Received November 25, 2019, accepted December 20, 2019, date of publication January 27, 2020, date of current version February 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969849

Incremental Cluster Validity Indices for Online Learning of Hard Partitions: Extensions and Comparative Study

LEONARDO ENZO BRITO DA SILVA[1,2], **(Member, IEEE),**
NIKLAS MAX MELTON[1], **(Member, IEEE), AND**
DONALD C. WUNSCH II[1], **(Fellow, IEEE)**

[1]Applied Computational Intelligence Laboratory, Missouri University of Science and Technology, Rolla, MO 65409, USA
[2]CAPES Foundation, Ministry of Education of Brazil, Brasília 70040-020, Brazil

Corresponding author: Leonardo Enzo Brito da Silva (leonardoenzo@ieee.org)

This research was sponsored by the Missouri University of Science and Technology Mary K. Finley Endowment and Intelligent Systems Center; the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brazil (CAPES)-Finance code BEX 13494/13-9; the U.S. Dept. of Education Graduate Assistance in Areas of National Need program; the Army Research Laboratory (ARL) and the Lifelong Learning Machines program from the DARPA/Microsystems Technology Office, and it was accomplished under Cooperative Agreement Number W911NF-18-2-0260. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ABSTRACT Validation is one of the most important aspects of clustering, particularly when the user is designing a trustworthy or explainable system. However, most clustering validation approaches require batch calculation. This is an important gap because of the value of clustering in real-time data streaming and other online learning applications. Therefore, interest has grown in providing online alternatives for validation. This paper extends the incremental cluster validity index (iCVI) family by presenting incremental versions of Calinski-Harabasz (iCH), Pakhira-Bandyopadhyay-Maulik (iPBM), WB index (iWB), Silhouette (iSIL), Negentropy Increment (iNI), Representative Cross Information Potential (irCIP), Representative Cross Entropy (irH), and Conn_Index (iConn_Index). This paper also provides a thorough comparative study of correct, under- and over-partitioning on the behavior of these iCVIs, the Partition Separation (PS) index as well as four recently introduced iCVIs: incremental Xie-Beni (iXB), incremental Davies-Bouldin (iDB), and incremental generalized Dunn's indices 43 and 53 (iGD43 and iGD53). Experiments were carried out using a framework that was designed to be as agnostic as possible to the clustering algorithms. The results on synthetic benchmark data sets showed that while evidence of most under-partitioning cases could be inferred from the behaviors of the majority of these iCVIs, over-partitioning was found to be a more challenging problem, detected by fewer of them. Interestingly, over-partitioning, rather then under-partitioning, was more prominently detected on the real-world data experiments within this study. The expansion of iCVIs provides significant novel opportunities for assessing and interpreting the results of unsupervised lifelong learning in real-time, wherein samples cannot be reprocessed due to memory and/or application constraints.

INDEX TERMS Clustering, validation, incremental cluster validity index (iCVI), adaptive resonance theory (ART), incremental (online) clustering algorithms, data streams.

I. INTRODUCTION

Cluster validation [1] is a fundamental topic in cluster analysis. It is crucial to assess the quality of partitions detected by clustering algorithms where no class label information is available. Moreover, different clustering solutions may be found by distinct algorithms, or even by the same algorithm subjected to different hyper-parameters or different presentation orders [2], [3]. *Cluster validity indices* (CVIs) function as evaluators of such solutions by computing some cluster quality measure based on (i) the degree of agreement between the output and a reference partition (*external* CVIs), or (ii) the data itself and the output partition information (*internal* CVIs). Numerous examples of such criteria have

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

been presented in the literature to evaluate partitions in offline mode; for comprehensive reviews and experimental studies, the interested reader may refer to [4]–[16].

Since 2018, *incremental cluster validity indices* (iCVIs) have been developed to track the effectiveness of online clustering methods over data streams [17]–[20]. These are online versions of sum-of-squares (SS)-based internal CVIs [21], which typically exhibit a trade-off between measures of compactness (a.k.a. dispersion or within-cluster scatter) and isolation (a.k.a. between-cluster separation) [2], [21]. To enable cluster validation in online applications, a recursive formulation of compactness was introduced in [17], [18]. This strategy was used to develop incremental versions of Davies-Bouldin [22] (iDB) and Xie-Beni [23] (iXB) in [17], [18] as well as incremental versions of two generalized Dunn's indices [24] (iGDs) in [25]. Particularly, the behavior of iXB and iDB were analyzed in both accurately and poorly partitioned data sets in [17], [18], whereas the studies in [19], [20] only investigated the behavior of iDB in cases in which the MU streaming clustering (MUSC) [26] accurately detected the structures present in the data. In this context, the contributions of our work [27] are two-fold:

1) Introducing 7 novel iCVIs. The incremental versions of Calinski-Harabasz [28], WB index [29], Pakhira-Bandyopadhyay-Maulik [30], and Silhouette [31] were realized by employing the incremental update of compactness developed in [17], [18]. The incremental versions of Negentropy Increment [32], [33], Representative Cross Information Potential and Representative Cross Entropy [34], [35] were realized using the incremental update of covariance matrices [36]. Finally, the incremental version of the Conn_Index [37], [38] was realized by storing co-activation counts of multiple prototypes generated using fuzzy adaptive resonance theory (ART)-based models [39], [40]. ART networks have been used in a multitude of applications [41], [42] and were chosen for their simple parameterization of cluster granularity and other useful properties [43], [44].
2) Performing a comparative study among 13 iCVIs in cases of correct, under- and over-partitioning on synthetic and real-world benchmark data sets. It is not the focus of this study to contrast the iCVIs' behavior associated with specific online clustering algorithms and their dynamics and, therefore, a framework that is as clustering algorithm agnostic as possible was used to define the data partitions.

To the best of our knowledge, this work provides the first comprehensive and systematic comparative study on iCVIs. The remainder of this paper is structured as follows: Section II provides a brief review of batch CVIs, iCVIs, as well as ART; Section III presents novel extensions of several other CVIs to the incremental family; Section IV details the set-up used in the numerical experiments; Section V describes and discusses the results; Section VI compares batch and incremental versions of the Conn_Index; and Section VII summarizes the findings of this paper.

II. BACKGROUND AND RELATED WORK

This section provides an overview of CVIs, iCVIs and ART neural networks used in this study.

A. BATCH CLUSTER VALIDITY INDICES (CVIS)

Consider a data set $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$ and its hard partition $\Omega = \{\omega_i\}_{i=1}^{k}$ of k disjointed clusters ω_i, such that $\bigcup_{i=1}^{k} \omega_i = \boldsymbol{X}$. In the following CVI overview, $\boldsymbol{v}_i$ is the prototype (centroid) for cluster ω_i, defined as:

$$\boldsymbol{v}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \boldsymbol{x}_j, \quad \boldsymbol{x}_j \in \omega_i, \tag{1}$$

k is the number of clusters, d is the dimensionality of the data ($\boldsymbol{x}_i \in I\!R^d$), and N and n_i are the cardinalities of a data set and cluster ω_i, respectively. Additionally, the geometric center of the data is given by:

$$\boldsymbol{\mu}_{data} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i, \tag{2}$$

and the compactness of cluster ω_i with respect to point $\boldsymbol{z}$ is:

$$CP_q^p(z, \omega_i) = \sum_{j=1}^{n_i} \|\boldsymbol{x}_j - \boldsymbol{z}\|_q^p, \quad \boldsymbol{x}_j \in \omega_i, \tag{3}$$

where $\|\cdot\|_q^p$ is the ℓ_q norm to the p^{th} power.

1) CALINSKI-HARABASZ (CH)

The CH index [28] is defined as:

$$CH = \frac{BGSS/(k-1)}{WGSS/(N-k)}, \tag{4}$$

where the between-group sum-of-squares (BGSS) and within-group sum of squares (WGSS) are computed as:

$$WGSS = \sum_{i=1}^{k} CP_2^2(\boldsymbol{v}_i, \omega_i), \tag{5}$$

$$BGSS = \sum_{i=1}^{k} n_i \|\boldsymbol{v}_i - \boldsymbol{\mu}_{data}\|_2^2. \tag{6}$$

This is an optimization-like criterion [11] such that larger values of CH indicate better clustering solutions.

2) WB-INDEX (WB)

The WB index [29] is related to CH as discussed in [21] and is given by:

$$WB = k \frac{WGSS}{BGSS}. \tag{7}$$

Smaller values of WB suggest better data partition quality.

3) DAVIES-BOULDIN (DB)

The DB index [22] averages the similarities R of each cluster i with respect to its maximally similar cluster $j \neq i$:

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i, \tag{8}$$

where:

$$R_i = \max_{i \neq j}\left(\frac{S_i + S_j}{M_{i,j}}\right), \tag{9}$$

$$S_l = \left[\frac{1}{n_l} CP_2^q(\boldsymbol{v}_l, \omega_l)\right]^{\frac{1}{q}}, \quad l = \{1, \ldots, k\}, \tag{10}$$

$$M_{i,j} = \left[\sum_{t=1}^{d} |v_{it} - v_{jt}|^p\right]^{\frac{1}{p}}, \quad p \geq 1. \tag{11}$$

The variables (p, q) are user-defined parameters, and S_l and $M_{i,j}$ (Minkowski metric) measure compactness and separation, respectively. Smaller values of DB indicate better clustering solutions.

4) XIE-BENI (XB)

The XB index [23] was originally designed to detect compact and separated clusters in fuzzy c-partitions. A hard partition version is given by the following ratio of compactness to separation [45], [46]:

$$XB = \frac{WGSS/N}{\min\limits_{i \neq j} \|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2}. \tag{12}$$

Smaller values of XB indicate better clustering solutions.

5) GENERALIZED DUNN'S INDICES (GDS)

The GDs [24] comprise a set of 17 variants of the original Dunn's index [47] devised to address sensitivity to noise in the latter. These CVIs are given by [25]:

$$GD_{rs} = \frac{\min\limits_{i \neq j}\left[\delta_r(\omega_i, \omega_j)\right]}{\max\limits_{k}\left[\Delta_s(\omega_k)\right]}, \tag{13}$$

where $\delta_r(\cdot)$ is a measure of separation, and $\Delta_s(\cdot)$ is a measure of compactness. The parameters r and s index the measures' formulations ($r \in \{1, \ldots, 6\}$ and $s \in \{1, 2, 3\}$). In particular, when employing Euclidean distance, the GD_{43} and GD_{53} variants are formulated using:

$$\delta_4(\omega_i, \omega_j) = \|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2, \tag{14}$$

$$\delta_5(\omega_i, \omega_j) = \frac{CP_2^1(\boldsymbol{v}_i, \omega_i) + CP_2^1(\boldsymbol{v}_j, \omega_j)}{n_i + n_j}, \tag{15}$$

$$\Delta_3(\omega_k) = \frac{2 \times CP_2^1(\boldsymbol{v}_k, \omega_k)}{n_k}. \tag{16}$$

Larger values of these GDs suggest better clustering partitions.

6) PAKHIRA-BANDYOPADHYAY-MAULIK (PBM)

Consider the I index [48] defined as:

$$I = \left(\frac{1}{k} \times \frac{E_1}{E_k} \times D_k\right)^p, \quad p \geq 1, \tag{17}$$

where:

$$E_1 = \sum_{i=1}^{N} ||\boldsymbol{x}_i - \boldsymbol{\mu}_{data}|_2, \tag{18}$$

$$E_k = \sum_{i=1}^{k} CP_2^1(\boldsymbol{v}_i, \omega_i), \tag{19}$$

$$D_k = \max_{i \neq j}\left(|\boldsymbol{v}_i - \boldsymbol{v}_j|_2\right). \tag{20}$$

The quantities E_k and D_k measure compactness and separation, respectively. This CVI comprises a trade-off among the three competing factors in Eq. (17): $\frac{1}{k}$ decreases with k, whereas both $\frac{E_1}{E_k}$ and D_k increase. By setting $p = 2$ in Eq. (17), the I index reduces to the PBM index [30]. Larger values of PBM indicate better clustering solutions.

7) SILHOUETTE (SIL)

The SIL index [31] is computed by averaging the silhouette coefficients sc_i across all data samples $\boldsymbol{x}_i$:

$$SIL = \frac{1}{N}\sum_{i=1}^{N} sc_i, \tag{21}$$

where:

$$sc_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{22}$$

$$a_i = \frac{1}{n_i - 1} CP_2^1(\boldsymbol{x}_i, \omega_i), \tag{23}$$

$$b_i = \min_{l, l \neq i}\left[\frac{1}{n_l} CP_2^1(\boldsymbol{x}_i, \omega_l)\right]. \tag{24}$$

The variables a_i and b_i measure compactness and separation, respectively. Larger values of SIL (close to 1) indicate better clustering solutions. To reduce computational complexity, some SIL variants, such as [49]–[52], use a centroid-based approach. The simplified SIL [49], [50] has been used successfully in clustering data streams processed in chunks, in which the silhouette coefficients are also used to make decisions regarding the centroids' incremental updates [53].

8) PARTITION SEPARATION (PS)

The PS index [54] was originally developed for fuzzy clustering; its hard clustering version is given by [55]:

$$PS = \sum_{i=1}^{k} PS_i, \tag{25}$$

where:

$$PS_i = \frac{n_i}{\max\limits_{j}(n_j)} - exp\left[-\frac{\min\limits_{i \neq j}\left(\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2\right)}{\beta_T}\right], \tag{26}$$

$$\beta_T = \frac{1}{k}\sum_{l=1}^{k} \|\boldsymbol{v}_l - \bar{\boldsymbol{v}}\|_2^2, \tag{27}$$

$$\bar{\boldsymbol{v}} = \frac{1}{k}\sum_{l=1}^{k} \boldsymbol{v}_l. \tag{28}$$

The PS index only comprises a measure of separation between prototypes. Although included in the batch CVI section, it can be readily used to evaluate the partitions identified by unsupervised incremental learners that model clusters using centroids (*e.g.*, [55]). Larger values of PS indicate better clustering solutions.

9) NEGENTROPY INCREMENT (NI)

The NI index [32], [33] measures the average normality of the clusters of a given partition Ω via negentropy [56], while avoiding the direct computation of the clusters' differential entropies. Unlike the other CVIs discussed so far, the NI is not explicitly constructed using measures of compactness and separation [13], [32], thereby being defined as:

$$NI = \frac{1}{2}\sum_{i=1}^{k} p_i \ln|\boldsymbol{\Sigma}_i| - \frac{1}{2}\ln|\boldsymbol{\Sigma}_{data}| - \sum_{i=1}^{k} p_i \ln p_i, \quad (29)$$

where $|\cdot|$ denotes the determinant. The probabilities (p) and covariance matrices ($\boldsymbol{\Sigma}$) are estimated as:

$$p_i = \frac{n_i}{N}, \quad (30)$$

$$\boldsymbol{\Sigma}_i = \frac{1}{n_i - 1}\sum_{\substack{j=1 \\ \boldsymbol{x}_j \in \omega_i}}^{n_i} (\boldsymbol{x}_j - \boldsymbol{v}_i)(\boldsymbol{x}_j - \boldsymbol{v}_i)^T, \quad (31)$$

$$\boldsymbol{\Sigma}_{data} = \frac{1}{N-1}\left(\boldsymbol{X}^T\boldsymbol{X} - N\boldsymbol{\mu}_{data}\boldsymbol{\mu}_{data}^T\right), \quad (32)$$

and the means $\boldsymbol{v}$ and $\boldsymbol{\mu}_{data}$ are estimated using Eqs. (1) and (2), respectively. Smaller values of NI indicate better clustering solutions.

10) REPRESENTATIVE CROSS INFORMATION POTENTIAL (RCIP)

Cluster evaluation functions (CEFs) based on cross information potential (CIP) [57], [58] have been used consistently in the literature to evaluate partitions and drive optimization algorithms searching for data structures [34], [35], [57], [58]; thus, this work includes these CEFs under the CVI category. Precisely, representative approaches [34], [35] replace the sample-by-sample estimation of Renyi's quadratic entropy [59] using the Parzen-window method [36] (original CIP [57], [58]) via prototypes and the statistics of their associated Voronoi polyhedrons. The rCIP was devised for prototype-based clustering (*i.e.*, two-step methods: vector quantization followed by clustering of the prototypes) [60]–[64]. The CEF used here is defined as [35]:

$$CEF = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} rCIP(\omega_i, \omega_j), \quad (33)$$

where:

$$rCIP(\omega_i, \omega_j) = \frac{1}{M_i M_j}\sum_{l=1}^{M_i}\sum_{m=1}^{M_j} G(\Delta\boldsymbol{v}_{l,m}, \boldsymbol{\Sigma}_{l,m}), \quad (34)$$

$$G(\Delta\boldsymbol{v}_{l,m}, \boldsymbol{\Sigma}_{l,m}) = \frac{e^{-\frac{1}{2}\Delta\boldsymbol{v}_{l,m}^T\boldsymbol{\Sigma}_{l,m}^{-1}\Delta\boldsymbol{v}_{l,m}}}{\sqrt{(2\pi)^d\,|\boldsymbol{\Sigma}_{l,m}|}}, \quad (35)$$

$\Delta\boldsymbol{v}_{l,m} = \boldsymbol{v}_l - \boldsymbol{v}_m$, $\boldsymbol{\Sigma}_{l,m} = \boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_m$, $\{\boldsymbol{v}_l, \Sigma_l\} \in \omega_i$, $\{\boldsymbol{v}_m, \Sigma_m\} \in \omega_j$, and M_i and M_j are the number of prototypes used to represent clusters ω_i and ω_j, respectively. The prototypes and covariance matrices are estimated using Eqs. (1) and (31), respectively. Smaller values of CEF indicate better clustering solutions. Recently, the information potential [65] measure has been used to define the state of a system when modeling and analyzing dynamic processes [66], [67].

11) CONN_INDEX

The Conn_Index [37], [38] was also developed for prototype-based clustering. It is formulated using the connectivity strength matrix (***CONN***), which is a symmetric square similarity matrix that represents local data densities between neighboring prototypes [68], [69]. Its $(i,j)^{th}$ entry is formally given by:

$$\boldsymbol{CONN}(i,j) = \boldsymbol{CADJ}(i,j) + \boldsymbol{CADJ}(j,i), \quad (36)$$

where the $(i,j)^{th}$ entry of the non-symmetric cumulative adjacency matrix (***CADJ***) corresponds to the number of samples for which $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are, simultaneously, the first and second closest prototypes (according to some dissimilarity measure $D(\cdot,\cdot)$, such as Euclidean distance), respectively:

$$\boldsymbol{CADJ}(i,j) = card(RF_{i,j}), \quad (37)$$

$$RF_{i,j} = \{\boldsymbol{x}_k \in RF_i : D(\boldsymbol{x}_k, \boldsymbol{v}_j) \leq D(\boldsymbol{x}_k, \boldsymbol{v}_l)\ \forall l \neq i\}, \quad (38)$$

$$RF_i = \left\{\boldsymbol{x}_k \in \boldsymbol{X} : D(\boldsymbol{x}_k, \boldsymbol{v}_i) \leq D(\boldsymbol{x}_k, \boldsymbol{v}_j)\ \forall j\right\}, \quad (39)$$

where $card(\cdot)$ is the cardinality operator. The Conn_Index is defined as:

$$Conn\_Index = Intra\_Conn \times (1 - Inter\_Conn), \quad (40)$$

where the intra-cluster (*Intra_Conn*) and inter-cluster (*Inter_Conn*) connectivities are:

$$Intra\_Conn = \frac{1}{k}\sum_{l=1}^{k} \boldsymbol{Intra}(\omega_l), \quad (41)$$

$$\boldsymbol{Intra}(\omega_l) = \frac{1}{n_l}\sum_{\substack{i,j \\ \boldsymbol{v}_i, \boldsymbol{v}_j \in \omega_l}}^{M} \boldsymbol{CADJ}(i,j), \quad (42)$$

$$Inter\_Conn = \frac{1}{k}\sum_{l=1}^{k} \max_{m, m\neq l}\left[\boldsymbol{Inter}(\omega_l, \omega_m)\right], \quad (43)$$

$$\boldsymbol{Inter}(\omega_l, \omega_m) = \frac{\sum_{\substack{i,j \\ \boldsymbol{v}_i \in \omega_l, \boldsymbol{v}_j \in \omega_m}}^{M} \boldsymbol{CONN}(i,j)}{\sum_{\substack{i,j \\ \boldsymbol{v}_i \in V_{l,m}}}^{M} \boldsymbol{CONN}(i,j)}, \quad (44)$$

$$V_{l,m} = \{\boldsymbol{v}_i : \boldsymbol{v}_i \in \omega_l, \exists \boldsymbol{v}_j \in \omega_m : \boldsymbol{CADJ}(i,j) > 0\}, \quad (45)$$

the variable M is the total number of prototypes, and $\boldsymbol{Inter}(\omega_l, \omega_m) = 0$ if $V_{l,m} = \{\emptyset\}$. Naturally, the quantities

TABLE 1. Cluster validity indices used in this study.

Cluster validity index (CVI)	Formulation[a]	Reference(s)
Calinski-Harabasz[b]	$CH = \frac{\left(\sum_{i=1}^{k} n_i \|\boldsymbol{v}_i - \boldsymbol{\mu}_{data}\|_2^2\right)/(k-1)}{\left[\sum_{i=1}^{k} CP_2^2(\boldsymbol{v}_i, \omega_i)\right]/(N-k)}$	[28]
WB-Index[b]	$WB = k \frac{\sum_{i=1}^{k} CP_2^2(\boldsymbol{v}_i, \omega_i)}{\sum_{i=1}^{k} n_i \|\boldsymbol{v}_i - \boldsymbol{\mu}_{data}\|_2^2}$	[21], [29]
Davies-Bouldin[b]	$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} \left(\frac{\left[\frac{1}{n_i} CP_2^q(\boldsymbol{v}_i, \omega_i)\right]^{\frac{1}{q}} + \left[\frac{1}{n_j} CP_2^q(\boldsymbol{v}_j, \omega_j)\right]^{\frac{1}{q}}}{\left[\sum_{t=1}^{d} \|v_{it} - v_{jt}\|^p\right]^{\frac{1}{p}}} \right)$	[22]
Xie-Beni[b]	$XB = \frac{\sum_{i=1}^{k} CP_2^2(\boldsymbol{v}_i, \omega_i)/N}{\min_{i \neq j} \|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2}$	[23]
Generalized Dunn's index 43[b]	$GD_{43} = \frac{\min_{i \neq j}\left[\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2\right]}{\max_k \left[\frac{2CP_2^1(\boldsymbol{v}_k, \omega_k)}{n_k}\right]}$	[24], [25]
Generalized Dunn's index 53[b]	$GD_{53} = \frac{\min_{i \neq j}\left[\frac{CP_2^1(\boldsymbol{v}_i, \omega_i) + CP_2^1(\boldsymbol{v}_j, \omega_j)}{n_i + n_j}\right]}{\max_k \left[\frac{2CP_2^1(\boldsymbol{v}_k, \omega_k)}{n_k}\right]}$	[24], [25]
PBM Index[b]	$PBM = \left[\frac{1}{k} \frac{\sum_{i=1}^{N} \|\boldsymbol{x}_i - \boldsymbol{\mu}_{data}\|_2}{\sum_{i=1}^{k} CP_2^1(\boldsymbol{v}_i, \omega_i)} \max_{i \neq j} \left(\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2\right) \right]^2$	[30], [48]
Silhouette[b]	$SIL = \frac{1}{N}\sum_{i=1}^{N} \left\{ \frac{\min_{l, l \neq i}\left[\frac{1}{n_l} CP_2^1(\boldsymbol{x}_i, \omega_l)\right] - \frac{1}{n_i - 1} CP_2^1(\boldsymbol{x}_i, \omega_i)}{\max\left(\frac{1}{n_i - 1} CP_2^1(\boldsymbol{x}_i, \omega_i), \min_{l, l \neq i}\left[\frac{1}{n_l} CP_2^1(\boldsymbol{x}_i, \omega_l)\right]\right)} \right\}$	[31]
Partition Separation	$PS = \sum_{i=1}^{k} \left\{ \frac{n_i}{\max_j(n_j)} - exp\left[- \frac{\min_{i \neq j}\left(\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2\right)}{\frac{1}{k}\sum_{l=1}^{k} \|\boldsymbol{v}_l - \frac{1}{k}\sum_{t=1}^{k} \boldsymbol{v}_t\|_2^2} \right] \right\}$	[54], [55]
Negentropy Increment[c]	$NI = \frac{1}{2}\sum_{i=1}^{k} p_i \ln\|\boldsymbol{\Sigma}_i\| - \frac{1}{2}\ln\|\boldsymbol{\Sigma}_{data}\| - \sum_{i=1}^{k} p_i \ln p_i$	[32], [33]
Representative Cross Information Potential[c]	$CEF = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} \left[\frac{1}{M_i M_j} \sum_{l=1}^{M_i}\sum_{m=1}^{M_j} \frac{e^{-\frac{1}{2}(\boldsymbol{v}_l - \boldsymbol{v}_m)^T (\boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_m)^{-1} (\boldsymbol{v}_l - \boldsymbol{v}_m)}}{\sqrt{(2\pi)^d \|\boldsymbol{\Sigma}_l + \boldsymbol{\Sigma}_m\|}} \right]$	[34], [35]
Conn_Index[d]	$Conn\_Index = \left(\frac{1}{k}\sum_{l=1}^{k} \boldsymbol{Intra}(\omega_l)\right)\left(1 - \frac{1}{k}\sum_{l=1}^{k} \max_{m, m \neq l} [\boldsymbol{Inter}(\omega_l, \omega_m)]\right)$	[37], [38]

[a] $\boldsymbol{x}_i$ represents the ith sample of a data set $\boldsymbol{X}$, k represents the number of clusters, N represents the data set cardinality, n_i represents the number of samples in cluster ω_i, $\boldsymbol{v}_i$ represents a prototype of cluster ω_i (computed using Eq. (1)), and $\boldsymbol{\mu}_{data}$ represents the data mean (computed using Eq. (2)).
[b] The compactness CP_q^p is defined using Eq. (3).
[c] The a priori probabilities (p) and covariance matrices ($\boldsymbol{\Sigma}$) are computed using Eqs. (30) through (32).
[d] The $\boldsymbol{Intra}(\cdot)$ and $\boldsymbol{Inter}(\cdot,\cdot)$ are defined using Eqs. (42) and (44), respectively.

Intra_Conn and *Inter_Conn* measure compactness and separation, respectively. Larger values of the Conn_Index (close to 1) indicate better clustering solutions.

12) SUMMARY

Table 1 summarizes the formulations of all the CVIs discussed in this section. Most of these are suited to data comprising cluster structures like hyperspheres and hyperellipses. The exceptions are the Conn_Index and rCIP, these are also appropriate for data encompassing more complex cluster structures given their multi-prototype representation nature.

B. INCREMENTAL CLUSTER VALIDITY INDICES (ICVIS)

The compactness and separation terms commonly found in CVIs are generally computed using data samples and prototypes, respectively [17], [19]. In order to handle online clustering application demands (*i.e.*, data streams), an incremental CVI (iCVI) formulation that recursively estimates the compactness term was introduced in [17], [18] in the context of fuzzy clustering.

Remark 1: Hereafter, the notation CP_q^p is simplified to CP. This notation was changed because only the squared Euclidean norm ($p = q = 2$) will be used for the compactness. Henceforth, the subscript of CP designates *cluster membership*.

Specifically, consider the hard clustering version of compactness for cluster i (*i.e.*, by setting the fuzzy memberships in [17], [18] to indicator functions):

$$CP_i = \sum_{\substack{j=1 \\ \boldsymbol{x}_j \in \omega_i}}^{n_i} \|\boldsymbol{x}_j - \boldsymbol{v}_i\|_2^2. \tag{46}$$

In such a case, when a new sample $\boldsymbol{x}$ is presented and encoded by cluster i, then its new compactness value becomes:

$$CP_i^{new} = \sum_{\substack{j=1 \\ \boldsymbol{x}_j \in \omega_i}}^{n_i^{new}} \|\boldsymbol{x}_j - \boldsymbol{v}_i^{new}\|_2^2, \tag{47}$$

where:

$$n_i^{new} = n_i^{old} + 1, \tag{48}$$
$$\boldsymbol{v}_i^{new} = \boldsymbol{v}_i^{old} + (\boldsymbol{x} - \boldsymbol{v}_i^{old})/n_i^{new}, \tag{49}$$

and:

$$N^{new} = N^{old} + 1. \tag{50}$$

The compactness in Eq. (47) can be updated incrementally as [17], [18]:

$$CP_i^{new} = CP_i^{old} + \|\boldsymbol{z}_i\|_2^2 + n_i^{old}\|\Delta\boldsymbol{v}_i\|_2^2 + 2\Delta\boldsymbol{v}_i^T\boldsymbol{g}_i^{old}, \tag{51}$$

where:

$$\boldsymbol{z}_i = \boldsymbol{x} - \boldsymbol{v}_i^{new}, \tag{52}$$
$$\Delta\boldsymbol{v}_i = \boldsymbol{v}_i^{old} - \boldsymbol{v}_i^{new}, \tag{53}$$

and vector $\boldsymbol{g}$, which is formally defined as:

$$\boldsymbol{g}_i = \sum_{j=1}^{n_i} (\boldsymbol{x}_j - \boldsymbol{v}_i), \tag{54}$$

is incrementally updated at each iteration using:

$$\boldsymbol{g}_i^{new} = \boldsymbol{g}_i^{old} + \boldsymbol{z}_i + n_i^{old}\Delta\boldsymbol{v}_i. \tag{55}$$

Using such incremental formulation, the following iCVIs were derived (their hard partition counterparts are shown here [25]):

1) incremental Xie-Beni (iXB) [17], [18]:

$$XB^{new} = \frac{1}{N^{new}} \times \frac{\sum_{i=1}^{k^{new}} CP_i^{new}}{\min_{i \neq j}\left(\|\boldsymbol{v}_i^{new} - \boldsymbol{v}_j^{new}\|_2^2\right)}, \tag{56}$$

2) incremental Davies-Bouldin (iDB) [17], [18]:

$$DB^{new} = \frac{1}{k^{new}} \sum_{i=1}^{k^{new}} \max_{j, j \neq i} \left(\frac{\frac{CP_i^{new}}{n_i^{new}} + \frac{CP_j^{new}}{n_j^{new}}}{\|\boldsymbol{v}_i^{new} - \boldsymbol{v}_j^{new}\|_2^2} \right), \tag{57}$$

3) incremental generalized Dunn's indices (iGDs) [25]:

$$GD_{43}^{new} = \frac{\min_{i \neq j}\left(\|\boldsymbol{v}_i^{new} - \boldsymbol{v}_j^{new}\|_2\right)}{\max_k \left(\frac{2CP_k^{new}}{n_k^{new}}\right)}, \tag{58}$$

$$GD_{53}^{new} = \frac{\min_{i \neq j}\left(\frac{CP_i^{new} + CP_j^{new}}{n_i^{new} + n_j^{new}}\right)}{\max_k \left(\frac{2CP_k^{new}}{n_k^{new}}\right)}. \tag{59}$$

Note that only one prototype $\boldsymbol{v}$, counter n and compactness CP are updated after each input presentation. If a new cluster emerges, then $k^{new} = k^{old} + 1$, and its compactness CP and vector $\boldsymbol{g}$ are initialized as 0 and $\vec{\boldsymbol{0}}$ (because $\boldsymbol{v} = \boldsymbol{x}$), respectively.

C. ADAPTIVE RESONANCE THEORY (ART)

This study uses a neural network implementation of adaptive resonance theory (ART) [70] given its fast and stable online learning and its automatic category recognition capabilities. ART models encompass a rich history with many implementations well-suited to iCVI computation (see [44] for a comprehensive review on ART models); the models used for the experiments in this study are discussed next.

1) FUZZY ART

The fuzzy ART model [39] utilizes fuzzy logic [71] to bound data within hyper-boxes. For a normalized data set $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ ($\boldsymbol{x}_i \in I\!R^d, 0 \leq x_{i,j} \leq 1$, $j = \{1, \ldots, d\}$), the fuzzy ART algorithm, with parameters ($\alpha > 0$, $0 < \beta \leq 1$, $0 \leq \rho \leq 1$), is defined by:

$$\boldsymbol{I} = (\boldsymbol{x}, 1 - \boldsymbol{x}), \tag{60}$$
$$T_j = \frac{\|\min(\boldsymbol{I}, \boldsymbol{w}_j)\|_1}{\alpha + \|\boldsymbol{w}_j\|_1}, \tag{61}$$
$$\|\min(\boldsymbol{I}, \boldsymbol{w}_j)\|_1 \geq \rho\|\boldsymbol{I}\|_1, \tag{62}$$
$$\boldsymbol{w}_j^{new} = \boldsymbol{w}_j^{old}(1 - \beta) + \beta\min(\boldsymbol{I}, \boldsymbol{w}_j^{old}). \tag{63}$$

Equation (60) is the complement-coding function, which concatenates sample $\boldsymbol{x}$ and its complement to form an input vector $\boldsymbol{I}$ with dimension $2d$. Equation (61) is the activation function for each fuzzy ART category j, where $\|\cdot\|_1$ is the L_1 norm, $min(\cdot)$ is performed component-wise, and α is a tie-breaking constant. Each category is checked in a descending order of activation for validity against the vigilance parameter ρ using Eq. (62). If no valid category is found during training, then a new category is initialized using $\boldsymbol{I}$ as the new weight vector $\boldsymbol{w}$. Otherwise, the winning category is updated according to Eq. (63) using learning rate β.

2) FUZZY ARTMAP

In a fuzzy ARTMAP network [40], two fuzzy ART modules, A-side and B-side, are supplied with separate but dependent data streams. In classification settings specifically, these streams consist of data and class labels, respectively. Both ART modules cluster their inputs according to local topology and parameters while an inter-ART module enforces a surjective mapping of the A-side to the B-side, effectively learning the functional map of the A-side to the B-side categories. This model will be *required* to (i) extend the iCVI study to prototype-based CVIs such as the Conn_Index, and (ii) perform the experiments under a clustering agnostic framework (see Section V), in which the A-side categories represent cluster prototypes and are driven by the B-side true data partition labels (note that we follow a simplified fuzzy

ARTMAP design [72], in which the B-side is replaced by a stream of ground-truth class labels).

III. EXTENSIONS OF ICVIS

To compute the CVIs mentioned in Section II-A incrementally, employing one of the following approaches is sufficient:

1) The recursive computation of compactness developed in [17], [18] for the SS-based CVIs (CH, WB, PBM, and SIL).
2) The incremental computation of probabilities, means and covariance matrices for the information-theoretic (IT)-based CVIs (rCIP and NI). Naturally, if the clustering algorithm of choice already models the clusters using a priori probabilities, means and covariance matrices (such as Gaussian ART [73] and Bayesian ART [74]), then, similarly to PS, these CVIs can be readily computed.
3) The incremental building of a multi-prototype representation of clusters using a modified ART model while tracking the density-based connections between neighboring prototypes for the graph-based CVI (Conn_index). Specifically, the latter is accomplished by updating (incrementing and/or expanding) the ***CADJ*** and ***CONN*** matrices as clusters grow and/or are dynamically created.

In the following iCVIs' extensions (iCH, iWB, iPBM, iSIL, ircIP, iNI, and iConn_index), if a new cluster is formed after sample $\boldsymbol{x}$ is presented, then the total number of clusters is updated to $k^{new} = k^{old} + 1$ (otherwise $k^{new} = k^{old}$), and, unless otherwise noted, the variables associated with this new cluster are initialized as $n_{k^{new}}^{new} = 1$ (number of samples encoded), $\boldsymbol{v}_{k^{new}}^{new} = \boldsymbol{x}$ (prototype of this cluster), $CP_{k^{new}}^{new} = 0$ (initial compactness), and $\boldsymbol{g}_{k^{new}}^{new} = \vec{\mathbf{0}}$ (initial vector $\boldsymbol{g}$). Naturally, clusters that do not encode the presented sample remain with constant parameter values for the duration of that input presentation. Also note that, where necessary, the Euclidean norm is replaced with the squared Euclidean norm (*i.e.*, $|\cdot|_2^2$) to compute the compactness CP (as per [17], [18]). Finally, for iCVIs that require the computation of pairwise (dis)similarity between prototypes, the (dis)similarity matrix is kept in memory, where only the rows and columns corresponding to the prototype that is adapted are modified.

A. INCREMENTAL CALINSKI-HARABASZ INDEX (ICH)

The iCH computation is defined as:

$$CH^{new} = \frac{\sum_{i=1}^{k^{new}} SEP_i^{new}}{\sum_{i=1}^{k^{new}} CP_i^{new}} \times \frac{N^{new} - k^{new}}{k^{new} - 1}, \tag{64}$$

where:

$$SEP_i^{new} = n_i^{new} \|\boldsymbol{v}_i^{new} - \boldsymbol{\mu}_{data}^{new}\|_2^2. \tag{65}$$

Note that the variables $\{n_1, \ldots, n_k\}$, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, $\{CP_1, \ldots, CP_k\}$, $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k\}$, $\boldsymbol{\mu}_{data}, k, N$, and $\{SEP_1, \ldots, SEP_k\}$ are all kept in memory. These are updated using Eqs. (48) to (55), except for SEP, which is adapted using Eq. (65). The data mean $\boldsymbol{\mu}_{data}$ is updated like the prototypes $\boldsymbol{v}$ (*i.e.*, Eq. (49) using $\boldsymbol{\mu}_{data}$ in place of $\boldsymbol{v}$ and N in place of n).

B. INCREMENTAL WB INDEX (IWB)

The iWB computation is very similar to that of iCH:

$$WB^{new} = k^{new} \frac{\sum_{i=1}^{k^{new}} CP_i^{new}}{\sum_{i=1}^{k^{new}} SEP_i^{new}}, \tag{66}$$

and the same variable definitions previously mentioned apply.

C. INCREMENTAL PAKHIRA-BANDYOPADHYAY-MAULIK INDEX (IPBM)

The iPBM computation is defined as:

$$PBM^{new} = \left[\frac{\max_{i \neq j} \left(\|\boldsymbol{v}_i^{new} - \boldsymbol{v}_j^{new}\|_2^2 \right)}{\sum_{i=1}^{k} CP_i^{new}} \times \frac{CP_0^{new}}{k^{new}} \right]^2, \tag{67}$$

where CP_0 (compactness of the entire data) and $\sum_{i=1}^{k} CP_i^{new}$ correspond to E_1 and E_k, respectively. These are updated according to Eqs. (48) to (55), along with the remaining compactness variables. Only the pairwise distances with respect to the updated prototype need to be recomputed at any given iteration.

D. INCREMENTAL SILHOUETTE INDEX (ISIL)

The SIL index is inherently batch (offline), because it requires the entire data set to be computed (the silhouette coefficients are averaged across all data samples in Eq. (21)). To remove such a requirement and enable incremental updates, a hard version of the centroid-based SIL variant introduced in [51] is employed here, as is the squared Euclidean norm (*i.e.*, $\|\cdot\|_2^2$); this is done in order to employ the recurrent formulation of the compactness in Eq. (51). Consider the matrix $\boldsymbol{S}_{k \times k}$, where k prototypes $\boldsymbol{v}_i$ are used to compute the centroid-based SIL (instead of the N samples $\boldsymbol{x}_i$ - which, by definition, are discarded after each presentation in online mode). Define each entry $s_{i,j} = D(\boldsymbol{v}_i, \omega_j)$ (dissimilarity of $\boldsymbol{v}_i$ to cluster ω_j) of $\boldsymbol{S}_{k \times k}$ as:

$$s_{i,j} = \frac{1}{n_j} \sum_{\substack{l=1 \\ \boldsymbol{x}_l \in \omega_j}}^{n_j} \|\boldsymbol{x}_l - \boldsymbol{v}_i\|_2^2 = \frac{1}{n_j} CP(\boldsymbol{v}_i, \omega_j), \tag{68}$$

where $i = \{1, \ldots, k\}$ and $j = \{1, \ldots, k\}$. The silhouette coefficients can be obtained from the entries of $\boldsymbol{S}_{k\times k}$ as:

$$sc_i = \frac{\min\limits_{l,l\neq J}(s_{i,l}) - s_{i,J}}{\max\left[s_{i,J}, \min\limits_{l,l\neq J}(s_{i,l})\right]}, \boldsymbol{v}_i \in \omega_J. \tag{69}$$

where $a_i = s_{i,J}$ and $b_i = \min\limits_{l,l\neq J}(s_{i,l})$.

Remark 2: At first, when examining Eq. (68), one might be tempted to store a $k \times k$ matrix of compactness entries along with their accompanying k^2 vectors $\boldsymbol{g}$ (one for each entry) to enable incremental updates of each element of matrix of $\boldsymbol{S}_{k\times k}$; this approach, however, may lead to unnecessarily large memory requirements. A more careful examination shows that it is sufficient to simply redefine CP and $\boldsymbol{g}$ for each cluster i ($i = \{1, \ldots, k\}$) as:

$$CP_i = \sum_{\substack{j=1\\ \boldsymbol{x}_j\in\omega_i}}^{n_i} \|\boldsymbol{x}_j - \vec{\boldsymbol{0}}\|_2^2 = \sum_{\substack{j=1\\ \boldsymbol{x}_j\in\omega_i}}^{n_i} \|\boldsymbol{x}_j\|_2^2, \tag{70}$$

$$\boldsymbol{g}_i = \sum_{\substack{j=1\\ \boldsymbol{x}_j\in\omega_i}}^{n_i} \left(\boldsymbol{x}_j - \vec{\boldsymbol{0}}\right) = \sum_{\substack{j=1\\ \boldsymbol{x}_j\in\omega_i}}^{n_i} \boldsymbol{x}_j, \tag{71}$$

which is equivalent to fixing $\boldsymbol{v} = \vec{\boldsymbol{0}}$. Therefore, their incremental update equations become (as opposed to Eqs. (51) and (55)):

$$CP_i^{new} = CP_i^{old} + \|\boldsymbol{x}\|_2^2, \tag{72}$$

$$\boldsymbol{g}_i^{new} = \boldsymbol{g}_i^{old} + \boldsymbol{x}. \tag{73}$$

Using this trick, when a sample $\boldsymbol{x}$ is assigned to cluster ω_J, then the update equations for each entry $s_{i,j}$ of $\boldsymbol{S}_{k\times k}$ are given by Eq. (74), as shown at the bottom of this page. Note that the numerators of the expressions in Eq. (74) update the compactness "as if" the prototype has changed from $\vec{\boldsymbol{0}}$ to $\boldsymbol{v}^{new}$ at every iteration ($\Delta\boldsymbol{v} = -\boldsymbol{v}^{new}$). The remaining variables such as n, N and $\boldsymbol{v}$ are updated as previously described. This allows $\{CP_1, \ldots, CP_k\}$ and $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k\}$ to continue being stored similarly to the previous iCVIs, instead of a $k \times k$ matrix of compactness and the associated k^2 vectors $\boldsymbol{g}$.

Remark 3: In the case in which a new cluster ω_{k+1} is created following the presentation of sample $\boldsymbol{x}$, then a new column and a new row are appended to the matrix $\boldsymbol{S}_{k\times k}$. Unlike the other iCVIs, the compactness CP_{k+1} and vector $\boldsymbol{g}_{k+1}$ of this cluster are initialized as $\|\boldsymbol{x}\|_2^2$ and $\boldsymbol{x}$, respectively. Then, the entries of $\boldsymbol{S}_{k\times k}$ are updated using Eq. (75), as shown at the bottom of this page.

Following the incremental updates of the entries of $\boldsymbol{S}_{k\times k}$ (Eq. (74) or (75)), the silhouette coefficients (sc_i) are computed (Eq. (69)), and the iSIL is updated using Eq. (76), as shown at the bottom of this page.

E. INCREMENTAL NEGENTROPY INCREMENT (INI)

The iNI computation is defined as:

$$NI^{new} = \sum_{i=1}^{k} p_i^{new} \ln\left(\frac{\sqrt{|\boldsymbol{\Sigma}_i^{new}|}}{p_i^{new}}\right) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_{data}| \tag{77}$$

where $p_i^{new} = n_i^{new}/N^{new}$, and $\boldsymbol{\Sigma}_i^{new}$ is computed using the following recursive formula [36]:

$$\boldsymbol{\Sigma}^{new} = \frac{n^{new}-2}{n^{new}-1}\left(\boldsymbol{\Sigma}^{old} - \delta I\right) + \frac{1}{n^{new}}\left(\boldsymbol{x} - v^{old}\right)\left(\boldsymbol{x} - v^{old}\right)^T + \delta I. \tag{78}$$

The authors of this work set $\delta = 10^{-\frac{\epsilon}{d}}$ to avoid numerical errors, where ϵ is a user-defined parameter. If a new cluster is created, then $\boldsymbol{\Sigma} = \delta I$ and $|\boldsymbol{\Sigma}| = 10^{-\epsilon}$.

F. INCREMENTAL REPRESENTATIVE CROSS INFORMATION POTENTIAL (IRCIP) AND CROSS-ENTROPY (IRH)

Section V will show that using the representative cross-entropy rH for computing the CEF makes it easier to

$$s_{i,j}^{new} = \begin{cases} \frac{1}{n_j^{new}}\left(CP_j^{old} + \|\boldsymbol{z}_i\|_2^2 + n_j^{old}\|\boldsymbol{v}_i^{old}\|_2^2 - 2\boldsymbol{v}_i^{old\,T}\boldsymbol{g}_j^{old}\right) & , (i\neq J, j = J) \\ \frac{1}{n_j^{old}}\left(CP_j^{old} + n_j^{old}\|\boldsymbol{v}_i^{new}\|_2^2 - 2\boldsymbol{v}_i^{new\,T}\boldsymbol{g}_j^{old}\right) & , (i = J, j\neq J) \\ \frac{1}{n_j^{new}}\left(CP_j^{old} + \|\boldsymbol{z}_j\|_2^2 + n_j^{old}\|\boldsymbol{v}_j^{new}\|_2^2 - 2\boldsymbol{v}_j^{new\,T}\boldsymbol{g}_j^{old}\right) & , (i = J, j = J) \\ s_{i,j}^{old} & , (i\neq J, j\neq J) \end{cases} \tag{74}$$

$$s_{i,j}^{new} = \begin{cases} CP_{k+1} + \|\boldsymbol{v}_i^{old}\|_2^2 - 2\boldsymbol{v}_i^{old\,T}\boldsymbol{g}_{k+1} & , (i\neq k+1, j = k+1) \\ \frac{1}{n_j^{old}}\left(CP_j^{old} + n_j^{old}\|\boldsymbol{v}_i^{new}\|_2^2 - 2\boldsymbol{v}_i^{new\,T}\boldsymbol{g}_j^{old}\right) & , (i = k+1, j\neq k+1) \\ 0 & , (i = k+1, j = k+1) \\ s_{i,j}^{old} & , (i\neq k+1, j\neq k+1) \end{cases} \tag{75}$$

$$SIL^{new} = \frac{1}{k^{new}}\sum_{i=1}^{k^{new}} sc_i^{new}. \tag{76}$$

observe the behavior of the incremental clustering process (this corroborates a previous study in which rH was deemed more informative than rCIP for multivariate data visualization [75]):

$$rH(\omega_i, \omega_j) = -\ln\left[rCIP(\omega_i, \omega_j)\right], \tag{79}$$

$$CEF = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} rH(\omega_i, \omega_j). \tag{80}$$

Note that, as opposed to the rCIP-based CEF, larger values of rH-based CEF indicate better clustering solutions. Concretely, because the CEF only measures separation, then, as with iNI, it is only necessary to update the means and the covariance matrices online in order to construct the incremental CEF (iCEF). This is also done using Eqs. (49) and (78), respectively. The iCEFs based on rCIP and rH are hereafter referred to as ircIP and irH, respectively.

G. INCREMENTAL CONN_INDEX (ICONN_INDEX)

The Conn_Index is an inherently batch CVI formulated around the ***CADJ*** and ***CONN*** matrices. Each element (i, j) of the ***CADJ*** matrix requires a count of the samples in the data set with the first and second closest prototypes, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ respectively, while the symetric ***CONN*** matrix is equal to the sum of the ***CADJ*** matrix with its transpose. When clustering data online, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ may change for previously presented samples, as prototypes are continuously modified or created. However, for the purpose of building and incrementing ***CADJ*** and ***CONN*** matrices online (with only one matrix entry changing per sample presentation), it is assumed that the trends exhibited over time by the iConn_Index do not differ dramatically from its offline counterpart. Batch calculation can be eliminated entirely by keeping the values of Eqs. (42) and (44) in memory and updating only the entries that depend on prototypes $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$.

In this study, the multi-prototype cluster representation required by the Conn_Index was generated using a modified fuzzy ARTMAP, whose modules A and B were used for prototype and cluster definition, respectively. Module A of fuzzy ARTMAP was modified in such a way as to forcefully create two prototypes using the first two samples of every emerging cluster in module B. By enforcing this dynamic, each cluster always possesses at least two prototypes for the computation of the iConn_Index. This strategy addresses two problems: first, it allows ***CADJ*** to be created from the second presented sample and onward; second, it prevents cases in which well-separated clusters are strongly connected simply because one of them does not have another prototype to assume the role of the second winner. The main algorithmic steps of the modified simplified fuzzy ARTMAP are summarized in Algorithm 1.

Remark 4: Fuzzy ART neural networks represent prototypes by the categories' weight vectors $\boldsymbol{w}$ (see Section II-C). Thus, the highest-ranked resonant category (i.e., the one with the largest activation function values according to Eq. (61) that also satisfies the resonance criteria of Eq. (62)) constitutes the first winner. The second winning prototype for a sample ($\boldsymbol{w}_j$) is the highest activated A-side category when the first winning prototype ($\boldsymbol{w}_i$) has been removed from the A-side category set.

Upon receiving the very first sample input, we can only form a single viable cluster and prototype; therefore, we would not be able to calculate the iConn_Index. We remedy this by introducing a counter separate from the ***CADJ*** matrix. This counter is incremented to count the number of times a sample has been presented while only a single prototype exists, thus preserving these otherwise troublesome samples. Upon creation of the second prototype $\boldsymbol{w}_2$ in the fuzzy ARTMAP module A, the ***CADJ*** matrix will be incremented for the first time at element $(2, 1)$. At this point, the element $(1, 2)$ will be incremented by the value of the instance counter. When this instance-counting technique is combined with the forcible splitting of prototypes previously mentioned, the result is that all samples will be taken into account when computing the iConn_Index. For all subsequent samples, the instance counter will remain unused,

Algorithm 1 Modified simplified fuzzy ARTMAP

```
/* Initialization                                          */
w_1 := 1⃗ (first weight vector of cluster ω_1) ;
n_{ω_1} := 0 (number of samples of cluster ω_1) ;
C := 1 (number of categories in the network) ;
/* Learning                                                */
Present new sample and label pair (x, y);
n_{ω_y} := n_{ω_y} + 1 ;
J_i := {∅}, i ∈ {1, 2} (first and second winners) ;
Generate input I using Eq. (60);
Compute the activation functions (T's) of all categories
  using Eq. (61);
Sort the activation functions (T's) in descending order
  and store indices in vector I ;
for j:= 1 to C do
    J := I(j);
    if w_J satisfies Eq. (62) and category J does not
    map to any cluster y' ≠ y and n_{ω_y} ≠ 2 then
        Update w_J using Eq. (63);
        J_1 := J ;
        break ;
    end
end
if J_1 == {∅} then
    C := C + 1, J := C, w_J := I ;
    Map the new category J to cluster y ;
    J_1 := J, J_2 := I(1) ;
end
if J_1 == I(1) then
    J_2 := I(2) ;
else
    J_2 := I(1) ;
end
```

the ***CONN*** and ***CADJ*** incrementing will be streamlined, and the iConn_Index will be calculable.

Remark 5: The iConn_Index boundary conditions are listed below:

1) Cluster represented by a single prototype (singleton), e.g., immediately following the creation of a new cluster: the ***Intra*** entry for that cluster, given by Eq. (42), defaults to a value of 0 because $\boldsymbol{CADJ}(i, i) = 0\ \forall i$.
2) A single non-singleton cluster exists (i.e., a unique cluster represented by multiple prototypes): $\boldsymbol{Intra} = 1$ for this cluster.
3) Like the remaining iCVIs in this study, iConn_Index is not defined for a single cluster because ***Inter*** (Eq. (43)) cannot be computed.
4) Instead of the original constraint $\boldsymbol{CADJ}(i, j) > 0$ imposed by Eq. (45), the implementation of iConn_Index in this paper uses $\boldsymbol{CONN}(i, j) > 0$.

Note that items (1)-(3) arise directly from the Conn_Index definitions [38], whereas item (4) follows from the step-by-step illustrative example in [37]. For further clarity, the pseudo-code for the iConn_Index is provided in Algorithm 2.

Algorithm 2 iConn_Index

```
/* Initialization                                   */
CADJ := {∅} ;
CONN := {∅} ;
Inter := {∅};
Intra := {∅};
Inter_Conn := 0;
Intra_Conn := 0;
S_counter := 0;
/* iConn_Index computation                          */
while streaming samples do
    x:= new sample;
    Process x with an ART-based model to obtain the
    first w_i ∈ ω_k (i == J_1) and second w_j ∈ ω_l
    (j == J_2) best matching prototypes;
    if w_j = {∅} then
        S_counter := S_counter + 1;
    else if S_counter > 0 then
        CADJ(w_j, w_i) := CADJ(w_j, w_i) + S_counter;
        S_counter := 0;
    end
    if S_counter = 0 then
        CADJ(w_i, w_j) := CADJ(w_i, w_j) + 1;
        Update CONN using Eq. (36);
        Update Intra(ω_k) using Eq. (42);
        if ω_k ≠ ω_l then
            Update Inter(ω_k, ω_l) and Inter(ω_l, ω_k)
            using Eq. (44);
        else
            Update Inter(ω_k, ω_m), ∀m using Eq. (44);
        end
        Recompute Intra_Conn using Eq. (41);
        Recompute Inter_Conn using Eq. (43);
        Recompute Conn_Index using Eq. (40);
    end
end
```

IV. NUMERICAL EXPERIMENTS DESIGN

The behaviors of 13 iCVIs (namely iCH, iSIL, iPBM, iWB, iXB, iDB, iGD_{43}, iGD_{53}, PS, iNI, irCIP, irH, and iConn_Index) were analyzed using the benchmark data sets summarized in Table 2. These synthetic and real-world data sets are also depicted in the scatter plots shown in Fig. 1 and encompass a diverse set of properties, such as unbalanced classes, high dimensionality, levels of overlap and number of samples.

As in [17]–[20], [25], a natural ordering, *i.e.*, meaningful temporal information, is assumed. To emulate such scenarios, the samples were presented in a *cluster-by-cluster* fashion (samples within a given cluster were randomized), and thus, this experiment setup is suitable for change-point detection [25]. All iCVIs were subjected to the same 10 random orders of clusters (and order of samples within each cluster) per data set per experiment (see Sections V-A to V-C).

The following discussion relates to the data sets used in the experiments and the application of linear normalization. Normalization assumes knowledge of the minimum and maximum data statistics, as the vector quantization required by the iConn_Index is realized via fuzzy ARTMAP. Therefore, for consistency, all data sets were normalized to the unit cube $[0, 1]^d$. Additionally, note that the fuzzy ARTMAP dynamics were performed with the additional application of complement-coding [40]. Finally, note that this study does not employ multi-prototype representations for irCIP or irH, *i.e.*, $M_i = M_j = 1, \forall i, j$ in Eq. (34), because unlike iConn_Index, such representations are not mandatory for their computation. Moreover, in these experiments, $\epsilon = 12$ in Eq. (78) for the incremental computation of the covariance matrices used by irCIP, irH and iNI.

The experiments and statistical analysis were carried out using the MATLAB software environment and the scmamp R package [90], respectively. The source code of the CVIs/iCVIs, algorithms of the ART models and experiments as well as links to the figures depicting the iCVIs' behaviors across all experiments for the data sets listed in Table 2 are provided by the *iCVI MATLAB Toolbox* at the Applied Computational Intelligence Laboratory public GitHub repository[1].

V. A COMPARATIVE STUDY

This section discusses the behavior of the iCVIs in three general cases when assessing the quality of clustering solutions in real-time: (1) correct partitions, (2) under-partitions, and (3) over-partitions. It should be emphasized that the purpose

[1] https://github.com/ACIL-Group/iCVI-toolbox

TABLE 2. Summary of the data sets' characteristics.

Data set	# Samples	# Features	# Clusters	Reference(s)
Synthetic data sets				
A3[a]	7500	2	50	[76], [77]
Birch1[a]	100000	2	100	[76], [78]
Birch2[a]	100000	2	100	[76], [78]
Dim032[a]	1024	32	16	[76], [79]
Dim064[a]	1024	64	16	[76], [79]
Dim128[a]	1024	128	16	[76], [79]
Dim256[a]	1024	256	16	[76], [79]
Dim512[a]	1024	512	16	[76], [79]
Dim1024[a]	1024	1024	16	[76], [79]
S1[a]	5000	2	15	[76], [80]
S2[a]	5000	2	15	[76], [80]
S3[a]	5000	2	15	[76], [80]
S4[a]	5000	2	15	[76], [80]
Unbalance[a]	6500	2	8	[76], [81]
Aggregation[a]	788	2	7	[82]
D31[a]	3100	2	31	[83]
R15[a]	600	2	15	[76], [83]
Hepta[b]	212	3	7	[84]
Lsun[b]	400	2	3	[84]
Tetra[b]	400	3	4	[84]
Real-world data sets				
Isolet[c]	7797	617	26	[85]–[87]
MNIST[c]	70000	784	10	[86]–[88]

[a] *Clustering basic benchmark* and *Other clustering datasets*, available at http://cs.uef.fi/sipu/datasets.
[b] *Fundamental Clustering Problem Suite*, available at https://www.uni-marburg.de/fb12/arbeitsgruppen/datenbionik/data?language_sync=1.
[c] MATLAB processed data sets, available at http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html.

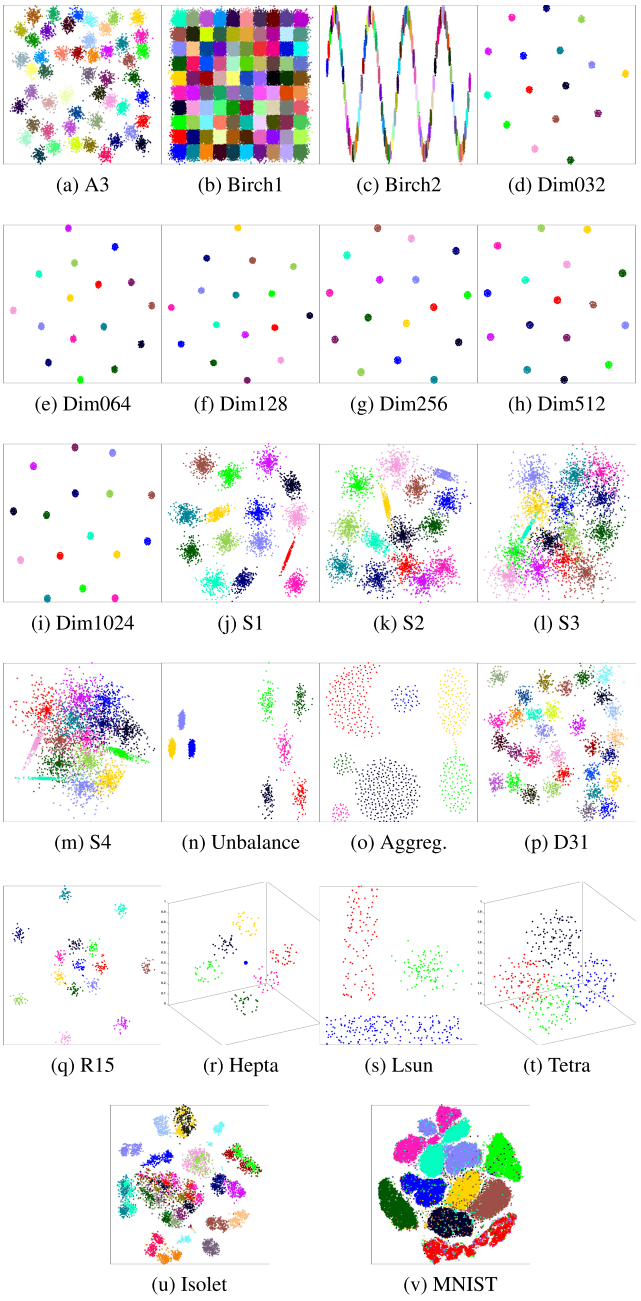


FIGURE 1. (a)-(t) Synthetic data sets. (u)-(v) Real-world data sets. High-dimensional data sets are shown using a 2-dimensional t-distributed stochastic neighbor embedding (t-SNE) [89] projection.

of this study is to observe the behavior of the iCVIs in these different scenarios in order to gain insight into their applicability; agnostic to the clustering algorithm. For that reason this analysis will not focus on evaluating the performance or capabilities of specific online clustering algorithms. Similar to [25], in each of these scenarios, we investigate the iCVIs' dynamics triggered by the two following events: (a) the creation of a new cluster (for scenarios (1) and (3)) or the merging of two clusters (for scenario (2)), and (b) the assignment of samples to the current (existing) cluster.

Note that this is not an exhaustive study of all possible permutations of clusters and samples (which is $k!$ for clusters, and $n_i!$ within each cluster i). Nonetheless, we seek to find typical behaviors that would allow the inference of specific problems that may arise during incremental unsupervised learning; iCVIs should help the practitioner to identify issues by yielding good values when correctly partitioning and bad values when problems occur. Particularly, the observations from case (1) are used as a reference behavior (or default) to which cases (2) and (3) are compared. The overarching goal is to observe the capabilities of the iCVIs in identifying anomalous behaviors caused by *synthetically* generated problems (under- and over-partitions).

A. CORRECT PARTITIONS

Assume that a suitable clustering algorithm was selected and optimally parameterized, thus yielding correct data partitions when presenting samples in a given cluster-by-cluster ordering. Again, the goal of this study is not to compare the merits of any particular incremental clustering algorithm; Therefore, in order to emulate the scenario previously described and make the experiments *clustering algorithm agnostic*, we simply *cluster* by merely classifying each sample based on their true labels and compute the iCVIs incrementally. This experimental setup relies on the assumption that, if there exists a subset of clustering algorithms that can perfectly cluster a given data set, then at each point in time they must make the same, and correct, sample assignment to clusters.

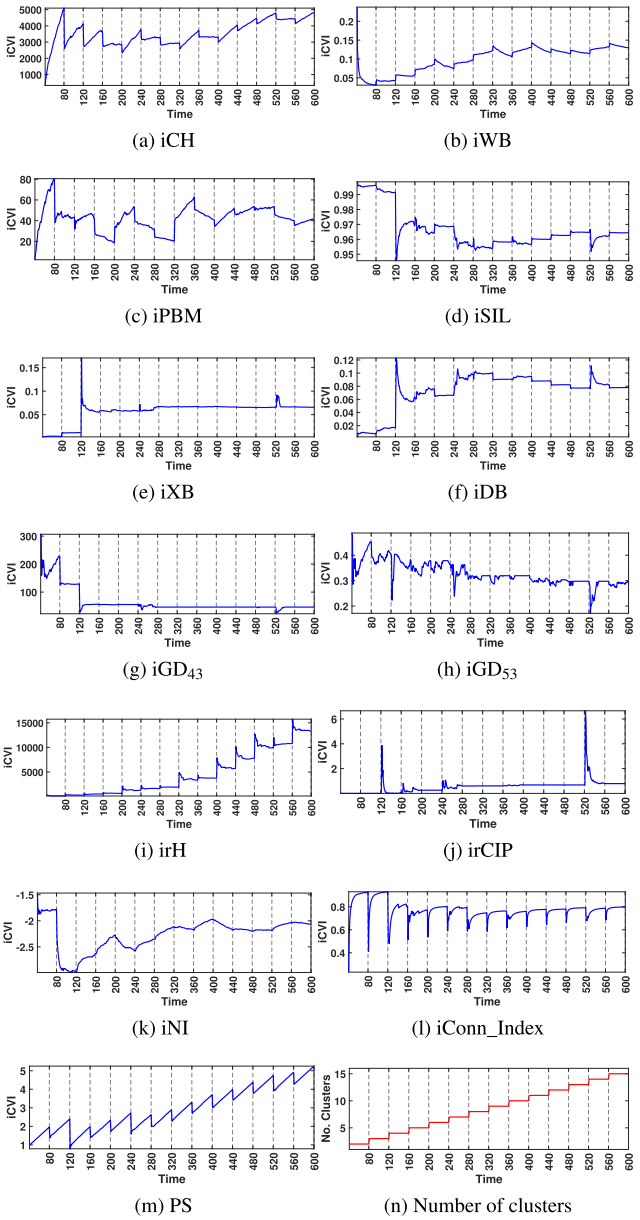


FIGURE 2. **(a)-(m) Behaviors of iCVIs (blue curves) when correctly partitioning the data set *R15*. (n) The number of clusters is depicted by the step-like red curve. Each discrete time instant (x-axis) corresponds to the presentation of one sample. The dashed vertical lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster, whereas samples after it belong to another.**

Furthermore, such correct assignments should be indicated by good iCVI values.

For brevity, Fig. 2 shows the behaviors of the iCVIs when correctly partitioning only the *R15* data set. Note that this figure depicts the behaviors of the iCVIs immediately following the creation of a second cluster because they usually cannot be computed for a single cluster. Note how iConn_Index behavior tends to follow an exponential of the form $A(1 - e^{-Bt})$ during the presentation of each cluster in well-behaved data sets. The response appears to approach the somewhat step-like behavior of its batch counterpart (see Section VI).

Fig. 2 also shows that sudden changes in many iCVI values follow the emergence of new clusters (as expected from previous studies [17]–[20], [25]). During the presentation of samples belonging to a particular cluster, different behaviors can be observed. To identify trends among the iCVIs in a principled manner, in each run of each data set, the following experimental data were collected:

1) The number of times the iCVI increased, decreased and remained constant immediately following the creation of a new cluster (hereafter referred to as immediate behavior).
2) The number of times the iCVI increased, decreased and remained constant during the assignment of samples to the current existing cluster (hereafter referred to as medium-term behavior). Particularly, in each time interval corresponding to the presentation of samples belonging to an existing cluster, a simple linear regression model [91] was fit and a t-test was performed for the first-order coefficient (slope). If the null hypothesis could be rejected under a 5% significance level, then we observed the sign of the first-order coefficient; it was counted as an increasing trend if positive and a decreasing trend if negative. Otherwise, if the t-test result was not deemed statistically significant, then the behavior was accounted for as constant (i.e., no iCVI change).

The experimental data from both (1) and (2) were then averaged across 10 runs for each data set. Next, both were analyzed by adapting the methodology discussed in [90], [92] to our problem. In particular,

1) We performed the Iman–Davenport version of Friedman's rank sum test to check the hypothesis that these trends are equally typical/probable.
2) If the null hypothesis of the previous test was rejected, then we proceeded with a post-hoc test (Bergmann–Hommel's method) to identify the most typical/probable trend.

This analysis was repeated for all iCVIs, and the results are summarized in Table 3. We emphasize that the behaviors listed in Table 3 are typical, *not exclusive*. The only iCVIs that consistently behaved following solely the trends shown in Table 3 (i.e., for all data sets, without exceptions) were the iCH, iWB, iConn_Index, iGD_{53}, irH and PS for experimental data (1). The iCVIs generally exhibited different trends, nonetheless, the ones that occurred frequently enough to be deemed as statistically significant are reported in Table 3.

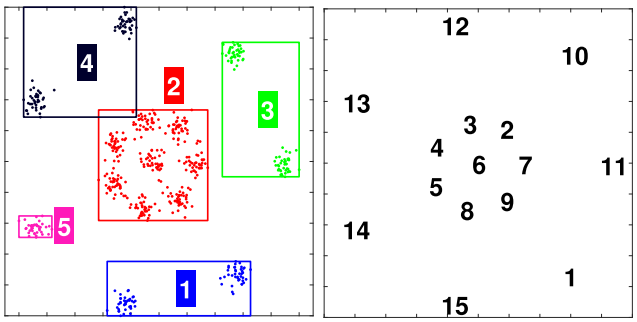
B. UNDER-PARTITIONS

Consider a scenario in which a suboptimal clustering algorithm is selected or an appropriate one is badly parameterized such that it yields an under-partition of the data set at hand. For instance, Fig. 3a shows an under-partition of the *R15* data set yielded by a fuzzy ART trained under a suboptimal parameter setting where clusters were presented in the order depicted in Fig. 3b. We are interested in how similar scenarios

TABLE 3. Summary of the iCVIs' behaviors when correctly, under-, and over-partitioning the synthetic data sets used in the experiments.

iCVI	Type	Optimality	Correct partition		Under-partition		Over-partition	
			immediate	medium-term[a]	immediate	medium-term[b]	immediate	medium-term[c]
iCH	SS-based	max-optimal	Decrease	Increase	Decrease	**Decrease**	Decrease	Increase
iPBM		max-optimal	Decrease	Increase	Decrease	**Decrease**	Decrease	-
iSIL		max-optimal	Increase	Decrease	Decrease	*Decrease*	**Decrease**	*Decrease*
iWB		min-optimal	Increase	Decrease	Increase	**Increase**	Increase	Decrease
iXB		min-optimal	Decrease	-	Increase	*Increase*	**Increase**	*Increase*
iDB		min-optimal	Decrease	Increase	Increase	*Increase*	**Increase**	*Increase*
iGD_{43}		max-optimal	Constant	Increase	-	**Decrease**	**Decrease**	**Decrease**
iGD_{53}		max-optimal	Decrease	Increase	-	**Decrease**	Decrease	Increase
PS	Separation-based	max-optimal	Decrease	Increase	-	**Decrease**	Decrease	Increase
iNI[d]	IT-based	min-optimal	Decrease	-	Increase	*Increase*	-	*Increase*
irCIP[d]		min-optimal	Increase	-	Increase	*Increase*	Increase	*Increase*
irH[d]		max-optimal	Increase	Decrease	Decrease	*Decrease*	Increase	Decrease
iConn_Index	Graph-based	max-optimal	Decrease	Increase	Decrease	Increase	Decrease	**Decrease**

[a] Behavior during the correct assignment of samples to the current cluster.
[b] Behavior during the merging of two clusters.
[c] Behavior during the split of a cluster.
[d] The experiments with these iCVIs did not include the high-dimensional synthetic data sets of *Dim512* and *Dim1024* given the issues associated with the reliable estimation of covariance matrices in high-dimensional spaces under small sample sizes.
Empty cells ("-"): these stand for inconclusive; no single trend stood out in a statistically significant manner (i.e., no statistical difference was observed among either the two highest-ranked behaviors or among all three behaviors).
Under-partition: behaviors in **boldface** are useful indicators of problems occurring during online clustering, as they differ from their respective counterparts when correctly partitioning the synthetic data sets used in this study; behaviors in *italic* are also useful indicators when they dominate the "natural" tendencies of their respective iCVIs.
Over-partition: behaviors in **boldface** are useful indicators of problems occurring during online clustering, as they differ from their respective counterparts when correctly partitioning the synthetic data sets used in this study; behaviors in *italic* are also useful indicators when very pronounced.



(a) Data partition (b) cluster-by-cluster order

FIGURE 3. (a) An under-partition of the data set *R15* by fuzzy ART. (b) Presentation order of the clusters.

would reflect in the iCVIs' behaviors (ideally they should yield poor values) and how strikingly these would deviate from the reference (*i.e.*, according to Table 3). Therefore, we *deliberately* under-partitioned each data set by randomly merging two close clusters, which were selected using a probability proportional to the Euclidean distance between their centroids. In particular, the probability of selecting clusters i and j for merging is given by:

$$p_{i,j} = \frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^6}{\sum_{m,n}^{\binom{k}{2}} \|\boldsymbol{v}_m - \boldsymbol{v}_n\|_2^6}, \tag{81}$$

where the 6^{th} power is used for contrast enhancement. After a cluster pair is selected, they are assigned the same label during the online computation of the iCVIs. It is reasonable to assume that a clustering algorithm might allocate samples from clusters closer together rather than those from clusters farther apart. Equation (81) is used to avoid repeatedly merging the same two closest clusters in all runs.

For brevity, Fig. 4 shows the behaviors of the iCVIs when under-partitioning only the *R15* data set. The gray shaded areas shown in these figures correspond to the exact time interval in which samples from different clusters were merged, and thus the total number of clusters remained constant. Note that the merged clusters were not necessarily consecutive, given that the sequence of clusters was randomized.

In order to identify under-partitioning trends among the iCVIs in a principled manner, the following data were collected during each run of each data set:

1) The number of times the iCVI increased, decreased and remained constant immediately following the first merged sample (hereafter referred to as immediate behavior).
2) The number of times the iCVI increased, decreased and remained constant during the incorrect assignment of samples, *i.e.*, during merging (hereafter referred to as medium-term behavior).

The procedures discussed in Section V-A were used to obtain the experimental data (2) and to perform the statistical comparison among trends. The results obtained and summarized in Table 3 show that:

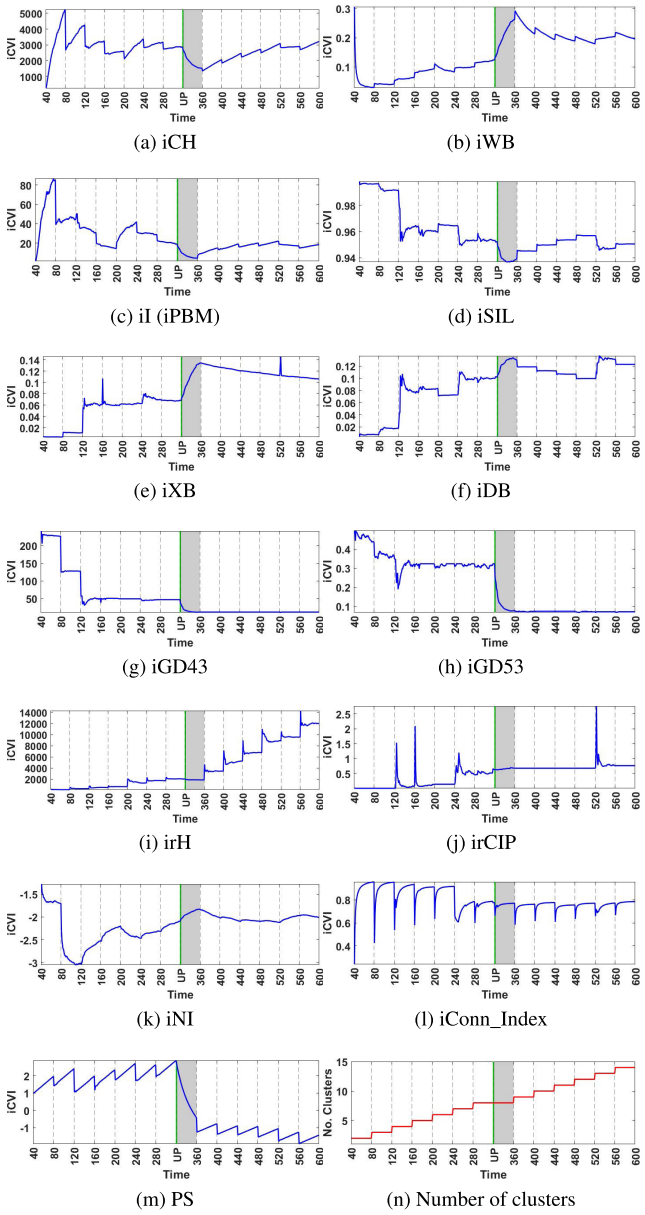


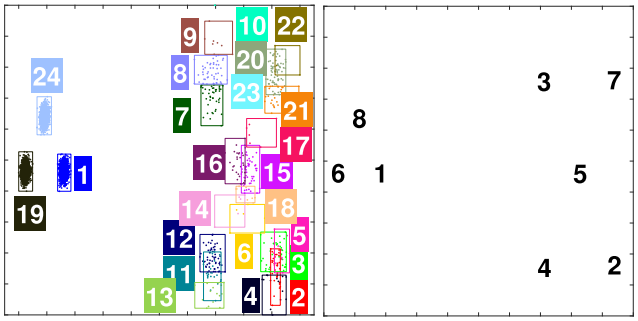
FIGURE 4. Each discrete time instant (x-axis) corresponds to the presentation of one sample of the data set *R15* during the under-partitioning experiment. The dashed vertical black lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster, whereas samples after it belong to another. The continuous vertical green lines indicate the instant in which the under-partition (UP) problem starts, and the samples delimited by the gray shaded interval are assigned to an existing cluster, instead of forming a new one. (a)-(m) Behaviors of iCVIs (blue curves). (n) Number of clusters (step-like red curve).

1) All iCVIs consistently worsened as the algorithm incorrectly agglomerated samples from different clusters (behavior during merging). The exception was the iConn_Index, for which an overall increasing trend was deemed statistically significant. Additionally, compared to the correct partition experiment *while under a constant number of clusters*, the iCH, iPBM, iWB, iGD_{43}, iGD_{53} and PS exhibited opposite behavior, which is a strong indication of the occurrence of this problem in the clustering process.
2) Immediately after starting to incorrectly merge clusters (*i.e.*, first merged sample), the performances of most iCVIs typically were accompanied by a change toward worse values under a constant number of clusters. The exceptions were iGD_{43}, iGD_{53} and PS, which did not exhibit a statistically significant immediate behavior across our experiments.
3) Although the trends exhibited by iSIL and iDB during merging are similar to the correct partition case under constant number of clusters, it remains possible to infer the under-partition issue because, in many cases, a sudden and pronounced worsening of these iCVIs was observed as a defining characteristic following such a problem. Many of these worsening trends during merging dominated the "natural" worsening tendencies of these iCVIs (*e.g.*, Figs. 4d and 4f). However, instances exist in which the challenge relies in patently identifying, without any external knowledge, how much the relative worsening would actually constitute a problem. The latter issue can potentially affect many iCVIs; for instance, it is also present in irCIP and irH. The statistically significant trend of the irH during merging is also the same as the ones for correct partitions. Although no direct comparison to the correct partitioning case is possible for iXB, a similar aggressive worsening behavior was observed in many cases (*e.g.*, Fig. 4e); thus, analogous conclusions and caveats apply.

In summary, a worsening iCVI trend *under a constant number of clusters* indicates that the clustering algorithm might be mistakenly grouping the samples under the same cluster umbrella, and thus should trigger the practitioner's attention. However, it is important to be cautious with respect to false positives. Even when a correct partition was retrieved in the experiments of Section V-A, some iCVIs exhibited large fluctuations while assigning samples of some data sets to their correct cluster (with the number of clusters remaining constant in that interval) as well as false negatives. It should be made clear that the behaviors listed in Table 3 are typical *not* exclusive. As a general recommendation, abrupt changes toward worse values of an iCVI under a constant number of clusters should be examined carefully. Also, as noted in [25], it is recommended to observe multiple iCVIs concurrently. This is especially important for reliable under-partition detection.

C. OVER-PARTITIONS

Consider a scenario in which a suboptimal clustering algorithm or parameterization is selected such that the data set at hand is over-partitioned. For instance, Fig. 5a shows an over-partition of the *unbalance* data set yielded by standard fuzzy ART (the clusters were presented in the order depicted in Fig. 5b), which is suboptimal given that the global vigilance parameter (ρ) assumes equally sized clusters. We are



(a) Data partition (b) cluster-by-cluster order

FIGURE 5. **(a) An over-partition of the data set *unbalance* by fuzzy ART. (b) Presentation order of the clusters.**

interested in how over-partition would reflect in the iCVIs' behaviors (ideally they should yield poor values) and how strikingly these would deviate from their expected behaviors when correct partitions are detected (*i.e.*, the "references" according to Table 3). Therefore, we *deliberately* over-partitioned each data set by splitting one of its clusters. A cluster was chosen for splitting with a probability proportional to its size, thus favoring the selection of large clusters. It is reasonable to assume that certain clustering algorithms, such as standard ART-based ones, would split large clusters according to their parameterization (e.g., the problem depicted in Fig. 5). A cluster size is measured by the smallest hyperrectangle that encloses all of its points. Thus, for cluster i, hyperrectangle size R_i is measured as [39]:

$$R_i = d - \| \bigwedge_{\boldsymbol{I}_j \in \omega_i} \boldsymbol{I}_j \|_1, \tag{82}$$

where $\boldsymbol{I}_j$ is the complement-coded version of $\boldsymbol{x}_j$ (see Eq. (60)). To avoid splitting large clusters with a small number of samples (n) and consequently permit a better observation of the iCVIs' behaviors during over-partition, if $n_i < 10$, then R_i was set to 0.

Naturally, some method must be employed to split a cluster. That is why the over-partition experiment is (unfortunately) not completely clustering algorithm agnostic; fuzzy ART was used to create the over-partition. Therefore, results might be somewhat biased toward fuzzy ART solutions. For clarity, the selected clusters were split only into two sub-clusters. In particular, for each selected cluster, that cluster's samples were shuffled and fed to fuzzy ART modules trained for 1 epoch (*i.e.*, online mode) with progressively larger vigilance parameter (ρ) values until a solution with 3 clusters was found, in which case the vigilance parameter sweep was stopped. The vigilance values for the fuzzy ART trained with that specific sample order were successively increased using:

$$\rho(t+1) = \frac{1}{C(t+1)}(\rho(0) + C(t+1) - 1), \tag{83}$$

$$C(t+1) = C(t) + \delta, \tag{84}$$

where $\rho(0) = 1 - \frac{R_s}{d}$, R_s is the size of the selected cluster s (Eq. (82)), $C(0) = 1$ and δ is the step size, which was set to 0.001. From Eqs. (83) and (84), the constraint on the subcluster sizes becomes increasingly more strict as ρ increases. For instance, $C = 2$ would correspond to a maximum category size equal to half the size of the selected cluster [93]. However, given the ordering effects, the value selected for C is not necessarily equal to 2, hence the necessity of the vigilance parameter sweep, which is defined following the strategy described in Eqs. (83) and (84). This process was repeated for 10 random orders, and the clustering solution that yielded the two most balanced subclusters was used in the over-partition experiment. This strategy was followed in order to (i) create a realistic over-partition case for that cluster in online unsupervised learning mode, (ii) facilitate the observation of over-partition behaviors and (iii) avoid the creation of singletons. The over-partition experiment then proceeds as in the previous sections, but using the fuzzy ART labels for the split cluster during the online computation of the iCVIs.

For brevity, Fig. 6 shows the iCVIs' behaviors when over-partitioning only the *R15* data set. The gray shaded areas shown in these figures correspond to the time interval in which samples belonging to the same cluster were split into two subclusters. Note that the subclusters' samples are randomly presented, *i.e.*, they are *not* presented in a subcluster-by-subcluster manner.

To identify over-partitioning trends among the iCVIs in a principled manner, in each run of each data set, the following data were collected:

1) The number of times the iCVI increased, decreased and remained constant immediately following a clusters' split (hereafter referred to as immediate behavior).
2) The number of times the iCVI increased, decreased and remained constant during the split of the large cluster (hereafter referred to as medium-term behavior).

The procedures discussed in Section V-A were used to obtain the experimental data (ii) and to perform the statistical comparison among trends. The results, which are summarized in Table 3, show that:

1) The typical behaviors of iCH, iWB, PS, iGD_{53} and irH are usually indiscernible from the ones expected when accurately partitioning during both the (incorrect) creations of new clusters and the presentation of samples belonging to the current cluster. Additionally, the iPBM only exhibited one typical behavior, namely for the creation of a new cluster event, which was again identical to the correct partition case. Therefore, these iCVIs do not seem suitable for identifying over-partitions.
2) The iSIL, iXB and iDB only deviated partially, *i.e.*, they deviated for one trend, particularly the creation of a new cluster when incorrectly splitting a cluster. Although the typical trends exhibited by iSIL and iDB during the cluster split are identical to the correct partition case, and no direct comparison for iXB is possible, these iCVIs values underwent a

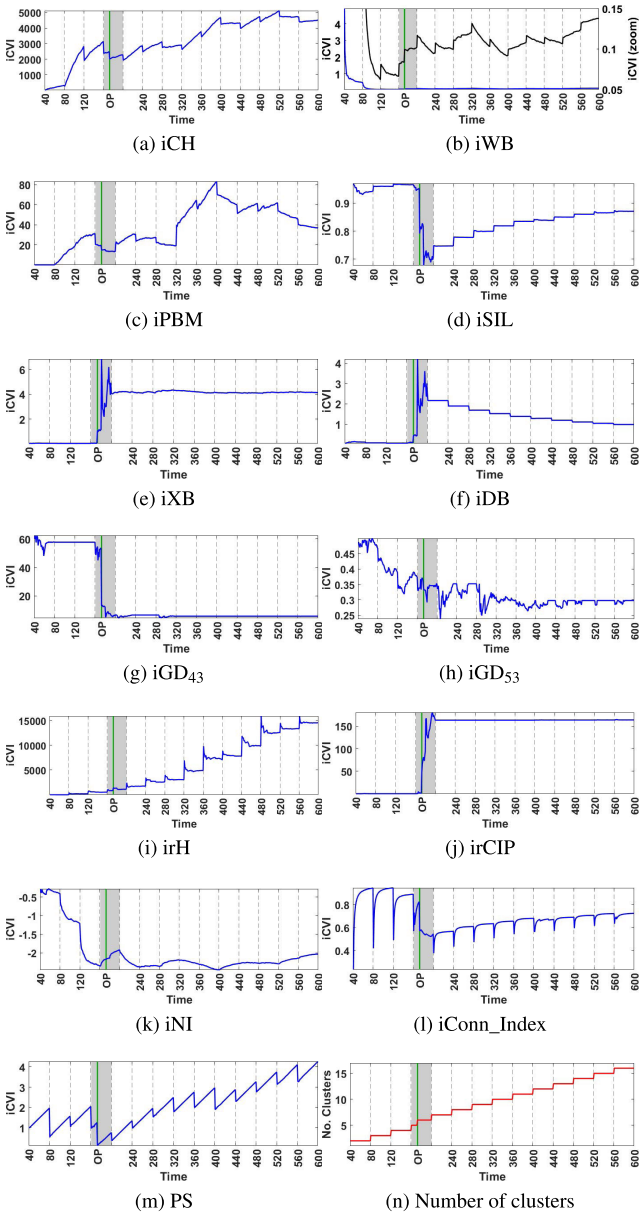


FIGURE 6. Each discrete time instant (x-axis) corresponds to the presentation of one sample of the data set *R15* during the over-partitioning experiment. The dashed vertical black lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster, whereas samples after it belong to another. The continuous vertical green lines indicate the instant in which the over-partition (OP) problem starts in the cluster delimited by the gray shaded interval. (a)-(m) Behaviors of iCVIs (blue curves). (n) Number of clusters (step-like red curve).

pronounced worsening in many data sets during the split (*e.g.*, Figs. 6d, 6e, and 6f). Similar to the discussion in the under-partitioning case, defining how much worsening would become a problem can be subjective, especially in borderline cases and with no additional information. Nevertheless, these iCVIs can indicate over-partitioning problems.

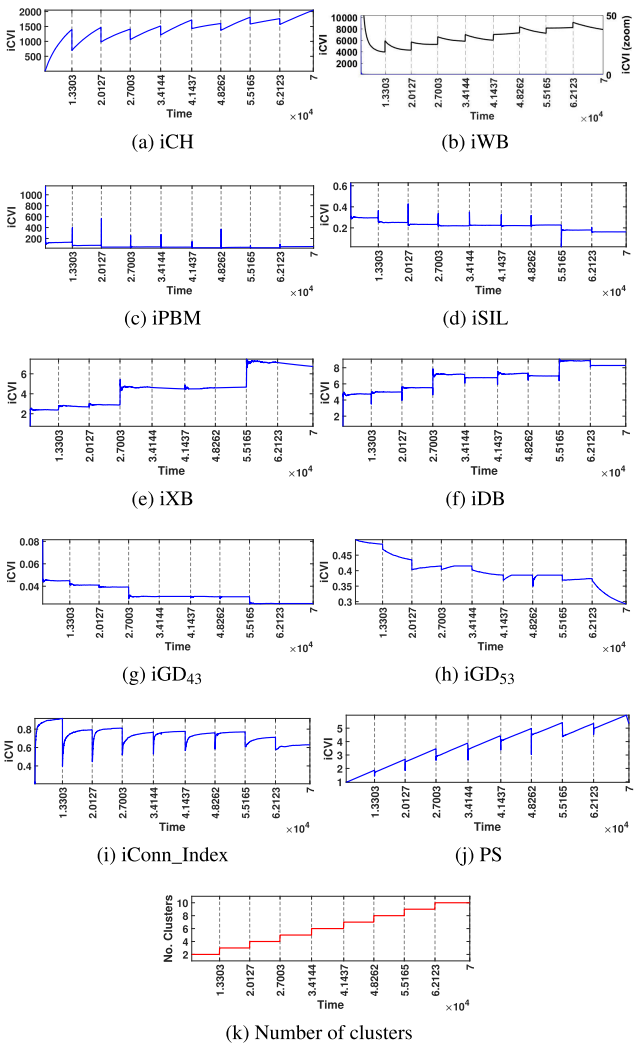
3) The irCIP exhibited the same typical trend following the presentation of the first sample of the second subcluster, and no direct comparison to the correct partition scenario is possible during split. However, as with the iSIL, iXB and iDB, the irCIP usually undergoes a noticeable worsening during the splitting of the cluster (*e.g.*, Fig. 6j).
4) The iGD_{43} and iConn_Index were the only iCVIs that exhibited trends opposite to their correct-partition experiment counterparts, thereby providing a strong indication of over-partition over time. Moreover, when clustering well-behaved data sets such as *dim032* through *dim1024*, the iConn_Index does not follow its characteristic exponential curve (expected from correct partitions) after the erroneous creation of a new cluster and subsequent incorrect assignment of samples. In turn, the iGD_{43} was the only iCVI that exhibited opposite tendencies for both the emergence of a new cluster and the posterior assignment of samples.

In summary, six out of thirteen iCVIs (iCH, iPBM, iWB, iGD_{53}, PS and irH) did not provide distinctive insights to enable definitive detection of the over-partition problems. In this scenario, without additional a priori information (*e.g.*, the cardinality of clusters) to indicate a premature partition, these iCVIs were unable to patently identify over-partition based on their immediate and/or medium-term behaviors. On the other hand, five iCVIs (iSIL, iXB, iDB, irCIP, and iNI) hinted at over-partition through their immediate behaviors and/or a considerable worsening of their medium-term behaviors. For these iCVIs, the medium-term behavior was either the same as the correct partition scenario, or a direct comparison was not possible. Finally, two iCVIs (iGD_{43} and iConn_Index) showed tendencies opposite to what was expected during the split, thus providing the strongest evidence for this particular problem.

Note that although there is a natural order for the presentation of clusters, the presentation of samples within each cluster is random. Consequently, samples of the over-partitioned cluster are *not* presented in a subcluster-by-subcluster manner. This adds another layer of complexity and thus makes this problem even more challenging. Also note that the vast majority of behaviors are typical, not deterministic, so we strongly recommend that the practitioner observe several iCVIs concurrently in order to accurately detect over-partition and to limit false positives/negatives.

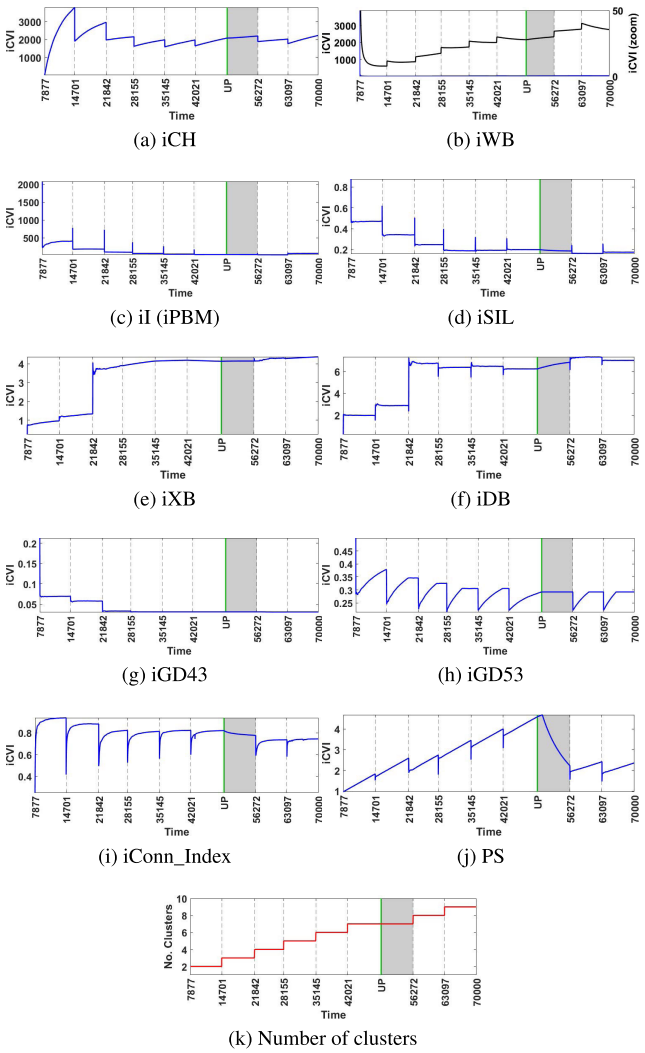
D. EXPERIMENTS WITH REAL-WORLD DATA SETS

In light of the results obtained for the synthetic data sets, in this section we analyze the scenarios of correct, under- and over-partition performed with the real-world data sets of *MNIST* and *Isolet*. The experiments were carried out under the same settings previously described. The discussion in this section is based on the observation of frequency of occurrence of each trend for these two data sets across 10 runs. For brevity, Figs. 7 through 9 illustrate the iCVIs' behaviors under correct, under-, and over-partition only for the *MNIST* data set. The iNI, irCIP and irH were not observed, given the issues associated with the reliable estimation of covariance matrices in high-dimensional spaces.



(a) iCH (b) iWB (c) iPBM (d) iSIL (e) iXB (f) iDB (g) iGD_{43} (h) iGD_{53} (i) iConn_Index (j) PS (k) Number of clusters

FIGURE 7. (a)-(j) Behaviors of iCVIs (blue curves) when correctly partitioning the data set *MNIST*. (k) The number of clusters is depicted by the step-like red curve. Each discrete time instant (x-axis) corresponds to the presentation of one sample. The dashed vertical lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster, whereas samples after it belong to another.



(a) iCH (b) iWB (c) iI (iPBM) (d) iSIL (e) iXB (f) iDB (g) iGD43 (h) iGD53 (i) iConn_Index (j) PS (k) Number of clusters

FIGURE 8. Each discrete time instant (x-axis) corresponds to the presentation of one sample of the data set *MNIST* during the under-partitioning experiment. The dashed vertical black lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster whereas samples after it belong to another. The continuous vertical green lines indicate the instant in which the under-partition (UP) problem starts; the samples delimited by the gray shaded interval are assigned to an existing cluster, instead of forming a new one. (a)-(j) Behaviors of iCVIs (blue curves).

All iCVIs followed the tendencies described in Table 3 when correctly partitioning both real-world data sets, with the exception of iPBM and iGD_{43}. The former did not consistently follow either expected trend, whereas the latter only followed the immediate behavior trend. Regarding under-partition scenarios, the iPBM, iSIL, iWB, iDB, iGD_{43} and PS consistently followed the trends listed in Table 3, whereas the iXB, iGD_{53} and iConn_Index behaved with varying degrees of agreement; notably, the medium-term behavior (merging interval) of iXB was consistent with the findings for the synthetic data sets. The iCH was the only iCVI that did not behave as expected. Interestingly, the constant trend was observed for both data sets regarding the immediate behavior of iGD_{43} and iGD_{53}. Finally, for the over-partition experiment, the iCH and iGD_{53} followed their expected trends. The remaining iCVIs only partially exhibited the trends presented in Table 3. In particular, iSIL, iXB, iDB and iGD_{43} were only consistent with their medium-term behavior, whereas iWB, iConn_Index and PS were only consistent with their immediate behavior.

Interestingly, for the real-world data sets studied, over-partition was prominently detected by more iCVIs than under-partition. The latter issue was only patently flagged by PS. Regarding over-partitions, the most visually useful iCVIs were iXB, iDB, iGD_{43} and iConn_Index. Although an increasing trend was observed for the latter, as opposed to the synthetic data set findings, the behavior following a cluster split usually does not follow the familiar exponential curve; instead, a sharp drop generally follows the split, with a small improvement/recovery afterwards. This behavior suggests that there might be an issue with the clustering solution. In such cases, a further challenge lies in discriminating

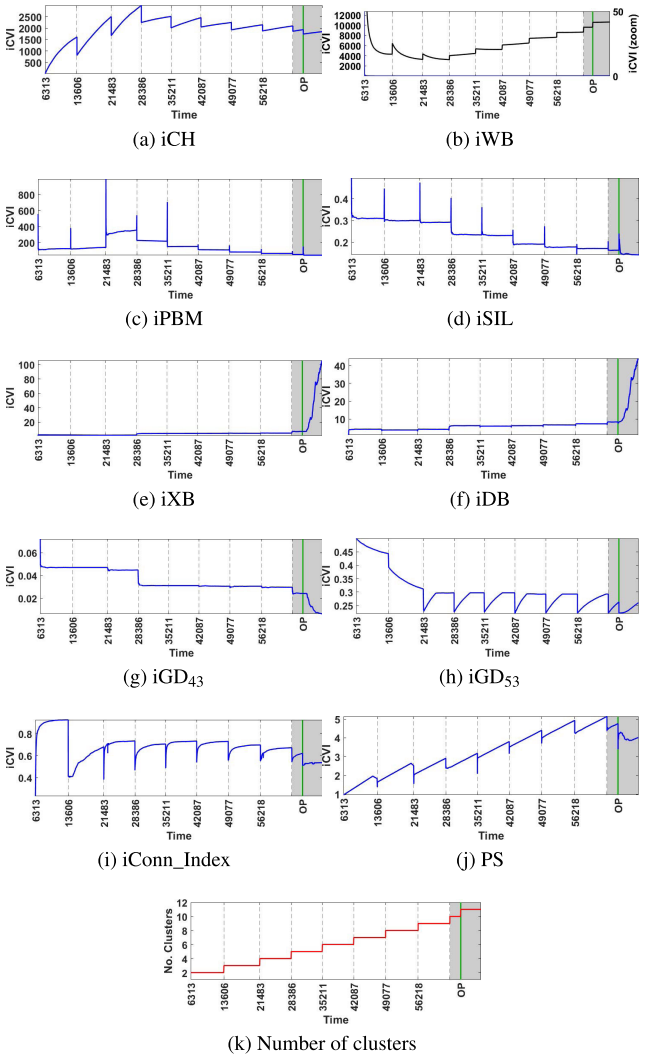


FIGURE 9. Each discrete time instant (x-axis) corresponds to the presentation of one sample of the data set *MNIST* during the over-partitioning experiment. The dashed vertical black lines delimit consecutive clusters (ground truth), *i.e.*, samples before a dashed line belong to one cluster, whereas samples after it belong to another. The continuous vertical green lines indicate the instant in which the over-partition (OP) problem starts in the cluster delimited by the gray shaded interval. (a)-(j) Behaviors of iCVIs (blue curves). (k) Number of clusters (step-like red curve).

between correct clusters that naturally do not follow an exponential curve during their evolution and/or determining a drop threshold that would constitute a problem. Similarly, note that iGD_{43} also decreases following the correct creation of some new clusters; therefore, discriminating among these two events may also be a challenge in some instances.

Finally, note that disagreements with the synthetic data set trends listed in Table 3 are to be expected, as those behaviors are typical but not unique. Such variance encourages the observation of several iCVIs to reliably make inferences about the quality of partitions of streaming data in real-time.

E. VISUAL INTERPRETATION POWER

In this section we examine a practical aspect of the iCVIs, namely their visual interpretation power, in terms of clear hints to problems occurring during the online clustering process; these include but not are limited to, substantial variations of their values (on a global scale) over time. Briefly, a useful iCVI behavior should make problems easier to spot. To accomplish this, we visually inspected the iCVIs' curves to gauge their usefulness in detecting the under- and over-partition issues that were artificially generated and intentionally inserted to the experiments described in the previous sections. The results of such visual inspection are summarized in Table 4.

For under-partition problems, the PS index consistently provided visually striking cues for both synthetic and real-world benchmark data sets. Moreover, it was the most robust CVI for increasing levels of cluster overlap (*S1* to *S4* data sets) and numbers of samples/clusters (*Birch1* and *Birch2*). The iGD_{43}, iGD_{53}, iXB, iDB, iSIL, iPBM, iCH and iWB (to a lesser extent) were also visually informative for the synthetic data sets, in which tendencies associated with this problem were clearly observable.

Regarding the over-partition problem, the iXB and iDB clearly flagged all over-partition issues. These were followed in their success by iGD_{43}, iSIL, iConn_Index and irCIP, which were also able to flag the majority of cases. As previously discussed, a potential challenge associated with iConn_Index consists of determining in which cases not following an exponential behavior during the evolution of sample assignments should signal a problem, and in which cases a cluster does not naturally follow such a function, as this characteristic is used to detect problems. Similarly, the caveat for iGD_{43} is related to determining whether its value reduction is associated with an over-partition problem or the correct emergence of a new cluster; in some correct partition instances, the creation of new clusters was followed by a reduction in this iCVI value.

VI. A CLOSER LOOK AT ICONN_INDEX

When evaluated over time, most iCVIs discussed in this study yield the same values as their batch counterparts (*e.g.*, the recursive formulation of compactness used in SS-based iCVIs is an exact computation, not an approximation [17], [18]). The iConn_Index is an exception, and thus is the subject of deeper analysis in this section. To obtain the batch Conn_Index values, all first-ranked and second-ranked fuzzy ART prototypes were recomputed after the presentation of each sample based on their activation function values.

For illustration purposes, Figs. 10 through 12 show the evolution of both the Conn_Index and iConn_Index for data set *R15* in all three scenarios described in Section V in one of the ten experiments. Moreover, they show a simple linear regression plot of Conn_Index and iConn_Index, as well as the final prototypes (hyperrectagles) and their connectivity visualization (CONNvis [69]). These figures show that iConn_Index smoothly follows the overall trends of its batch counterpart (with Pearson correlation coefficients [94] of 0.80, 0.74 and 0.94 for correct, under- and over-partition scenarios, respectively). The batch Conn_Index has a more jagged behavior

TABLE 4. The iCVIs that conveyed the most visually informative behavior across the experiments are indicated by a black dot symbol "·" for each data set. The dash symbol "-" indicates that the iCVI was not computed for the corresponding data set.

data set	Under-partition													Over-partition											
	iCH	iPBM	iSIL	iWB	iXB	iDB	iGD_{43}	iGD_{53}	PS	iNI	irCIP	irH	iConn_Index	iCH	iPBM	iSIL	iWB	iXB	iDB	iGD_{43}	iGD_{53}	PS	iNI	irCIP	irH	iConn_Index
Synthetic																										
A3	·	·	·	·	·	·	·	·	·							·		·	·	·				·		
Birch1							·	·	·							·		·	·	·				·		
Birch2									·		·							·	·	·				·		
Dim032	·	·	·	·	·	·	·	·	·							·		·	·	·						·
Dim064	·	·	·	·	·	·	·	·	·							·		·	·	·						·
Dim128	·	·	·	·	·	·	·	·	·		·					·		·	·	·				·		·
Dim256	·	·	·	·	·	·	·	·	·		·					·		·	·	·				·		·
Dim512	·	·	·	·	·	·	·	·	·	-	-	-				·		·	·	·			-	-	-	·
Dim1024	·	·	·	·	·	·	·	·	·	-	-	-				·		·	·	·			-	-	-	·
S1	·	·	·		·	·	·	·	·							·		·	·	·				·		·
S2	·		·		·	·	·	·	·							·		·	·	·				·		·
S3	·	·					·	·	·							·		·	·	·				·		
S4									·							·		·	·	·				·		
Unbalance	·	·	·		·	·			·	·						·		·	·	·				·		·
Aggregation			·		·	·				·						·		·	·					·		·
D31	·		·		·	·	·	·	·							·		·	·	·				·		·
R15	·	·	·	·	·	·	·	·	·	·						·		·	·	·				·		·
Hepta	·	·	·	·	·	·	·	·	·	·	·					·		·	·	·				·		·
Lsun	·	·	·		·	·	·			·	·					·		·	·	·			·	·		·
Tetra	·	·	·	·	·	·	·	·	·							·		·	·	·			·	·		·
Real World																										
Isolet									·	-	-	-						·	·				-	-	-	·
MNIST									·	-	-	-						·	·	·		·	-	-	-	·

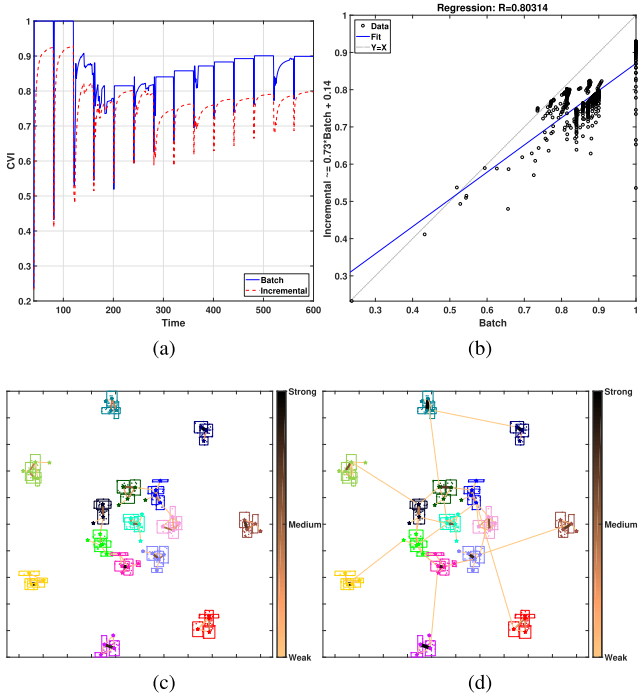


FIGURE 10. (a) Behaviors of Conn_Index (continuous blue line) and iConn_Index (dashed red line) when correctly partitioning the *R15* data set. (b) Regression plot between Conn_Index and iConn_Index in (a). The A-side categories of fuzzy ARTMAP and CONNvis (thicker and darker lines indicate stronger connections) generated with the (c) batch and (d) incremental *CONN* matrices.

and many plateaus. Additionally, note the faint but permanent connections between several different clusters; these are an artifact of the online learning process in that the second closest prototype of a sample that originated a new cluster always belongs to another existing cluster.

Table 5 reports the correlation coefficients and the mean square errors between the incremental and batch versions of the Conn_Index for all data sets averaged across the 10 experiments. For the majority of them, the average correlation between both Conn_Index versions is above (a) 0.75 (correct partitions), (b) 0.70 (under-partitions) and (c) 0.85 (over-partitions). Moreover, for most of the data sets, the average mean square error is below 0.02 in all scenarios. Some exceptions include the data sets *Aggregation*, *Lsun* and *D31* for the correct, under- and over-partition scenarios, respectively. These have smaller correlation coefficients. Therefore, the effect of quantization level within module A of fuzzy ARTMAP on the similarity of the batch and incremental implementations was investigated. This was accomplished by varying the vigilance parameter ρ_A of module A in the closed interval $[\rho_{min}, \rho_{max}]$, where ρ_{max} is the value listed in Table 5 for the respective data set, and $\rho_{min} = \min_i \left(1 - \frac{R_i}{d}\right)$ (i.e., ρ_{min} is computed based on the largest cluster of a given data set, see Eq. (82)), because the interval $[0, \rho_{min}]$ would yield identical results (the vigilance test would always be satisfied). Note that larger values of ρ_A produce finer granularity of clusters' prototypes.

The correlation coefficients and mean squared errors (averaged across 10 runs) depicted in Fig. 13 show that carefully tuning the vigilance parameter (granularity level) may improve the average correlation (from 0.5962 to 0.7977 when

TABLE 5. Vigilance parameter (ρ) values used in this study and the average correlation coefficient (R_{avg}) and mean square error (MSE_{avg}) between the incremental and batch Conn_Index curves.

Data set	Vigilance	Correct partition		Under-partition		Over-partition	
	ρ	R_{avg}	MSE_{avg}	R_{avg}	MSE_{avg}	R_{avg}	MSE_{avg}
A3	0.900	0.9511	0.0172	0.9681	0.0125	0.9559	0.0088
Birch1	0.920	0.9387	0.0020	0.8724	0.0026	0.8830	0.0028
Birch2	0.960	0.9869	0.0188	0.9890	0.0195	0.9885	0.0148
Dim032	0.930	0.7682	0.0040	0.7014	0.0068	0.9286	0.0045
Dim064	0.950	0.7654	0.0037	0.6996	0.0084	0.9299	0.0046
Dim128	0.960	0.7855	0.0028	0.7277	0.0056	0.9282	0.0040
Dim256	0.975	0.7838	0.0031	0.6999	0.0072	0.9296	0.0054
Dim512	0.980	0.7852	0.0028	0.7312	0.0044	0.9224	0.0046
Dim1024	0.988	0.7854	0.0030	0.7573	0.0071	0.9252	0.0044
S1	0.900	0.7379	0.0100	0.6884	0.0059	0.8592	0.0060
S2	0.900	0.7675	0.0409	0.7197	0.0402	0.8676	0.0249
S3	0.950	0.9166	0.0098	0.9107	0.0136	0.9083	0.0094
S4	0.950	0.8894	0.0185	0.8810	0.0101	0.8647	0.0153
Unbalance	0.880	0.6699	0.0033	0.7660	0.0034	0.9354	0.0078
Aggregation	0.750	0.5962	0.0114	0.6736	0.0214	0.9133	0.0077
D31	0.900	0.8448	0.0109	0.8977	0.0167	0.8337	0.0140
R15	0.950	0.7037	0.0179	0.7011	0.0154	0.8357	0.0197
Hepta	0.800	0.8089	0.0109	0.7821	0.0134	0.9267	0.0116
Lsun	0.900	0.7882	0.0550	0.5792	0.0350	0.9140	0.0328
Tetra	0.800	0.7459	0.0286	0.7101	0.0155	0.9307	0.0208
Isolet	0.500	0.8072	0.0167	0.7648	0.0167	0.8494	0.0117

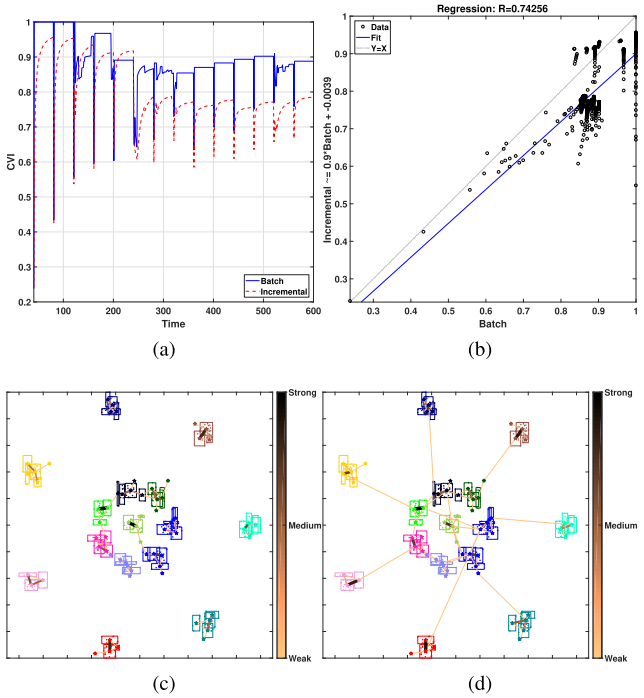


FIGURE 11. (a) Behaviors of Conn_Index (continuous blue line) and iConn_Index (dashed red line) when under-partitioning the *R15* data set. (b) Regression plot between Conn_Index and iConn_Index in (a). The A-side categories of fuzzy ARTMAP and CONNvis (thicker and darker lines indicate stronger connections) generated with the (c) batch and (d) incremental *CONN* matrices.

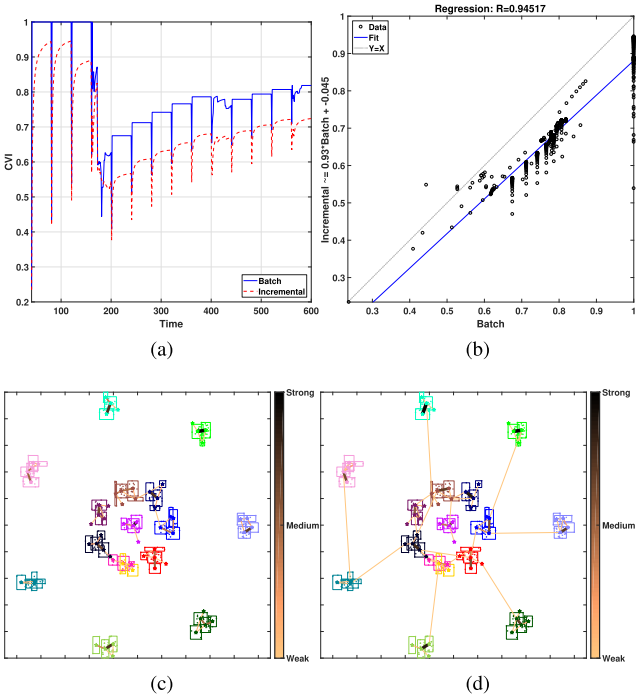


FIGURE 12. (a) Behaviors of Conn_Index (continuous blue line) and iConn_Index (dashed red line) when over-partitioning the *R15* data set. (b) Regression plot between Conn_Index and iConn_Index in (a). The A-side categories of fuzzy ARTMAP and CONNvis (thicker and darker lines indicate stronger connections) generated with the (c) batch and (d) incremental *CONN* matrices.

correctly partitioning the *Aggregation* data set, 0.5792 to 0.6810 when under-partitioning the *Lsun* data set, and 0.8337 to 0.9609 when over-partitioning the *D31* data set); however, its effect on this iCVIs' visual interpretation power when clustering data streams requires further investigation. All these results support the original assumption, stated in Section III-G, that both versions of the Conn_Index would behave similarly. Therefore, iConn_Index is suitable for

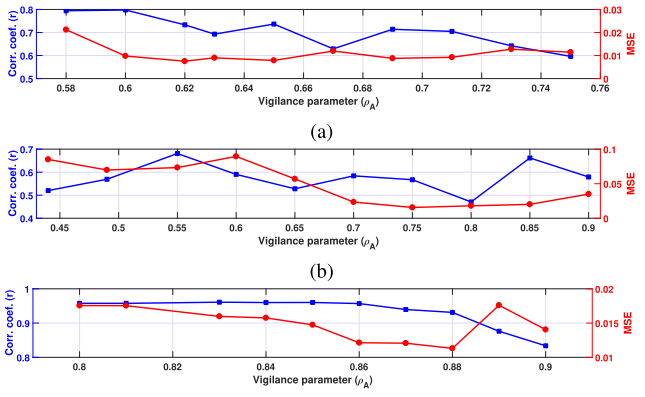


FIGURE 13. Correlation coefficients and MSEs between the batch and incremental versions of Conn_Index as a function of the vigilance parameter in module A of fuzzy ARTMAP. The values shown were averaged across 10 runs for (a) *Aggregation* data set under correct partition experiment, (b) *Lsun* data set under under-partition experiment, and (c) *D31* data set under over-partition experiment.

assessing the partitions generated by incremental clustering methods.

VII. CONCLUSION

This paper presented novel incremental versions of seven cluster validity indices (CVIs), namely, incremental Calinski-Harabasz (iCH), incremental Pakhira-Bandyopadhyay-Maulik (iPBM), incremental Silhouette (iSIL), incremental Negentropy Increment (iNI), incremental Representative Cross Information Potential (irCIP), incremental Cross Entropy (irH), and incremental Conn_Index (iConn_Index). These and previously developed incremental cluster validity indices (iCVIs) are essential tools that allow the practitioner to assess the quality of partitions in data streams. By definition, data streams require real-time processing of incoming samples because iterating over the entire data set is either prohibitive or unsuitable for the application.

Using an experimental framework made to be as agnostic as possible to any clustering algorithm as well as synthetic and real-world benchmark data sets, the dynamics of 13 iCVIs were analyzed in three different clustering scenarios: correct, under- and over-partitioning. Specifically, a thorough comparative study was performed among the presented iCVIs, the Partition Separation (PS), the incremental Xie-Beni (iXB), the incremental Davies-Bouldin (iDB) and the incremental generalized Dunn's indices 43 and 53 (iGD_{43} and iGD_{53}) in order to observe how these iCVIs are affected by the aforementioned problems and thus provide guidelines to aid the practitioner in identifying when these occur during online unsupervised learning. Additionally, it was shown that, although not equal to its batch counterpart, the iConn_Index follows the same general trends.

As expected from previous studies, most iCVIs underwent abrupt changes following the creation of a new cluster. However, when samples from an existing cluster were presented, each iCVI exhibited a particular behavior which was taken as a reference to compare the cases of under- and over-partitioning a data set. Most iCVIs detected under-partitioning of the synthetic data sets during the incremental clustering process, whereas only a subset of them provided insight to indicate over-partitioning problems. Interestingly, the opposite was observed for the real-world data sets. According to these findings, if the practitioner is expecting under-partitioning, the PS index can be particularly useful for detecting of this type of problem, as well as the following iCVIs: iCH, iPBM, iSIL, iWB, iXB, iDB, iGD_{43} and iGD_{53}. On the other hand, if over-partitioning issues are of concern, then we recommend iXB, iDB, iGD_{43}, iSIL, iConn_Index and irCIP. In any case, we corroborate the recommendations of previous studies regarding iCVIs: like their batch counterparts, it is good practice to observe a number of iCVIs at any given time rather than relying on the assessment of only one. It is expected that the observations from this study, as well as the iCVIs' MATLAB toolbox package provided, will assist in incremental clustering applications such as data streams.

ACKNOWLEDGMENT

The authors would like to thank J. M. Keller and his coauthors for providing an early copy of reference [20]. They also like to thank G. Olbricht for the helpful discussions on the statistical analysis performed in this study.

REFERENCES

[1] A. D. Gordon, "Cluster validation," in *Data Science, Classification, and Related Methods*, C. Hayashi, K. Yajima, H.-H. Bock, N. Ohsumi, Y. Tanaka, and Y. Baba, Eds. Tokyo, Japan: Springer, 1998, pp. 22–39.

[2] R. Xu, J. Xu, and D. C. Wunsch, "A comparison study of validity indices on swarm-intelligence-based clustering," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 4, pp. 1243–1256, Aug. 2012.

[3] L. E. Brito da Silva and D. C. Wunsch, "A study on exploiting VAT to mitigate ordering effects in Fuzzy ART," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 2351–2358.

[4] R. Dubes and A. K. Jain, "Validity studies in clustering methodologies," *Pattern Recognit.*, vol. 11, no. 4, pp. 235–254, Jan. 1979.

[5] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, Jun. 1985.

[6] J. C. Bezdek, W. Q. Li, Y. Attikiouzel, and M. Windham, "A geometric approach to cluster validity for normal mixtures," *Soft Comput.*, vol. 1, no. 4, pp. 166–179, Dec. 1997.

[7] E. Dimitriadou, S. Dolničar, and A. Weingessel, "An examination of indexes for determining the number of clusters in binary data sets," *Psychometrika*, vol. 67, no. 1, pp. 137–159, Mar. 2002.

[8] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Cluster validity methods: Part I," *SIGMOD Rec.*, vol. 31, no. 2, pp. 40–45, Jun. 2002.

[9] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Clustering validity checking methods: Part II," *SIGMOD Rec.*, vol. 31, no. 3, pp. 19–27, Sep. 2002.

[10] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets Syst.*, vol. 158, no. 19, pp. 2095–2117, Oct. 2007.

[11] L. Vendramin, R. J. G. B. Campello, and E. R. Hruschka, "Relative clustering validity criteria: A comparative overview," *Stat. Analy Data Mining*, vol. 3, no. 4, pp. 209–235, 2010.

[12] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Jan. 2010.

[13] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, Jan. 2013.

[14] R. Xu and D. C. Wunsch, II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[15] R. Xu and D. C. Wunsch, II, *Clustering*. Hoboken, NJ, USA: Wiley, 2009.
[16] J. Hämäläinen, S. Jauhiainen, and T. Kärkkäinen, "Comparison of internal clustering validation indices for prototype-based clustering," *Algorithms*, vol. 10, no. 3, p. 105, Sep. 2017.
[17] M. Moshtaghi, J. C. Bezdek, S. M. Erfani, C. Leckie, and J. Bailey, "Online cluster validity indices for streaming data," Jan. 2018, *arXiv:1801.02937v1*. [Online]. Available: https://arxiv.org/abs/1801.02937v1
[18] M. Moshtaghi, J. C. Bezdek, S. M. Erfani, C. Leckie, and J. Bailey, "Online cluster validity indices for performance monitoring of streaming data clustering," *Int. J. Intell. Syst.*, vol. 34, no. 4, pp. 541–563, Apr. 2019.
[19] O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Analysis of streaming clustering using an incremental validity index," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2018, pp. 1–8.
[20] O. A. Ibrahim, Y. Wang, and J. M. Keller, "Analysis of incremental cluster validity for big data applications," *Int. J. Unc. Fuzz. Knowl. Based Syst.*, vol. 26, no. 2, pp. 47–62, Dec. 2018.
[21] Q. Zhao and P. Fränti, "WB-index: A sum-of-squares based index for cluster validity," *Data Knowl. Eng.*, vol. 92, pp. 77–89, Jul. 2014.
[22] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI–1, no. 2, pp. 224–227, Apr. 1979.
[23] X. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
[24] J. Bezdek and N. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, no. 3, pp. 301–315, Jun. 1998.
[25] O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Evaluating evolving structure in streaming data with modified Dunn's indices," *IEEE Trans. Emerg. Topics Comput. Intell.*, to be published.
[26] O. A. Ibrahim, J. Shao, J. M. Keller, and M. Popescu, "A temporal analysis system for early detection of health changes," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2016, pp. 186–193.
[27] L. E. Brito da Silva, "Neuroengineering of clustering algorithms," Ph.D. dissertation, Missouri Univ. Sci. Technol., Rolla, MO, USA, 2019.
[28] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stats.-Simul. Comp. LSSP*, vol. 3, no. 1, pp. 1–27, 1974.
[29] Q. Zhao, M. Xu, and P. Fränti, "Sum-of-squares based cluster validity index and significance analysis," in *Adaptive and Natural Computing Algorithms*, M. Kolehmainen, P. Toivanen, and B. Beliczynski, Eds. Berlin, Germany: Springer, 2009, pp. 313–322.
[30] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, Mar. 2004.
[31] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
[32] L. F. Lago-Fernández and F. Corbacho, "Normality-based validation for crisp clustering," *Pattern Recognit.*, vol. 43, no. 3, pp. 782–795, Mar. 2010.
[33] L. F. Lago-Fernández and F. Corbacho, *Using Negentropy Increment to Determine Number Clusters*. Berlin, Germany: Springer, 2009, pp. 448–455.
[34] D. Araújo, A. D. Neto, and A. Martins, "Representative cross information potential clustering," *Pattern Recognit. Lett.*, vol. 34, no. 16, pp. 2181–2191, Dec. 2013.
[35] D. Araújo, A. D. Neto, and A. Martins, "Information-theoretic clustering: A representative and evolutionary approach," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4190–4205, Aug. 2013.
[36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, 2000.
[37] K. Taşdemir and E. Merényi, "A new cluster validity index for prototype based clustering algorithms based on inter- and intra-cluster density," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 2205–2211.
[38] K. Taşdemir and E. Merényi, "A validity index for prototype-based clustering of data sets with complex cluster structures," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 41, no. 4, pp. 1039–1053, Aug. 2011.
[39] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Netw.*, vol. 4, no. 6, pp. 759–771, Jan. 1991.
[40] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.
[41] L. Meng, A.-H. Tan, and D. C. Wunsch, II, *Adaptive Resonance Theory in Social Media Data Clustering: Roles, Methodologies, and Applications*. Cham, Switzerland: Springer, 2019.
[42] K. Al-Jabery, T. Obafemi-Ajayi, G. Olbricht, and D. C. Wunsch, II, *Computational Learning Approaches to Data Analytics in Biomedical Applications*. Amsterdam, The Netherlands: Elsevier, 2019.
[43] D. C. Wunsch, "ART properties of interest in engineering applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2009, pp. 3380–3383.
[44] L. E. Brito Da Silva, I. Elnabarawy, and D. C. Wunsch, "A survey of adaptive resonance theory neural network models for engineering applications," *Neural Netw.*, vol. 120, pp. 167–203, Dec. 2019.
[45] J.-C. Lamirel and P. Cuxac, "New Quality indexes for optimal clustering model identification with high dimensional data," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 855–862.
[46] J.-C. Lamirel, N. Dugue, and P. Cuxac, "New efficient clustering quality indexes," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3649–3657.
[47] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973.
[48] S. Bandyopadhyay and U. Maulik, "Nonparametric genetic clustering: Comparison of validity indices," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 31, no. 1, pp. 120–125, Feb. 2001.
[49] E. Hruschka, L. De Castro, and R. Campello, "Evolutionary algorithms for clustering gene-expression data," in *Proc. 4th IEEE Int. Conf. Data Mining (ICDM)*, Mar. 2005, pp. 403–406.
[50] E. R. Hruschka, R. J. Campello, and L. N. De Castro, "Evolving clusters in gene-expression data," *Inf. Sci.*, vol. 176, no. 13, pp. 1898–1927, Jul. 2006.
[51] M. Rawashdeh and A. Ralescu, "Center-wise intra-inter silhouettes," in *Scalable Uncertainty Management*, E. Hüllermeier, S. Link, T. Fober, and B. Seeger, Eds. Berlin, Germany: Springer, 2012, pp. 406–419.
[52] J. M. Luna-Romera, M. del Mar Martínez-Ballesteros, J. García-Gutiérrez, and J. C. Riquelme-Santos, "An approach to silhouette and Dunn clustering indices applied to big data in spark," in *Advances in Artificial Intelligence*, O. Luaces, J. A. Gámez, E. Barrenechea, A. Troncoso, M. Galar, H. Quintián, and E. Corchado, Eds. Cham, Switzerland: Springer, 2016, pp. 160–169.
[53] J. D. A. Silva and E. R. Hruschka, "A support system for clustering data streams with a variable number of clusters," *ACM Trans. Auton. Adapt. Syst.*, vol. 11, no. 2, pp. 1–26, Jul. 2016.
[54] M.-S. Yang and K.-L. Wu, "A new validity index for fuzzy clustering," in *Proc. 10th IEEE Int. Conf. Fuzzy Syst.*, vol. 1, Dec. 2001, pp. 89–92.
[55] E. Lughofer, "Extensions of vector quantization for incremental clustering," *Pattern Recognit.*, vol. 41, no. 3, pp. 995–1011, Mar. 2008.
[56] P. Comon, "Independent component analysis, A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
[57] E. Gokcay and J. Principe, "A new clustering evaluation function using Renyi's information potential," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 6, Jun. 2000, pp. 3490–3493.
[58] E. Gokcay and J. C. Principe, "Information theoretic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 158–171, Feb. 2002.
[59] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 1. Berkeley, CA, USA: Univ. of California Press, 1961, pp. 547–561.
[60] M. Cottrell and P. Rousset, "The kohonen algorithm: A powerful tool for analysing and representing multidimensional quantitative and qualitative data," in *Biological and Artificial Computation: From Neuroscience to Technology*, J. Mira, R. Moreno-Díaz, and J. Cabestany, Eds. Berlin, Germany: Springer, 1997, pp. 861–871.
[61] G. Karypis, E.-H. Han, and V. Kumar, "Chameleon: Hierarchical clustering using dynamic modeling," *Computer*, vol. 32, no. 8, pp. 68–75, Aug. 1999.
[62] E. W. Tyree and J. Long, "The use of linked line segments for cluster representation and data reduction," *Pattern Recognit. Lett.*, vol. 20, no. 1, pp. 21–29, Jan. 1999.
[63] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 586–600, May 2000.
[64] L. N. F. Ana and A. K. Jain, "Robust data clustering," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, p. 2.
[65] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*, 1st ed. New York, NY, USA: Springer-Verlag, 2010.
[66] A. Oliveira, A. D. Neto, and A. Martins, "An analysis of information dynamic behavior using autoregressive models," *Entropy*, vol. 19, no. 11, p. 612, Nov. 2017.

[67] A. G. Oliveira, A. Martins, and A. D. Neto, "Information state: A representation for dynamic processes using information theory," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
[68] K. Taşdemir and E. Merényi, "Data topology visualization for the Self-Organizing Maps," in *Proc. 14th Eur. Symp. Artif. Neural Netw. (ESANN)*, Apr. 2006, pp. 277–282.
[69] K. Taşdemir and E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 549–562, Apr. 2009.
[70] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comput. Vis., Graph., Image Process.*, vol. 36, nos. 2–3, p. 396, Nov. 1986.
[71] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
[72] T. Kasuba, "Simplified fuzzy ARTMAP," *AI Expert*, vol. 8, no. 11, pp. 18–25, Nov. 1993.
[73] J. R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Netw.*, vol. 9, no. 5, pp. 881–897, Jul. 1996.
[74] B. Vigdor and B. Lerner, "The Bayesian ARTMAP," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1628–1644, Nov. 2007.
[75] L. E. Brito da Silva and D. C. Wunsch, "An information-theoretic-cluster visualization for self-organizing maps," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2595–2613, Jun. 2018.
[76] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," *Appl. Intell.*, vol. 48, no. 12, pp. 4743–4759, Dec. 2018.
[77] I. Kärkkäinen and P. Fränti, "Dynamic local search algorithm for the clustering problem," Dept. Comput. Sci., Univ. Joensuu, Joensuu, Finland, Tech. Rep. A-2002-6, 2002.
[78] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: A new data clustering algorithm and its applications," *Data Mining Knowl. Discovery*, vol. 1, no. 2, pp. 141–182, Jun. 1997.
[79] P. Franti, O. Virmajoki, and V. Hautamaki, "Fast agglomerative clustering using a k-nearest neighbor graph," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1875–1881, Nov. 2006.
[80] P. Fränti and O. Virmajoki, "Iterative shrinking method for clustering problems," *Pattern Recognit.*, vol. 39, no. 5, pp. 761–775, May 2006.
[81] M. Rezaei and P. Franti, "Set matching measures for external cluster validity," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2173–2186, Aug. 2016.
[82] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, pp. 1–30, Mar. 2007.
[83] C. Veenman, M. Reinders, and E. Backer, "A maximum variance cluster algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1273–1280, Sep. 2002.
[84] A. Ultsch, "Clustering with SOM: U*C," in *Proc. Workshop Self-Organizing Maps (WSOM)*, 2005, pp. 75–82.
[85] D. Dua and C. Graff. *UCI Machine Learning Repository*. Accessed: 2017. [Online]. Available: http://archive.ics.uci.edu/ml
[86] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[87] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *VLDB J.*, vol. 20, no. 1, pp. 21–33, Feb. 2011.
[88] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Mar. 1998.
[89] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[90] B. Calvo and G. Santafé, "Scmamp: Statistical comparison of multiple algorithms in multiple problems," *R J.*, vol. 8, no. 1, p. 248, Feb. 2019.
[91] M. Kutner, C. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. New York, NY, USA: McGraw-Hill, 2004.
[92] B. Trawiński, M. Smętek, Z. Telec, and T. Lasota, "Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms," *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 4, pp. 867–881, Dec. 2012.
[93] M. Tscherepanow, *TopoART: A Topology Learning Hierarchical ART Network*. Berlin, Germany: Springer, 2010, pp. 157–167.
[94] L. J. Bain and M. Engelhardt, *Introduction to Probability and Mathematical Statistics*, 2nd ed. Pacific Grove, CA, USA: Brooks/Cole, 1992.

LEONARDO ENZO BRITO DA SILVA (Member, IEEE) received B.S. degree in electrical engineering and the M.S. degree in electrical and computer engineering from the Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil, in 2011 and 2013, respectively, and the Ph.D. degree in computer engineering from the Missouri University of Science and Technology, Rolla, MO, USA, in 2019. His current research interests include clustering algorithms, cluster validation, data visualization, adaptive resonance theory, self-organizing maps, information theory, and evolutionary computation.

NIKLAS MAX MELTON (Member, IEEE) received the B.S. degree in aerospace engineering from the Missouri University of Science and Technology, in 2017, where he is currently pursuing the Ph.D. degree in computer science. He is a GAANN Fellow of the Missouri University of Science and Technology. His current research interests include clustering algorithms, adaptive resonance theory, adaptive dynamic programming methods, deep learning, and computer vision.

DONALD C. WUNSCH II (Fellow, IEEE) received the B.S. degree in applied mathematics from the University of New Mexico, Albuquerque, NM, USA, the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, University of Washington, Seattle, and the Executive MBA from Washington University in St. Louis. He is currently the Mary K. Finley Missouri Distinguished Professor with the Missouri University of Science and Technology (Missouri S&T), Rolla, Missouri, USA. He is the Director of the Applied Computational Intelligence Laboratory, Multidisciplinary Research Group. Earlier employers were Texas Tech University, Lubbock, TX, USA, Boeing, Seattle, WA, USA, Rockwell International, Albuquerque, NM, USA, and the International Laser Systems, Albuquerque, NM, USA. He did his Jesuit Core Honors Program in Seattle University. His key research contributions are in Unsupervised and Reinforcement Learning and their applications. He is a previous International Neural Networks Society (INNS) President, INNS Fellow, NSF CAREER Awardee, the 2015 INNS Gabor Award Recipient, and the 2019 Ada Lovelace Service Award Recipient. He served as the IJCNN General Chair, and on several Boards, including the St. Patrick's School Board, the IEEE Neural Networks Council, INNS, and the University of Missouri Bioinformatics Consortium, Chaired the Missouri S&T Information Technology and the Computing Committee as well as the Student Design and Experiential Learning Center Board. He has produced 22 Ph.D. recipients in computer engineering, electrical engineering, systems engineering and computer science.