# Incorporating Association Patterns into Manifold Clustering for Enabling Predictive Analytics[1]

Bon K. Sy
SIPPA Solutions
Queens College & Graduate Center
CUNY, Queens, NY, USA
e-mail: bsy@sippasolutions.com

Jin Chen
SIPPA Solutions
Queens, NY, USA
e-mail: desmood0231@gmail.com

Rebecca Horowitz
Dept. of Computer Science
Queens College, CUNY
Queens, NY, USA
e-mail: rebecca.horowitz@outlook.com

*Abstract*—**The goal of this research is to develop a predictive analytics technique based on manifold clustering of mixed data type. In this research, we explore the concept of statistically significant association patterns to induce an initial partition on data for deriving manifolds. Manifolds are hyperplanes embedded in low dimensions. The advantage of this novel technique is a bootstrap on data clusters that reveals statistical associations from the information-theoretic perspective. As an illustration, the proposed technique is applied to a real data set of diabetes patients. An assessment on the proposed technique is performed to investigate the effect of bootstrap based on association patterns. Results of the preliminary study demonstrate the feasibility of applying the proposed technique to real-world data.**

***Keywords-Manifold clustering; Association patterns; Predictive analytics***

## I. INTRODUCTION

Prediction techniques such as linear regression and Principal Component Analysis (PCA) [1] rely on the linearity of the data of Real in the dimensions that the data reside. These techniques work well when the data distribution exhibits linearity. On the other hand, information-theoretic based techniques such as ID3 [2], utilize an entropy reduction concept for deriving a decision tree that maximizes information gain in each traversal step of the decision tree. Such a technique, although exponential with respect to the enumeration of the finite discrete types of multi-dimensional data, can be effective when the data distribution lends itself to rapid pruning of impossible cases.

A hyperplane is defined by a cluster of a data subset, which is not necessarily linear. Manifold clustering provides a means to discover data subsets that can be projected onto hyperplanes embedded in low dimensions. In contrast to techniques such as PCA, manifold clustering does not rely on the linearity of the data subset. However, manifold clustering techniques, such as Spectral Clustering [3], suffer from two limitations. First, it is sensitive to the initial seeding of the clusters and often requires a 2-phase approach [4]. Second, it cannot handle a data set composed of data with mixed data types, for example, data of Real (e.g., floating point) and data of finite discrete type (e.g., binary valued data).

This research addresses these two limitations via the following approach:

1. Data of Real will be discretized via an entropy approach that optimizes the trade-off between information loss and the granularity of the discrete representation of the information carried by the data of Real. The discrete representation of the data of Real then enables the discovery of statistically significant association patterns, which will be detailed at a later section.
2. Each of the statistically significant association patterns then induces an initial cluster for aggregating data within the proximity that characterizes the hyperplane embedded in the low dimension. This initial cluster then serves as initial seeding for clustering when applying techniques such as spectral clustering and allows a semantic interpretation from the information-theoretic perspective.

In section II a brief overview on the state-of-state on manifold clustering is presented. In section III the notations and definitions required for presenting this manifold-based predictive analytics technique will be given. The algorithmic steps will be detailed in section IV. A use case of this approach on a real-world pilot study for assisting a diabetes individual on self-health management will be described in section V, followed by the results obtained from applying the proposed approach in section VI. This paper will then be concluded with insights on future research direction.

## II. STATE-OF-THE-ART MANIFOLD CLUSTERING

Many researchers have proposed clustering algorithms to avoid the issue of linearity, but each comes with its own disadvantages. K-means, suggested by MacQueen in 1967 is

---

[1] This research is conducted under the support of NSF 1831214 to SIPPA Solutions LLC, with a subcontract to CUNY underwritten by SIPPA Solutions LLC.

978-1-7281-5584-5/19/$31.00 ©2019 IEEE
DOI 10.1109/CSCI49370.2019.00243

one of the most popular algorithms due to its *O(n²)* complexity. The algorithm consists of selecting the first $k$ data points to be the centers of $k$ clusters and finding the minimum arithmetic mean between each data point and the $k$ clusters [5]. However, k-means breaks down in higher dimensions. Zhang and Kwok suggest using an applied Nyström method to approximate the eigen-decomposition with low rank kernel matrices [6]. Alternatively, Wang et. al suggest using local adaptive learning to perform graph embedding and k-means simultaneously, thereby reducing dimensionality [7]. Both algorithms decrease run-time of typical clustering methods, but do not address the information lost in the dimension reduction process. Our proposed approach allows one to investigate the effect of dimension reduction on information loss from an information-theoretic perspective, as well as from a reconstruction error perspective during the projection of a data point to a hyperplane of a cluster.

Recent clustering research focuses on minimizing dimensionality without losing meaning in the data. Gong et. al propose using a structured sparse k-means algorithm to reduce the randomness of clusters. In doing so, they used Laplacian smoothing to exploit the correlation information among features, thereby improving clustering accuracy and retaining meaning [8]. Faivishevsky and Goldberger take a different approach by combining spectral clustering with a nonparametric information theoretic clustering algorithm to retain information via mutual information measure [9]. However, their algorithm assumes that the conditional density of each cluster follows a Gaussian distribution. Our approach differs from Faivishevsky's in that our approach does not make an assumption on Gaussian distribution, but rather the assumption on asymptotic convergence of mutual information measure towards Chi-Square as Kullback [10] has proven for the low-dimension, and extensible to high dimension [11].

Other researchers consider alternative methods for manifold clustering. Ge et. al suggest a geometrically local embedding (GLE) process that reduces dimensionality by assigning clusters according to geometric distance in the higher dimension. After finding optimal reconstruction weights, the algorithm filters for outliers and the manifold is mapped to a lower dimension [12]. Although GLE is effective, the procedure is computationally slow and may be too extensive for large datasets. While there is a large body of knowledge to advance research on manifold clustering, it remains scarce on their practical applications in the real world. This research intends to contribute to a better understanding of our proposed approach to solve a real-world problem on assisting diabetes patients to better self-manage their disease condition using predictive analytics to assess their behavior readiness.

## III. NOTATION, DEFINITION, AND PROBLEM FORMULATION

Let $\boldsymbol{X} = \{\boldsymbol{X}_i \mid \boldsymbol{X}_i = [X_i^1 \quad \cdots \quad X_i^n]^T \in \mathbb{R}^n \; for \; i = 1, \dots, N\}$ be a data set of Real.

Let $\boldsymbol{Y} = \{\boldsymbol{Y}_i \mid \boldsymbol{Y}_i = [Y_i^1 \quad \cdots \quad Y_i^n]^T \in \mathbb{Z}^n \; for \; Y_i^j = 0, \dots, K-1 < N; \; i = 1, \dots, K \le N\}$ be a data set of discrete non-negative integers.

Let $\boldsymbol{M} = \{M_k \mid M_k \subseteq \boldsymbol{X} \; for \; k = 1, \dots, |\boldsymbol{M}|\}$ be the set of $|\boldsymbol{M}|$ manifold clusters.

Let $F: X^j \to Y^j$ be a one-on-one bijective mapping function that defines the discretization of the multivariate data set $\boldsymbol{X}$.

Let $S(M_k) = \{P_j^{k,o} \mid Given \; M_k, \; P_j^{k,o} = (val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o})$ for $j = 1, \dots, |S(M_k)|\}$. $P_j^{k,o}$ is an $o^{th}$ $(2 \le o \le n)$ order statistically significant association pattern [11] when $Pr(val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o}) > \alpha$ for some predefined threshold $\alpha$, and $MI(val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o}) \to \chi^2$ as defined below:

$$MI(val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o}) \to \left(\frac{1}{Pr(val_{j,1}^{k,o} \; val_{j,2}^{k,o} \dots val_{j,o}^{k,o})}\right)\left(\frac{\chi^2}{2N}\right)^{\left(\frac{\hat{E}}{E\prime}\right)^{o/2}} \tag{1}$$

where $MI(val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o}) = Log_2 \frac{Pr(val_{j,1}^{k,o}, \dots, val_{j,o}^{k,o})}{Pr(val_{j,1}^{k,o}) Pr(val_{j,2}^{k,o}) \dots Pr(val_{j,o}^{k,o})}$

and

$N$ is the sample size
$\chi^2$ is the Pearson chi-square defined as $(o_i - e_i)^2 / e_i$
$\hat{E}$ is the expected entropy measure
$E'$ is the maximum possible entropy

### *A. Definition 1*

The scope coverage $SC(P_j^{k,o})$, with respect to a set $\boldsymbol{Y}$, is defined as a subset of $\boldsymbol{Y}$ in which the semantic interpretation of the existence of $P_j^{k,o}$ is always true.

*Example*

Let $\boldsymbol{Y} = \{ [d1{:}\,0, \; d2{:}\,0, \; d3{:}\,0, \; d4{:}\,0],$
$[d1{:}\,0, \; d2{:}\,0, \; d3{:}\,0, \; d4{:}\,1], \dots, [d1{:}\,1, \; d2{:}\,1, \; d3{:}\,1, \; d4{:}\,1]\}$; i.e., $|\boldsymbol{Y}| = 16$.
Let $P_j^{k,o} = [d1{:}\,1, \; d3{:}\,0]$.
Then, $SC(P_j^{k,o}) = \{[d1{:}\,1, \; d2{:}\,0, \; d3{:}\,0, \; d4{:}\,0],$
$[d1{:}\,1, \; d2{:}\,0, \; d3{:}\,0, \; d4{:}\,1], [d1{:}\,1, \; d2{:}\,1, \; d3{:}\,0, \; d4{:}\,0],$
$[d1{:}\,1, \; d2{:}\,1, \; d3{:}\,0, \; d4{:}\,1]\}$

### *B. Definition 2*

The membership function $f(P_q^{k,o}, P_j^{k\prime,o\prime}) \to [0,1]$ is defined by the geometric mean measure below:

$$f\left(P_q^{k,o}, {P_j}^{k',o'}\right) = \sqrt{\frac{\left|SC(P_q^{k,o}) \cap SC\left(P_j^{k\prime,o\prime}\right)\right|}{\left|SC(P_q^{k,o})\right|} \times \frac{\left|SC(P_q^{k,o}) \cap SC\left(P_j^{k\prime,o\prime}\right)\right|}{\left|SC\left(P_j^{k\prime,o\prime}\right)\right|}} \tag{2}$$

$P_j^{k\prime,o\prime}$ is a member of the *k<sup></sup>*th manifold induced by $P_q^{k,o}$ when $k = ArgMax_{q,k}\, f\left(P_q^{k,o},\ P_j^{k\prime,o\prime}\right)$; where $P_q^{k,o}$ is a statistically significant association pattern including the $k^{th}$ manifold.

*Example*

Let's assume $P_1^{1,2} = [d1\colon 1,\ d3\colon 0]$
$P_1^{2,4} = [d1\colon 0, \mathrm{d}2\colon 1\ d3\colon 1\ \mathrm{d}4\colon 1]$
$P_2^{2,3} = [d1\colon 0,\ d3\colon 1, d4\colon 1]$
$P_1^{k\prime,3} = [d1\colon 1, d3\colon 0, d4\colon 1]$

The following terms could be derived based on the definitions:

$$
\begin{aligned}
SC\left(P_1^{1,2}\right) &= \{[d1\colon 1,\ d2\colon 0,\ d3\colon 0\ d4\colon 0],\\
&\qquad [d1\colon 1,\ d2\colon 0,\ d3\colon 0,\ d4\colon 1],\\
&\qquad [d1\colon 1,\ d2\colon 1,\ d3\colon 0,\ d4\colon 0],\\
&\qquad [d1\colon 1,\ d2\colon 1,\ d3\colon 0,\ d4\colon 1]\}\\
SC\left(P_1^{2,4}\right) &= [d1\colon 0,\ d2\colon 1,\ d3\colon 1\ d4\colon 1]\\
SC\left(P_2^{2,3}\right) &= \{[d1\colon 1,\ d2\colon 0,\ d3\colon 1\ d4\colon 1],\\
&\qquad [d1\colon 1,\ d2\colon 1,\ d3\colon 1,\ d4\colon 1]\}\\
SC\left(P_1^{k\prime,3}\right) &= \{[d1\colon 1,\ d2\colon 0,\ d3\colon 0\ d4\colon 1],\\
&\qquad [d1\colon 1,\ d2\colon 1,\ d3\colon 0,\ d4\colon 1]\}
\end{aligned}
$$

$$
\begin{aligned}
f\left({P_1}^{1,2}, P_1^{k\prime,3}\right) &= \sqrt{\left(\tfrac{2}{4}\right)\left(\tfrac{2}{2}\right)}\\
f\left({P_1}^{2,4}, P_1^{k\prime,3}\right) &= \sqrt{\left(\tfrac{0}{2}\right)\left(\tfrac{0}{2}\right)}\\
f\left({P_2}^{2,3}, P_1^{k\prime,3}\right) &= \sqrt{\left(\tfrac{0}{2}\right)\left(\tfrac{0}{2}\right)}
\end{aligned}
$$

$P_3^{k\prime,3}$ is member of the *1<sup></sup>*st manifold induced by $P_1^{1,2}$ because $ArgMax_{q,k}\, f\left(P_q^{k,o},\ P_3^{k\prime,3}\right) = 1$.

## IV. MANIFOLD CLUSTERING BASED ON ASSOCIATION PATTERNS

The significance of the proposed technique is the ability to provide a semantic meaning on the clusters based on association patterns. Specifically, each cluster is a collection of data that are "closest" to a statistically significant association pattern in terms of semantic similarity as measured by the membership function. By referencing the definition of statistically significant association pattern [11], such a pattern manifests a frequent occurrence as defined by the support measure exceeding a predefined threshold, as well as an inter-relationship among the underlying variables of the pattern.

Manifold clustering based on association patterns is comprised of four tasks. First, deriving the corresponding discrete data representation of a given set of data of Real. Second, identifying the statistically significant association patterns of the discrete mapping. Third, assigning each discretely represented data point of real to a cluster based on the evaluation of the membership function against every statistically significant association pattern. Fourth, deriving the data clustering on manifold by minimizing reconstruction error.

### *4.1) Deriving Discrete Data Representation of Data of Real*

Consider a discrete variable *Y* of *N* possible states, the entropy of a system defined by *Y*:

$$
\begin{aligned}
H_N(P_1 \ldots P_N) &= \textstyle\sum_{i=1}^{N} -Pr\ (Y = y_i)\ Log_2 Pr\ (Y = y_i)\\
&= \textstyle\sum_{i=1}^{N} -P_i Log_2 P_i \qquad (3)
\end{aligned}
$$

It can be shown that the following equality holds [11]:

$$
\begin{aligned}
H_N(P_1 \ldots P_N) &= H_{N-1}(P_1 + P_2, P_3\ \ldots\ P_N)\\
&\quad + (P_1 + P_2)\ H_2\left(\tfrac{P_1}{P_1+P_2} + \tfrac{P_2}{P_1+P_2}\right) \qquad (4)
\end{aligned}
$$

In the quantization process, combining two terms will reduce the number of terms by one, while resulting in an information loss amounting to the second term on the right-hand side of equation (4). The quantization of a data set of Real proposed in this paper will utilize the above entropy equations to incrementally combine terms until it reaches the inflection point where there is a change of direction on the rate of change of information loss. The details of the algorithm are shown below.

Let $\boldsymbol{X} = \{\boldsymbol{X}_i \mid \boldsymbol{X}_i = [X_i^1 \quad \cdots \quad X_i^n]^T \in \mathbb{R}^n\ for\ i = 1, \ldots, N\}$ be a data set of Real. For each dimension *j = 1,…,n* of ***X***, perform the following steps for the data of the $j^{th}$ dimension:

Step 1: Order $X_i^j$ in an ascending order. Create a bin for each term in $X^j$. Treat each bin as a state of a discrete variable of ***Y*** and associate a value for a bin equal to the mean of its term(s). In other words, $Y^j$ is a discrete variable of *N* states. If the values of $X_i^j$ are all different, the initial distribution of $Y^j$ is then even and the probability of $Y_i^j$ is equal to *1/N*.

Step 2: Initialize an iteration count *C = 1*. Derive the entropy $H_N(P_1 \ldots P_N)$ and record it as $H_N^C$.

Step 3: Increment the iteration count by *1*. Identify two adjacent bins, *l* and *l+1* in the ordered list where the difference between the mean of the terms in the $l^{th}$ and $(l+1)^{th}$ bins is the smallest. Merge the two adjacent $l^{th}$ and $(l+1)^{th}$ bins via arithmetic mean and update the probability distribution of $Y^j$. Re-derive the entropy $H_{N-1}^{C+1}$. Record the information loss $I^{C+1}$ (i.e., the 2nd term in equation (4)) from combining the two terms.

Step 4: Repeat step 3 until the direction on the rate of change of $I^{C+k}$ is changed. When this occurs at the $k^{th}$ iteration, the following result is obtained:

$$
\begin{aligned}
\boldsymbol{Y} = \{\boldsymbol{Y}_i \mid \boldsymbol{Y}_i = [Y_i^1 \quad \cdots \quad Y_i^n]^T \in \mathbb{Z}^n\ for\ Y_i^j &=\\
0, \ldots, K-1 < N;\ i = 1, \ldots, K \le N\}&
\end{aligned}
$$

A one-on-one bijective mapping function $F\colon X^j \rightarrow Y^j$ can then be defined for discretizing ***X*** to ***Y***.

### *4.2) Algorithm for Deriving Data Clustering on Manifolds*

Given $\boldsymbol{X}$, $\boldsymbol{Y}$, $F$, and a predefined error threshold δ, the algorithm for the proposed manifold clustering based on statistically significant association patterns is shown below:

Step 1: Based on $\boldsymbol{Y}$, derive the set of statistically significant association patterns $S(M_k)$.

Step 2: Define $|S(M_k)|$ disjoint clusters such that initially each cluster has one and only one statistically significant association pattern. Let $W$ be the set of clusters "*holding*" the data points in $\boldsymbol{X}$; i.e., $W = \{\boldsymbol{X}^{n,j} | \boldsymbol{X} = \cup_j \boldsymbol{X}^{n,j} \text{ for } k = 1, \ldots, |S(M_k)|\}$.

Step 3: Partition $\boldsymbol{X}$ by assigning each data point $X_i$ to the cluster $\boldsymbol{X}^{n,j}$ if $ArgMax_{q,k}\, f\left(F(X_i), P_q^{k,o}\right) = k$; where $P_q^{k,o}$ is a pattern that defines the cluster $\boldsymbol{X}^{n,j}$. If $f\left(F(X_i), P_q^{k,o}\right)$ is zero in all cases, $X_i$ is assigned to a non-semantic cluster *NS*.

Step 4: Let $S = \{S_j \mid j = 1, \ldots, |S(M_k)|\}$ be the set of subspaces corresponding to the clusters defined in step 2. Repeat the following for each $j$ where the corresponding cluster has more than one element:

Let $\boldsymbol{D}^{n,j} = \{d_k^{n,j} \mid k = 1, \ldots, |\boldsymbol{X}^{n,j}|\}$ be the data set of the cluster $\boldsymbol{X}^{n,j}$. The subspace $S_j$ corresponding to $\boldsymbol{X}^{n,j}$ is then derived based on the following:

Step 4.1: Derive the mean vector and variance matrix of $\boldsymbol{D}^{n,j}$ for each $j = 1, \ldots, |S(M_k)|$ i.e., $\boldsymbol{A}^{n,j} = \frac{1}{|\boldsymbol{D}^{n,j}|} \sum_{k=1}^{|\boldsymbol{D}_k^{n,j}|} (d_k^{n,j} - \mu^{n,j})\,(d_k^{n,j} - \mu^{n,j})^T$ where $\mu^{n,j} = \frac{1}{|\boldsymbol{D}^{n,j}|} \sum_{k=1}^{|\boldsymbol{D}^{n,j}|} (d_k^{n,j})$.

Step 4.2: Conduct eigen decomposition on $\boldsymbol{A}^{n,j}$ to obtain the eigenvector matrix $\boldsymbol{Q}^{n,j}$ and the eigenvalue matrix $\Lambda^{n,j}$ such that $\boldsymbol{A}^{n,j} = (\boldsymbol{Q}^{n,j}) \Lambda^{n,j} (\boldsymbol{Q}^{n,j})^{-1}$.

Step 4.3: Let $P \leq n$ be the number of non-zero eigenvalues obtained in step 4.2. Sort the $P$ eigenvalues and define a cut-point based on some pre-defined criteria to split the corresponding eigenvectors into $P'$ leading eigenvectors and $n - P'$ remaining eigenvectors.

Step 4.4: Use the eigenvectors in $\boldsymbol{Q}^{n,j}$ that correspond to $P'$ leading eigenvalues in the sorted array to define the local coordinate frame for the subspace $S_j$, and rewrite $\boldsymbol{Q}^{n,j} = \left[W^{P',j}\; W^{n-P',j}\right]$.

Step 4.5: The projection error of mapping a data point $d_k^{n,j}$ to the subspace $S_j$ defined by the local coordinate frame is $e = \left(W^{n-P',j}\right)^T \left(d_k^{n,j} - \mu^{n,j}\right)$; where $W^{n-P',j}$ is a $n$ by *(n-P')* matrix. Or the square-magnitude projection error of $d_k^{n,j}$ to the subspace $S_j$ is then equal to $Err\left(d_k^{n,j}, S_j\right) = \left(d_k^{n,j} - \mu^{n,j}\right)^T \left(W^{n-P',j}\right)\left(W^{n-P',j}\right)^T \left(d_k^{n,j} - \mu^{n,j}\right)$.
Calculate the total error: $\sum_j \sum_k Err(d_k^{n,j}, S_j)$.

Step 4.6: Repeat steps 4.4 and 4.5 with a new $P'$ (leading eigenvectors) that is one less; i.e., $P' - 1$. Record the total error.

Step 4.7: Compute the total reconstruction error ratio of two successive rounds in step 4.6; i.e., (total reconstruction error using *P'-q-1* leading eigenvectors)/(total reconstruction error using *P'-q* leading eigenvector) where *q = 0,...,P'-2*.

Step 4.8: Finalize the local coordinate frame for the subspace $S_j$ with a dimension *P'-q* when the error ratio in step 4.7 is the largest for the given *q*.

Step 5: Merge two or more clusters that do not involve *NS*. If there are clusters with only one data point, these clusters will take the priority; then repeat step 4. Retain the solution with a lower total error.

Step 6: Repeat step 5 until the total error is below the predefined error threshold δ, or the algorithm reaches the maximum number of iterations allowed.

One noteworthy observation on step 5 of algorithm 4.2 above is that a merged cluster will be characterized by not one, but multiple statistically significant association patterns. The meaning of a data point will now be its closeness to some association pattern in a merged cluster, in terms of the semantic interpretation defined by the scope coverage and the membership function.

## V. APPLICATION OF MANIFOLD CLUSTERING TO PREDICTIVE ANALYTICS

The manifold clustering based on association patterns is applied to personalized health coaching for understanding its practicality. A pilot study on engaging individuals on self-health management using the SIPPA Health Informatics Platform was conducted. The aim is to understand the feasibility of affecting behavior change toward a healthy lifestyle through a behavior model grounded on the Theory of Planned Behavior [13]. In this study, a validated survey instrument [14] is used to discover the behavior readiness measure of an individual to engage in actionable health activities. Behavior readiness measure is a vector of Real characterizing four behavior attributes [motivation, intention, attitude, ownership]. On a daily basis, actionable health recommendations which range from daily advice on healthy diet and setting goals on physical activities, to self-monitoring of vitals such as blood glucose/pressure readings, were sent to each individual.

In this research, a preliminary experiment was conducted to evaluate the feasibility of applying the proposed manifold clustering for predictive analytics to identify behavior characteristics accounting for compliance to daily messages.

In the pilot study, daily messages are sent via push notifications to a subject's mobile device. A subject is then asked to provide feedback on each message in terms of "like" (equivalent to useful), "dislike" (not useful), and "dismiss" (neutral). If a daily message is actionable such as self-monitoring of one's glucose level, the subject is expected to carry out the self-monitoring activities.

This paper will focus on illustrating the predictive analytics based on proposed manifold clustering to identify non-trivial subgroups of pilot participants who are responsive to the daily messages. The testbed for this preliminary study is a sample collection of data from 53 individuals for the behavior attributes; among them eight have participated in the pilot for about three months. The average number of days of participation of the eight is 96.43 days.

A compliance index was derived for measuring the average responsiveness of a subject to the push notifications over a subject participation period. The self-monitoring compliance index is defined as the (average) number of self-monitoring performed per day divided by the physician recommended number of self-monitoring per day according to clinical guidelines and the diabetes condition of an individual. Similarly, the daily wisdom compliance index is defined as the number of responses to daily wisdom divided by the number of daily wisdom notifications. Daily wisdom consists of healthy tips from a pool of over 100 messages; e.g., "Getting enough sleep is critical to keeping stress under control." This is in addition to a push notification that could carry an actionable message such as "It's time to self-monitor your glucose level and sync the reading to your personal health record."

The training data set is used to derive the manifold spaces and identify the manifolds that are induced by statistically significant association patterns, thereby defining the clusters of the pilot participants. The spanning space is a 6-dimensional space composed of behavior constructs [motivation, intention, attitude, ownership], together with the self-monitoring compliance index and the daily wisdom compliance index.

## VI. EXPERIMENTAL RESULTS, ANALYSIS AND DISCUSSION

Behavior readiness measure is a 1x4 vector of Real composed of behavior constructs [motivation, intention, attitude, ownership]. There are 53 such vectors, and each vector is discretized using algorithm 4.1. The vectors of discrete values are used to discover statistically significant association patterns based on equation 1. Using a support measure threshold 0.2, twenty-five statistically significant association patterns are found and shown in table 1.

One would expect 25 manifold clusters when there are 25 statistically significant association patterns; but, the number of non-empty manifold clusters can be fewer. For example, pattern 25 [ownership:1 self- monitoring:1 daily-wisdom:1] is a special case under the scope covered by pattern 9 [ownership:1 daily-wisdom:1]. In addition, there may be clusters with only one data point when the data set is sparse, as in our case. As a result, only six manifold clusters are non-empty, and four of them contain only one data point. The manifolds containing only one data point are merged with others as depicted in algorithm 4.2 step 5:

TABLE I. STATISTICALLY SIGNIFICANT ASSOCIATION PATTERNS

Statistically Significant Association Patterns for Motivation (M), Intention (I), Attitude (A), Ownership (O), Self-Monitoring (S), and Daily Wisdom (D). Percentage of patterns identified as statistically significant: 20/1296 = 1.5%

|  | M | I | A | O | S | D |
|---|---|---|---|---|---|---|
| 1 |  |  | 2 | 1 |  |  |
| 2 |  | 1 |  | 1 |  |  |
| 3 |  | 2 | 0 |  |  |  |
| 4 |  | 1 | 1 |  |  |  |
| 5 | 2 |  |  | 1 |  |  |
| 6 | 2 |  | 1 |  |  |  |
| 7 | 2 |  | 1 | 1 |  |  |
| 8 | 2 | 1 |  |  |  |  |
| 9 |  |  |  | 1 |  | 1 |
| 10 |  |  | 2 |  |  | 1 |
| 11 |  |  | 1 |  |  | 2 |
| 12 |  | 1 |  |  |  | 1 |
| 13 | 1 |  |  |  |  | 1 |
| 14 |  |  |  | 1 | 1 |  |
| 15 |  |  |  | 1 | 0 |  |
| 16 |  |  | 0 |  | 1 |  |
| 17 |  |  | 1 | 1 | 1 |  |
| 18 |  | 1 |  |  | 1 |  |
| 19 |  | 2 |  |  | 1 |  |
| 20 |  | 1 |  | 1 | 1 |  |
| 21 | 2 |  |  |  | 1 |  |
| 22 | 0 |  |  |  | 1 |  |
| 23 | 2 |  |  | 1 | 1 |  |
| 24 |  |  |  |  | 1 | 1 |
| 25 |  |  |  | 1 | 1 | 1 |

*D-motivation: {0,1,2} , D-intention = {0,1,2,3}, D-attitude={0,1,2}, D-ownership {0,1,2}, D-daily-wisdom-engagement: {0,1,2,3}, D-self-monitoring:{0,1,2}

TABLE II. MANIFOLDS (AFTER MERGING CLUSTERS OF ONLY ONE DATA POINT WITH OTHERS)

| | Eigenvectors | Discarded | Reconstr. Error |
|---|---|---|---|
| Cluster 2,3 | 6 | 5 | 0.0807 |
| Cluster 1,4,6 | 6 | 4 | 0.0381 |
| Cluster 5 | 6 | 4 | 0.0241 |
| Total | 18 | 13 | 0.1429 |

TABLE III. MANIFOLDS (AFTER MERGING CLUSTERS IN STEP 6)

| | Eigenvectors | Discarded | Reconstr. Error |
|---|---|---|---|
| Cluster 1,2,3,4,6 | 6 | 2 | 0.0032 |
| Cluster 5 | 6 | 4 | 0.0241 |
| Total | 12 | 4 | 0.0273 |

## VII. CONCLUSION

This paper presents a novel manifold clustering approach based on the concept of statistically significant association patterns. A preliminary study based on real world data was applied to better understand the proposed approach for predicting hidden nonlinear sub-population of diabetes patients. While this paper demonstrates the feasibility of applying the proposed approach to real world data, the order sequence effect of merging special case of a cluster of singletons is not yet well understood. Furthermore, dimension reduction based on the ratio of incremental change of the total reconstruction errors, in addition to the membership function, is just one out of other possible criteria for deriving the manifold clusters. Future research opportunities exist regarding these two aspects, which will be the focus of our future research.

## VIII. REFERENCES

[1] Boehmke, B., & Greenwell, B. (2019). *Hands-On Machine Learning with R*. Chapman and Hall/CRC, ISBN 9781138495685.

[2] Quinlan, J.R. (1986), Induction of Decision Trees. *Machine Learning*, **9**(1):81-106, 1986.

[3] Kak A., (2018). *Low Dimensional Manifold in a HIgh-Dimensional Measurement Space.* Data on Manifolds Tutorial, Purdue University.

[4] Luxburg, V. (2007). U. Stat Comput, 17: 395. https://doi.org/10.1007/s11222-007-9033-z

[5] Macqueen, J. B., & Lee, H. B. (1980). A K-Means Cluster Analysis Computer Program with Cross-Tabulations and Next-Nearest-Neighbor Analysis. *Educational and Psychological Measurement*, *40*(1), 133–138.

[6] Zhang, K., & Kwok, J. T. (2010). Clustered Nyström Method for Large Scale Manifold Learning and Dimension Reduction. *IEEE Transactions on Neural Networks*, *21*(10), 1576–1587.

[7] Wang, X.-D., Chen, R.-C., Zeng, Z.-Q., Hong, C.-Q., & Yan, F. (2019). Robust Dimension Reduction for Clustering with Local Adaptive Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(3), 657–669.

[8] Gong, W., Zhao, R., & Grünewald, S. (2018). Structured sparse K-means clustering via Laplacian smoothing. *Pattern Recognition Letters*, *112*, 63–69.

[9] Faivishevsky, L., & Goldberger, J. (2012). An unsupervised data projection that preserves the cluster structure. *Pattern Recognition Letters*, *33*(3), 256–262.

[10] Kullback, S., (1959). *Information Theory and Statistics*. Wiley and Sons, New York.

[11] Sy, B., & Gupta A. (2004). *Information-Statistical Data Mining: Warehouse Integration with Examples of Oracle Basics.* eBook ISBN: 978-1-4419-9001-3, DOI: 10.1007/978-1-4419-9001-3, Springer.

[12] Ge, S. S., He, H., & Shen, C. (2012). Geometrically local embedding in manifolds for dimension reduction. *Pattern Recognition*, 45(4), 1455–1470.

[13] Sy, B. (2017). SEM Approach for TPB: Application to Digital Health Software and Self-Health Management. *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1660-1665.

[14] Roblin, N., Little, M.A., & McGuire, H. (2004). Diabetes Self-Efficacy Questionnaire (DSEQ) Outcome Measurement for Diabetes Education. (Last access: 10/7/2019. http://dnig.rnao.ca/sites/dnig/files/DSEQ_Manual_0.pdf)