World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:12, No:03, 2018

Optimized Preprocessing for Accurate and Efficient Bioassay Prediction with Machine Learning Algorithms

Authors : Jeff Clarine, Chang-Shyh Peng, Daisy Sang

Abstract : Bioassay is the measurement of the potency of a chemical substance by its effect on a living animal or plant tissue. Bioassay data and chemical structures from pharmacokinetic and drug metabolism screening are mined from and housed in multiple databases. Bioassay prediction is calculated accordingly to determine further advancement. This paper proposes a four-step preprocessing of datasets for improving the bioassay predictions. The first step is instance selection in which dataset is categorized into training, testing, and validation sets. The second step is discretization that partitions the data in consideration of accuracy vs. precision. The third step is normalization where data are normalized between 0 and 1 for subsequent machine learning processing. The fourth step is feature selection where key chemical properties and attributes are generated. The streamlined results are then analyzed for the prediction of effectiveness by various machine learning algorithms including Pipeline Pilot, R, Weka, and Excel. Experiments and evaluations reveal the effectiveness of various combination of preprocessing steps and machine learning algorithms in more consistent and accurate prediction.

Keywords : bioassay, machine learning, preprocessing, virtual screen
Conference Title : ICMISE 2018 : International Conference on Machine Intelligence and Systems Engineering
Conference Location : Paris, France
Conference Dates : March 15-16, 2018