Assessment of Several Regression Based Machine Learning Approaches in Runoff Prediction

Sadegh Sadeghi Tabas
Department of Civil Engineering
CUID: C93338799
sadeghs@clemson.edu

Project Proposal

Runoff prediction from meteorological data and historical runoff observations provides the basic information for the management of water resources, the design of hydropower plants and the planning of irrigation schemes. They also provide an important backbone to reduce the damages and casualties from floods, which are among the most frequent and destructive natural hazards. In this project I am going to evaluate a number of machine learning methods including support vector machine, multi-layer perceptron, lasso, ridge, and vanilla linear regression methods in prediction of sequential flow rate values (daily runoff simulation) based on a set of collected runoff factors for a watershed located at South Carolina, US. To do so, I am going to use the Camel daily meteorological and flow observations datasets which are provided by Newman et. al (2015) and are publicly available at the *UCAR website* [1]. The dataset contains seven variables including precipitation(mm), solar radiation(W/m^2), snow water equivalent (mm), minimum temprature (C), maximum temprature (C), vapor pressure (Pa) and streamflow (ft^3/s). As it mentioned the data time scale is daily and all the data represented in real numbers available from 1980 to 2014. In order to predict runoff at time step (t) (the real values of runoff not the probability), I am going to input all meteorological forcing data (the first six variables) at time step (t) as well as observed runoff at time step (t-1) to the models. In order to evaluate the mentioned method, I'll divide the available dataset to two periods, I'll use the first part of the dataset in order to train different methods and after that I'll test the mentioned models using the rest of the available data (the second part of the dataset). Results of the fitted models will be validated and tested using Nash – Sutcliffe Efficiency (NSE), Kling – Gupta Efficiency (KGE), Root Mean Squared Error (RMSE) and R-squared criteria.

The project Category: Application

References

[1] A. Newman; K. Sampson; M. P. Clark; A. Bock; R. J. Viger; D. Blodgett, 2014. A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO: UCAR/NCAR. https://dx.doi.org/10.5065/D6MW2F4D

Submitted for review for the course CSPS8420 - Advanced Machine Learning .