CS670/470 Term Project Phase 3: Advanced Machine Learning in Time Series Analysis Using 1D Convolutional Networks(1D-CNN)

Yong Zhuang, Wei Ding, UMass-Boston 05/07/2019

1 Educational Goal

Design and implement advanced machine learning algorithms for time series analysis; Practice how to apply the 1D Convolutional Neural Networks(1D-CNN) for time series analysis.

2 Details

Project Goal: Using state-of-the-art machine learning libraries to apply the 1D-CNN for streamflow forecasting.

Project Data: Twenty-one years from 1996 to 2016, daily streamflow of the Ganges river in India.

Due Date: 3:00 pm, May 21, 2019

Programming language: Python.

Package and framework: scikit-learn, statsmodels, pandas, numpy, colab, Keras.

Keras: is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. More details at https://keras.io/

3 Tasks

3.1 Task 1

Read the online lecture notes at:
https://colab.research.google.com/drive/1OHLxpSsECno79ls3Xc2sT7_xxbfxV5d7
to learn the workflow of time series analysis and the principle of convolutional neural networks.

3.2 Task 2

Load the streamflow data set from http://kdl.cs.umb.edu/CS670/data/Ganges_1996_2016.csv. Create samples by splitting up the sequence into time windows as follows:

$\mathbf{X}$	y
$[S_{1996-01-01}, ..., S_{1996-01-20}]$	$S_{1996-01-21}$
$[S_{1996-01-02}, ..., S_{1996-01-21}]$	$S_{1996-01-22}$
$[S_{1996-01-03}, ..., S_{1996-01-22}]$	$S_{1996-01-23}$
$\vdots$	

Where $S_{1996-01-01}$ means the streamflow on 1996-01-01. Split all samples into 3 parts, (1996 – 2009) for training, (2010 – 2011) for validation, and (2012 – 2016) for testing, use the date of y for splitting.

3.3 Task 3

Create your convolutional neural network model using the Keras libraries and train your model on the training set. Use the validation set for unbiased evaluation of your model fit on the training dataset while tuning your models' hyperparameters. After training, evaluate your model on the test set and calculate the root mean square errors (RMSE). Draw a curve to show the original data and the predictions.

3.4 Task 4 for undergraduate students

Every undergraduate student should write a conclusion (up to 1 page) to discuss your understanding of all the models you used in term project (phase 1, 2, 3).

3.5 Task 4 for graduate students

Every graduate student should write a short paper(up to 2 pages) for this term project, which includes an introduction, models description (for all the models you used in phase 1, 2, 3), experiments, and conclusion.

4 Submission Requirements

1. One team only need to submit one solution (colab notebook).
2. Share your team colab notebook with teaching.yong@gmail.com.
3. Every student should print your conclusion(for undergraduate students) or short paper(for graduate students) and submit it in class.