A neuro-fuzzy model with multimodal fusion for time series forecasting

Marvin Querales
Instituto de Estadística, Universidad de Valparaíso

Rodrigo Salas-Fuentes
Escuela de Ingeniería Civil Biomédica, Universidad de Valparaíso

Harvey Rosas
Instituto de Estadística, Universidad de Valparaíso

Yerel Morales
Escuela de Ingeniería Civil, Universidad de Valparaíso

Héctor Allende-Cid
Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso

Abstract

Time series is a sequence of data that is collected at a regular and specific time period, and their forecasting is important because many problems that are related to prediction include a time component. Recently, neuro-fuzzy systems have been developed to forecast time series. This structure combines the capability of neural networks to learn from processes and the capability of interpolative reasoning under linguistic information pertaining to numerical variables. Moreover, to obtain a good forecasting, learning structures have to consider that information can be obtained from different types of instruments, experimental designs and other types of sources. This multimodal data could be fused to extract more features and to get a good description of the event or to obtain a global model to describe a phenomenon. In this work we propose to develop a new structure combining a neuro-fuzzy model and a multimodal data fusion approach to improve the time series forecasting. Our model was developed taking the general structure of stacking regression, considering as first level a Self-Organization Neuro-Fuzzy Inference System (SONFIS) and as second level to fuse data, a multiple linear regression. Data from the Diguillín river located in Ñuble region and average daily runoff and average daily rainfall recorded from years 2000 to 2018, were collected from the Chilean directorate of water resources (DGA). We considered two sources of data: Fundo Atacalco station and Las Trancas station. For Fundo Atacalco station, the experimental results showed a good adjustment in the first level of the model using the last 3 years as validation set, which improved when only the last year was used as validation test. In addition, Nash-Sutcliffe efficiency and percent Bias indicate that the algorithm is promising model. In Las Trancas Station, we obtained a poor performance all cases. Nevertheless, when we combined both results in the second level of the stacking regression model, we got an improved performance and better metrics than obtained in single sources of data. Therefore, the proposed architecture is a good for time series forecasting using multimodal data. We expect extend this model to other areas.