

## A buffer-based online clustering for evolving data stream

*Md. Kamrul Islama<sup>a</sup>, Md. Manjur Ahmed<sup>b</sup>, Kamal Z. Zamli<sup>a</sup>*

<sup>a</sup>Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang,  
Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia

<sup>b</sup>Department of Computer Science and Engineering, University of Barisal, Kornokathi,  
Barisal-8200, Bangladesh

### ABSTRACT

Data stream clustering plays an important role in data stream mining for knowledge extraction. Numerous researchers have recently studied density-based clustering algorithms due to their capability to generate arbitrarily shaped clusters. However, most of the algorithms are either fully offline, hybrid online/offline, or cannot handle the property of evolving data stream. Recently, a fully online clustering algorithm for evolving data stream called CEDAS was proposed. However, similar to other density-based clustering algorithms, CEDAS requires predefining the global optimal radius of micro-clusters, which is a difficult task; in addition, an erroneous choice deteriorates cluster performance. Moreover, the algorithm ignores the presence of temporarily irrelevant micro-clusters, which may be relevant in the future. In this study, we present a fully online density-based clustering algorithm called buffer-based online clustering for evolving data stream (BOCEDS). This algorithm recursively updates the micro-cluster radius to its local optimal. It also introduces a buffer for storing irrelevant micro-clusters and a fully online pruning method for extracting the temporarily irrelevant micro-cluster from the buffer. In addition, BOCEDS proposes an online micro-cluster energy-updating function based on the spatial information of the data stream. Experimental results are compared with those of CEDAS and other alternative hybrid online/offline density-based clustering algorithms, and BOCEDS proves its superiority over the other clustering algorithms. The sensitivity of clustering parameters is also measured. The proposed algorithm is then applied to real-world weather data streams to demonstrate its capability to detect changes in data stream and discover arbitrarily shaped clusters. The proposed BOCEDS can be available in <https://sites.google.com/view/md-manjur-ahmed> and <https://sites.google.com/view/kamrul-just>.

**KEYWORDS:** Density-based clustering; Evolving data stream; Arbitrarily shaped cluster; Clustering graph