A framework for initialising a dynamic clustering algorithm: ART2-A

Simon J. Chambers, Ian H. Jarman
and Paulo J. G. Lisboa
School of Computing and
Mathematical Sciences
Liverpool John Moores University
Liverpool, L3 3AF, UK

Abstract—Algorithms in the Adaptive Resonance Theory (ART) family adapt to structural changes in data as new information presents, making it an exciting candidate for dynamic online clustering of big health data. Its use however has largely been restricted to the signal processing field. In this paper we introduce an adaptation of the ART2-A method within a separation and concordance (SeCo) framework which has been shown to identify stable and reproducible solutions from repeated initialisations that also provides evidence for an appropriate number of initial clusters that best calibrates the algorithm with the data presented. The results show stable, reproducible solutions for a mix of real-world heath related datasets and well known benchmark datasets, selecting solutions which better represent the underlying structure of the data then using a single measure of separation. The scalability of the method and it's facility for dynamic online clustering makes it suitable for finding structure in big data.

I. INTRODUCTION

The ability of modern computers to collect and store information has lead to increased challenges for those whose role it is to make use of this information. This is particularly the case for medical data which is collected and stored but rarely exploited and the result is that much of it lies unused [1]. It is important therefore that strategies and methods are developed which allow clinicians and other healthcare professionals to make the best use of this resource where possible. This puts the healthcare field at the forefront of big data analysis, where it is critical that robust techniques are developed to deal with the need to inform clinical and public health decision making.

Traditional clustering techniques such as k-means are not necessarily appropriate for big data which is assessed by the application of the 5 Vs:

- Volume
- Velocity
- Variety
- Veracity
- Value

in such a context the static nature of k-means and other hierarchical clustering methods mean that once the initial clusters are identified it becomes unwieldy to update such a model for new data as it arrives. To better approach this a dynamic partitioning algorithm which continually updates it's prototypes and identifies new groupings of observations as they occur is important, especially for big data and health where new observations are constantly arriving and may even be discarded immediately after use mean that iterative clustering techniques such as k-means are inappropriate choices.

The use of unsupervised clustering techniques is therefore vital in the context of big medical data as these form the groundwork of any large scale exploratory analysis for understanding of the structure of the data. An appropriate example for big healthcare data is the UK Hospital Episode Statistics (HES) data maintained by the NHS [2] which is a massive dataset of patient hospital visits in the UK over an extended time frame. The data is made available to researchers healthcare professionals alike and is regularly updated with over 125 million records annually, meaning that it is a candidate for the use of dynamic clustering algorithms which allow for updates rather than requiring a complete re-run of any analysis.

The Adaptive Resonance Theory (ART) family of algorithms are neural networks designed to mimic the way the human brain recognises patterns [3], with the ART2 [4] method specifically designed to handle continuous analogue inputs. The motivation for the use of the ART family of algorithms in health data is to handle the stability-plasticity dilemma, which relates to how a neural network can learn about new patterns in data without being forced to discard information previously stored. This is a problem which affects many artificial and biological neural learning systems [5]. The ability of ART methods to dynamically update and identify new prototypes as data structure changes over time gives it an advantage over static methods such as k-means.

The ART2-A method is a derivative of ART2 and models the essential dynamics whilst improving computation efficiency and allowing for dynamic learning rates and prototype growth [6]. Surprisingly it has been often ignored in favour of other better understood and perhaps unsuitable clustering techniques such as k-means. It's usage has largely been limited to signal processing datasets.

Unlike in ART2 [4], for it's successor ART-2A there is no constraint on the number of prototypes and it can dynamically increase the number of groups. Importantly once the initial training phase is complete the method can continue to identify new structure and allocate prototypes to new data as it arrives, continually assessing whether the existing prototypes are appropriate choices. This plasticity means that the method is highly suitable for situations in which the underlying structure

of the data is unknown or subject to change over time.

The model is initialisation dependent as the order of input presentation affects the outcome, although less so than the standard ART2 model [6]. This variability of solutions does mean that a strategy is required for ensuring that the solution is both robust and reproducible. To address this shortcoming, a previously introduced separation and concordance framework (SeCo) [7] which uses measures of separation and concordance to produce a map of results. This map allows for the identification of a robust solution whilst also giving a visualisation of the structure of the solution space giving potential guidance of an appropriate number of prototypes.

This paper introduces an adaptation of the SeCo framework [7] for ART-2A where the dependence on presentation order is used rather than the initial selection of centres as previously to explore the solution space. The new framework is evaluated using a number of publicly available datasets and one synthetic breast cancer dataset to demonstrate the potential for a health and big data scenario. The results section presents an in-depth review of the synthetic breast cancer data and an overview of the results for the other datasets.

II. METHODS

The proposed framework comprises two components, the ART-2A algorithm which is incorporated into the SeCo framework where the separation and concordance map is produced.

There are two methods for ART-2A described, the first of which is the standard ART2-A method proposed by Carpenter and Grossberg [6] followed by a variation on ART-2A [8], both of which produce an output set of two cluster indices for each row M_i and $m \times n$ matrix of observations M and have the stability-plasticity characteristics fundamental to ART-2. A filter is applied to these solutions where prototypes having less than a proportion r of n observations assigned will be discarded and the data points allocated to the nearest alternative prototype. This has the effect of stabilising the number of prototypes produced for a given vigilance parameter ρ over multiple re-initialisations and thus giving fewer but better populated indices. These results are used in the SeCo framework method described here to identify stable and reproducible solutions.

A. ART-2A: Method 1

ART-2A is a derivative of ART-2 which approximates the dynamic nature of the original whilst improving the computational efficiency by an order of magnitude.

The ART2-A algorithm takes as it's input a matrix M with m observations of n features, in addition to three control parameters ρ the vigilance parameter, α a threshold variable satisfying the constraint (5) and β which controls the learning rate of the method. A fourth parameter θ is referenced in the paper as a means of suppressing noisy signals in the data, but is unused here. The matrix P represents the prototypes and is initialised as an empty set and populated as the method progresses.

Each element of the matrix M is presented to the algorithm in sequence as a non-uniform vector I_0, and transformed with each step being followed for each presentation of the vector to the algorithm over a given number of iterations. Prototypes are dynamically updated throughout using the learning function.

1) Presentation:

$$I = \mathcal{N}\mathcal{F}_0\mathcal{N}I_0 \tag{1}$$

where

$$\mathcal{N}x \equiv \frac{x}{||x||} \tag{2}$$

and

$$(\mathcal{F}_0 x)_i \equiv \begin{cases} x_i & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

subject to the constraint

$$0 < \theta \leq \frac{1}{\sqrt{m}} \tag{4}$$

where eqs. (1) to (4) imply that I is nonzero and normalised to unit euclidean length.

$$0 < \alpha \leq \frac{1}{\sqrt{m}} \tag{5}$$

2) Choice and Resonance: The input vector I is compared to each of prototypes p_j in P using the function (6). ART type functions use what are called committed and uncommitted prototypes, where a committed prototype is a prototype which has been previously created, and an uncommitted one is one that hasn't been selected yet. When an uncommitted prototype is selected, it creates a new category to which subsequent presentations of input vectors can be compared.

$$T_i = \begin{cases} \alpha \sum_i I_i & \text{when } p_j uncommitted \\ I \cdot p_j & \text{when } p_j committed \end{cases} \tag{6}$$

To start all the nodes are uncommitted, and the first data point presented forms the first prototype, meaning that the result is dependent upon the order in which data is presented. For each T_j calculated the maximal value is selected, and this becomes a committed node. This node j is tested (7) with constraints as in (8) if it is a previously committed node, if uncommitted, then the choice remains.

$$T_j \geq \rho^* \tag{7}$$

$$0 \leq \rho^* \leq 1 \tag{8}$$

When T_j is a previously committed node, the value T_j is equivalent to the cosine distance between the two vectors.

If the selected prototype does not pass the constraint in (8) then the value j is reset to the value of an uncommitted prototype.

3) Learning: Once a prototype has been selected by the above method, then the prototype is updated depending on whether it is a committed or uncommitted node. If a previously committed node then the update is performed as in (9) otherwise is it set as $P_j = I$.

$$vP_j^{new} = \mathcal{N}((1 - \beta)P_j^{old} + \beta \mathcal{N}I) \tag{9}$$

with

$$0 \leq \beta \leq 1 \tag{10}$$

It is important to note that the choice of learning parameter is significant as the extreme of each will result in different behaviour. Setting β to 1 will result in ART-2A having fast-learn properties whereas the other extreme will result in learning more like a leader algorithm, with the prototype remaining fixed after commitment. Small values of β will result in a slow-learning rate.

This method will produce varying numbers of prototypes with varying separation for each reinitialisation with the same given parameters, meaning some decision on which solution is most appropriate is necessary.

B. ART-2A: Method 2

The second method for ART-2A [8] uses the same normalisation functions as does the first with L2 Normalisation occurring through eqs. (1) to (3) differing in the way choice and resonance occur. The use of the parameters (11)is also slightly different from the originally proposed method [6] with the α parameter having a slightly different purpose and the β parameter having a changed constraint.

As with Method A, P, which contains the prototype vectors is initialised as empty and populated over time, and the vector I represents the normalised pattern under presentation.

1) Parameters:

α positive number $\leq \frac{1}{\sqrt{m}}$
β small positive number
θ normalisation parameter, having $0 < \theta \leq \frac{1}{\sqrt{m}}$ (11)
ρ vigilance parameter, having $0 < \rho \leq 1$

2) Choice and Resonance: Choice of prototype is determined by finding the prototype vector P_j which maximises (12), as before any ties at this point are solved arbitrarily.

$$P_j \cdot I \tag{12}$$

Following this, P_j is tested to see if it is sufficiently similar to the presented data point (13), if it fails this test, then a new prototype is created as before with P_{new} set to I.

$$P_j \cdot I \geq \alpha \sum_i I_i \tag{13}$$

If the prototype is similar then it is tested to ensure that the value is greater than the vigilance parameter $P_j \cdot I \geq \rho$, if failing then a new prototype is again created. If the prototype matches then the method proceeds to the learning stage.

3) Learning: During the learning stage, the prototype is updated to be a combination of the existing prototype and the presented data vector, using β to control the update rate (14).

$$P_j = \frac{(1 - \beta)P_j + \beta I}{||(1 - \beta)P_j + \beta I||} \tag{14}$$

This second method has a similar result to the original method, however it also has the benefit that the α parameter provides additional controls over the sparsity of the solution, enforcing a minimum cosine distance below which a new prototype is created, even if an otherwise matching prototype exists.

C. Comparison of methods

The two methods indicated here for ART-2A based clustering are similar in that they both use the cosine similarity measure to identify how similar partitions are to existing prototypes and determine which is the most appropriate match. The main divergence is the use of the α parameter between each, Method A uses the alpha parameter in combination with the cosine similarity measure to determine whether the best match is a new prototype or an existing one, whereas Method B applies a test first to see whether a new prototype is the good option, and if so creates one rather than continuing to the match and resonance step. This latter approach produces a sparser clustering with more numerous prototypes than the first. Both will produce prototypes when ρ is set to zero.

In the case of this study Method B has been used throughout although Method A could also have been used in it's place.

D. Framework

The stabilisation framework proposed here is a variation on a method previously used for k-means [7] and uses a dual measure approach to cluster solutions stability, namely Separation and Concordance. The separation measure chosen here is the total within cluster sum of squares (SSQ) (15) and is used in combination with Cramers' V statistic [9], a statistical score measuring the strength of the association between discrete variables shown in equation (17). However any other suitable score could be used, such as the Adjusted Rand Index [10].

$$\underset{s}{\arg\min} \sum_{i=1}^{k} \sum_{x_j \in S_i} ||x_j - \mu_i||^2 \tag{15}$$

$$\chi^2 = \sum_{p=1}^{P} \sum_{q=1}^{Q} \frac{(O_{pq} - E_{pq})^2}{E_{pq}} \tag{16}$$

$$C_v = \sqrt{\frac{\chi^2}{N \times \min(P - 1, Q - 1)}} \tag{17}$$

For a given dataset the following process is followed in applying the ART2-A method

1) Perform ART2-A algorithm
2) Calculate the within Cluster Sum of Squares
3) Repeat Steps 1–2 until sufficient number of solutions for each generated number of prototypes
4) Select the top 10% of these solutions by the Within Cluster Sum of Squares.
5) Calculate the stability measure, the pairwise Cramers V statistic for each combination of these solutions.
6) Calculate the median within group Cramers' V statistic for each solution
7) Produce the Separation Concordance (SeCo) map of these solutions.

The selection of the top 10% of solutions is important as it ensures that only well separated solutions are considered when looking at those results which are stable. Step 3 requires waiting until a sufficient number of solutions is found for each different prototype number group, so n=3, n=4 and n=5 prototypes for example. As previously [7] 500 was used as a sufficient number to select the top 10% from.

The pairwise Cramers' V concordance of each of these solutions is calculated within groups of solutions identified by having the same number of prototypes, and the median of these is taken as a representative value of the overall stability of the solution.

Finally the SeCo map is generated as a visualisation of the relative performance of the different groups allowing the user to gauge performance in selecting an appropriate solution for use, the separation of solutions is on the y-axis and the internal concordance on the x-axis. Previous results show that choosing solutions towards the right-most edge is the most effective strategy in obtaining stable, reproducible solutions [7].

III. DATASETS

Four datasets were used to test the performance of the proposed stability measure for ART2-A against the best separated solution, these are the Wine[11] and Olive Oil[12] datasets, and a synthetic protein expression dataset derived from a real world breast cancer dataset.

The characteristics of these datasets are as follows:

Wine Comprising 178 observations of 13 properties resulting from the chemical analysis of wines grown in the same region of Italy yet derived from three cultivars. The attributes are Alcohol, Malic Acid, Ash, Alkalinity of Ash, Magnesium, Total Phenols, Flavonoids, Non-flavonoid phenols, Proanthocyanins, Colour Intensity, Hue, OD280/OD315 of diluted wine and Proline.

Olive Oil Taken from 572 samples of olive oil derived from nine regions of Italy and comprising the normalized concentrations of eight fatty acids within the oil. These fatty acids are Palmitic, Palmitoleic, Stearic, Oleic, Linoleic, Linolenic, Arachidic and Eisenoic. The nine regions of Italy are themselves grouped into three areas, North (Umbria, East and West Liguria),

Sardinia (Inland and Coastal Sardinia) and from the South (North and South Apulia, Calabria and Sicily).

Synthetic Breast Cancer This dataset has 5000 observations in three dimensions with the points randomly sampled from a mixture of ten multivariate normal distributions. Parameter settings were chosen to replicate a real world breast cancer dataset[13] with data projected onto three principle separation axes[14] and comprising separated and contiguous cohorts.

Thyroid Data A dataset looking at thyroid function[15], [16] containing 215 observations with 5 features - T3resin, Thyroxin, Triiodothyronine, Thyroidstimulating, TSH_value - with three target classes, Normal, Hyperthyroidism and Hypothyroidism.

Prior to use all the datasets were scaled such that $\bar{x} = 0$ and $s = 1$.

IV. RESULTS

Application of the framework to each of the three datasets results in a Separation Concordance (SeCo) map being produced for each. As with previous applications of the method[7] the map indicates solutions which are of interest for the user. The maps are produced using parameters of $\rho = 0.95$, $\alpha = \frac{0.5}{sqrt(m)}$ and $\beta = \frac{0.5}{sqrt(m)}/$, where m is the number of columns specific to the data set. The latter parameter is a balanced option in terms of the learning rate, with that particular alpha having a reasonably broad matching parameter. The vigilance parameter being set at $\rho = 0.95$ was a deliberate choice based on evidence that the separation of solutions being produced by a range of parameters.

A. Choice of ρ

The choice of an appropriate value for the vigilance parameter is of importance and strategies have been developed which attempt to effectively deal with this[17], however the evidence presented here indicates that a higher vigilance will yield better solutions than one which is lower even for the same number of clusters.

Figure 1 shows box plots of the separation measure (within cluster sum of squares) for each set of results with similar number of prototypes, within the set of solutions produced using a given vigilance parameter. As can be seen the higher the vigilance parameter it is often the case that the variance of the within cluster sum of squares is lower for solutions with the same number of prototypes and the overall separation is higher also. For the lower vigilance parameters, the expected low number of cohorts is observed with only one or two different classes being generated for $\rho \in 0.7 \ldots 0.8$ which for the most part show that there are separation of the results is broadly comparable, however the large number of outliers on these indicate that the solutions are quite different.

As the vigilance parameter increases beyond 0.8 a rapid increase in the overall separation of the solutions is observed, albeit with a corresponding rise in the variability of separation, until reaching the chosen value of $\rho = 0.95$ where the solutions exhibit has low variability of separation with few outlying
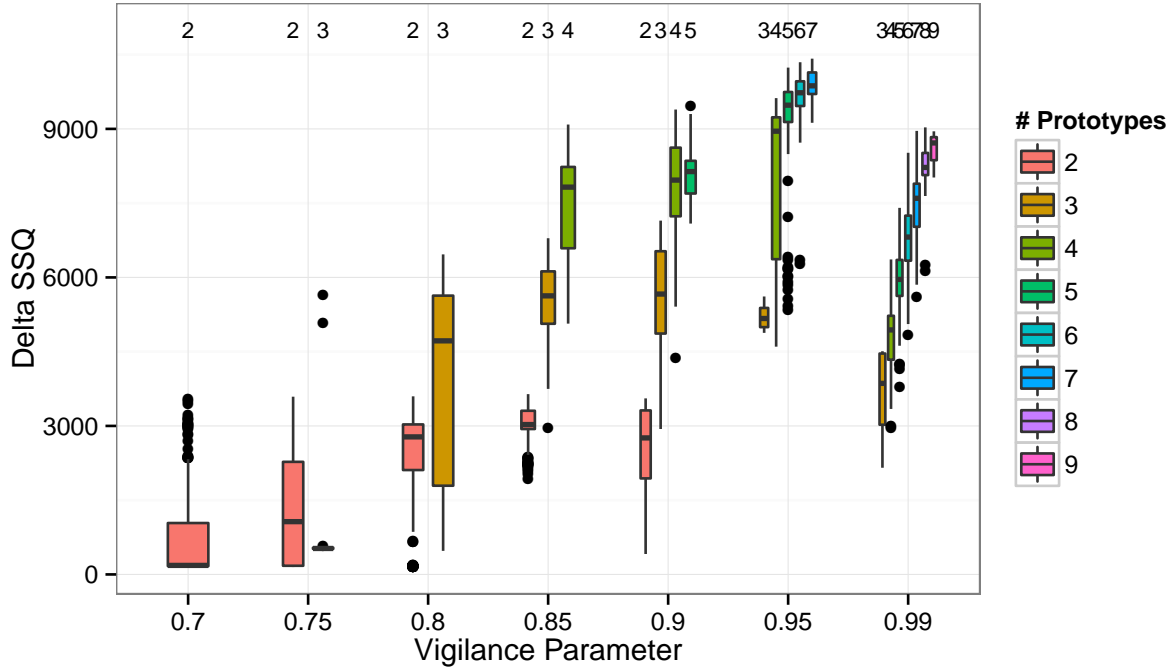
Fig. 1. Box plots showing stability for the Synthetic Dataset using within cluster sum of squares and showing poorer clustering performance for lower values of ρ

solutions. This would seem to confirm the choice of vigilance parameter as being prudent.

B. SeCo Map

The Separation Concordance (SeCo) maps represent stability on the y-axis in the form of the Within Cluster Sum of Squares (transformed to aid in visualisation[7]) and the internal consistency of the solutions on the y-axis. An ideal solution is one which is both well separated and self-consistent with the other produced solutions, meaning that it will be both robust and reproducible. This property allows repeated application of the framework to the same dataset to produce similar results consistently. Interpreting the map therefore means looking at the right hand edge of the plot where the solutions are most stable, and looking for a group of solutions along that edge.

Using the SeCo map to evaluate the solutions generated by the ART2-A algorithm we get the plot shown in Figure 2 which shows the results for the Synthetic Breast Cancer dataset having six different cluster numbers $k \in 2 \ldots 7$. It is known that the data was generated using an initial 10 clusters, but that it is not possible to recover the true structure as a result of a combination of sparse and mixed clusters.

Immediately it can be seen that there are two main clouds of solutions, one for $k = 3$ and one for $k \in 4 \ldots 7$, with $k = 2$ showing a grouping with poor separation and poor overall concordance. The solution set for the group $k = 3$ show a consistent block of solutions on the right hand side with another set of less consistent solutions on the left. The most interesting grouping if for those with four prototypes as these have a dense cloud of solutions on the right hand edge

of the map all having good separation, these results are stable and highly reproducible.

The remaining solutions for $k \in 5 \ldots 7$ exhibit increased cluster separation but correspondingly a marked decrease in the internal stability of the solutions, this indicates that although the ART2-A algorithm is able to separate the observations into tight groupings, it is not able to do so in a robust manner; meaning that a choice of solutions here is likely to result in an unrepresentative partitioning of the data. The SeCo map therefore indicates that the best choice of solution would be that with the highest internal consistency for $k = 4$ as these have better overall separation and stability than the $k = 3$ block.

It is worth noting here that strong indications are given through the SeCo map as to the appropriate number of partitions for the data, this is particularly important in the context of the ART2-A algorithm as there is no direct mechanism, as with methods like k-means, for specifying the number of clusters, and the differing partitions are generated using the same tuning parameters. This means that in an exploratory data analysis the map provides useful insights to the structure of the data.

TABLE I.   SUMMARY OF RESULTS FOR DIFFERENT DATASETS

	ρ	Framework (k)	Separation (k)
Wine	0.95	0.907 (3)	0.903 (8)
Olive Oil	0.95	0.905 (5)	0.823 (8)
Synthetic	0.95	0.862 (4)	0.748 (7)
Thyroid	0.95	0.742 (3)	0.742 (3)

Application of the method to the three remaining datasets give the summary shown in Table I where the solution chosen
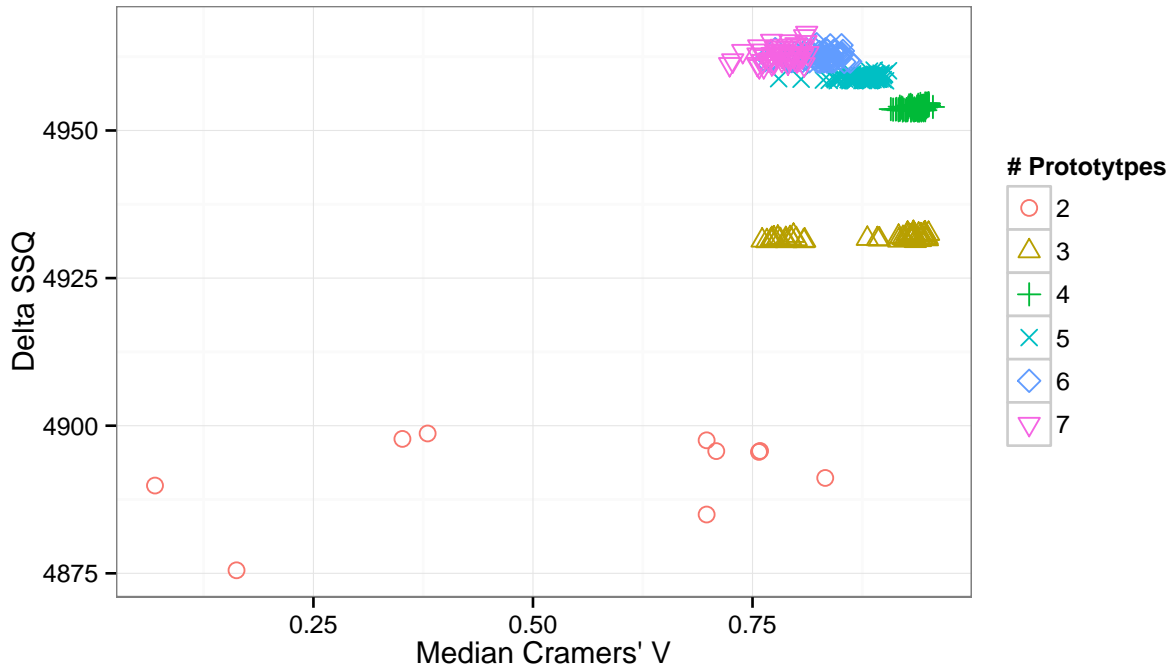
Fig. 2. Separation Concordance Map of Synthetic data, $\alpha = 0.7$ and $\beta = 0.5$

by the framework for each dataset is compared against the solution with the best separation overall for that dataset, using the Cramers' V of those partitions with the reference partitions for the data. The number of clusters in each solution is indicated in brackets with the Wine data having 3 underlying groups, the Olive Oil having 9 groups, Thyroid 3 groups, and the Synthetic data having 10.

Partitions generated for the Wine dataset are the most similar with both the Framework approach and the best Separated solution having broadly comparable solutions at around 0.9 concordance, this is to be expected as the Wine dataset is relatively easy to partition being comprised of three reasonably distinct groups. The next most complicated dataset the Olive Oil where again the Framework approach works well having a concordance of about 0.9 again, performing slightly better than the k-Means version of the framework on the same data which obtained a concordance of about 0.83, which is equivalent to that selected by the best separated solution by for ART2-A.

The thyroid dataset shows that the best separated solution and the solution chosen by the framework both have the same concordance with the underlying solution as each other, with both having three partitions also. Closer examination of the results revealed that both solutions produced the same partition of the data, having a Cramers' V concordance of 1. The most notable results though are for the Synthetic dataset where the dual measure approach obtained a concordance with the underlying solution of ~0.86 with the best separated solution getting ~0.75.

The competing best separated partition was obtained by looking at the complete result set and selecting the single best separated result from here, rather than looking at the best separated result for the same number of partitions as identified

by the framework. This was a deliberate choice as in the absence of the framework or any additional measures there is no way to appropriately identify that number of cohorts as being an appropriate choice, so the reasonable approach was to simply select the overall most separated. It would not be appropriate to select from the same value of k as this is already a self selecting group of highly consistent solutions.

These results confirm the original hypothesis that a dual measure approach to selecting a solution is preferable to considering just a separation measure alone being equivalent or better in each case for the three datasets of varying complexity examined here. However one of the important features of the framework approach is that the solutions are robust and reproducible, so repeated application of the approach should yield results that are equivalent when comparing with the reference partitions. It is also important to comment that the solutions generated here are produced using the same vigilance parameter, despite having differing numbers of cohorts within the solution space, and absent the use of a dual measure approach to evaluating the solutions there is little indication as to which distinct number of partitions best represents the structure of the data.

To test whether the method is both stable and reproducible a check was performed by applying the framework against the data repeatedly, each time selecting the best solution and comparing it with the reference partition, the results of which are shown in Figure 3. The concordance with the reference partition is shown on the y-axis with the x-axis being meaningless, the data points simply being plotted in descending order from left to right. It can be seen here that although there is a slight drop in the concordance with the actual solutions of ~0.02 the solutions are broadly comparable,
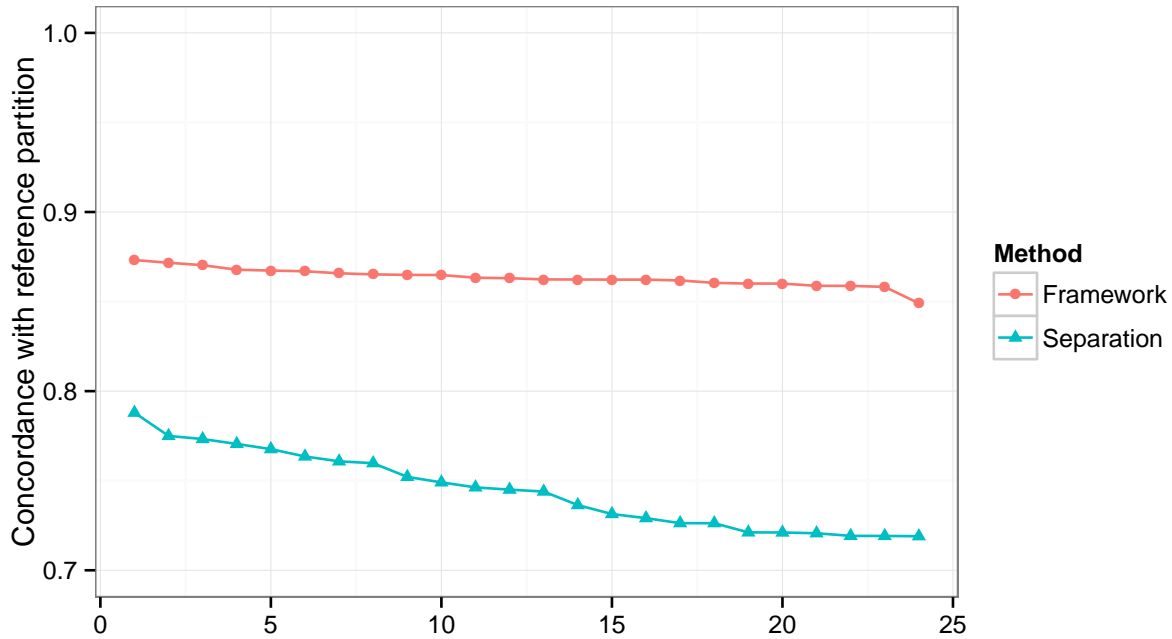
Fig. 3. Stability analysis of 24 runs of the ART2-A framework on the Synthetic dataset

whereas the clusters selected by SSQ alone perform worse against the reference solution with more variability.

V. CONCLUSION

An initialisation framework for ART2-A is introduced to assist in the identification of well separated and reproducible solutions using repeated random reinitialisations of the algorithm and tested on three real world datasets and one Synthetic Breast Cancer Dataset for which the ground truth is known. It has been shown previously that the use of a dual measure approach to evaluate the suitability of solutions from a clustering algorithm without using any external reference is a robust strategy for dealing with initialisation dependence [7].

The results here have shown that the method presented here provides a credible solution to the problem of initialisation dependence for the ART2-A algorithm and allows for the robust identification of reproducible solutions and also to assist in the selection of an appropriate number of solutions to represent the underlying structure of the data. The results have clearly shown that for the same initial parameters that there is a substantial variability in the resultant partitions of the data and that simply selecting a solution based on separation will not result in a reliably good partition. Using a dual measure approach to initialisation of this algorithm provides a good starting position to using the algorithm in a big data environment, allowing for robust partitions to be used to cluster incoming data whilst still allowing for the possibility of new cohorts being identified over time and in such a context ART2-A provides a realistic alternative to hierarchical, k-means or other similar Expectation-Maximisation like algorithms.

The earlier discussion mentioned the 5 "V"s of big data, and how they were commonly used to define what is meant by such. The use of ART-2A to cluster data allows for a dynamic clustering where Velocity, Volume and Value are of particular concern, given the ability of this family of algorithms to adapt to new prototypes over time and where learning these new cohorts is important. Healthcare is one such field, but is widely applicable to clustering scenarios outside of this.

Identifying data which does not fit existing prototypes is not enough however, and whilst this work provides a good starting point for clustering of big data in healthcare, there is a need to ensure that such new prototypes are in fact signal and not noise, and further work is required to evaluate the performance of the method in actual big data scenarios and benchmark the effect of introducing new cohorts into a previously trained structure.

The results indicate that ART-2A within the framework approach excels at identifying gross structure within these datasets however the fine structure is still elusive as shown by the choice of 4 cohorts for the Synthetic Breast Cancer data where a previous study using k-means has identified eight with equivalent concordance to the reference partition. A further step might be to include a hierarchical application of this approach to explore the existence of substructure within the membership of these groups.

REFERENCES

[1] B. Saha, D.-S. Pham, D. Phung, and S. Venkatesh, "Clustering Patient Medical Records via Sparse Subspace Representation," in *Advances in knowledge discovery and data mining : 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013 : proceedings*, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds. Springer Berlin / Heidelberg, 2013, ch. 11, pp. 123–134. [Online]. Available: http://hdl.handle.net/10536/DRO/DU:30053591

[2] H. . S. C. I. Centre, "Hospital Episode Statistics," 2014. [Online]. Available: http://www.hscic.gov.uk/hes

[3] S. Grossberg, "Adaptive Resonance Theory: how a brain learns to consciously attend, learn, and recognize a changing world." *Neural networks : the official journal of the International Neural Network Society*, vol. 37, pp. 1–47, Jan. 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23149242

[4] G. a. Carpenter and S. Grossberg, "ART 2: self-organization of stable category recognition codes for analog input patterns." *Applied optics*, vol. 26, no. 23, pp. 4919–30, Dec. 1987. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20523470

[5] M. Mermillod, A. Bugaiska, and P. Bonin, "The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects." *Frontiers in psychology*, vol. 4, no. August, p. 504, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3732997&tool=pmcentrez&rendertype=abstract

[6] G. a. Carpenter, S. Grossberg, and D. B. Rosen, "ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, vol. 4, no. 4, pp. 493–504, Jan. 1991. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0893608091900457

[7] P. J. Lisboa, T. A. Etchells, I. H. Jarman, and S. J. Chambers, "Finding reproducible cluster partitions for the k-means algorithm," *BMC Bioinformatics*, vol. 14, no. Suppl 1, p. S8, 2013. [Online]. Available: http://www.biomedcentral.com/1471-2105/14/S1/S8

[8] S. I. Gallant, "No Title," in *Neural Network Learning and Expert Systems*, 1993, ch. Chapter 7, pp. 147–149.

[9] H. Cramer, *Mathematical Methods of Statistics. (PMS-9).* Princeton University Press, Mar. 1999.

[10] D. Steinley, "Properties of the Hubert-Arabie adjusted Rand index." *Psychological methods*, vol. 9, no. 3, pp. 386–96, Sep. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15355155

[11] M. Forina, R. Leardi, C. Armanino, and S. Lanteri, "PARVUS: An Extendible Package for Data Exploration, Classification and Correlation," *Journal of Chemometrics*, vol. 4, no. 2, pp. 191–193, Mar. 1990. [Online]. Available: http://doi.wiley.com/10.1002/cem.1180040210

[12] M. Forina and E. Tiscornia, "Pattern recognition methods in the prediction of Italian olive oil origin by their fatty acid content," *Annali di Chimica*, vol. 72, pp. 143–155, 1982.

[13] D. M. Abd El-Rehim, G. Ball, S. E. Pinder, E. Rakha, C. Paish, J. F. R. Robertson, D. Macmillan, R. W. Blamey, and I. O. Ellis, "High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses." *International journal of cancer. Journal international du cancer*, vol. 116, no. 3, pp. 340–50, Sep. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15818618

[14] P. Lisboa, I. Ellis, a. Green, F. Ambrogi, and M. Dias, "Cluster-based visualisation with scatter matrices," *Pattern Recognition Letters*, vol. 29, no. 13, pp. 1814–1823, Oct. 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167865508001682

[15] A. Frank and A. Asuncion, "UCI Machine Learning Repository," 2010. [Online]. Available: http://archive.ics.uci.edu/ml

[16] M. Kousarrizi, F. Seiti, and M. Teshnehlab, "An experimental comparative study on thyroid disease diagnosis based on feature subset selection and classification," *International Journal of Electrical . . .*, no. February, 2012. [Online]. Available: http://www.ijens.org/vol_12_i_01/126001-8989-ijecs-ijens.pdf

[17] J. He, A.-H. Tan, and C.-L. Tan, "Modified ART 2A growing network capable of generating a fixed number of nodes." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 15, no. 3, pp. 728–37, May 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15384559

[18] J. Luo and D. Chen, "An Enhanced ART2 Neural Network for Clustering Analysis," *Proceedings of the 1st International ICST Conference on Forensic Applications and Techniques in Telecommunications, Information and Multimedia*, pp. 1–5, 2008. [Online]. Available: http://eudl.eu/doi/10.4108/wkdd.2008.2693

[19] S. Du, "Influence on ART2 Clustering Algorithm with Different Adjusting Learning Rate," *. . . in Information Science and Management Engineering*, vol. 2, pp. 30–34, 2012. [Online]. Available: http://www.jcisme.org/paperInfo.aspx?paperid=13370

[20] D. Soria, J. M. Garibaldi, F. Ambrogi, A. R. Green, D. Powe, E. Rakha, R. D. Macmillan, R. W. Blamey, G. Ball, P. J. G. Lisboa, T. a. Etchells, P. Boracchi, E. Biganzoli, and I. O. Ellis, "A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients." *Computers in biology and medicine*, vol. 40, no. 3, pp. 318–30, Mar. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20106472

[21] J. He, A. Tan, and C. Tan, "ART-C: A neural architecture for self-organization under constraints," *. . . of International Joint Conference on Neural . . .*, 2002.

[22] R. Xu and D. Wunsch, "Survey of clustering algorithms." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 16, no. 3, pp. 645–78, May 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18252358

[23] K.-L. Du, "Clustering: a neural network approach." *Neural networks : the official journal of the International Neural Network Society*, vol. 23, no. 1, pp. 89–107, Jan. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19758784