Review of: "A Novel Framework for Concept Drift Detection using Autoencoders for Classification Problems in Data Streams"

Edwin Lughofer[1]

1 Johannes Kepler University Linz

Potential competing interests: No potential competing interests to declare.

Well done! - an appealing paper which suggests the usage of the reconstruction error trend line of an autoencoder to detect drifts in streams. This functionality and usage of the autoencoder enables the possibility to detect drifts fully in an unsupervised manner, hence no target labels are required. This is essential, as typically labeling efforts (feedback from experts) are costly and often not possible, especially in an on-line manner, where the processing speed is often high. And that autoencoders work well for providing a good measure of signal reconstruction, is well-known in the neural network community. To my best knowledge, autoencoders have not been used for such purpose of drift detection, thus i see a high significant novelty and innovation aspect of the paper, and therefore recommend its acceptance due to the convincing evaluation results, where the authors provide insights how the methods works in which (drift) circumstances.

Some minor issues to be integrated:

1.) The authors should also include a discussion on the following related works of drift detection:

Self-Adaptive and Local Strategies for a Smooth Treatment of Drifts in Data Streams, Evolving Systems, vol. 5 (4), pp. 239--257, 2014

Scaffolding Type-2 Classifier for Incremental Learning under Concept Drifts, Neurocomputing, vol. 191, pp. 304--329, 2016

2.) Number of batches N: how do you set this ideally in advance or what is the sensitivity of your method with respect to this parameter?

3.) Algorithm 2, Step 1.3 : the threshold seems to be calculated as m+3*sigma - why is a three sigma area and not a 2-sigma area used (being the most conventional choice in statistical process control)?