7-2016

Self-regulated incremental clustering with focused preferences

Di WANG

Ah-hwee TAN

Singapore Management University, ahtan@smu.edu.sg

Self-Regulated Incremental Clustering with Focused Preferences

Di Wang† and Ah-Hwee Tan‡
†Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY) and
‡School of Computer Science and Engineering
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
Email: {wangdi, asahtan}@ntu.edu.sg

***Abstract*—Due to their online learning nature, incremental clustering techniques can handle a continuous stream of data. In particular, various incremental clustering techniques based on Adaptive Resonance Theory (ART) have been shown to have low computational complexity in adaptive learning and are less sensitive to noisy information. However, parameter regularization in existing ART clustering techniques is applied either on different features or on different clusters exclusively. In this paper, we introduce Interest-Focused Clustering based on Adaptive Resonance Theory (IFC-ART), which self-regulates the vigilance parameter associated with each feature and each cluster. As such, we can incorporate the domain knowledge of the data set into IFC-ART to focus on certain preferences during the self-regulated clustering process. For performance evaluation, we use a real-world data set, named American Time Use Survey (ATUS), which records nearly 160,000 telephone interviews conducted with U.S. residents from 2003 to 2014. Specifically, we conduct case studies to explore three types of interesting relationship, focusing on the wage, age, and provision of elderly care, respectively. Experimental results show that the performance of IFC-ART is highly competitive and stable when compared with two well-established clustering techniques and three ART models. In addition, we highlight the important and unexpected findings observed from the clusters discovered.**

I. Introduction

Among all the data analysis models, clustering techniques are still the major learning methods used in unsupervised learning, which handles unlabelled data [1]. There is no universally agreed upon definition on clustering [2]. However, "most researchers describe a cluster by considering the internal homogeneity and the external separation, i.e., patterns in the same cluster should be similar to each other, while patterns in different clusters should not" [3].

Adaptive Resonance Theory (ART) [4] is a well-established self-organizing neural network model. Associations among the low-level input patterns are encoded in the high-level category field based on the resonance effect. Specifically, inspired by how human brains perceive the environment, ART involves a bottom-up processing of the external information and a top-down modulation of the internal knowledge (see Section III-A). Based on the different number of input fields in use and their corresponding usage, ART models can perform various types of learning [5], such as unsupervised [6], [7], reinforcement [8], [9], and supervised learning [10], [11].

ART has been widely employed in various clustering techniques in the literature. ARTMAP [12] uses a match tracking strategy (see Section III-B) to regulate the vigilance parameter (see Section III-A) during the clustering process. However, ARTMAP requires the presence of class labels of the data, which makes it unsuitable for unsupervised learning. ART under Constraint (ART-C) [13], [14] also regulates the vigilance parameter, but requires a user-defined number of clusters *a priori*. Generalized Heterogeneous Fusion ART (GHF-ART) [6] regulates both the vigilance and the contribution parameters (see Section III-A) associated with each input field to adaptively tune its relative importance. Hybrid Integration ART (HI-ART) [7] regulates the vigilance parameter associated with each formed cluster to adaptively tune its cluster boundary. In this paper, we propose Interest-Focused Clustering based on Adaptive Resonance Theory (**IFC-ART**), which extends GHF-ART and HI-ART in that the vigilance parameter associated with each input field and each formed cluster is self-regulated during the autonomous clustering process.

The major advantage of IFC-ART lies in that we can incorporate domain knowledge into the autonomous clustering process. Please note that by domain knowledge, we refer to some basic understanding of the data set, such as the nature of certain features, rather than some heuristic assumptions, such as the number of clusters. Before applying any learning model to a data set, we inevitably require some basic understanding of the data set, such as the number of features, whether their values are continuous or categorical, the value range of each feature, whether there are missing values, etc. Such basic understanding plays a critical role when we pre-process the data set and more importantly, enables us to further identify certain interesting features to be focused on during the data mining process. For example, instead of categorizing the personal income into rigidly defined wage groups, we may simply apply IFC-ART and pre-set its vigilance parameter associated with wage larger than those associated with the other features. As such, although all parameters are regulated in a generic way during the autonomous clustering process, we can still roughly control the different level of interest that we desire in various features. Therefore, IFC-ART removes the rigid restriction on the categorization of the sub-groups in the input features but still regulates the formation of data clusters according to our interests or preferences.

For performance evaluation, we apply IFC-ART on a real-world data set, named American Time Use Survey (**ATUS**)

[15]. ATUS is a federally administered on-going survey on the national scale on how U.S. residents spend their time on various activities. It is sponsored by the U.S. Bureau of Labor Statistics and conducted by the U.S. Census Bureau. The subjects participated in the survey were randomly selected from a large number of U.S. households and were interviewed via telephone. During the interviews, the subjects reported on how much time they spent on all the activities they performed from 4am on the previous day to 4am on the interview day. ATUS categorizes all activities into 17 types, which can be further expanded into over 400 sub-types. Besides the extremely detailed activity information, ATUS also consists of a wide range of demographic information, such as age, gender, education attainment, occupation, income, marital status, information about other occupants in the household, etc. ATUS has been studied in various aspects (see Section II-B). In this paper, we focus on three interesting case studies (see Section IV), namely (a) the relationship among wage, age, and education attainment, (b) the relationship among age and time spent on various activities, and (c) the relationship among age and time spent on child and elderly care.

The rest of the paper is organized as follows. Section II reviews the related work. Section III introduces the dynamics of ART and the details of IFC-ART. Section IV presents the case studies conducted on the ATUS data set. Section V concludes this paper and proposes future work.

II. Related Work

In this paper, we mainly focus on applying IFC-ART on the ATUS data set for performance evaluation and knowledge discovery. Therefore, in this section, we review the related clustering techniques and the prior studies on ATUS.

A. Closely Related Clustering Techniques

Clustering techniques may differ in many ways and can be categorized using various criteria. For example, based on the formation process, they can be broadly divided into the incremental and partitioning types. Generally speaking, incremental clustering techniques start with forming the first cluster using the first datum and subsequently decide for all the remaining data whether to form new clusters or to add them into existing ones based on certain similarity assessments. According to their nature, incremental clustering techniques such as ART-C [13], [14], ECM [16], and DIC [17] are often used for online learning. On the other hand, partitioning clustering techniques such as K-means [18], DBSCAN [19], and GARSC [20] treat all data as a whole and then divide them into clusters, which may only be used for offline learning.

Adaptive Resonance Theory (ART) [4] is a generic self-organizing neural model comprising two layers of neural fields connected by bidirectional conditional links. ART clustering techniques have been shown to have low computational complexity [6], [7] in online adaptive learning [13], [14] and are less sensitive to noisy information [6], [7]. However, ART-C [13], [14] requires a predefined number of clusters *a priori*, which is often heuristically assigned. Both GHF-ART [6] and HI-ART [7] do not require such empirical knowledge of the data sets and the parameters in use can be adaptively tuned along the clustering process. GHF-ART is mainly used for co-clustering of multimodal features and its parameters are regulated according to each feature. HI-ART is mainly used for homogeneous clustering of large scale data and its parameters are regulated according to each formed cluster. In this paper, we propose a new member IFC-ART to the family of ART clustering techniques, which regulates the vigilance parameter associated with each feature and each cluster. As such, we can incorporate the domain knowledge grasped from the basic understanding of the nature of the data set into the autonomous clustering process to maintain the focus on the identified interests or preferences.

B. Prior Studies on ATUS

Due to its large scale (159,900+ respondents) and large number of features (500+), ATUS [15] has attracted researchers from various backgrounds and has been studied in various aspects. Among all the academic studies on ATUS, first of all, some are specific to wage. In terms of the subjective well-being (reported in another data set), which is "connected to how people spend their time" [21], Kahneman et al. found that "people with greater income tend to devote relatively more of their time to work, compulsory non-work activities, and active leisure and less of their time to passive leisure activities" [21]. In terms of the relationship between paid work and housework, Hersch found that "housework has a negative relation with wages for both women and men" [22]. In terms of the relationship between wage and emotion, Kushlev et al. found that "higher income is associated with experiencing less daily sadness, but has no bearing on daily happiness" [23].

Secondly, some studies are specific to age. Zick et al. studied the difference in the amount of time spent on various physical activities among the subjects aged between 15 and 29 [24]. In terms of the relationship between social connections and the subjective well-being of the elderly, Kang and Michael found that "the number of daily social contact was positively associated with the self-rated health of older adults" (65+) [25]. In terms of subjective tiredness, Dolan and Kudrna found that elderly (65+) are "almost one point less tired" than people aged between 15 and 24 on a self-reported 0–6 scale [26].

Thirdly, some studies are specific to health. Most of these studies focused on how different behavioral or lifestyle factors influence the subjects' general health, or a specific aspect of health, such as sleep [27]. In terms of the amount of healthcare required, Russell et al. found that "the prevalence of health-care related activities rose with age" [28]. In terms of physical well-being, Hamermesh found that "eating meals more frequently is associated with lower BMI and better self-reported health, as is grazing more frequently" [29]. In addition, Podor and Halliday found that "better health is associated with more time allocated towards production on the market and at home, but less consumption of leisure" [30].

Some other interesting studies on ATUS are the analyses on the activity patterns of certain demographic groups, such

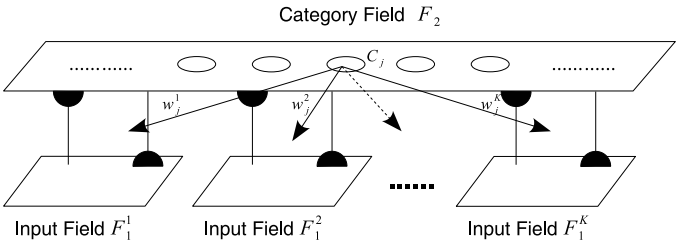

Fig. 1. Network structure of IFC-ART for associating K input fields.

as (a) how does a mother spend her time among various activities [31]; (b) how does the education attainment of the parents affect the amount of time they spend with their children [32]; and (c) what are the most frequently performed activities in each intensity-different category [33].

From the literature review, we observe that most existing studies on ATUS consider no more than three years of the collected data, despite now ATUS assembles twelve years of data. In our experiments, we use all available data across all years and apply IFC-ART for knowledge discovery.

III. Interest-Focused Clustering Based on Adaptive Resonance Theory

As depicted in Fig. 1, IFC-ART consists of a high-level category field and K number of low-level feature channels (or input fields), where K varies according to different clustering tasks and different data sets. Each committed code in the category field represents certain learned association among the input features in its weight vectors. The fundamental operations of IFC-ART follow the standard dynamics of **Fusion-ART** [5], which are introduced in the following subsection.

A. Dynamics of Fusion Adaptive Resonance Theory

Before we present the dynamics of Fusion-ART [5], we introduce all terms involved in its operations as follows.

Input vectors: Let $\mathbf{I}^k = (I_1^k, I_2^k, \ldots, I_L^k)$ denote the input vector, where I_l^k denotes input l to channel k, for $l = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, K$, where L denotes the length of $\mathbf{I}^k$ and K denotes the total number of input fields.

Input fields: Let F_1^k denote an input field that receives $\mathbf{I}^k$ and let $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_L^k)$ denote the activation vector of F_1^k receiving $\mathbf{I}^k$. Please note that normalization is performed on $\mathbf{I}^k$ to obtain $\mathbf{x}^k$, such that $\mathbf{x}_l^k \in [0, 1]$. If fuzzy ART operations (see (1) and (3)) [34] are used, $\mathbf{x}^k$ is further augmented with a complement vector $\overline{\mathbf{x}}^k$, where $\overline{x}_l^k = 1 - x_l^k$. This augmentation is named complement coding, which is applied to prevent the code proliferation problem [34]. As such, by applying fuzzy ART learning (see (3)), the learned codes in the category field represent more generalized associations [34].

Category field: Let F_2 denote the category field and let $\mathbf{y} = (y_1, y_2, \ldots, y_J)$ denote the activation vector of F_2, where J denotes the number of codes in F_2. Please note that there are always $J-1$ committed (learned) codes and one uncommitted (Jth) code in F_2. If ART learns from scratch, initially there is only one uncommitted code in F_2.

Weight vectors: Let $\mathbf{w}_j^k$ denote the weight vector of the jth code C_j in F_2 for learning the input patterns in F_1^k, where $j = 1, 2, \ldots, J$. In terms of clustering, $\mathbf{w}_j$ represents the template used to characterise the jth cluster.

Parameters: The dynamics of Fusion-ART are regulated by the parameters associated with each input field, namely choice parameters $\alpha^k > 0$, learning rate parameters $\beta^k \in [0, 1]$, contribution parameters $\gamma^k \in [0, 1]$, where $\sum \gamma^k = 1$, and vigilance parameters $\rho_j^k \in [0, 1]$. Please note that as aforementioned, in IFC-ART, the vigilance parameters are associated with each feature and each cluster.

As briefly introduced in Section I, ART involves a bottom-up processing of the external information and a top-down modulation of the internal knowledge. Specifically, the bottom-up processing consists of the *code activation* and *code competition* processes and the top-down modulation consists of the *template matching*, *template learning*, and *knowledge readout* processes. All these five processes are introduced as follows.

Code activation: Given $\{\mathbf{x}^k|_{k=1}^K\}$, for each F_2 code j, the corresponding activation T_j is computed as follows:

$$T_j = \sum_k \gamma^k \frac{|\mathbf{x}^k \wedge \mathbf{w}_j^k|}{\alpha^k + |\mathbf{w}_j^k|}, \tag{1}$$

where the fuzzy AND operation $\wedge$ is defined by $p_i \wedge q_i \equiv min(p_i, q_i)$ and the norm $|.|$ is defined by $|\mathbf{p}| \equiv \sum_i p_i$.

Code competition: Given $\{T_j|_{j=1}^J\}$, the F_2 code with the highest activation value is named the winner, which is indexed at j^*, where $j^* = \arg\max_j T_j$.

Template matching: Given the winner code C_{j^*}, the match between the input pattern and the weight vector of C_{j^*} is computed as follows:

$$M_{j^*}^k = \frac{|\mathbf{x}^k \wedge \mathbf{w}_{j^*}^k|}{|\mathbf{x}^k|}. \tag{2}$$

If C_{j^*} satisfies the vigilance criteria such that $\forall M_{j^*}^k \geq \rho_{j^*}^k$, a resonance occurs in which leads to the subsequent learning or readout process. Otherwise, a mismatch reset occurs in which $T_{j^*} \leftarrow 0$ until a resonance occurs at another F_2 code. This template matching process is guaranteed to end, because either a committed code that satisfies the vigilance criteria will be identified or an uncommitted one, whose weights are all 1s that definitely satisfies the criteria, will be recruited to encode the new input pattern. Once an uncommitted code is recruited, a new uncommitted code will be autonomously added in F_2. Thus, ART self-organizes its network structure.

Template learning: If learning is required, once found C_{j^*} that satisfies the vigilance criteria, its corresponding weight vectors are updated by the following learning rule:

$$\mathbf{w}_{j^*}^{k(\text{new})} = (1 - \beta^k)\mathbf{w}_{j^*}^{k(\text{old})} + \beta^k(\mathbf{x}^k \wedge \mathbf{w}_{j^*}^{k(\text{old})}). \tag{3}$$

Knowledge readout: If readout is required, C_{j^*} presents its weight vectors to the input fields, such that $\mathbf{x}^{k(\text{new})} = \mathbf{w}_{j^*}^k$.

In terms of clustering, the dynamics of Fusion-ART can be summarized as follows. Based on the similarity measures (see (1)), a winner cluster can be identified. If the input pattern

satisfies the vigilance criteria of the winner cluster (see (2)), it will be added into the identified cluster (see (3)). Otherwise, Fusion-ART will select another winner until the vigilance criteria are satisfied and learn accordingly. At the end of the autonomous clustering process, each committed code in the category field represents one formed cluster.

B. IFC-ART for Clustering with Focused Preferences

From the introduction of how Fusion-ART performs clustering, we can find that the autonomous clustering process is affected by the vigilance parameters to a great extent, which regulate the boundaries or regions of the clusters. For graphical demonstrations of the cluster boundaries in ART, readers may refer to [7]. To adaptively regulate the vigilance parameters while maintaining the focus on the interesting or preferred features, we introduce IFC-ART, which self-regulates the vigilance parameter associated with each feature and each cluster under a generic framework.

The dynamics of IFC-ART is summarized in Algorithm 1. Comparing to Fusion-ART, IFC-ART incorporates two more procedures to enable the self-regularization of the vigilance parameters during the autonomous clustering process, one refers to the identification of interesting features to incorporate domain knowledge and the other is called *match tracking* to reduce the overlap between cluster boundaries.

First of all, based on the basic understanding of the data set and the clustering task, we can identify the set of interesting features *IF* among the K number of input fields of IFC-ART. Then the remaining ones form the set of normal features *NF*, such that $|IF|+|NF| = K$. Given the initial vigilance parameter value $\rho_0 \in [0, 1]$, the vigilance of an uncommitted code, which always has the largest index in F_2, is computed as follows:

$$\rho_J^k = \begin{cases} \min\{(1+\phi)\rho_0, 1\}, & \text{if } k \in \textit{IF}, \\ \rho_0, & \text{otherwise}, \end{cases} \tag{4}$$

where ϕ refers to the magnifying parameter and $\phi \in (0, 1)$. As such, the difference between the interesting or preferred features and the normal ones is preserved during the initial formation of all clusters.

Furthermore, during the clustering process, the vigilance parameters are also self-regulated according to the learned weight vectors and the input patterns. Specifically, when a mismatch reset occurs during template matching, other than setting the activation value T_{j^*} to 0 during the presence of the current input pattern, for every input field that violates the vigilance criterion, the corresponding vigilance parameter is updated using the following equation:

$$\rho_{j^*}^{k(\text{new})} = M_{j^*}^k + \delta, \tag{5}$$

where δ is a significantly small number and $\delta > 0$. The regularization of the vigilance parameters using (5) is named match tracking [7], [12], [35]. The rational of match tracking is to minimize the conflict between the clusters to a minimum degree of δ. Specifically, matching tracking is applied when a mismatch reset occurs, which means that although the winner code j^* has the maximum activation value in response to the presented input pattern, j^* fails to fulfil the vigilance criteria. As such, to minimize the risk of incorrect categorization and to well maintain the cluster boundaries, we shrink the corresponding vigilance value according to the match value $M_{j^*}^k$ between the input pattern and the weight vector (see (2)). Match tracking has been shown to be effective as a regularization strategy used for clustering [7]. Therefore, Fusion-ART with match tracking, along with Fusion-ART and Fusion-ART with interesting features, are selected as the benchmarking ART models when conducting the case studies.

It is easy to identify from Algorithm 1 that the complexity of IFC-ART is $O(N \cdot J \cdot K)$, where N denotes the number of data samples in the data set. In Section IV of this paper, we report the execution time of IFC-ART and all the other benchmarking models for comparisons.

Algorithm 1 Interest-Focused Clustering based on ART

Require: $\rho_0 \in [0, 1]$, $\phi \in (0, 1)$, $\delta > 0$, $\alpha^k > 0$, $\beta^k \in [0, 1]$, $\gamma^k \in [0, 1]$, and $\sum \gamma^k = 1$, where $k = 1, 2, \ldots, K$ and K denotes the number of input fields

Compute ρ_1^k for the uncommitted code in F_2 {see (4)}
Initialize IFC-ART with K, α^k, β^k, γ^k, and ρ_1^k
for all $\{\mathbf{I}^k|_{k=1}^K\}$ in the given data set **do**
  obtain $\{\mathbf{x}^k, \overline{\mathbf{x}}^k|_{k=1}^K\}$, such that $\mathbf{x}_l^k \in [0, 1]$, where $l = 1, 2, \ldots, |\mathbf{I}^k|$, and present the pattern to F_1^k
  for all C_j in F_2, where $j = 1, 2, \ldots, J$ and J denotes the number of clusters in F_2 **do**
    compute T_j {see (1)}
  end for
  loop
    identify j^*, such that $j^* = \arg\max_j T_j$
    compute $M_{j^*}^k$ {see (2)}
    if $\forall M_{j^*}^k \geq \rho_{j^*}^k$ **then**
      exit loop
    else
      $T_{j^*} \leftarrow 0$
      for all $M_{j^*}^k < \rho_{j^*}^k$ **do**
        $\rho_{j^*}^{k(\text{new})} \leftarrow M_{j^*}^k + \delta$ {see (5)}
      end for
    end if
  end loop
  if $j^* = J$ {winner is an uncommitted cluster} **then**
    $\mathbf{w}_J^k = \{\mathbf{x}^k, \overline{\mathbf{x}}^k\}$
    $J \leftarrow J + 1$ {create a new uncommitted cluster}
    $\mathbf{w}_J^k \leftarrow (1, 1, \ldots, 1)$ and $\overline{\mathbf{w}}_J^k \leftarrow (1, 1, \ldots, 1)$
    assign ρ_J^k {see (4)}
  else
    update $\mathbf{w}_{j^*}^k$ {see (3)}
  end if
end for

IV. Case Studies on American Time Use Survey

As introduced in Sections I and II-B, the American Time Use Survey (ATUS) records 159,937 interviews conducted

with U.S. residents from 2003 to 2014. We select this real-world data set to assess the performance of IFC-ART.

In terms of the parameters used by all ART models in all the experiments, we vary the value of ρ_0 (see (4)) from 0.1 to 0.9 with an increment of 0.1 for performance comparisons. The other parameters are set as follows: $\alpha^k = 0.01$, which is mainly used to avoid NaN in (1); $\beta^k = 0.5$, which is the median between the fastest learning (i.e., $\beta^k = 1$) and no learning (i.e., $\beta^k = 0$); and $\gamma^k = 1/K$, which equally assigns the contribution of each feature given that the importance of the identified interesting features has already been reflected on their vigilance parameters. Moreover, for IFC-ART, we simply assign ϕ in (4) to γ^k and assign δ in (5) to 0.001.

Because there are no class labels given in the ATUS data set, we cannot evaluate the performance using accuracy and entropy types of measures. For the internal evaluation of the clustering results, we select the Davies–Bouldin Index (DBI) [36] as the measuring metric, which is defined by

$$DBI = \frac{1}{J}\sum_{i=1}^{J}\max_{i\neq j}\left(\frac{\sigma_i + \sigma_j}{d(m_{C_i}, m_{C_j})}\right), \tag{6}$$

where σ_i denotes the average distance of all elements in the ith cluster to its centroid m_{C_i} and $d()$ computes the Euclidean distance between the two vectors. DBI combines the measure of both intra-cluster similarity (numerator of the max term in (6)) and inter-cluster similarity (denominator of that term). A smaller DBI value suggests better performance.

For the ART benchmarking models, other than Fusion-ART (see Section III-A), we also select **Fusion-ART-wif** and **Fusion-ART-wmt**, which denote Fusion-ART with interesting features (i.e., Step 16 of Algorithm 1 is skipped) and Fusion-ART with match tracking (i.e., Step 1 of Algorithm 1 is skipped). For classical clustering techniques, we select K-means [37] and X-means [38] for comparisons. For the number of clusters pre-requested by K-means, we assign $J-1$, J, and $J+1$ to K, where J is the number of clusters obtained by IFC-ART when achieving the best DBI. Similarly, we assign the minimum number of clusters to $J-1$ and the maximum to $J+1$ in X-means. However, due to its characteristics, X-means may form less number of clusters than $J-1$. Then we run K-means and X-means for nine times with different initializations and record their best DBI, respectively. The reason why we did not select density based clustering technique such as DBSCAN [19] for comparisons is because it always forms a single cluster consisting of all data in all three case studies, even with a high epsilon value of 0.9 and the minimum number of data in one cluster being set to 1. Although we run the K-means and X-means algorithms implemented in WEKA [39], in this paper, we still compare their execution time with that of the ART models, which are implemented in JAVA, given that we run all the models on the same computer.

A. Relationship among Wage, Age, and Education Attainment

It is a common stereotype that people who earn more money are better educated and probably older. Based on the real-world cases recorded in ATUS, we study the relationship among wage, age, and education attainment and we select wage as the only interesting feature.

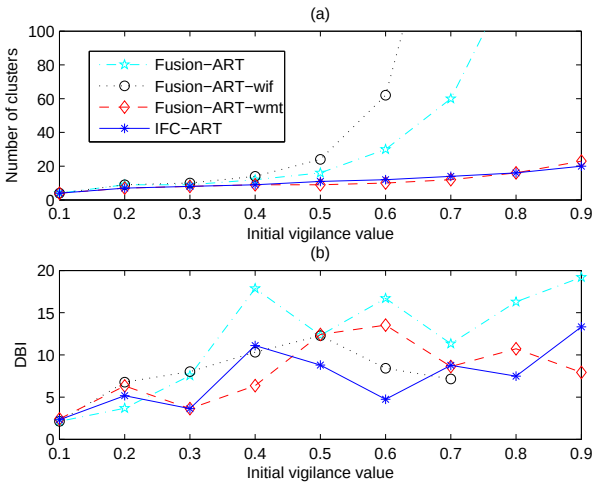

Fig. 2. Performance comparisons among the four ART models on the wage study: (a) number of clusters and (b) DBI.

TABLE I
PERFORMANCE COMPARISONS ON WAGE STUDY.

Model	Best DBI	J	Time (s)	Mean DBI
K-means	2.66	4	0.17	–
X-means	2.94	3	2.15	–
Fusion-ART	**2.13** (ρ_0=0.1)	4	0.09	11.90 (6.26)
Fusion-ART-wif	**2.13** (ρ_0=0.1)	4	0.09	–
Fusion-ART-wmt	2.41 (ρ_0=0.1)	4	0.09	8.00 (3.77)
IFC-ART	2.32 (ρ_0=0.1)	4	0.10	**7.26** (3.62)

ATUS records the weekly wages of the respondents if they are employees (excluding self-employments). Therefore, after removing all the not-applicable entries, we have 88,598 data samples in this case study. Furthermore, because the interviews were conducted over twelve years, we adjust the wages according to the yearly Commercial Price Index (CPI) [40], so that all wages after pre-processing are levelled on year 2014. The age of the respondents ranges from 15 to 85 and the education attainment ranges from less than first grade to doctorate degree in step-wise definitions. Therefore, for ART models, the inputs are normalized according to the minimum and maximum values in each feature. The performance comparisons of the ART models are shown in Fig. 2. Please note that for Fusion-ART-wif, we do not plot when ρ_0 = 0.8 and 0.9, because DBI is returned as "infinity" and J is extremely large (35,047 and 47,649, respectively). Moreover, without match tracking, the number of clusters obtained by Fusion-ART and Fusion-ART-wif increases exponentially when $\rho_0 > 0.5$.

Table I summarizes the best performance of all models and the averaged performance of ART models. Please note that DBI of every model is computed on the original feature values. The reason why X-means has a significantly longer execution time is because it is the only iterative clustering technique benchmarked in this paper. All the others only

TABLE II
CLUSTER REPRESENTATIONS OF IFC-ART ON THE WAGE STUDY.

ID	# data	Education attainment	Age	Weekly wage
1	5,298	College but no degree	29.98 (15.90)	346.19 (354.90)
2	14,678	Bachelor's degree	43.28 (10.46)	1736.84 (875.42)
3	24,961	High school diploma	37.21 (13.74)	649.97 (406.65)
4	43,661	Occupational degree	45.43 (11.33)	915.69 (598.73)

run for one iteration, i.e., for incremental ART models, the presented results show their capability to deal with a stream of data that are fed in an online manner for only once. For performance evaluation, although IFC-ART does not achieve the best DBI, it achieves the best averaged DBI.

The best clustering results obtained by IFC-ART is presented in Table II. The first cluster shown in Table II can be interpreted as "on average, those 30 year-old people who went to college but did not graduate with a degree earn 346 dollars per week". Because we select wage as the interesting feature, among the four obtained clusters, values in wage are relatively better separated when compared to those in the other two features. Clusters 2 and 4 show that the high earners do have higher education attainment and are relatively older. People in cluster 3 have the lowest education attainment, however, due to the age difference, they earn more than those in cluster 1, who have slightly higher education attainment on average.

B. Relationship among Age and Time Spent on Activities

Due to the completeness of ATUS, all the 159,937 data samples record the detailed activities performed by the respondents on the day before the interview. In this paper, for all the amount of time spent on various activities, we normalize them by 1440, which is the total number of minutes in one day. As such, all those features are levelled on the same scale.

In terms of the broad categories of the various activities used in this case study, we follow the categorization used in [21], namely work, compulsory, and leisure activities. Specifically, work related activities include "work & work-related activities", "education", and "travelling", self-compulsory activities include "personal care activities", "household activities", "consumer purchases", "government services & civic obligations", and "eating & drinking", and leisure activities include "socializing, relaxing & leisure", "sports, exercise & recreation", "religious & spiritual activities", and "volunteer activities". Other than the time spent on these three types of activities, age is also used in this case study as the only interesting feature.

The performance comparisons among the ART models are illustrated in Fig. 3 and comparisons among all models are presented in Table III. It took us by surprise that both K-means and X-means achieve much better DBI than the ART models, unlike in the other two case studies (see Tables I and V). One possible reason is that both K-means and X-means only optimize the intra-cluster distance without considering the inter-cluster distance and the overlapping of cluster boundaries. Therefore, in this particular case study, their optimization strategy may well fit the distribution of this big-volume data set and thus they obtain better DBI. Moreover, among all

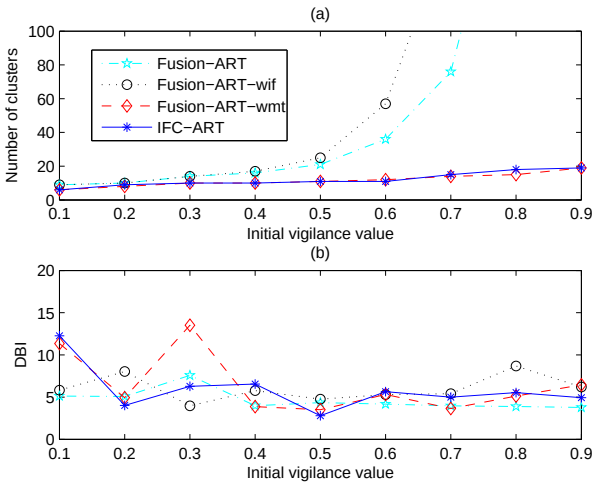

Fig. 3. Performance comparisons among the four ART models on the age study: (a) number of clusters and (b) DBI.

TABLE III
PERFORMANCE COMPARISONS ON AGE STUDY.

Model	Best DBI	J	Time (s)	Mean DBI
K-means	1.02	11	1.15	–
X-means	**0.96**	10	19.68	–
Fusion-ART	3.78 (ρ_0=0.9)	1381	10.61	**4.66** (1.20)
Fusion-ART-wif	3.96 (ρ_0=0.3)	14	0.17	5.58 (1.26)
Fusion-ART-wmt	3.51 (ρ_0=0.5)	11	0.17	6.41 (3.58)
IFC-ART	2.80 (ρ_0=0.5)	11	0.17	5.88 (2.65)

TABLE IV
CLUSTER REPRESENTATIONS OF IFC-ART ON THE AGE STUDY.

ID	# data	Age	Work	Compulsory	Leisure
1	4,252	51.52 (8.47)	503.68 (241.48)	624.34 (174.12)	242.23 (165.06)
2	9,204	31.07 (8.06)	413.83 (275.95)	708.11 (195.54)	229.28 (157.80)
3	3,433	39.28 (12.13)	834.75 (177.58)	480.84 (140.03)	98.04 (115.33)
4	51,082	60.66 (10.95)	191.66 (221.61)	891.64 (194.60)	296.16 (146.81)
5	21,543	64.18 (11.92)	54.82 (68.61)	674.20 (117.72)	679.93 (127.58)
6	21,549	27.94 (7.99)	178.98 (181.03)	705.52 (141.65)	485.23 (160.33)
7	44,473	33.72 (8.42)	303.31 (257.28)	837.18 (203.15)	204.39 (134.04)
8	9	46.89 (11.67)	100.00 (100.69)	91.22 (55.31)	57.78 (71.92)
9	4,387	58.04 (6.64)	592.17 (134.16)	634.43 (110.38)	186.35 (111.38)
10	4	21.00 (1.73)	18.5 (24.63)	95.5 (60.85)	1326.00 (79.16)
11	1	85 (–)	870 (–)	570 (–)	0.00 (–)

Note: Clusters 8, 10, and 11 are correctly identified outliers.

the ART models, although IFC-ART achieves the best DBI, Fusion-ART and Fusion-ART-wif perform better in terms of the averaged DBI. This is because when ρ_0 is large, both the two models generate an unnecessarily large number of clusters, which leads to the non-proportionally decrease in DBI (i.e., the $\frac{1}{J}$ coefficient in (6)). For example, when Fusion-ART obtains the best DBI, it generates 1,381 clusters (see Table III).

The best clustering results obtained by IFC-ART is presented in Table IV. The first cluster shown in Table IV can be interpreted as "on average, people aged around 51.5 spend 504, 624, and 242 minutes per day on work-related, compulsory, and leisure activities, respectively". Please note that the three features on time spent do not always add up to 1440 minutes because there are other activities not included in this study, such as telephone calls, time spent on enjoying

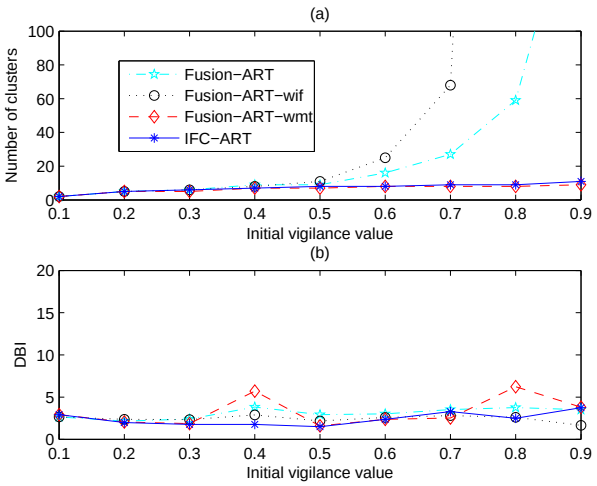

Fig. 4. Performance comparisons among the four ART models on the elderly care study: (a) number of clusters and (b) DBI.

TABLE V
PERFORMANCE COMPARISONS ON ELDERLY CARE STUDY.

Model	Best DBI	J	Time (s)	Mean DBI
K-means	3.58	8	0.09	–
X-means	2.51	5	0.57	–
Fusion-ART	2.16 (ρ_0=0.2)	5	0.06	3.09 (0.60)
Fusion-ART-wif	1.66 (ρ_0=0.5)	11	0.07	2.46 (0.38)
Fusion-ART-wmt	1.58 (ρ_0=0.5)	7	0.06	3.23 (1.69)
IFC-ART	**1.50** (ρ_0=0.5)	8	0.07	**2.43** (0.76)

others' services, and unlabelled activities. It is interesting to learn that people around 40 year-old (see cluster 3) spend the most amount of time on their work and has very little time for leisure activities. On the contrary, people around 64 year-old (see cluster 5) spend very little time on their work and very long time on leisure activities. More encouragingly, IFC-ART correctly identifies some outliers, such as the 85 year-old elder (see cluster 11) still spent more than half a day on work and no time on leisure activity, but a small group of young people (see cluster 10) spent nearly all their time on leisure activities.

C. Relationship among Elderly Care, Child Care and Age

ATUS started to record the time of the respondents spent on elderly care from 2011 (47,899 data samples). In this case study, we investigate the relationship between the different amount of time spent by people in all ages on child ($<$ 15 year-old) care and elderly ($\geq$ 65 year-old) care and we select the time spent on elderly care as the only interesting feature.

The performance comparisons among the ART models are illustrated in Fig. 4 and comparisons among all models are presented in Table V. It is reassuring to learn that IFC-ART performs best in terms of both the best and averaged DBI.

The best clustering results obtained by IFC-ART is presented in Table VI. The first cluster shown in Table VI can be interpreted as "on average, people aged around 64.4 spend 26.3 and 7.3 minutes on child care and elderly care in one day, respectively". As expected, we may observe from Table VI that

TABLE VI
CLUSTER REPRESENTATIONS OF IFC-ART ON THE ELDERLY CARE STUDY.

ID	# data	Age	Child care	Elderly care
1	21,576	64.42 (9.92)	26.32 (103.41)	7.28 (43.29)
2	19,060	33.78 (10.28)	74.55 (144.28)	4.75 (42.23)
3	1,672	44.04 (7.91)	733.82 (118.31)	10.13 (66.44)
4	138	65.64 (10.53)	5.26 (30.83)	861.14 (115.48)
5	5,449	34.11 (6.57)	546.26 (199.82)	0.69 (7.28)
6	2	27.50 (2.50)	238.5 (238.5)	1092.5 (27.50)
7	1	23 (–)	1020 (–)	0 (–)
8	1	67 (–)	0 (–)	1175 (–)

Note: Clusters 6, 7, and 8 are correctly identified outliers.

people spent much less time on elderly care than child care. In particular, only 141 people (see clusters 4, 6, and 8) out of 47,899 (approximately 0.3%) spent a long period of time on elderly care. When looking into the date set, we find that these 141 people provided the elderly care either as a family member or as a professional caregiver. Moreover, out of the 47,899 respondents, only 8,517 (17.8%) ever spent time on elderly care on the day before the interview and only 1,114 (2.3%) lived in the same household with the elderly. All these findings suggest that there is a vast amount of need to support the independence of the elderly [41].

D. Discussions on Performance Comparisons

Based on the results of the case studies, match tracking is shown to be an effective method to self-regulate the number of obtained clusters. In contrast, Fusion-ART models without match tracking often generate an unnecessarily large number of clusters. The fact that IFC-ART outperforms Fusion-ART-wmt in terms of both the best and averaged DBI in all the three case studies (see Tables I, III, and V) suggests that by introducing the interesting features merely based on the basic understanding of the data set, the intrinsic structure of the data set can be better discovered. In summary, IFC-ART performs better than the other ART models, because it always generates a reasonable number of clusters across different initial vigilance values and always performs better than Fusion-ART-wmt in terms of DBI.

V. CONCLUSION

In this paper, we introduce a self-regulated incremental clustering technique named IFC-ART, which performs better than the other ART models and the classical ones (for majority cases). All the initial parameters used by IFC-ART are set to standard or intuitive values and IFC-ART self-regulates the vigilance parameters associated with each input field and each formed cluster during the autonomous clustering process. By introducing the interesting features to the cluster formation process merely based on the basic understanding of the data set, not only we may better control the type of knowledge we tend to discover but also we improve the clustering performance (comparing the performance of IFC-ART with that of Fusion-ART-wmt).

Going forward, on one hand, we will further investigate and improve the dynamics of IFC-ART. On the other hand, we will design and conduct more case studies using ATUS.

ACKNOWLEDGMENT

This work is supported in part by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 (Grant No. RG 137/14) and National Research Foundation, Prime Ministers Office, Singapore under its IDM Futures Funding Initiative and Competitive Research Grant (Grant No. NRF-CRP8-2011-05). We thank Sok-Wei Kwan for the pre-processing of the ATUS data set.

REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley, 2001.

[2] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. Wiley, 2011.

[3] R. Xu and D. Wunsch II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[4] G. A. Carpenter and S. Grossberg, "Adaptive Resonance Theory," in *The Handbook of Brain Theory and Neural Networks*. MIT Press, 2003, pp. 87–90.

[5] A.-H. Tan, G. A. Carpenter, and S. Grossberg, "Intelligence through interaction: Towards a unified theory for learning," in *Proceedings of International Symposium on Neural Networks*, 2007, pp. 1094–1103.

[6] L. Meng, A.-H. Tan, and D. Xu, "Semi-supervised heterogeneous fusion for multimedia data co-clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2293–2306, 2014.

[7] L. Meng, A.-H. Tan, and D. Wunsch, "Adaptive scaling of cluster boundaries for large-scale social media data clustering," *IEEE Transactions on Neural Networks and Learning Systems*, 2015, in press.

[8] D. Wang and A.-H. Tan, "Creating autonomous adaptive agent in a real-time first-person shooter computer game," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 2, pp. 123–138, 2015.

[9] T.-H. Teng, A.-H. Tan, and J. Zurada, "Self-organizing neural networks integrating domain knowledge and reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 889–902, 2015.

[10] A.-H. Tan, "Adaptive resonance associative map," *Neural Networks*, vol. 8, no. 3, pp. 437–446, 1995.

[11] J. Zhou and S. Bennett, "A supervised learning network based on Adaptive Resonance Theory," *International Journal of Neural Systems*, vol. 8, no. 2, pp. 239–246, 1997.

[12] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural networks*, vol. 4, no. 5, pp. 565–588, 1991.

[13] J. He, A.-H. Tan, and C.-L. Tan, "ART-C: A neural architecture for efficient on-line clustering under constraints," in *Proceedings of International Joint Conference on Neural Networks*, 2002, pp. 2550–2555.

[14] ——, "Modified ART 2A growing network capable of generating a fixed number of nodes," *IEEE Transactions on Neural Networks*, vol. 5, no. 3, pp. 728–737, 2004.

[15] ATUS, American Time Use Survey. [Online]. Available: http://www.bls.gov/tus/

[16] Q. Song and N. Kasabov, "ECM: A novel on-line, evolving clustering method and its applications," in *Proceedings of Conference on Artificial Neural Networks and Expert Systems*, 2001, pp. 87–92.

[17] D. Wang, C. Quek, and G. S. Ng, "MS-TSKfnn: Novel Takagi-Sugeno-Kang fuzzy neural network using art like clustering," in *Proceedings of International Joint Conference on Neural Networks*, 2004, pp. 2361–2366.

[18] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.

[20] D. Wang, G. S. Ng, and C. Quek, "A novel hybrid intelligent system: Genetic algorithm and rough set incorporated neural fuzzy inference system," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2008, pp. 2546–2553.

[21] D. Kahneman, A. B. Krueger, D. Schkade, N. Schwarz, and A. A. Stone, "Would you be happier if you were richer? A focusing illusion," *Science*, vol. 312, pp. 1908–1910, 2006.

[22] J. Hersch, "Home production and wages: Evidence from the American Time Use Survey," *Review of Economics of the Household*, vol. 7, no. 2, pp. 159–178, 2009.

[23] K. Kushlev, E. W. Dunn, and R. E. Lucas, "Higher income is associated with less daily sadness but not more daily happiness," *Social Psychological and Personality Science*, vol. 6, no. 5, pp. 483–489, 2015.

[24] C. D. Zick, K. R. Smith, B. B. Brown, J. X. Fan, and L. Kowaleski-Jones, "Physical activity during the transition from adolescence to adulthood," *Journal of Physical Activity and Health*, vol. 4, pp. 125–137, 2007.

[25] H. Kang and Y. L. Michael, "Social integration: How is it related to self-rated health?" *Advances in Aging Research*, vol. 2, no. 1, pp. 10–20, 2013.

[26] P. Dolan and L. Kudrna, "More years, less yawns: Fresh evidence on tiredness by age and other factors," *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, vol. 70, no. 4, pp. 576–580, 2013.

[27] M. Basner, K. M. Fomberstein, F. M. Razavi, S. Banks, J. H. William, R. R. Rosa, and D. F. Dinges, "American Time Use Survey: Sleep time and its relationship to waking activities," *Sleep*, vol. 30, no. 9, pp. 1085–1095, 2007.

[28] L. B. Russell, Y. Ibuka, and K. G. Abraham, "Health-related activities in the American Time Use Survey," *Medical Care*, vol. 45, no. 7, pp. 680–685, 2007.

[29] D. S. Hamermesh, "Incentives, time use and BMI: The roles of eating, grazing and goods," *Economics and Human Biology*, vol. 8, no. 1, pp. 2–15, 2010.

[30] M. Podor and T. J. Halliday, "Health status and the allocation of time," *Health Economics*, vol. 21, no. 5, pp. 514–527, 2012.

[31] J. Kimmel and R. Connelly, "Mothers' time choices: Caregiving, leisure, home production, and paid work," *Journal of Human Resources*, vol. 42, no. 3, pp. 643–681, 2007.

[32] J. Guryan, E. Hurst, and M. Kearney, "Parental education and parental time with children," *Journal of Economic Perspectives*, vol. 22, no. 3, pp. 23–46, 2008.

[33] C. Tudor-Locke, W. D. Johnson, and P. T. Katzmarzyk, "Frequently reported activities by intensity for U.S. adults: The American Time Use Survey," *American Journal of Preventive Medicine*, vol. 39, no. 4, pp. e13–e20, 2010.

[34] 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.

[35] 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 Transaction on Neural Networks*, vol. 3, no. 5, pp. 698–713, 1992.

[36] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[37] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *Journal of the Royal Statistical Society, Series C*, vol. 28, no. 1, pp. 100–108, 1979.

[38] D. Pelleg and A. W. Moore, "X-means: Extending K-means with efficient estimation of the number of clusters," in *Proceedings of International Conference on Machine Learning*, 2000, pp. 727–734.

[39] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[40] CPI, Consumer Price Index. [Online]. Available: http://www.bls.gov/cpi/

[41] D. Wang, B. Subagdja, Y. Kang, A.-H. Tan, and D. Zhang, "Towards intelligent caring agents for aging-in-place: Issues and challenges," in *Proceedings of IEEE Symposiumon on Computational Intelligence for Human-Like Intelligence*, 2014, pp. 1–8.