Hyper-heuristic framework for sequential semi-supervised classification based on core clustering

ABSTRACT

Existing stream data learning models with limited labeling have many limitations, most importantly, algorithms that suffer from a limited capability to adapt to the evolving nature of data, which is called concept drift. Hence, the algorithm must overcome the problem of dynamic update in the internal parameters or countering the concept drift. However, using neural network-based semi-supervised stream data learning is not adequate due to the need for capturing quickly the changes in the distribution and characteristics of various classes of the data whilst avoiding the effect of the outdated stored knowledge in neural networks (NN). This article presents a prominent framework that integrates each of the NN, a meta-heuristic based on evolutionary genetic algorithm (GA) and a core online-offline clustering (Core). The framework trains the NN on previously labeled data and its knowledge is used to calculate the error of the core online-offline clustering block. The genetic optimization is responsible for selecting the best parameters of the core model to minimize the error. This integration aims to handle the concept drift. We designated this model as hyper-heuristic framework for semi-supervised classification or HH-F. Experimental results of the application of HH-F on real datasets prove the superiority of the proposed framework over the existing state-of-the art approaches used in the literature for sequential classification data with evolving nature.

Keyword: Hyper-heuristic; Extreme learning machine; Genetic algorithm; Online clustering; Off-line clustering; Evolving stream data; Semi-supervised classification