NO2_RNN2-Copy1

November 30, 2017

```
In [1]: import numpy as np
        import pandas as pd
        import os
        import matplotlib.pyplot as plt
        import tensorflow as tf
        from keras.backend.tensorflow_backend import set_session
        config = tf.ConfigProto()
        config.gpu_options.per_process_gpu_memory_fraction = 0.3
        set_session(tf.Session(config=config))
Using TensorFlow backend.
In [2]: os.listdir('data')
Out[2]: ['AllNO2_QH.csv',
         'AllPM_QH.csv',
         'Env_QH.csv',
         'GradientTemp_15minDataSet.csv',
         'micro_sud3.pkl',
         'micro_sud3_normalized.pkl',
         'Patm_15minDataSet.csv',
         'pickles']
In [3]: df = pd.read_pickle('data/micro_sud3_normalized.pkl')
        df = df.reset_index()
        def split_dataframe(dataframe, percent):
            nb_rows = int(np.floor(percent * len(dataframe)))
            return dataframe[:nb_rows], dataframe[nb_rows:]
        def dataframe_to_xy(df, look_back):
            i = look_back
            while True:
                sequence = df.iloc[i - look_back:i]
                yield np.array(sequence[['NO2_61FD', 'NO2_61F0', 'NO2_61EF', 'temp', 'rh',\
                                          'tgrad', 'pressure', 'pluvio']]).reshape(look_back, 1
```

```
i += 1
                                       if i == len(df):
                                                 i = look_back
                   def dataframe_to_xy_test(df, look_back):
                             X_{\text{test}}, y_{\text{test}} = [], []
                              i = look back
                              while i < len(df):
                                       sequence = df.iloc[i - look_back:i]
                                       X_test.append(np.array(sequence[['NO2_61FD', 'NO2_61F0', 'NO2_61EF', 'temp', ':
                                                                                                       'tgrad', 'pressure', 'pluvio']]).reshape(look_back, 1
                                       y_test.append(np.array(df.iloc[i]['NO2_ref']))
                                       i += 1
                                       if i == len(df):
                                                 break
                             return np.array(X_test), np.array(y_test)
                   df_test, df_train = split_dataframe(df, 0.5)
                   df_valid, df_test = split_dataframe(df_test, 0.5)
                   X_train = dataframe_to_xy(df_train, 10)
                   X_valid = dataframe_to_xy(df_valid, 10)
                   X_test, y_test = dataframe_to_xy_test(df_test, 10)
In [4]: def dataframe_to_xy_no_generator(df, look_back):
                             X, y = [], []
                              i = look_back
                              while i < len(df):
                                       sequence = df.iloc[i - look_back:i]
                                       \label{eq:continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous_continuous
                                                                                                       'tgrad', 'pressure', 'pluvio']]).reshape(look_back, 1
                                       y.append(np.array(df.iloc[i]['NO2_ref']))
                                       if i == len(df):
                                                 break
                             X = np.array(X)
                              y = np.array(y)
                             X = X.reshape((X.shape[0], X.shape[1], X.shape[3]))
                             return X, y
In [5]: X_train, y_train = dataframe_to_xy_no_generator(df_train, 24)
                   X_valid, y_valid = dataframe_to_xy_no_generator(df_valid, 24)
                   X_test, y_test = dataframe_to_xy_no_generator(df_test, 24)
In [6]: from keras.layers import SimpleRNN, Dense, LSTM, GRU
                    from keras.models import Sequential
```

```
def simple_rnn_model(nb_units, input_dim, loss='mean_squared_error', optimizer='adam')
          model = Sequential()
          model.add(SimpleRNN(nb_units, input_shape=input_dim))
          model.add(Dense(nb_units, activation='relu'))
          model.add(Dense(1, kernel_initializer='normal'))
          model.compile(loss=loss, optimizer=optimizer)
          model.summary()
          return model
       def lstm_model(nb_units, input_dim, loss='mean_squared_error', optimizer='adam'):
          model = Sequential()
          model.add(LSTM(nb_units, input_shape=input_dim))
          model.add(Dense(nb_units, activation='relu'))
          model.add(Dense(1, kernel_initializer='normal'))
          model.compile(loss=loss, optimizer=optimizer)
          model.summary()
          return model
       def gru_model(nb_units, input_dim, loss='mean_squared_error', optimizer='adam'):
          model = Sequential()
          model.add(GRU(nb_units, input_shape=input_dim))
          model.add(Dense(nb_units, activation='relu'))
          model.add(Dense(1, kernel_initializer='normal'))
          model.compile(loss=loss, optimizer=optimizer)
          model.summary()
          return model
In [7]: model = simple_rnn_model(32, X_train.shape[1:])
Layer (type) Output Shape
                                        Param #
______
simple_rnn_1 (SimpleRNN)
                        (None, 32)
                                                 1312
                         (None, 32)
dense_1 (Dense)
                                                 1056
dense_2 (Dense) (None, 1)
_____
Total params: 2,401
Trainable params: 2,401
Non-trainable params: 0
In [8]: early_stopping = EarlyStopping(monitor='val_loss', verbose=1, mode='auto', patience=10
       history = model.fit(X_train, y_train, batch_size=32, epochs=5000, validation_data=(X_value)
```

from keras.callbacks import EarlyStopping

```
Train on 1103 samples, validate on 539 samples
Epoch 1/5000
Epoch 2/5000
Epoch 3/5000
Epoch 4/5000
Epoch 5/5000
Epoch 6/5000
Epoch 7/5000
Epoch 8/5000
Epoch 9/5000
Epoch 10/5000
Epoch 11/5000
Epoch 12/5000
Epoch 13/5000
Epoch 14/5000
Epoch 15/5000
Epoch 16/5000
Epoch 17/5000
Epoch 18/5000
Epoch 19/5000
Epoch 20/5000
Epoch 21/5000
Epoch 22/5000
Epoch 23/5000
Epoch 24/5000
```

```
Epoch 25/5000
Epoch 26/5000
Epoch 27/5000
Epoch 28/5000
Epoch 29/5000
Epoch 30/5000
Epoch 31/5000
Epoch 32/5000
Epoch 33/5000
Epoch 34/5000
Epoch 35/5000
Epoch 36/5000
Epoch 37/5000
Epoch 38/5000
Epoch 39/5000
Epoch 40/5000
Epoch 41/5000
Epoch 42/5000
Epoch 43/5000
Epoch 44/5000
Epoch 45/5000
Epoch 46/5000
Epoch 47/5000
Epoch 48/5000
```

```
Epoch 49/5000
Epoch 50/5000
Epoch 51/5000
Epoch 52/5000
Epoch 53/5000
Epoch 54/5000
Epoch 55/5000
Epoch 56/5000
Epoch 57/5000
Epoch 58/5000
Epoch 59/5000
Epoch 60/5000
Epoch 61/5000
Epoch 62/5000
Epoch 63/5000
Epoch 64/5000
Epoch 65/5000
Epoch 66/5000
Epoch 67/5000
Epoch 68/5000
Epoch 69/5000
Epoch 70/5000
Epoch 71/5000
```

Epoch 72/5000

```
Epoch 73/5000
Epoch 74/5000
Epoch 75/5000
Epoch 76/5000
Epoch 77/5000
Epoch 78/5000
Epoch 79/5000
Epoch 80/5000
Epoch 81/5000
Epoch 82/5000
Epoch 83/5000
Epoch 84/5000
Epoch 85/5000
Epoch 86/5000
Epoch 87/5000
Epoch 88/5000
Epoch 89/5000
Epoch 90/5000
Epoch 91/5000
Epoch 92/5000
Epoch 93/5000
Epoch 94/5000
Epoch 95/5000
```

Epoch 96/5000

```
Epoch 97/5000
Epoch 98/5000
Epoch 99/5000
Epoch 100/5000
Epoch 101/5000
Epoch 102/5000
Epoch 103/5000
Epoch 104/5000
Epoch 105/5000
Epoch 106/5000
Epoch 107/5000
Epoch 108/5000
Epoch 109/5000
Epoch 110/5000
Epoch 111/5000
Epoch 112/5000
Epoch 113/5000
Epoch 114/5000
Epoch 115/5000
Epoch 116/5000
Epoch 117/5000
Epoch 118/5000
Epoch 119/5000
```

Epoch 120/5000

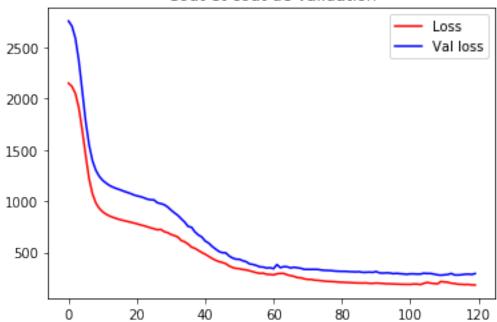
```
In [9]: y_pred = model.predict(X_test)
    plt.title('Coût et coût de validation')
    line1,=plt.plot(history.history['loss'], label="Loss", linestyle='-', color='r')
    line2,=plt.plot(history.history['val_loss'], label="Val loss", linestyle='-', color='b
    first_legend = plt.legend(handles=[line1, line2], loc=1)

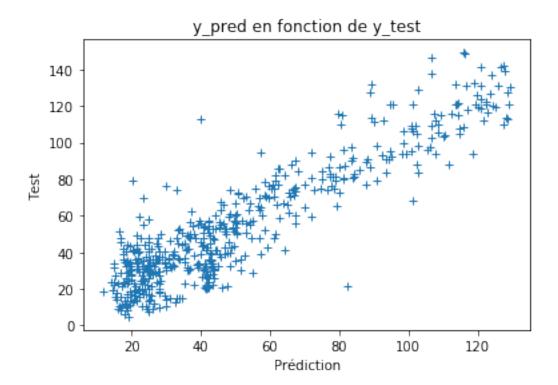
    plt.show()

    plt.title('y_pred en fonction de y_test')

    plt.plot(y_pred[:], y_test[:], '+')
    plt.ylabel('Test')
    plt.xlabel('Prédiction')
    plt.show()
```

Coût et coût de validation





In [10]: model = lstm_model(32, X_train.shape[1:])

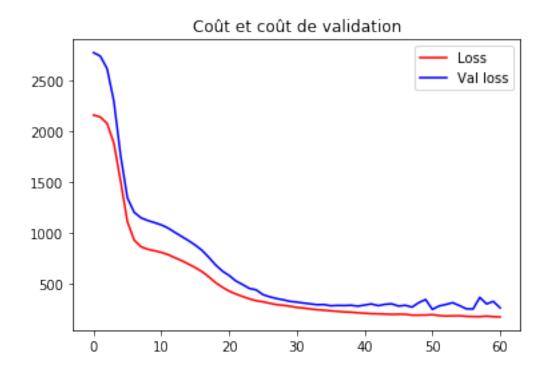
Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 32)	5248
dense_3 (Dense)	(None, 32)	1056
dense_4 (Dense)	(None, 1)	33

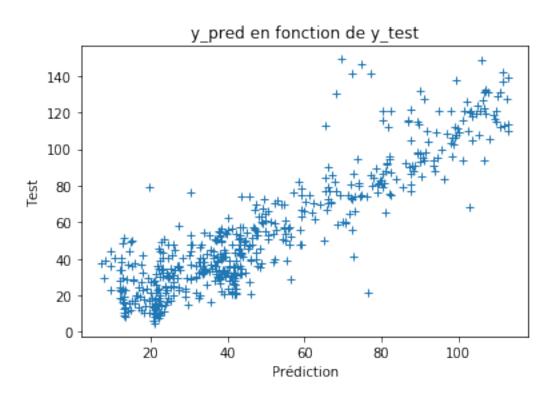
Total params: 6,337 Trainable params: 6,337 Non-trainable params: 0

```
Epoch 3/5000
Epoch 4/5000
Epoch 5/5000
Epoch 6/5000
Epoch 7/5000
Epoch 8/5000
Epoch 9/5000
Epoch 10/5000
Epoch 11/5000
Epoch 12/5000
Epoch 13/5000
Epoch 14/5000
Epoch 15/5000
Epoch 16/5000
Epoch 17/5000
Epoch 18/5000
Epoch 19/5000
Epoch 20/5000
Epoch 21/5000
Epoch 22/5000
Epoch 23/5000
Epoch 24/5000
Epoch 25/5000
Epoch 26/5000
```

```
Epoch 27/5000
Epoch 28/5000
Epoch 29/5000
Epoch 30/5000
Epoch 31/5000
Epoch 32/5000
Epoch 33/5000
Epoch 34/5000
Epoch 35/5000
Epoch 36/5000
Epoch 37/5000
Epoch 38/5000
Epoch 39/5000
Epoch 40/5000
Epoch 41/5000
Epoch 42/5000
Epoch 43/5000
Epoch 44/5000
Epoch 45/5000
Epoch 46/5000
Epoch 47/5000
Epoch 48/5000
Epoch 49/5000
Epoch 50/5000
```

```
Epoch 51/5000
Epoch 52/5000
Epoch 53/5000
Epoch 54/5000
Epoch 55/5000
Epoch 56/5000
Epoch 57/5000
Epoch 58/5000
Epoch 59/5000
Epoch 60/5000
Epoch 61/5000
Epoch 00061: early stopping
In [12]: y_pred = model.predict(X_test)
   plt.title('Coût et coût de validation')
   line1,=plt.plot(history.history['loss'], label="Loss", linestyle='-', color='r')
   line2,=plt.plot(history.history['val_loss'], label="Val_loss", linestyle='-', color='
   first_legend = plt.legend(handles=[line1, line2], loc=1)
   plt.show()
   plt.title('y_pred en fonction de y_test')
   plt.plot(y_pred[:], y_test[:], '+')
   plt.ylabel('Test')
   plt.xlabel('Prédiction')
   plt.show()
```





In [13]: model = gru_model(32, X_train.shape[1:])

```
Layer (type)
             Output Shape
                           Param #
______
              (None, 32)
gru 1 (GRU)
                           3936
       _____
              (None, 32)
dense 5 (Dense)
                           1056
dense_6 (Dense) (None, 1)
                           33
______
Total params: 5,025
Trainable params: 5,025
Non-trainable params: 0
```

history = model.fit(X_train, y_train, batch_size=32, epochs=5000, validation_data=(X_ Train on 1103 samples, validate on 539 samples Epoch 1/5000 Epoch 2/5000 Epoch 3/5000 Epoch 4/5000 Epoch 5/5000 Epoch 6/5000 Epoch 7/5000 Epoch 8/5000 Epoch 9/5000 Epoch 10/5000 Epoch 11/5000 Epoch 12/5000 Epoch 13/5000 Epoch 14/5000

In [14]: early_stopping = EarlyStopping(monitor='val_loss', verbose=1, mode='auto', patience=1

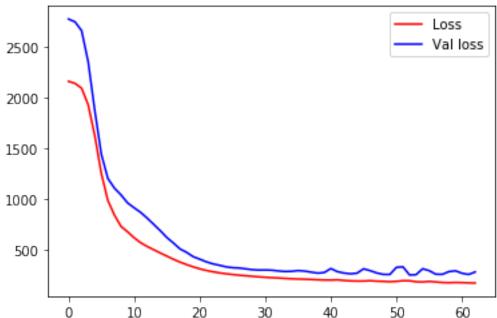
Epoch 15/5000

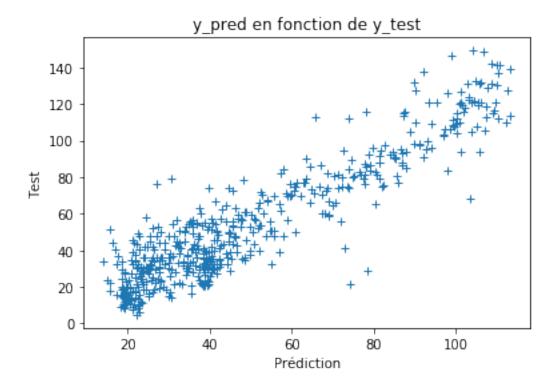
```
Epoch 16/5000
Epoch 17/5000
Epoch 18/5000
Epoch 19/5000
Epoch 20/5000
Epoch 21/5000
Epoch 22/5000
Epoch 23/5000
Epoch 24/5000
Epoch 25/5000
Epoch 26/5000
Epoch 27/5000
Epoch 28/5000
Epoch 29/5000
Epoch 30/5000
Epoch 31/5000
Epoch 32/5000
Epoch 33/5000
Epoch 34/5000
Epoch 35/5000
Epoch 36/5000
Epoch 37/5000
Epoch 38/5000
Epoch 39/5000
```

```
Epoch 40/5000
Epoch 41/5000
Epoch 42/5000
Epoch 43/5000
Epoch 44/5000
Epoch 45/5000
Epoch 46/5000
Epoch 47/5000
Epoch 48/5000
Epoch 49/5000
Epoch 50/5000
Epoch 51/5000
Epoch 52/5000
Epoch 53/5000
Epoch 54/5000
Epoch 55/5000
Epoch 56/5000
Epoch 57/5000
Epoch 58/5000
Epoch 59/5000
Epoch 60/5000
Epoch 61/5000
Epoch 62/5000
```

Epoch 63/5000

Coût et coût de validation





In []: