The validation scheme is based on <u>seq2seq-rnn-with-gru</u>, and cleaned data is from <u>data-without-drift</u> and Kalman filter is from https://www.kaggle.com/teejmahal20/single-model-lgbm-kalman-filter and the added feature is from wavenet-with-1-more-feature. I also used ragnar's data in this version clean-kalman. The Wavenet is based on https://github.com/philipperemy/keras-tcn, https://github.com/peustr/wavenet and https://github.com/basveeling/wavenet and also https://www.kaggle.com/wimwim/wavenet-lstm. If any refrence is not mentioned it was not intentional, please add them in comments. Previous versions were mainly based on https://www.kaggle.com/wimwim/wavenet-lstm In [1]: !pip install --no-warn-conflicts -q tensorflow-addons In [2]: from tensorflow.keras.layers import (TimeDistributed, Dropout, BatchNormalization, Flatten, Convolution 1D, Activation, Input, Dense, LSTM, Lambda, Bidirectional, Add, AveragePooling1D, Multiply, GRU, GRUCell, LSTMCell, SimpleRNN Cell, SimpleRNN, TimeDistributed, RNN, RepeatVector, Conv1D, MaxPooling1D, Concatenate, GlobalAveragePool ing1D, UpSampling1D) from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping, Callback, ReduceLROnPlateau, Lea rningRateScheduler from tensorflow.keras.losses import binary crossentropy, categorical crossentropy, mean squared error from tensorflow.keras.optimizers import Adam, RMSprop, SGD from tensorflow.keras.utils import Sequence, to_categorical from tensorflow.keras import losses, models, optimizers from tensorflow.keras import backend as K import tensorflow as tf from typing import List, NoReturn, Union, Tuple, Optional, Text, Generic, Callable, Dict from sklearn.metrics import f1_score, cohen_kappa_score, mean_squared_error from logging import getLogger, Formatter, StreamHandler, FileHandler, INFO from sklearn.model selection import KFold, GroupKFold from tqdm import tqdm_notebook as tqdm from contextlib import contextmanager from joblib import Parallel, delayed from IPython.display import display from sklearn import preprocessing import tensorflow_addons as tfa import scipy.stats as stats import random as rn import pandas as pd import numpy as np import scipy as sp import itertools import warnings import time import pywt import os import gc warnings.simplefilter('ignore') warnings.filterwarnings('ignore') pd.set option('display.max columns', 1000) pd.set option('display.max rows', 500) %matplotlib inline In [3]: EPOCHS=120 NNBATCHSIZE=20 BATCHSIZE = 4000SEED = 321SELECT = True SPLITS = 5LR = 0.001fe config = [(True, 4000), In [4]: def init logger(): handler = StreamHandler() handler.setLevel(INFO) handler.setFormatter(Formatter(LOGFORMAT)) fh handler = FileHandler('{}.log'.format(MODELNAME)) fh handler.setFormatter(Formatter(LOGFORMAT)) logger.setLevel(INFO) logger.addHandler(handler) logger.addHandler(fh_handler) In [5]: @contextmanager def timer(name : Text): t0 = time.time()yield logger.info(f'[{name}] done in {time.time() - t0:.0f} s') COMPETITION = 'ION-Switching' logger = getLogger(COMPETITION) LOGFORMAT = '% (asctime)s % (levelname)s % (message)s' MODELNAME = 'WaveNet' In [6]: def seed everything(seed : int) -> NoReturn : rn.seed(seed) np.random.seed(seed) os.environ['PYTHONHASHSEED'] = str(seed) tf.random.set_seed(seed) seed everything(SEED) In [7]: def read data(base: os.path.abspath) -> Tuple[pd.DataFrame, pd.DataFrame, pd.DataFrame]: train = pd.read csv('/kaggle/input/clean-kalman/train clean kalman.csv', dtype={'time': np.float32, 'signal': np.float32, 'open channels':np.int32}) test = pd.read_csv('/kaggle/input/clean-kalman/test_clean_kalman.csv', dtype={'time': np.float32, 'signal': np.float32}) sub = pd.read csv('/kaggle/input/liverpool-ion-switching/sample submission.csv', dtype={'time': np return train, test, sub In [8]: def batching(df : pd.DataFrame, batch_size : int) -> pd.DataFrame : df['group'] = df.groupby(df.index//batch size, sort=False)['signal'].agg(['ngroup']).values df['group'] = df['group'].astype(np.uint16) return df In [9]: def reduce mem usage(df: pd.DataFrame, verbose: bool = True) -> pd.DataFrame: numerics = ['int16', 'int32', 'int64', 'float16', 'float32', 'float64'] start mem = df.memory usage().sum() / 1024**2for col in df.columns: col type = df[col].dtypes if col_type in numerics: c min = df[col].min() c max = df[col].max()if str(col_type)[:3] == 'int': if (c min > np.iinfo(np.int32).min and c max < np.iinfo(np.int32).max):</pre> df[col] = df[col].astype(np.int32) elif (c_min > np.iinfo(np.int64).min and c max < np.iinfo(np.int64).max):</pre> df[col] = df[col].astype(np.int64) else: if (c min > np.finfo(np.float16).min and c max < np.finfo(np.float16).max):</pre> df[col] = df[col].astype(np.float16) elif (c_min > np.finfo(np.float32).min and c_max < np.finfo(np.float32).max):</pre> df[col] = df[col].astype(np.float32) df[col] = df[col].astype(np.float64) end mem = df.memory usage().sum() / 1024**2reduction = (start mem - end mem) / start mem msg = f'Mem. usage decreased to {end mem:5.2f} MB ({reduction * 100:.1f} % reduction)' if verbose: print(msg) return df In [10]: def lag_with_pct_change(df : pd.DataFrame, shift_sizes : Optional[List]=[1, 2], add_pct_change : Optional[bool]=False, add pct change lag : Optional[bool]=False) -> pd.DataFrame: for shift_size in shift_sizes: df['signal_shift_pos_'+str(shift_size)] = df.groupby('group')['signal'].shift(shift_size).filln a(0)df['signal shift neg '+str(shift size)] = df.groupby('group')['signal'].shift(-1*shift size).fi llna(0) if add pct change: df['pct change'] = df['signal'].pct change() if add_pct_change_lag: for shift size in shift sizes: df['pct change shift pos '+str(shift size)] = df.groupby('group')['pct change'].shift(s df['pct change shift neg '+str(shift size)] = df.groupby('group')['pct change'].shift(-1*shift size).fillna(0) return df In [11]: def run feat enginnering(df : pd.DataFrame, create all data feats : bool, batch size : int) -> pd.DataFrame: df = batching(df, batch_size=batch_size) if create_all_data_feats: df = lag_with_pct_change(df, [1, 2, 3], add_pct_change=False, add pct change lag=False) df['signal 2'] = df['signal'] ** 2 return df In [12]: def feature selection(df : pd.DataFrame, df_test : pd.DataFrame) -> Tuple[pd.DataFrame , pd.DataFrame, List]: use cols = [col for col in df.columns if col not in ['index', 'group', 'open_channels', 'time']] df = df.replace([np.inf, -np.inf], np.nan) df_test = df_test.replace([np.inf, -np.inf], np.nan) for col in use_cols: col mean = pd.concat([df[col], df test[col]], axis=0).mean() df[col] = df[col].fillna(col mean) df test[col] = df test[col].fillna(col mean) gc.collect() return df, df test, use cols In [13]: def augment(X: np.array, y:np.array) -> Tuple[np.array, np.array]: X = np.vstack((X, np.flip(X, axis=1)))y = np.vstack((y, np.flip(y, axis=1)))return X, y In [14]: def run cv model by batch(train : pd.DataFrame, test : pd.DataFrame, splits : int, batch col : Text, feats : List, sample submission: pd.DataFrame, nn epochs : int, nn batch size : int) -> NoReturn: seed everything(SEED) K.clear session() config = tf.compat.v1.ConfigProto(intra op parallelism threads=1,inter op parallelism threads=1) sess = tf.compat.v1.Session(graph=tf.compat.v1.get default graph(), config=config) tf.compat.v1.keras.backend.set session(sess) oof_ = np.zeros((len(train), 11)) $preds_{=} = np.zeros((len(test), 11))$ target = ['open channels'] group = train['group'] kf = GroupKFold(n splits=5) splits = [x for x in kf.split(train, train[target], group)] new splits = [] for sp in splits: new split = [] new split.append(np.unique(group[sp[0]])) new split.append(np.unique(group[sp[1]])) new_split.append(sp[1]) new splits.append(new split) tr = pd.concat([pd.get dummies(train.open channels), train[['group']]], axis=1) tr.columns = ['target '+str(i) for i in range(11)] + ['group'] target cols = ['target '+str(i) for i in range(11)] train_tr = np.array(list(tr.groupby('group').apply(lambda x: x[target_cols].values))).astype(np.flo at32) train = np.array(list(train.groupby('group').apply(lambda x: x[feats].values))) test = np.array(list(test.groupby('group').apply(lambda x: x[feats].values))) for n_fold, (tr_idx, val_idx, val_orig_idx) in enumerate(new_splits[0:], start=0): train x, train y = train[tr idx], train tr[tr idx] valid_x, valid_y = train[val_idx], train_tr[val_idx] **if** n fold < 2: train_x, train_y = augment(train_x, train_y) gc.collect() shape_ = (None, train_x.shape[2]) model = Classifier(shape) cb lr schedule = LearningRateScheduler(lr schedule) cb prg = tfa.callbacks.TQDMProgressBar(leave epoch progress=False, leave overall progress=False, show_epoch_progress=False, show_overall_progress=True) model.fit(train x, train y, epochs=nn epochs, callbacks=[cb prg, cb lr schedule, MacroF1(model, valid x,valid y)], batch size=nn batch size,verbose=0, validation data=(valid x, valid y)) preds f = model.predict(valid x) $\texttt{f1_score} = \texttt{f1_score} (\texttt{np.argmax}(\texttt{valid_y, axis=2}).\texttt{reshape} (-1), \quad \texttt{np.argmax}(\texttt{preds_f, axis=2}).\texttt{reshape} (-1), \\$ e(-1), average = 'macro') logger.info(f'Training fold {n fold + 1} completed. macro f1 score : {f1 score :1.5f}') preds f = preds f.reshape(-1, preds f.shape[-1]) oof_[val_orig_idx,:] += preds f te preds = model.predict(test) te_preds = te_preds.reshape(-1, te_preds.shape[-1]) preds += te preds / SPLITS fl_score_ =fl_score(np.argmax(train_tr, axis=2).reshape(-1), np.argmax(oof_, axis=1), average = 'm acro') logger.info(f'Training completed. oof macro f1 score : {f1 score :1.5f}') sample submission['open channels'] = np.argmax(preds , axis=1).astype(int) sample_submission.to_csv('submission.csv', index=False, float_format='%.4f') display(sample_submission.head()) np.save('oof.npy', oof_) np.save('preds.npy', preds) return In [15]: **def** lr schedule(epoch): **if** epoch < 40: lr = LRelif epoch < 50:</pre> lr = LR / 3elif epoch < 60:</pre> lr = LR / 6elif epoch < 75:</pre> lr = LR / 9elif epoch < 85:</pre> lr = LR / 12elif epoch < 100:</pre> lr = LR / 15else: lr = LR / 50return lr In [16]: class Mish(tf.keras.layers.Layer): def init (self, **kwargs): super(Mish, self). init (**kwargs) self.supports masking = True def call(self, inputs): return inputs * K.tanh(K.softplus(inputs)) def get config(self): base config = super(Mish, self).get config() return dict(list(base config.items()) + list(config.items())) def compute_output_shape(self, input_shape): return input shape def mish(x): return tf.keras.layers.Lambda(lambda x: x*K.tanh(K.softplus(x)))(x) from tensorflow.keras.utils import get custom objects from tensorflow.keras.layers import Activation get custom objects().update({'mish': Activation(mish)}) In [17]: import tensorflow as tf from tensorflow.keras.layers import Layer from tensorflow.keras import initializers from tensorflow.keras import regularizers from tensorflow.keras import constraints class Attention(Layer): """Multi-headed attention layer.""" def init (self, hidden size, num heads = 8, attention dropout=.1, trainable=True, name='Attention'): if hidden size % num heads != 0: raise ValueError ("Hidden size must be evenly divisible by the number of heads.") self.hidden_size = hidden_size self.num heads = num heads self.trainable = trainable self.attention dropout = attention dropout self.dense = tf.keras.layers.Dense(self.hidden size, use bias=False) super(Attention, self).__init__(name=name) def split heads(self, x): """Split x into different heads, and transpose the resulting value. The tensor is transposed to insure the inner dimensions hold the correct values during the matrix multiplication. x: A tensor with shape [batch size, length, hidden size] Returns: A tensor with shape [batch size, num heads, length, hidden size/num heads] with tf.name_scope("split_heads"): batch size = tf.shape(x)[0]length = tf.shape(x)[1]# Calculate depth of last dimension after it has been split. depth = (self.hidden_size // self.num_heads) # Split the last dimension x = tf.reshape(x, [batch_size, length, self.num_heads, depth]) # Transpose the result **return** tf.transpose(x, [0, 2, 1, 3]) def combine heads(self, x): """Combine tensor that has been split. x: A tensor [batch_size, num_heads, length, hidden size/num heads] Returns: A tensor with shape [batch size, length, hidden size] with tf.name_scope("combine_heads"): batch size = tf.shape(x)[0]length = tf.shape(x)[2]x = tf.transpose(x, [0, 2, 1, 3]) # --> [batch, length, num heads, depth]return tf.reshape(x, [batch_size, length, self.hidden_size]) def call(self, inputs): """Apply attention mechanism to inputs. inputs: a tensor with shape [batch size, length x, hidden size] Attention layer output with shape [batch size, length x, hidden size] # Google developper use tf.layer.Dense to linearly project the queries, keys, and values. q = self.dense(inputs) k = self.dense(inputs) v = self.dense(inputs) q = self.split heads(q) k = self.split heads(k)v = self.split heads(v)# Scale q to prevent the dot product between q and k from growing too large. depth = (self.hidden_size // self.num_heads) q *= depth ** -0.5logits = tf.matmul(q, k, transpose b=True) # logits += self.bias weights = tf.nn.softmax(logits, name="attention_weights") if self.trainable: weights = tf.nn.dropout(weights, 1.0 - self.attention_dropout) attention output = tf.matmul(weights, v) attention output = self.combine heads(attention output) attention output = self.dense(attention output) return attention_output def compute_output_shape(self, input_shape): return tf.TensorShape(input shape) In [18]: | def categorical_focal_loss(gamma=2.0, alpha=0.25): Implementation of Focal Loss from the paper in multiclass classification $loss = -alpha*((1-p)^gamma)*log(p)$ Parameters: alpha -- the same as wighting factor in balanced cross entropy gamma -- focusing parameter for modulating factor (1-p) Default value: gamma -- 2.0 as mentioned in the paper alpha -- 0.25 as mentioned in the paper def focal_loss(y_true, y_pred): # Define epsilon so that the backpropagation will not result in NaN # for 0 divisor case epsilon = K.epsilon() # Add the epsilon to prediction value #y pred = y pred + epsilon # Clip the prediction value y_pred = K.clip(y_pred, epsilon, 1.0-epsilon) # Calculate cross entropy cross_entropy = -y_true*K.log(y_pred) # Calculate weight that consists of modulating factor and weighting factor weight = alpha * y_true * K.pow((1-y_pred), gamma) # Calculate focal loss loss = weight * cross entropy # Sum the losses in mini batch loss = K.sum(loss, axis=1)return loss return focal loss In [19]: def WaveNetResidualConv1D(num_filters, kernel_size, stacked_layer): def build residual block(l input): resid_input = l_input for dilation rate in [2**i for i in range(stacked_layer)]: l sigmoid conv1d = Conv1D(num filters, kernel size, dilation rate=dilation rate, padding='same', activation='sigmoid')(l_input) l tanh conv1d = Conv1D(num filters, kernel size, dilation rate=dilation rate, padding='same', activation='mish')(l input) l_input = Multiply()([l_sigmoid_conv1d, l_tanh_conv1d]) l_input = Conv1D(num_filters, 1, padding='same')(l_input) resid input = Add()([resid input ,l input]) return resid input return build residual block def Classifier(shape): num_filters_ = 16 kernel size = 3stacked_layers_ = [12, 8, 4, 1] l input = Input(shape=(shape)) x = Conv1D(num filters , 1, padding='same')(l input) x = WaveNetResidualConv1D(num_filters_, kernel_size_, stacked_layers_[0])(x) x = Conv1D(num_filters_*2, 1, padding='same')(x) x = WaveNetResidualConv1D(num_filters_*2, kernel_size_, stacked_layers_[1])(x) x = Conv1D(num filters *4, 1, padding='same')(x)x = WaveNetResidualConv1D(num_filters_*4, kernel_size_, stacked_layers_[2])(x) x = Conv1D(num_filters_*8, 1, padding='same')(x) $x = WaveNetResidualConv1D(num_filters_*8, kernel_size_, stacked_layers_[3])(x)$ l output = Dense(11, activation='softmax')(x) model = models.Model(inputs=[l_input], outputs=[l_output]) opt = Adam(lr=LR) opt = tfa.optimizers.SWA(opt) model.compile(loss=losses.CategoricalCrossentropy(), optimizer=opt, metrics=['accuracy']) return model In [20]: def Classifierx(shape): inp = Input(shape=(shape)) x = Bidirectional(GRU(256, return sequences=True))(inp)x = Attention(512)(x)x = TimeDistributed(Dense(256, activation='mish'))(x)x = TimeDistributed(Dense(128, activation='mish'))(x) out = TimeDistributed(Dense(11, activation='softmax', name='out'))(x) model = models.Model(inputs=inp, outputs=out) opt = Adam(lr=LR) opt = tfa.optimizers.SWA(opt) model.compile(loss=losses.CategoricalCrossentropy(), optimizer=opt, metrics=['accuracy']) return model In [21]: class MacroF1 (Callback): def init (self, model, inputs, targets): self.model = model self.inputs = inputs self.targets = np.argmax(targets, axis=2).reshape(-1) def on epoch end(self, epoch, logs): pred = np.argmax(self.model.predict(self.inputs), axis=2).reshape(-1) score = f1 score(self.targets, pred, average="macro") print(f' F1Macro: {score:.5f}') In [22]: def normalize(train, test): train input mean = train.signal.mean() train input sigma = train.signal.std() train['signal'] = (train.signal-train input mean)/train input sigma test['signal'] = (test.signal-train input mean)/train input sigma return train, test In [23]: | def run everything(fe config : List) -> NoReturn: not feats cols = ['time'] target col = ['open_channels'] init logger() with timer(f'Reading Data'): logger.info('Reading Data Started ...') base = os.path.abspath('/kaggle/input/liverpool-ion-switching/') train, test, sample submission = read data(base) train, test = normalize(train, test) logger.info('Reading and Normalizing Data Completed ...') with timer(f'Creating Features'): logger.info('Feature Enginnering Started ...') for config in fe config: train = run_feat_enginnering(train, create_all_data_feats=config[0], batch_size=config[1]) test = run feat enginnering(test, create all data feats=config[0], batch size=config[1]) train, test, feats = feature selection(train, test) logger.info('Feature Enginnering Completed ...') with timer(f'Running Wavenet model'): logger.info(f'Training Wavenet model with {SPLITS} folds of GroupKFold Started ...') run cv model by batch (train, test, splits=SPLITS, batch col='group', feats=feats, sample submis sion=sample_submission, nn_epochs=EPOCHS, nn batch size=NNBATCHSIZE) logger.info(f'Training completed ...')

