The R^2 is choosen from 60%,70%,80%,90%, the head number of multi-head attention is chosen from 8,16, the dimension of the output of multi-head attention is chosen as 256, 512.

For FEDformer, we set the mode M in the decomposition block to 64, the number of orthogonal basis k is set to 3, the fixed kernel size in the average pooling layer is set to 24.

For Informer, we set the number of attention heads h to 16 and the stride in the max pooling to 2. The length of encoder's input sequence and decoder's start token is chosen from 24,48,96,168,336,480,720 for ETTh1,WTH and Stock1 and stock 2 dataset. Also, 24,48,96,192,288,480,672 for ETTm1, ETTm2, NDBC and synthetic datasets.

For Autoformer, the hyper-parameter c of Autocorrelation is set to 2. The numbers of encoder layers and decoder layers are set to 2 and 1, respectively.

For Reformer, the number of layers is set to 3. The embedding size is set to 1024 in Reformer and 512 for the other approaches. R^2 is set to 80% for the truncation of the high component in Periormer.

All experiments are conducted in PyTorch on a single NVIDIA GeForce RTX 3060 Ti GPU.