TopoBARTMAP: Biclustering ARTMAP with or without Topological Methods in a Blood Cancer Case Study

Raghu Yelugam*, Leonardo Enzo Brito da Silva*†, and Donald C. Wunsch II*
*Applied Computational Intelligence Laboratory, Missouri University of Science and Technology, Rolla, MO, USA
†CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil
Email: ry222@umsystem.edu, {leonardoenzo, wunsch}@ieee.org

***Abstract*—Biclustering is a special case of subspace clustering that has become viable in several domains. Particularly, in genomic data analysis, biclustering has been used to identify conditions under which a subset of genes are highly co-expressed, while topological data analysis has been used to analyze disease-specific subgroups, evolution, and disease progression. In this work, we combine biclustering with topological data analysis to achieve the best of both methods. We present TopoBARTMAP - produced by hybridizing BARTMAP, an adaptive resonance theory (ART)-based biclustering method, with TopoART, a topology learning ART network - in order to identify topological associations between biclusters. TopoBARTMAP outperformed both TopoART and BARTMAP in the experimental analysis on six benchmark blood cancer data sets. In some cases, BARTMAP may nevertheless be preferred due to implementation simplicity.**

I. Introduction

Biclustering, at it's crux, is the problem of finding a subset of samples with high association across a subset of features. It has produced significant results in many different fields, starting with the application to gene expression data analysis [1]. Over the past decade, several algorithms were proposed under co-clustering [2], biclustering [3], [4], and subspace clustering [5]. Biclustering applied to gene expression data is used to identify subsets of conditions under which subsets of genes are highly co-expressed. Such identification is essential for gaining insights into regulatory networks [6] and gene-disease associations. However, biclustering can fail to completely uncover gene regulatory networks [7] due to the inability to identify functional associations between genes within a bicluster and to group functionally related genes that might not co-express significantly.

In an attempt to mitigate this problem, topological data analysis is becoming popular for gene expression data [8], [9]. In general, topological analysis extends clustering to identify local relationships in the data. Depending on the proximity of clusters, the identified clusters represented by cluster points are connected to generate a high-level graphical representation of the underlying shape of the space. Since topological methods discover the geometric structure within the data, these methods are sensitive to large and small scale patterns, invariant to noise, independent of the coordinate system, and can produce a compressed representation of data [8], [10], [11]. In practice, topological data analysis methods, such as Mapper [12], had proven to be effective in identifying subgroups of cancers [13], understanding genome dynamics [14] and cellular differentiation [15], and disease progression [16], which would otherwise be difficult to detect using traditional clustering methods applied to biomedical research [17].

In this work, we take a step towards combining topological data analysis with biclustering. To do so, we combine BARTMAP [18] with TopoART [19] to produce TopoBARTMAP, which can simultaneously identify the structure within the data and the biclusters. This algorithm, while identifying inter-bicluster relationships, is resilient to noise due to the prototype pruning mechinism of TopoART. While the literature contains a few topological biclustering algorithms, such as BiTM [20], wBiTM [21], and SKB [7], to the best of our knowledge TopoBARTMAP is the first adaptive resonance theory (ART) [22] based topological biclustering algorithm. Amongst the other methods, the BiTM (Biclustering using Topological Maps) identifies biclusters by simultaneously clustering rows and columns by identifying a discrete topology. The biclustering method wBiTM, based on BiTM, is a feature group weighting method that uses topological maps for biclustering. Both BiTM and wBiTM produce topological maps that enable data visualisation. SKB, Skeleton Biclustering, is capable of identifying and mining missing genes in the biclusters while building inter-bicluster and intra-bicluster (relationships among genes within a bicluster) relationship skeletons. This method uses hierarchical biclustering and gene ontology annotations.

The remainder of this paper is organized as follows: we start with a brief review of related ART models [22] in Section II, which is necessary to contextualize the main contribution of this paper introduced in Section III, which is the TopoBARTMAP model. The details concerning the experimental design are presented in Section IV, while Section V reports and discusses the biclustering results on six benchmark blood cancer data sets. Although in our experimentation we compare TopoBARTMAP only with one biclustering algorithm, BARTMAP, in other papers [18], [23], [24] BARTMAP was already compared with other ART-based and non-ART-based methods and found to be superior. Finally, Section VI summarizes the findings of this paper.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

II. Background

This section presents an overview of Fuzzy ART [25], TopoART [19] and BARTMAP [18]. We refer the reader to [26]–[30] for a comprehensive treatment of ART models and their applications.

A. Fuzzy ART

This section exists to make the paper self-contained. The reader familiar with Fuzzy ART can skip to Section II.B. Readers new to ART will gain useful context from this section but will need the cited papers to understand ART well.

Several winner-take-all neural models have been proposed to perform supervised and unsupervised learning based on ART. Of them, Fuzzy ART, described in [26], extends the binary ART-1 by using fuzzy operations and is capable of recognising stable codes (prototypes) in response to real or binary-valued inputs. A typical ART-based learning model consists of two interacting fields: F1 and F2. Field F1 receives input as a vector X which is normalized and complement coded, i.e., if $x = [x_1, x_2, ..., x_d]$ denotes the normalized sensory input vector with d features, then $X = [x, 1 - x^c]$, where $x_i^c = 1 - x_i$. It is customary to consider each dimension of the input X as a node. Field F2 consists of several prototypes (neurons) - one per node - which are used to categorize the input pattern. The field F1 generates bottom-up sensory information which produces activity across each prototype in field F2 using bottom-up memory traces. The highly active prototype produces a top-down expectation signal based on the top-down memory traces. At the orientation subsystem, this expectation signal is compared with the sensory information produced by field F1 for similarity; this is known as vigilance criterion. If the similarity is greater than a certain threshold called the vigilance value, resonance ensues between the highly active prototype and field F1; the corresponding memory traces are updated. In any other case, the highly active prototype is shut down, and a search is done to find the next active prototype that produces a matching expectation. If no such prototype is found, a new node is created to register the current input, and this prototype is said to be committed.

Bottom-up memory traces can be represented using top-down memory traces, which essentially is a key feature of Fuzzy ART. In this work, we represent the bottom-up memory traces as a vector for the j^{th} node in field F2 as weights $W_j = [W_1^j, W_2^j, ..., W_{2d}^j]$. Initially, field F2 consists of uncommitted prototypes. Instead of using two different memory traces, Fuzzy ART has traces or weights (W) connecting F1 to F2 and F2 to F1, i.e., the prototypes are directly compared with the input signal. Because the input, X, to Fuzzy ART is normalized and complement coded, which is implicitly assumed in the models derived from Fuzzy ART, all the uncommitted node weights are initially set to a value of 1. During the input presentation, field F2 has an uncommitted neuron along with committed neurons. A measure of activity is computed using the category choice function defined by

$$T_j = \frac{|X \wedge W_j|}{\alpha + |W_j|}, \tag{1}$$

where $|\cdot|$ is the ℓ_1 norm, $\wedge$ is the fuzzy AND operation defined by $(X \wedge W_j)_i = min(x_i, W_i^j)$, $X \in \Re^{2d}, W \in \Re^{2d}$, and $\alpha > 0$ helps in breaking the tie when more than one prototype is a fuzzy subset of the input pattern. A winning neuron J is selected using winner-take-all rule. A category match is measured using a match function defined by

$$M_J = \frac{|X \wedge W_J|}{|X|}. \tag{2}$$

If M_J is greater than the vigilance parameter, ρ, it meets the vigilance criterion. In case of failure to satisfy, this winning neuron is shut off and the next winning neuron is tested against the vigilance criterion. The search process continues until the vigilance criterion is satisfied. When a neuron is found that satisfies the vigilance criterion, its weight vector or prototype is updated using the learning law defined as

$$W_J = \beta(W_J \wedge X) + (1 - \beta)W_J, \tag{3}$$

where $0 < \beta \leq 1$ is the learning rate. The output of the network y^{F2} is set for each i^{th} node as

$$y_i^{F2} = \begin{cases} 0, & \text{if } i \neq J \\ 1, & \text{if } i = J \end{cases}. \tag{4}$$

In the case of fast learning, when an uncommitted prototype is selected, β is set to a value of 1. Each prototype learned this way summarizes all the input patterns it is associated with and thus represents either a cluster or cluster point. Due to complement coding, it is possible to estimate the size of the cluster and region of space that each prototype summarizes [31].

B. TopoART

Topological ART, or TopoART [19], is a Fuzzy ART-based hierarchical clustering model that learns topologies present within the input data. All rules used for prototype learning of Fuzzy ART remain the same in TopoART. However, TopoART uses a counter, n_i, for each i^{th} prototype that records the number of samples it summarized. After τ input presentations (time steps), the prototypes with a counter less than the threshold ϕ, called candidate prototypes, are pruned from field F2. Hence, the prototypes with $n_i \geq \phi$ are called permanent prototypes. This pruning mechanism renders TopoART robust to noise. Further, along with the best matching prototype J_1, a second highly active prototype J_2 satisfying the vigilance criterion is selected (namely, the second-best). This prototype's weights are updated using the same update rule defined by Eq. (3), but with a smaller learning rate, i.e.,

$$\beta_{J_2} < \beta_{J_1}. \tag{5}$$

This procedure of identifying and updating the second-best prototype's weights increases noise robustness by making growth of prototypes in pertinent areas of the input space more likely [19].

To learn topologies, an edge is established between prototypes J_1 and J_2 (if J_2 can be found), creating a topological structure with prototypes acting as nodes. While edges formed

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

with candidate prototypes are removed when these prototypes are pruned, edges between permanent prototypes remain stable. Each one of the permanent prototypes is assigned with a label depending on its topological association with others, i.e, all the connected permanent prototypes are assigned with the same label. These labels are recorded in a vector - $\overrightarrow{C} =< l_i : l_i \in \{1,2,...,k\} >$, where k is the number of topologies identified and l_i is the label for the i^{th} permanent prototype in field F2. With the aid of this vector, one can identify the topology to which an input datum belongs. Unlike Fuzzy ART, the clusters are represented by these topologies rather than the prototypes, which enables the identification of arbitrarily shaped clusters. The input patterns that correspond to the deleted prototypes are reclassified by associating them to the permanent prototypes that yield the strongest responses. This activity is computed using the following rule [19]:

$$T_j = 1 - \frac{|(X \wedge W_j) - W_j|}{|X|}. \tag{6}$$

Further, TopoART achieves hierarchical clustering by cascading multiple Fuzzy ART modules, with each having an increasing vigilance parameter value. The vigilance parameters across the modules in the hierarchy are computed using the following rule,

$$\rho_{next} = \frac{1}{2}(1 + \rho_{previous}), \tag{7}$$

where ρ_{next} and $\rho_{previous}$ are the vigilance values for the next and previous modules respectively. The module at the lowest hierarchical level receives all the input patterns, whereas subsequent modules receive an input pattern if and only if that pattern activates a permanent prototype in the previous module. When new data is presented, inference is made using Eq. (6) and the output y^{F2} for each module of TopoART is computed with Eq. (4) using only permanent prototypes.

C. BARTMAP

Biclustering ARTMAP, or BARTMAP [18], is an ART-based two-way clustering method that identifies biclusters present in the data adaptively without the explicit requirement of prior knowledge of the number of biclusters. We refer the reader to [32] and [33] for detailed expositions on ARTMAP and Fuzzy ARTMAP as BARTMAP is inspired by the theory of Fuzzy ARTMAP [33]. Here, we present a brief treatment of BARTMAP while retaining pertinent details. In doing so, we use the following notation. We represent the gene expression data matrix as $G = (F, O)$, where $F = \{f_1, ..., f_N\}$ represents the set of N features (or rows) and $O = \{o_1, ..., o_M\}$ represents the set of M observations (or columns). An element $g_{ij} \in G$ represents the intensity of feature i in observation j. In using this notation, which we did for ease of generalization, we considered each gene as a feature and each sample as an observation throughout the exposition.

In construction, BARTMAP consists of two Fuzzy ART modules, ARTa and ARTb, connected via an inter-ART module. While the ARTb module receives features as input, the ARTa module receives observations as input. In essence, BARTMAP learns a mapping between the clusters found by the ARTa and ARTb modules across observations and features, respectively, such that local relationships are preserved and captured by the produced biclusters. The learning in BARTMAP progresses in two steps. The first step is to identify clusters across one of the dimensions, let us say the row dimension, of the matrix. This is achieved by using the ARTb module as a standard Fuzzy ART, thus resulting in k_f feature prototypes or clusters $\{F_i | i = 1, ..., k_f\}$. In the next step, observations (or columns) are presented to the ARTa module. At this stage, when a new observation o_k is presented, the potential associable cluster is identified using winner-take-all. Eventhough the winner qualifies via the vigilance test, at this point, it might not represent the observation unless the winner corresponds to an uncommitted prototype. If the winner corresponds to an uncommitted prototype, its weights are updated. However, if the winner is a committed prototype, then the current sample o_k is associated to it if and only if the sample exhibits a similar trend across at least one feature cluster F_i, as do the other member observations that are associated to the winning prototype. This similarity is verified using the average Pearson correlation coefficient that is computed as

$$\eta(o_k, O_J, F_i) = \frac{1}{M_J} \sum_{l=1}^{M_J} r(o_k, o_l, F_i), \tag{8}$$

where,

$$r(o_k, o_l, F_i) = \frac{\sum_{t=1}^{N_i} (g_{o_k f_{it}} - \overline{g}_{o_k F_i})(g_{o_l f_{it}} - \overline{g}_{o_l F_i})}{\sqrt{\sum_{t=1}^{N_i} (g_{o_k f_{it}} - \overline{g}_{o_k F_i})^2} \sqrt{\sum_{t=1}^{N_i} (g_{o_l f_{it}} - \overline{g}_{o_l F_i})^2}}, \tag{9}$$

$$\overline{g}_{o_k F_i} = \frac{1}{N_i} \sum_{t=1}^{N_i} g_{o_k f_{it}}, \tag{10}$$

$$\overline{g}_{o_l F_i} = \frac{1}{N_i} \sum_{t=1}^{N_i} g_{o_l f_{it}}, \tag{11}$$

and o_k is a new observation, o_l belongs to observation cluster $O_J = \{o_1, \ldots, o_{M_J}\}$ with M_J observations, and the set $F_i = \{f_{i1}, \ldots, f_{iN_i}\}$ represents the i^{th} feature cluster with N_i features. The observation o_k becomes a member of cluster O_J if and only if $\eta(o_k, O_J, F_i)$ is greater than a threshold η_{th} for at least one of the feature clusters. Cluster O_J is represented by ARTa's winning prototype J; therefore, if the previously mentioned constraint is satisfied, then its weights are updated using Eq. (3). If the test fails, match tracking ensues [32], i.e, the winning prototype is inhibited by increasing the vigilance value ρ_a of ARTa sufficiently above its base value. With the inhibition, a search for the next winning prototype ensues. Once a suitable cluster is identified or a new cluster is created, the vigilance value ρ_a of ARTa is reset to its base value.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

III. TopoBARTMAP

This section discusses the architectural novelty introduced in this paper. To understand the topological associations between the features (genes) and observations of a given gene expression matrix, $G = (F, O)$, BARTMAP is modified to produce topological biclusters. This is achieved by replacing both Fuzzy ART modules of BARTMAP with TopoART modules. The resulting biclustering method, TopoBARTMAP, is more robust to noise and can identify associations better than BARTMAP. While the construction and much of the training procedure remain the same, i.e., the training proceeds in two steps with a similarity/association test using average Pearson correlation, there are subtle differences rendered by the nuances in the training of TopoART.

Just as in the training of BARTMAP, one of the dimensions of matrix G, features F for example, are presented to TopoARTb during the first stage, resulting in identifying k_f prototypes $\{F_i | i = 1, ..., k_f\}$ and k topologies or clusters formed by these prototypes. During the second stage of training, the observations are presented to TopoARTa. Similar to BARTMAP, a winning prototype is identified and verified for a match using the vigilance test (Eq. (2)) and associability with the similarity test (Eqs. (8) to (11)). If the presented observation exhibits the same trend as the other observations summarized by the winner J_1 across at least one of TopoARTb's k_f prototypes, it is associated with the winner. Otherwise, match tracking ensues and the search proceeds until either another representative prototype is found or an uncommitted prototype is assigned. Further, a second-best prototype J_2 is identified using winner-take-all and verified for the match. An edge is established if and only if the presented observation exhibits a trend similar to the other observations represented by the second winner across at least one of TopoARTb's k_f prototypes. If the similarity test fails, then the current second winner is inhibited and the next winner is searched for. This procedure terminates once a second winner passing the vigilance and similarity tests is found, or all the prototypes are inhibited. If a qualifying second-best prototype exists, then its weights are updated using Eq. (3). The pruning of candidate prototypes and reclassification of the observations represented by the pruned prototypes are conducted without any changes to TopoART. The learning results in the identification of topologies, which represent biclusters, enabling TopoBARTMAP to identify arbitrarily shaped biclusters. Similar to TopoART, the permanent prototypes are used for making an inference when new data are presented. The TopoBARTMAP architecture is depicted in Fig. 1, and its working procedure is summarized by Algorithm 1.

IV. Experimentation

To verify and compare the performance of TopoBARTMAP, we used six benchmark blood cancer data sets brought together by [34]. The corresponding information about the data sets, number of features, number of observations, and number of classes per data set is available in Table I. Of these data sets, Alizadeh-V1 and Alizadeh-V2 [35] have a

Algorithm 1: TopoBARTMAP

Input : G, {TopoART a and b parameters}, η_{th}, δ
Output : $y^{(O)}$ observation classes

/* Notation */
O_j : j^{th} cluster of observations.
F_i : i^{th} cluster of features.
$\eta(o_k, O_j, F_i)$: correlation between o_k and O_j across F_i.
/* Training */

for $f \in F$ **do** // TopoARTb
  Train TopoARTb using $< \rho_b, \phi_b, \tau_b, \alpha_b, \beta_{J_1}, \beta_{J_2} >$
end
Initialize time step counter t_a to 0
for $o_k \in O$ **do** // TopoARTa
  Increment t_a
  Compute $T_j \ \forall j$ using Eq. (1)
  Find first winner: $J_1 = \arg\max_j (T_j)$
  Compute M_{J_1} using Eq. (2)
  if $M_{J_1} \geq \rho_a$ **then**
    Compute $\eta(o_k, O_{J_1}, F_i) \ \forall \ F_i$ using Eq. (8)
    if $\exists F_i | \eta(o_k, O_{J_1}, F_i) \geq \eta_{th}$ **then**
      Update W_{J_1} using Eq. (3), $n_{J_1} = n_{J_1} + 1$
      reset ρ_a
      Find second winner: $J_2 = \arg\max_{j, j \neq J_1} (T_j)$
      if $(M_{J_2} \geq \rho_a) \wedge \exists F_i [\eta(o_k, O_{J_2}, F_i) \geq \eta_{th}]$ **then**
        Update W_{J_2} using Eq. (3)
        Establish a link between J_1 and J_2
      else if $\forall F_i \forall J_2 (\eta(o_k, O_{J_2}, F_i) < \eta_{th})$ **then**
        Stop the search
      else
        Shut down J_2
        Go to step 15
      end
  else if $(\neg\exists(M_{J_1} \geq \rho_a)) \parallel (\rho_a \geq 1)$ **then**
    $W_{new} = o_k$, $n_{new} = 1$
    reset ρ_a
  else
    Shut down J_1, $\rho_a = \rho_a + \delta$
    Go to step 8
  end
  if $t_a \bmod \tau_a = 0$ **then**
    Remove clusters with counter less than ϕ_a
  end
end
Identify and label the topologies using the connected prototypes
/* Inference */
for $o_l \in O$ **do**
  Compute $T_j \ \forall j$ using Eq. (6)
  $y_l = \arg\max_j (T_j)$ and Cluster $= l_{y_l}$
end

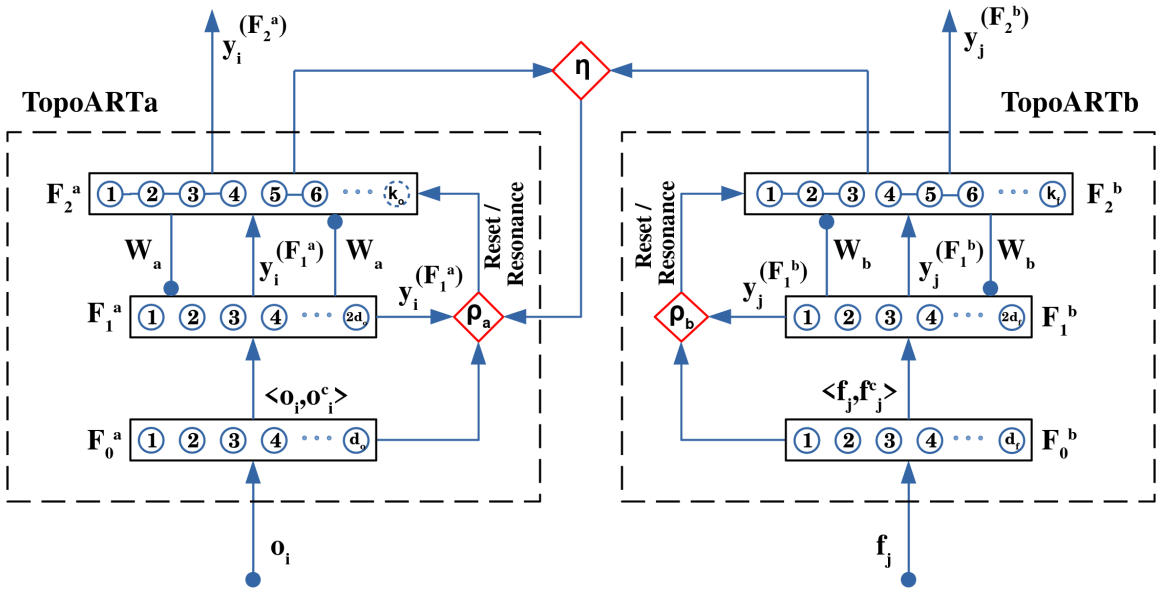

Fig. 1. Block diagram of TopoBARTMAP. TopoBARTMAP consists of two TopoART modules whose clusters are related by correlation test module. While TopoARTb is shown to have learned k_f prototypes, TopoARTa is shown to have learned k_o prototypes (note that prototype labeled k_o is a temporary prototype).

TABLE I
DATA SETS USED FOR EXPERIMENTAL ANALYSIS [4]. HERE, SL.NO, N_o, N_f, AND N_c REFER TO SERIAL NUMBER, NUMBER OF OBSERVATIONS, FEATURES, AND CLASSES WITHIN OBSERVATIONS, RESPECTIVELY.

Sl.no	Data set	N_o	N_f	N_c
1	Alizadeh-V1 [35]	42	1095	2
2	Alizadeh-V2 [35]	62	2093	3
3	Alizadeh-V3 [35]	62	2093	4
4	Armstrong-V1 [37]	72	1081	2
5	Armstrong-V2 [37]	72	2194	3
6	Shipp-V1 [38]	77	798	2

The data sets are available at https://github.com/padilha/biclustlib.

randomized order of observations, while the rest of the data sets have observations ordered according to the classes. We used MATLAB (Statistics and Machine Learning Toolbox) and the Cluster Validity Analysis Platform toolbox [36] to conduct all the experiments. To enhance the reproducibility of this research, the MATLAB source code is available at the Applied Computational Intelligence Laboratory GitHub repository[1].

To investigate the performance of TopoBARTMAP on each of these data sets, we ran tests with a Genetic Algorithm (GA) [39] to optimize parameters $< \rho_a,\ \beta^a_{J_2},\ \phi_a,\ \tau^{\%}_a,\ \rho_b,\ \beta^b_{J_2},\ \phi_b,\ \tau^{\%}_b,\ \eta >$ with the Adjusted Rand Index [40] as the performance measure, where $\tau^{\%}_{-}$ represents the percentage of the total number of samples presented to each module. To measure the biclustering performance, we considered topologies found by TopoARTa as biclusters. Although topologies can be retrieved from the TopoARTb module, for the correlation test, individual prototypes rather than topologies were considered. Further, the vectors $< 0.0, 0.0, 0.0, 0.1, 0.0, 0.0, 0.0, 0.1, 0.0 >$ and $< 0.95, 1.0, 5.0, 0.3, 0.95, 1.0, 5.0, 0.3, 0.99 >$ were used as lower and upper bounds for parameters with the aforementioned order during optimization, along with the following constraints:

$$\phi_a - \tau^{\%}_a N_o \geq 0, \tag{12}$$

and

$$\phi_b - \tau^{\%}_b N_f \geq 0, \tag{13}$$

where $N_o = \#\{observations\}$ and $N_f = \#\{features\}$. Note that $\tau_a = \lfloor \tau^{\%}_a N_o \rceil$ and $\tau_b = \lfloor \tau^{\%}_b N_f \rceil$. The GA was run 10 times for 25 generations with a population size of 200 agents. Throughout the experiments, β_a and β_b were set to 1, i.e., fast learning is used with the value of choice parameter α set to 0.001, match tracking step size set to 0.01, and data presented for one epoch.

BARTMAP, TopoART, and Fuzzy ART were run on the same data sets in order to compare their performance with TopoBARTMAP's. For BARTMAP, $< \rho_a,\ \rho_b,\ \eta >$ were optimized using GA, with the lower and upper bounds as given by vectors $< 0.0, 0.0, 0.0 >$ and $< 0.95, 0.95, 0.99 >$, respectively. For TopoART, $< \rho,\ \beta_{J_2},\ \phi,\ \tau^{\%} >$ were optimized using GA, with the lower and upper bounds for these parameters given by vectors $< 0.0, 0.0, 0.0, 0.1 >$ and $< 0.95, 1.0, 5.0, 0.3 >$, respectively, as well as the inequality constraint

$$\phi - \tau^{\%} N_o \geq 0, \tag{14}$$

where $\tau^{\%}$ and τ are defined as in TopoBARTMAP. Note that unlike the standard TopoART network, which uses two modules for hierarchical purposes, here only one module was trained, and performance was measured using topologies identified by TopoART. While these other methods were run with GAs similar to TopoBARTMAP, Fuzzy ART was run with a parameter sweep. In particular, for running Fuzzy ART

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

on each of the six data sets, ρ was varied amongst 4960 values uniformly sampled from $[0,1]$, which corresponds to roughly the same amount of fitness evaluations of the GAs used to optimize the other ART networks. For all of these methods, fast learning was used, with the value of choice parameter α set to 0.001, and data were presented for one epoch. For BARTMAP, the match tracking step size was set to 0.01, similar to TopoBARTMAP.

All agglomerative clustering algorithms (which includes online methods), such as most unsupervised ART networks, are prone to ordering effects (see [41]–[43] and the references therein for further discussion and methods to mitigate order dependency). Therefore, once the optimal parameters that produce the best Adjusted Rand Index on each of the six data sets were determined, tests were conducted to verify the hyper-parameter sensitivity of each method. To do this, each data set is randomized with respect to columns and rows separately to produce ten different data sets. Performance on each of these data sets was measured using the Adjusted Rand Index with model parameters set to the optimal values for the corresponding original data set. Later using GA with corresponding aforementioned settings, TopoBARTMAP and BARTMAP had their parameters optimized for the data sets produced by randomization to identify the sensitivity to order of presentation. As TopoART and Fuzzy were not sensitive to changes in feature order, using the respective settings mentioned previously, parameters were optimized for TopoART using GA, and Fuzzy ART's ρ was varied to measure the performance on data sets with randomization in order of observations.

V. Results and Discussion

We compare the performance of TopoBARTMAP, BARTMAP, TopoART, and Fuzzy ART on the aforementioned experiments in this section.

Table II shows the clustering performance of the contenders measured using the Adjusted Rand Index. For the original data sets, TopoBARTMAP outperforms the other three algorithms on all six data sets. While not being a biclustering algorithm, TopoART outperforms BARTMAP on four out of six data sets, and its performance is comparable with TopoBARTMAP on the Armstrong-V1 [37] and Shipp-V1 [38] data-sets. On the other hand, BARTMAP outperforms TopoART but loses to TopoBARTMAP on Alizadeh-V1 and Alizadeh-V3 [35]. We attribute TopoART's better performance over BARTMAP to the topological learning and order of observation presentation. Of all the algorithms compared, Fuzzy ART fails to identify better clusters and consequently performs poorly.

The results of experimentation on data sets with a randomized order of observations and features with parameters fixed to optimal parameters are summarized in the Tables III and IV, respectively. Unsurprisingly, none of the contenders, TopoBARTMAP, BARTMAP, TopoART, or Fuzzy ART, perform better with the order of observation presentation changed. Since TopoART and Fuzzy ART are impervious to the randomization of the features, the results summarized in Table II are valid for this case. Due to this, the performance of TopoBARTMAP and BARTMAP are juxtaposed in Table IV, where BARTMAP wins four out of six times in experimentation with randomized features. These tests reveal that TopoBARTMAP is more sensitive to parameterization than are BARTMAP or TopoART.

The results of order sensitivity analyses run with GA optimization (and parametric sweep for Fuzzy ART) are summarized in Tables V and VI. While the performance of each of the algorithms differs noticeably, TopoBARTMAP still outperforms the others. In contrast to the original-order experiments, in the randomized cases, BARTMAP outperforms TopoART on all randomized observation order data sets. To estimate the statistical significance of the performance differences on these data sets, a t-test was conducted among TopoBARTMAP, BARTMAP, and TopoART. The results indicate that the performances of TopoBARTMAP and TopoART, and of BARTMAP and TopoART are significantly different at 0.01 significance on all the data sets. However, the performance difference between TopoBARTMAP and BARTMAP varied with the data set under consideration. For data sets generated from Alizadeh-V1, Armstrong-V1, and Armstrong-V2, the performance differed significantly at 0.05 significance. For data sets generated from Alizadeh2000v1, Alizadeh2000v3, and Shipp2002v1 the performances are significantly different at 0.075, 0.10, and 0.15 significance levels, respectively. For the randomized feature order experiments, when run with GA, the results remained close to the original order experiments. TopoBARTMAP outperforms BARTMAP in all these experiments.

VI. Conclusion

Our goal was to integrate BARTMAP with topological data analysis to improve the effectiveness of biclustering and robustness to noise, and to uncover complex relationships present in the gene expression data sets. We experimentally demonstrated that combining TopoART with BARTMAP, which we call TopoBARTMAP, results in an algorithm that outperforms BARTMAP. TopoBARTMAP is more robust to order of presentation than either TopoART or BARTMAP. It is also more robust to noise and finds arbitrarily shaped biclusters.

The experiments on parameter sensitivity demonstrate that optimal TopoBARTMAP hyperparameters for one presentation order might not work for different presentation orders. Further, this behavior raises a question about the possibility of TopoBARTMAP being subject to overfitting, which requires further investigation. Due to the rigidness concerning parameters and its higher computational cost than BARTMAP, we recommend using TopoBARTMAP when data are known to have complex structures that are otherwise not easily identifiable. In any other case, we recommend using BARTMAP.

Acknowledgment

This research was sponsored by the Missouri University of Science and Technology Mary K. Finley Endowment and Intelligent Systems Center; the Coordenação de Aperfeiçoamento

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

TABLE II

BICLUSTERING RESULTS ON ORIGINAL DATA SETS WITH PERFORMANCE MEASURED USING ADJUSTED RAND INDEX. BEST PERFORMANCES ARE REPORTED IN BOLD.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	Fuzzy ART
1	Alizadeh-V1	**0.4407**	0.4292	0.1899	0.0922
2	Alizadeh-V2	**1.0000**	0.8952	0.9186	0.8147
3	Alizadeh-V3	**0.5521**	0.5079	0.4852	0.4055
4	Armstrong-V1	**1.0000**	0.5565	**1.0000**	0.3379
5	Armstrong-V2	**1.0000**	0.8270	0.9583	0.6281
6	Shipp-V1	**1.0000**	0.9443	**1.0000**	0.1583

TABLE III

BICLUSTERING RESULTS ON RANDOMIZED ORDER OF OBSERVATIONS DATA SETS RUN WITH PARAMETERS FIXED TO OPTIMAL PARAMETERS FOUND FOR ORIGINAL DATA SET. MEAN ADJUSTED RAND INDEX IS USED AS METRIC FOR COMPARISON. BEST PERFORMANCES ARE REPORTED IN BOLD.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	FuzzyART
1	Alizadeh-V1	0.0091 ± 0.0662	0.0119 ± 0.0099	**0.0132 ± 0.0125**	-0.0056 ± 0.0003
2	Alizadeh-V2	**0.4327 ± 0.3773**	0.0419 ± 0.0304	0.0183 ± 0.0084	-0.0025 ± 0.0005
3	Alizadeh-V3	**0.2576 ± 0.1501**	0.1237 ± 0.0871	0.0220 ± 0.0146	-0.0079 ± 0.0000
4	Armstrong-V1	0.0284 ± 0.0536	**0.0455 ± 0.0579**	0.0075 ± 0.0073	0.0153 ± 0.0006
5	Armstrong-V2	0.0015 ± 0.0034	**0.1117 ± 0.0485**	0.0093 ± 0.01401	0.0135 ± 0.0005
6	Shipp-V1	0.0116 ± 0.0260	0.0259 ± 0.0195	**0.0277 ± 0.0150**	0.0066 ± 0.0002

TABLE IV

BICLUSTERING RESULTS ON RANDOMIZED ORDER OF FEATURES DATA SETS RUN WITH PARAMETERS FIXED TO OPTIMAL PARAMETERS FOUND FOR ORIGINAL DATA SET. MEAN ADJUSTED RAND INDEX IS USED AS METRIC FOR COMPARISON. BEST PERFORMANCES ARE REPORTED IN BOLD.

Sl.no	DataSet	TopoBARTMAP	BARTMAP
1	Alizadeh-V1	0.0179 ± 0.0234	**0.0666 ± 0.0641**
2	Alizadeh-V2	0.3407 ± 0.4660	**0.8952 ± 0.0000**
3	Alizadeh-V3	0.2747 ± 0.1987	**0.2927 ± 0.0440**
4	Armstrong-V1	**0.3989 ± 0.3215**	0.2881 ± 0.1632
5	Armstrong-V2	**1.0000 ± 0.0000**	0.8152 ± 0.0000
6	Shipp-V1	0.6312 ± 0.5079	**0.9443 ± 0.0000**

TABLE V

BICLUSTERING RESULTS ON RANDOMIZED ORDER OF OBSERVATION DATA SETS RUN WITH GA OPTIMIZATION FOR TOPOBARTMAP, BARTMAP, AND TOPOART AND VARIATION OF ρ FOR FUZZY ART. ADJUSTED RAND INDEX IS USED AS METRIC FOR COMPARISON. BEST PERFORMANCES ARE REPORTED IN BOLD.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	Fuzzy ART
1	Alizadeh-V1	**0.5893 ± 0.1120**	0.4485 ± 0.0853	0.1208 ± 0.0114	0.0658 ± 0.0012
2	Alizadeh-V2	**0.9826 ± 0.0237**	0.8836 ± 0.0401	0.3529 ± 0.0689	0.1166 ± 0.0013
3	Alizadeh-V3	**0.6006 ± 0.0754**	0.5708 ± 0.0896	0.1727 ± 0.0250	0.0954 ± 0.0005
4	Armstrong-V1	**0.9068 ± 0.0330**	0.7537 ± 0.0478	0.1567 ± 0.0204	0.1147 ± 0.0022
5	Armstrong-V2	**0.9149 ± 0.0482**	0.8142 ± 0.0776	0.1992 ± 0.0191	0.0990 ± 0.0021
6	Shipp-V1	**0.5297 ± 0.0902**	0.4116 ± 0.0660	0.2875 ± 0.0538	0.0328 ± 0.0000

TABLE VI

BICLUSTERING RESULTS ON RANDOMIZED ORDER OF FEATURE DATA SETS RUN WITH GAS. PERFORMANCE MEASURED USING ADJUSTED RAND INDEX. BEST PERFORMANCES ARE REPORTED IN BOLD.

Sl.no	DataSet	TopoBARTMAP	BARTMAP
1	Alizadeh-V1	**0.5580 ± 0.1023**	0.5210 ± 0.0646
2	Alizadeh-V2	**0.9799 ± 0.0000**	0.8952 ± 0.0000
3	Alizadeh-V3	**0.5890 ± 0.0501**	0.5630 ± 0.0700
4	Armstrong-V1	**1.0000 ± 0.0000**	0.5640 ± 0.0364
5	Armstrong-V2	**0.9840 ± 0.0219**	0.8380 ± 0.0061
6	Shipp-V1	**1.0000 ± 0.0000**	0.9550 ± 0.0249

de Pessoal de Nível Superior - Brazil (CAPES) - Finance code BEX 13494/13-9; 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.

REFERENCES

[1] Y. Cheng and G. M. Church, "Biclustering of expression data." in *Ismb*, vol. 8, no. 2000, 2000, pp. 93–103.

[2] G. Govaert and M. Nadif, *Co-clustering: models, algorithms and applications*. John Wiley & Sons, 2013.

[3] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: a survey," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, no. 1, pp. 24–45, 2004.

[4] V. A. Padilha and R. J. Campello, "A systematic comparative evaluation of biclustering techniques," *BMC bioinformatics*, vol. 18, no. 1, p. 55, 2017.

[5] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *Acm Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[6] H. Wang, W. Wang, J. Yang, and P. S. Yu, "Clustering by pattern similarity in large data sets," in *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 394–405.

[7] J. Chen, L. Ji, W. Hsu, K.-L. Tan, and S. Y. Rhee, "Exploiting domain knowledge to improve biological significance of biclusters with key missing genes," in *2009 IEEE 25th International Conference on Data Engineering*. IEEE, 2009, pp. 1219–1222.

[8] A. Patania, F. Vaccarino, and G. Petri, "Topological analysis of data," *EPJ Data Science*, vol. 6, no. 1, p. 7, 2017.

[9] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[10] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson, "Extracting insights from the shape of complex data using topology," *Scientific reports*, vol. 3, p. 1236, 2013.

[11] S. Lockwood and B. Krishnamoorthy, "Topological features in cancer gene expression data," in *Pacific Symposium on Biocomputing Co-Chairs*. World Scientific, 2014, pp. 108–119.

[12] G. Singh, F. Mémoli, and G. E. Carlsson, "Topological methods for the analysis of high dimensional data sets and 3d object recognition." in *SPBG*, 2007, pp. 91–100.

[13] M. Nicolau, A. J. Levine, and G. Carlsson, "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival," *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011.

[14] M. Perkins and K. Daniels, "Visualizing dynamic gene interactions to reverse engineer gene regulatory networks using topological data analysis," in *2017 21st International Conference Information Visualisation (IV)*, July 2017, pp. 384–389.

[15] A. H. Rizvi, P. G. Camara, E. K. Kandror, T. J. Roberts, I. Schieren, T. Maniatis, and R. Rabadan, "Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development," *Nature biotechnology*, vol. 35, no. 6, p. 551, 2017.

[16] X. Ma, L. Gao, and K. Tan, "Modeling disease progression using dynamics of pathway connectivity," *Bioinformatics*, vol. 30, no. 16, pp. 2343–2350, 04 2014.

[17] R. Xu and D. C. Wunsch, "Clustering algorithms in biomedical research: A review," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 120–154, 2010.

[18] R. Xu and D. C. Wunsch II, "BARTMAP: A viable structure for biclustering," *Neural Networks*, vol. 24, no. 7, pp. 709–716, 2011.

[19] M. Tscherepanow, "TopoART: A topology learning hierarchical ART network," in *International Conference on Artificial Neural Networks*. Springer, 2010, pp. 157–167.

[20] A. Chaibi, M. Lebbah, and H. Azzag, "A new bi-clustering approach using topological maps," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013, pp. 1–7.

[21] T. Sarazin, M. Lebbah, H. Azzag, and A. Chaibi, "Feature group weighting and topological biclustering," in *International Conference on Neural Information Processing*. Springer, 2014, pp. 369–376.

[22] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer vision, graphics, and image processing*, vol. 37, no. 1, pp. 54–115, 1987.

[23] S. Kim, "Novel approaches to clustering , biclustering algorithms based on adaptive resonance theory and intelligent control," Ph.D. dissertation, Missouri University of Science and Technology, 2016.

[24] I. Elnabarawy, D. C. Wunsch, and A. M. Abdelbar, "Biclustering ARTMAP collaborative filtering recommender system," in *2016 International Joint Conference on Neural Networks (IJCNN)*, July 2016, pp. 2986–2991.

[25] 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 Networks*, vol. 4, no. 6, pp. 759 – 771, 1991.

[26] G. A. Carpetner, S. Grossberg, and D. B. Rosen, "Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns," Boston University Center for Adaptive Systems and Department of Cognitive ..., Tech. Rep., 1991.

[27] G. A. Carpenter and S. Grossberg, *Adaptive resonance theory*. Springer, 2010.

[28] S. Grossberg, "Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world," *Neural Networks*, vol. 37, pp. 1–47, 2013.

[29] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch II, "A survey of adaptive resonance theory neural network models for engineering applications," *Neural Networks*, vol. 120, pp. 167 – 203, 2019.

[30] D. C. Wunsch II, "Admiring the Great Mountain: A Celebration Special Issue in Honor of Stephen Grossberg's 80th Birthday," *Neural Networks*, vol. 120, pp. 1 – 4, 2019, special Issue in Honor of the 80th Birthday of Stephen Grossberg.

[31] R. Xu and D. Wunsch, *Clustering*. John Wiley & Sons, 2008, vol. 10.

[32] G. A. Carpenter, S. Grossberg, J. H. Reynolds *et al.*, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network."

[33] 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 Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, Sep. 1992.

[34] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: a comparative study," *BMC bioinformatics*, vol. 9, no. 1, p. 497, 2008.

[35] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[36] K. Wang, B. Wang, and L. Peng, "CVAP: validation for cluster analyses," *Data Science Journal*, pp. 0 904 220 071–0 904 220 071, 2009.

[37] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. J. Korsmeyer, "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, no. 1, pp. 41–47, 2002.

[38] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub, "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," *Nature Medicine*, vol. 8, no. 1, pp. 68–74, 2002.

[39] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, 2nd ed. Springer Publishing Company, Incorporated, 2015.

[40] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[41] L. E. Brito da Silva and D. C. Wunsch II, "A study on exploiting VAT to mitigate ordering effects in Fuzzy ART," in *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 2351–2358.

[42] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch II, "Dual vigilance fuzzy adaptive resonance theory," *Neural Networks*, vol. 109, pp. 1 – 5, 2019.

[43] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch II, "Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence," *Neural Networks*, vol. 121, pp. 208 – 228, 2020.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE