



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *The 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC2018), Miyazaki, Japan, Oct. 7-10, 2018.*

Citation for the original published paper:

Ali Hamad, R., Järpe, E., Lundström, J. (2018)

Stability analysis of the t-SNE algorithm for human activity pattern data

In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-38442>

Stability analysis of the t-SNE algorithm for human activity pattern data

Rebeen Ali Hamad

Intelligent Systems and Digital Design *Intelligent Systems and Digital Design*
Halmstad University, Sweden Halmstad University, Sweden
rebeen.ali_hamad@hh.se eric.jarpe@hh.se

Eric Järpe

Jens Lundström

JeCom Consulting
Halmstad, Sweden

jens.lundstrom@jecom-consulting.com

I. ABSTRACT

Health technological systems learning from and reacting on how humans behave in sensor equipped environments are today being commercialized. These systems rely on the assumptions that training data and testing data share the same feature space, and residing from the same underlying distribution - which is commonly unrealistic in real-world applications. Instead, the use of transfer learning could be considered. In order to transfer knowledge between a source and a target domain these should be mapped to a common latent feature space. In this work, the dimensionality reduction algorithm t-SNE is used to map data to a similar feature space and is further investigated through a proposed novel analysis of output stability. The proposed analysis, Normalized Linear Procrustes Analysis (NLPA) extends the existing Procrustes and Local Procrustes algorithms for aligning manifolds. The methods are tested on data reflecting human behaviour patterns from data collected in a smart home environment. Results show high partial output stability for the t-SNE algorithm for the tested input data for which NLPA is able to detect clusters which are individually aligned and compared. The results highlight the importance of understanding output stability before incorporating dimensionality reduction algorithms into further computation, e.g. for transfer learning.

II. INTRODUCTION

Exploratory data analysis based on all available dimensions of a high-dimensional data set (HDD) is generally intractable. The underlying manifold structure could have a low intrinsic dimensionality and is therefore often explored using dimensionality reduction (DR) techniques. The intrinsic dimensionality of an HDD could be considered as the minimum number of parameters needed for explaining the observed properties of the data. Moreover, such parameters could preserve both global and local structure of the HDD in the lower-dimensional representative space [28]. Principal Component Analysis (PCA) [9] and multidimensional scaling (MDS) [30] techniques that linearly map an HDD into a lower dimensional representation are broadly used in business and marketing applications for DR and data visualization. PCA and MDS compute low-dimensional maps where dissimilar data points are far apart. However, keeping similar data points close together in a low-dimensional map is crucial for

an HDD that lies on or near a low-dimensional manifold, which is often difficult using linear mapping techniques [19]. Therefore PCA and MDS are not suitable for many complex and non-linear datasets [12]. Instead, a popular non-linear DR algorithm is the t-distributed Stochastic Neighbor Embedding (t-SNE) [19] commonly known for producing low-dimensional representations of the input data [5], [22] presumably close to the sought real low-dimensional manifold. In contrast to PCA and MDS, t-SNE as a non-linear DR method has the ability to cope with complex data sets that are likely to lie on a low-dimensional non-linear manifold [4]. t-SNE works in an unsupervised fashion, can utilize any distance metric and commonly adapts to both sparse and dense input data [16], better than ISOMAP [29] and kernel-PCA as it seems [27]. Algorithms for manifold learning, such as t-SNE are used across a broad range of information processing applications including immunology [3], and data compression [32].

Commonly many DR algorithms such as t-SNE are used for visualization of HDD data [23] or for further data processing such as when modelling of human activity patterns [18] by mapping data to a low-dimensional representation. A suitable representation of *how*, *when* and *where* humans performing activities in their own home opens up for various health technology applications such as systems for anomaly detection (e.g. falls) or tracking progression of diseases (e.g. early-warning of dementia). In this paper, data from a sensor equipped smart home is used.

Although t-SNE is presented to be a suitable method for data visualization, t-SNE has a few potential weaknesses [19]. One is the uncertainty of convergence. Due to the non-convex cost function there is no guarantee that the mapped output results are similar even for different runs of the algorithm given identical input data, especially since the initialization of the map points is randomized. Despite the non-deterministic setup of the algorithm, the visual interpretation of different runs is easily compared by humans and far more simple to perform than automatic machine-based comparisons which are necessary for accurate evaluation of the algorithm. The primary problem studied in this work is how to analyze the stability of the t-SNE algorithm output. The proposed approach utilizes comparisons of several output maps as a whole and partially by clustered low-dimensional data points. This is performed by using smart home data of human activity

patterns as input data.

Learned manifolds can be aligned to build a correspondence between different disparate data sets and thereby provide knowledge transfer across the data sets from different domains [34]. Accordingly, knowledge transfer using machine learning (i.e. transfer learning), is gained during the transition from one learned domain to another, aiming to improve learning in the target task by leveraging knowledge from the source task [31]. In our work, it is hypothesized that learned manifolds (by a data-driven model such as t-SNE) from disparate data sets could be used for transfer learning. Therefore, it is crucial to investigate the stability of t-SNE output in order to use this algorithm further for aligning manifolds for the purpose of transfer learning.

Random Forests (RF) are commonly used as a one-class classifier, e.g. in order to model human behaviour patterns in an unsupervised fashion from sensor data collected in a smart home environment [18]. Then the proximity matrix from RF from such a model is fed to the t-SNE algorithm in order to map human behaviour patterns to a lower dimensional manifold. However, since t-SNE is a stochastic algorithm and since there is a large variance of t-SNE maps, a thorough analysis of the stability is required before applying TL.

The contribution of this work is the development of methods and tools for studying t-SNE output stability on smart home data used for modelling human activity patterns. The long-term goal of this research is to achieve automatic knowledge transfer between related data sets from different smart homes using manifold comparisons.

The rest of the paper is organized as follows. In Section 2, related work will be described and in Section 3, methods for testing stability of t-SNE for different runs is proposed. In Section 4, experiment results will be presented and discussed. Finally, the findings and opportunities of further research will be summarized in Section 5, Conclusions.

III. RELATED WORK

Dimensionality reduction techniques have been widely used in many application domains to map HDD onto a low-dimensional manifold in order to produce a meaningful and visualizable representation. In [34] manifold alignment are used to construct connections (low-dimensional mappings) between different but related data sets by aligning their underlying learned manifolds for transferring knowledge across the data sets. Several comparative studies outlining the various DR techniques have been addressed in the literature [32], [33], [14]. In various research studies a set of quality assessment criteria has been considered based on local and global geometry preservation concepts [21], [7], [35]. However, these studies are mostly based on artificial data sets and the assessment of ability to find a good representation often relies on visual interpretation. To our knowledge an exhaustive empirical investigation analysis of the stability of non-linear DR techniques has not been carried out. Moreover, several studies have been done for investigating the performance of

the non-linear DR methods in artificial and real tasks [2], [15], [24].

The stability of unsupervised DR techniques was studied by Garcia et. al who studied the parameter and data variations on several artificial data sets [5]. The study concluded by visual inspection that parameter variations in the resulting embeddings did not render instability.

Moreover, Laplacian Eigenmaps (LE) and Local Linear Embedding (LLE) were tested for stability when small or minor parameter variations which led to the conclusion that local methods (LE and LLE) are more likely to be affected by small modifications in parameter variations and therefore less stable than t-SNE.

Khoder et al performed a comparative study [11] to investigate the stability of unsupervised dimensionality reduction techniques using perturbed data of large images. The authors presented a new method for measuring the stability of non-linear and linear methods based on the noise variance at various scales. Results showed that PCA and MDS are limited by their linear character or are difficult to use when working on HDD because of their complexity.

A method for comparing DR techniques in terms of loss of quality with the aim to preserve the geometry of data sets has also been proposed [6]. Results revealed that the best results on all data sets are obtained by t-SNE.

Moreover, the accuracy of non-linear DR methods has been under review using real and synthetic data sets [32]. The experiment results show that non-linear DR methods perform well on the preferred synthetic data set, while this strong performance is not proved to extend to the real data sets.

Consequently, to the best of our knowledge, the stability analysis of t-SNE using manifold alignment on partial data points in the output maps has not been attempted before. Stability of low-dimensional manifolds plays a key role to align manifolds properly for the purpose of transfer learning. Thus, the contribution of this paper is important for transferring knowledge in a multi-smart home environment based on aligning manifolds.

IV. METHODS & PROPOSED APPROACH

To measure the stability of t-SNE output an approach based on partially aligning t-SNE maps is proposed. The following steps (see Algorithm 1) constitutes the method for computing stability measures for a set of t-SNE maps. Firstly, T maps are produced by repeatedly computing t-SNE output maps with random initialization of low-dimensional data points given identical HDD input, X (see lines 2-4). Secondly, the resulting maps are pair-wise aligned into a modified target space, creating T^2 alignments (where $\frac{1}{2}T^2 - T$ are unique alignments due to the commutativity of the align operator: $\text{align}(\text{map_list}[i], \text{map_list}[j]) = \text{align}(\text{map_list}[j], \text{map_list}[i])$). The variability of these alignments is then evaluated by two measures of point-cloud alignment: mean disparity (\bar{d}_{ij} , where disparity is the sum of squared pair-wise differences between observations) and the

estimated mean probability of obtaining the true corresponding data point within the aligned five nearest neighbors, p^{5NN} .

Algorithm 1 Compute t-SNE output stability

```

1: Input: HDD,  $X$ 
2: for  $i \leftarrow 0$  to  $T$  do
3:   map_list  $\leftarrow$  compute_tSNE( $X$ )
4: end for
5: for  $i \leftarrow 0$  to  $T$  do
6:   for  $j \leftarrow 0$  to  $T$  do
7:     ( $am_i, am_j$ )  $\leftarrow$  align(map_list[ $i$ ], map_list[ $j$ ])
8:      $p_{ij}^{5NN} \leftarrow$  prob_5NN( $am_i, am_j$ )
9:      $d_{ij} \leftarrow$  disparity( $am_i, am_j$ )
10:  end for
11: end for
12: Output:  $\overline{p^{5NN}}, \overline{d_{ij}}$ 

```

Besides giving an explanation of the t-SNE algorithm this section describes three different methods of alignment: Procrustes Analysis (PA), Local Procrustes Analysis (LPA) and the proposed extension Normalized Local Procrustes Analysis (NLPA).

A. t-SNE

t-SNE, introduced by van der Maaten and Hinton [19], is a nonlinear dimensionality reduction algorithm that maps high-dimensional data-points into a lower dimensional space. t-SNE utilizes embedding, which is constructed such that data-points in the vicinity of each other (i.e. similar data-points) in a high-dimensional space will remain in the vicinity by embed-points in a lower-dimensional space. Mainly, the t-SNE technique consists of two phases. Firstly, a joint probability over pairs of the data-points will be computed so that similar data points from the original (high-dimensional) data set have a large probability of being picked by each other for the embedding space. This results in dissimilar data-points to having a smaller probability of being picked. Accordingly, t-SNE is preserving the local geometry of the original high-dimensional data [19]. Secondly, over the map-points a probability distribution will be determined by t-SNE that fits data-point positions in the map in order to minimize the Kullback-Leibler divergence between both high and low dimensional distributions. Furthermore, t-SNE algorithm has originated from Stochastic neighbor Embedding (SNE) [8] and aimed to alleviate the main SNE challenges related to the thin tails of the normal distribution resulting in a data representation where even dissimilar data-points are *crushed* together which is known as the crowding problem [19]. t-SNE use a heavy-tailed distribution (Student-t distribution) to compute the similarity between two points. Moreover a symmetric version of the SNE cost function is implemented by t-SNE with simpler gradients.

B. Aligning by Procrustes Analysis

PA is one of the most popular rigid shape analysis algorithms. PA applies translation, scaling and rotation to two identically sized data sets in a multivariate Euclidean space to

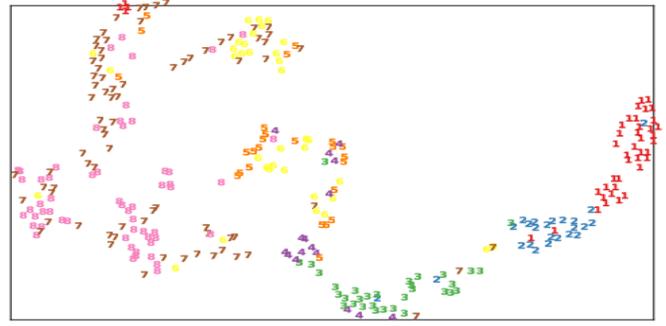


Fig. 1: t-SNE map for human behaviour patterns data

find the optimal alignment and to minimize the disparity [25]. Algorithm 2 shows the PA process. Firstly, PA translates the data sets to their origin. Secondly, PA normalizes the data sets using the Frobenius norm. Finally, PA rotates the second dataset to fit the first dataset in order to minimize the disparity. In this work PA is used to align the two-dimensional manifolds of the smart home dataset produced by t-SNE. PA performs well for data sets which are linear transformations of each other. However, PA works poorly on aligning non-linear smart home maps produced by t-SNE.

Algorithm 2 Procrustes Analysis (PA)

```

1: Input:  $M_1, M_2$        $M_1, M_2$  are the input t-SNE maps
2:  $M_1 \leftarrow M_1 - \mu(M_1)$ 
3:  $M_2 \leftarrow M_2 - \mu(M_2)$   translate both data sets to their origin
4:  $M_1 \leftarrow M_1 / \|M_1\|_F$   scaling of  $M_1$  and  $M_2$ 
5:  $M_2 \leftarrow M_2 / \|M_2\|_F$   where  $\|\cdot\|_F$  denotes Frobenius norm
6:  $M_2 \leftarrow$  Rotation( $M_1, M_2$ )
   rotation  $M_2$  with respect to  $M_1$  to minimize disparity
7: disparity  $\leftarrow \sum_{i=0}^n (M_{1i} - M_{2i})^2$ 
   measure the dissimilarity between the two data sets
8: Output:  $M_1, M_2, \text{disparity}, \|M_1\|_F, \|M_2\|_F, \mu(M_1)$ 

```

C. Aligning by Local Procrustes Analysis

LPA was introduced by [20] to non-linearly align manifolds by using locally linear mappings. This algorithm comprises two main steps. Firstly, it follows a divisive approach to cluster datasets. The algorithm starts by considering a cluster of all data points and keeps on splitting into two sub-clusters recursively and terminates if the diversity of a cluster is below a predetermined threshold. Secondly, at each stage PA is applied to all clusters in the first data set and the corresponding cluster in the second dataset to compute disparity. If the disparity falls short of the threshold, the clustering process stops for these clusters at this stage. LPA uses K-means to create clusters. K-means is a non-deterministic algorithm that gives different results for different runs in terms of number of data points in each cluster and centroids location. For the analysis in this paper, K-means is replaced by a deterministic clustering algorithm, *hierarchical agglomerative clustering*, that plays a significant role in having a proper stability investigation of the t-SNE which is a non-deterministic algorithm. Figure 2 shows

the process of clustering data. At initialization, each point cloud (which in this paper is the t-SNE map) is assigned to one cluster and the disparity of Procrustes Analysis is computed for the entire two maps. If the disparity is greater than a threshold, then the cluster will be divided into two sub-clusters if both clusters have at least two data points each. For this work, this condition has been modified to make sure that there are at least two distinct data points in order to meet the criteria of applying PA. This process will be applied to the sub-clusters for better manifold alignment with respect to disparity. Cluster alignments that offer a disparity lower than the threshold will not be further clustered. It is worth noting that disparities of LPA and PA cannot be fairly compared. This is due to the fact that the space of the data-clusters for the LPA and the space of the entire dataset for PA are in normalized for each cluster and map respectively.

For instance, ten disparities will be obtained if ten data-clusters in the first low-dimensional dataset are compared with their corresponding data-clusters in the second dataset. On the other hand, only one disparity will be obtained by applying PA on the two low dimensional dataset. Therefore, it is not reasonable to compare the disparities obtained using the PA and LPA. In this paper LPA is only applied to create clusters. In the next section, we propose an extension for the LPA in order to have the same space with the PA for the sake of the disparity comparison.

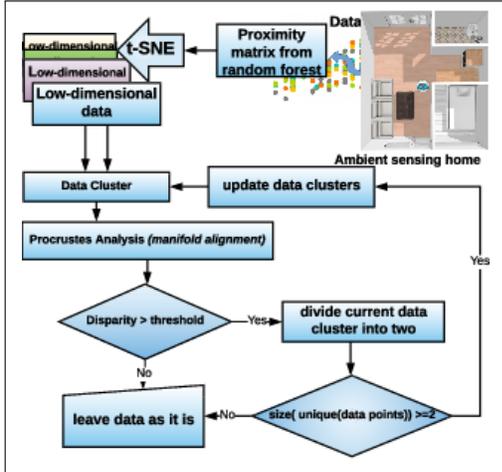


Fig. 2: Clustering data using LPA, this is from Halmstad Intelligent Home [17].

D. Aligning by Normalized Local Procrustes Analysis

In this work, LPA is used for creating clusters of t-SNE low-dimensional maps. We propose an extension to LPA which we call Normalized Local Procrustes Analysis (NLPA) in order to compare locally aligned clusters alignments to a complete map, which can be compared to PA as well as being used for stability analysis. Algorithm 3 shows the NLPA procedure. NLPA applies PA on each cluster and then normalizes the clusters so that the transformed data-points in each cluster are mapped back to the original space of the data after alignment.

Algorithm 3 Normalized Local Procrustes Analysis (NLPA)

```

1: Input:  $M_1, M_2$        $M_1, M_2$  are the input t-SNE maps
2:  $n_c \leftarrow n$       number of clusters created using LPA
3: for  $i \leftarrow 0$  to  $n_c$  do
4:    $\{M_1, M_2, -, \text{norm}_1, \text{norm}_2, \mu(M_1)\} \leftarrow \text{PA}(M_1, M_2)$ 
   call PA from Algorithm 2
5:    $M_1 \leftarrow M_1 \cdot \text{norm}_1 + \mu(M_1)$ 
6:    $M_2 \leftarrow M_2 \cdot \text{norm}_2 + \mu(M_1)$ 
7:    $\text{templist}_1 \leftarrow M_1$ 
8:    $\text{templist}_2 \leftarrow M_2$ 
9: end for
10:  $\text{Map}_1 \leftarrow \text{templist}_1$ 
11:  $\text{Map}_2 \leftarrow \text{templist}_2$ 
12:  $\text{disparity} \leftarrow \sum_{i=0}^n (\text{Map}_{1,i} - \text{Map}_{2,i})^2$ 
   dissimilarity between the two sets
13: Output: disparity

```

Normalization is done by multiplying the clusters with the norms of the aligned clusters that are produced by PA and then adding the mean of the first cluster, as shown in lines 5-6 of Algorithm 3. This normalization allocates the combined clusters of data-points to the same space as to original data. Finally, NLPA computes the disparity and estimated mean probability of obtaining the true corresponding data point within the aligned five nearest neighbors for the combined aligned clusters.

The changes of the proposed method NLPA compared to the LPA procedure can be summarized

- 1) Modification: Firstly the clustering algorithm is modified from k-means to agglomerative clustering.
- 2) Improvement: Secondly the creating clustering criteria is improved to have two distinct data points in each cluster and the threshold is minimized to render better alignment.
- 3) Extension: Finally, the NLPA is extended on LPA to normalize the transformed clusters in order to the combined clusters with NLPA and the whole dataset with PA have a same space.

V. EXPERIMENTAL SETUP

A. Data

Data from Halmstad Intelligent Home [17] (HINT) was acquired for this work. HINT is a sensor-equipped home able to capture occupancy, movement, and interactions. In this home, 8 activities were performed by 11 individuals. The data were generated by an incoming stream of binary events from 37 sensors of the home. The events are represented by a particular ID of the triggered sensor, the associated binary state, and a time-stamp of when the event occurred. The observations of the data set are equal to the number of time windows over the measurement time period for the 11 individuals (310 observations). One observation (over a time window of 30 seconds) holds R number of features where R is equal to the time resolution (1 second) within a moving window times the number of events. The data pre-processing also involves a convolution over time in order to create a

memory between sensor interactions over time (a process which has been demonstrated to be successful when modelling human behaviour [18]). Moreover are the observations being fed to a Random Forest which discriminates between the observations class and a class randomized from the data itself to train a one-class classifier. At last, the proximity matrix is extracted from the forests (a detailed explanation of the process can be found in [18]) and used as input for the t-SNE stability analysis.

Fig 1 is an example of a low-dimensional representation of human activities computed using t-SNE. The numbers in Fig 1 indicate the following activities 1. go to bed 2. use bathroom 3. prepare breakfast 4. leave house 5. get cold drink 6. office 7. get hot drink 8. prepare dinner.

B. Measurements & Parameter selection

To have an exhaustive analysis of the stability of the t-SNE, 100 low-dimensional data are mapped using 10 different configurations of t-SNE. Each configuration of t-SNE has a specific value of the t-SNE perplexity parameter from a set of numbers which are 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 as a typical value between 5 and 50 recommended in the original paper of the t-SNE [19]. The default perplexity parameter is 20. The perplexity parameter controls how far to look for neighbors around a data-point and is related to the width of the t-distribution used in t-SNE. Each t-SNE map is compared with the rest maps for each t-SNE configuration to compute disparity and probability of obtaining the correct correspondences observation within five nearest neighbors for the transformed data by PA and NLPA. Therefore, 10000 comparisons are conducted in the experiments for each t-SNE configuration. Figure 3 shows the experiments procedure of the t-SNE stability analysis. Lastly, PA and NLPA are applied on the low dimensional manifolds to compute disparities. The experimental results show that the smaller disparity threshold renders better results. Therefore, the threshold is decreased from 0.001 of LPA to 0.00001 for NLPA. Besides, for every point in the first input dataset, the correct correspondence observation is found within five nearest neighbors in the second input dataset to compute the probability of obtaining the correct correspondence observation within five nearest neighbors.

VI. RESULTS & DISCUSSION

Recently, t-SNE stability has gained a lot of interest for projecting high-dimensional data into a low-dimensional manifold with the aim of transferring knowledge using manifold alignment. However, since t-SNE is a stochastic algorithm and since there is a large variance of t-SNE maps, a thorough analysis of the stability is required before applying TL.

Exhaustive scenarios are considered for an investigation about the t-SNE stability through manifold alignment using PA and NLPA. Table I shows the estimated disparity and probability of obtaining the correct correspondence observation within five nearest neighbors of the PA and the NLPA methods respectively for different values of perplexity. It turns out that

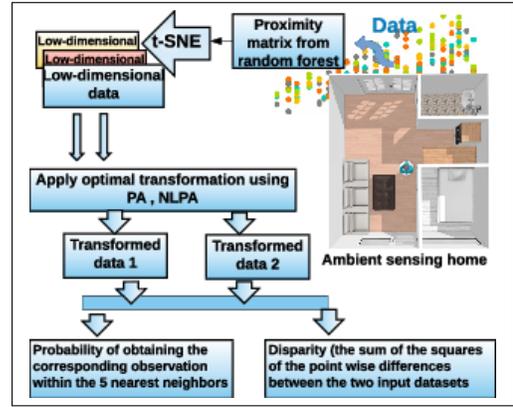


Fig. 3: Experiments process, ambient sensing home is from Halmstad Intelligent Home [17].

TABLE I: Estimated expected disparity and estimated probability of obtaining the correspondence observation within five nearest neighbors. After each estimate the standard error is given.

Perplexity	Disparity		Probability of obtaining the correspondence within the 5 nearest neighbors %	
	Mean (SE)			
	PA	NLPA	PA	NLPA
5	0.5954 (0.0206)	0.0010 (0.0002)	7.363 (0.0074)	98.541 (0.0099)
10	0.3135 (0.0248)	0.0006 (0.0001)	24.527 (0.0226)	98.541 (0.0099)
15	0.1097 (0.0167)	0.0007 (0.0001)	44.618 (0.0235)	98.476 (0.0099)
20	0.0291 (0.0022)	0.0006 (0.0001)	60.365 (0.0163)	98.450 (0.0099)
25	0.0296 (0.0026)	0.0007 (0.0001)	63.873 (0.0189)	98.485 (0.0099)
30	0.0447 (0.0058)	0.0009 (0.0002)	56.693 (0.0261)	98.435 (0.0099)
35	0.1483 (0.0152)	0.0029 (0.0005)	28.216 (0.0241)	98.126 (0.0098)
40	0.1648 (0.0129)	0.0046 (0.0006)	26.186 (0.0223)	97.869 (0.0098)
45	0.1699 (0.0123)	0.0049 (0.0006)	31.522 (0.0256)	97.637 (0.0098)
50	0.1501 (0.0116)	0.0056 (0.0006)	38.485 (0.0296)	97.744 (0.0098)

for all perplexity values considered the disparity values from using NLPA are less than 4% of the corresponding disparity value from using PA. In other words, the NLPA method is 25 times better than the PA method in terms of disparity. Also, the disparity from using PA decreases slightly for perplexity ranging from 5 to 20 while it increases for perplexity values from 25 to 50. The disparity values from using NLPA increases almost monotonically for all perplexity values considered.

With respect to the probability of obtaining the correct correspondence within the five nearest neighbors, the PA correct correspondence mean values commonly increase by increasing perplexity, especially as the perplexity value reaches 25. On the other hand, the probability of obtaining correspondence within the five nearest neighbors slightly decreases by increasing perplexity ranging from 5 to 50 for NLPA.

Figure 6 shows PA and NLPA histograms of disparity and probability of obtaining the correct correspondence observation within the five nearest neighbors where perplexity equals 20. Both the table and the histograms show that NLPA consistently outperforms PA for all t-SNE configurations, see Figures 4 and 5 respectively. Based on these results, it is concluded that t-SNE maps are stable locally for human behavior data that reflects the indicated property of t-SNE and which preserves local structure of data. Figure 7 shows the

alignment of t-SNE maps that has been used to investigate the t-SNE stability using PA and NLPA. The alignment maps in Figure 7 indicate that the t-SNE maps are stable locally compared to globally aligned t-SNE maps using PA.

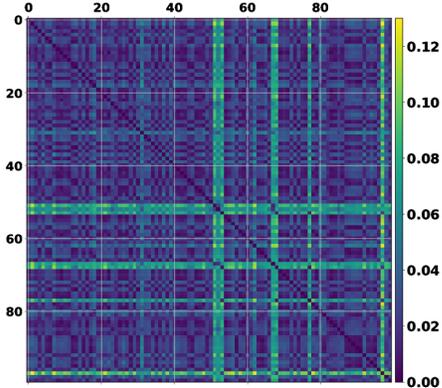


Fig. 4: PA disparity similarity matrix of size 100×100

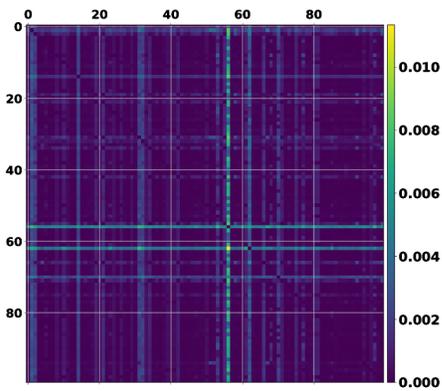


Fig. 5: NLPA disparity similarity matrix of size 100×100

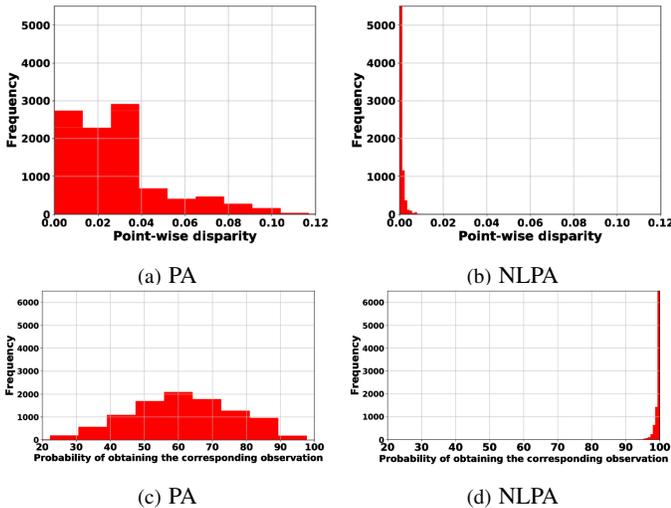


Fig. 6: PA and NLPA histograms of disparity and probability of obtaining the correspondence observation within five nearest neighbors where perplexity is equal to 20

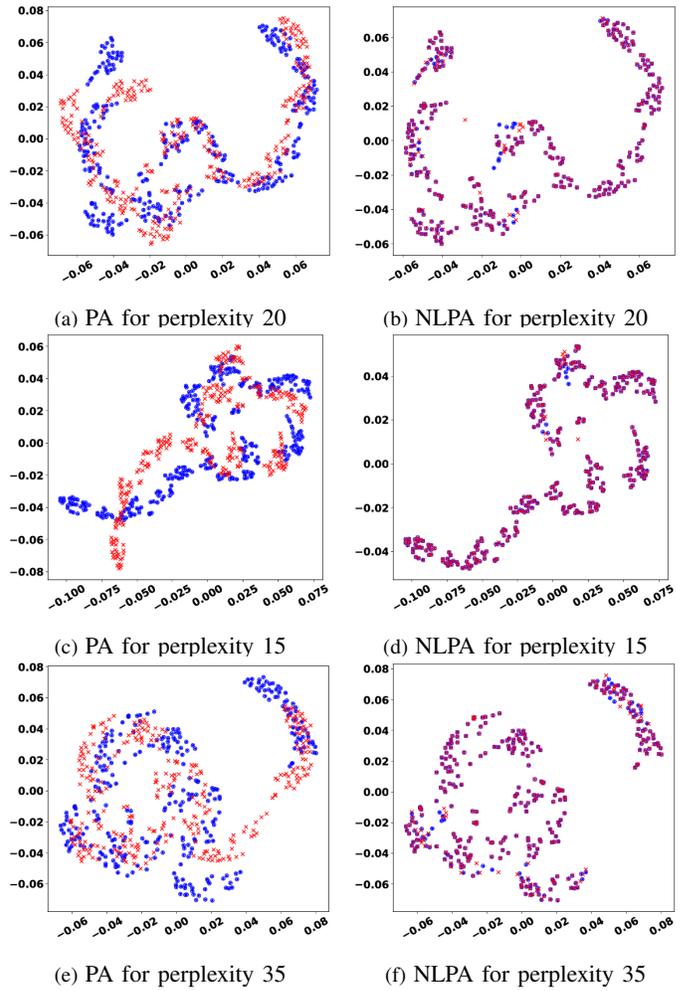


Fig. 7: Manifold alignment using PA and NLPA for three cases

VII. CONCLUSION

The t-SNE mapping stability of human activity patterns in smart homes via the analysis of reproducibility of low-dimensional manifolds is investigated. One could claim that any two data sets could be aligned via a non-linear mapping function with enough degrees of freedom. However, this study aims at analyzing parts of a map in order to investigate the stability of t-SNE. Therefore, the choice of linear and local transformations gives human intuition about the stability of t-SNE. Procrustes Analysis (PA) is used for linearly aligning low-dimensional manifolds in order to compute disparity and correct correspondence observation within the five nearest neighbors. An extension to Local Procrustes Analysis called Normalized Local Procrustes Analysis (NLPA) is proposed to non-linearly align manifolds by using locally linear mappings. Experiments show that the disparity from using NLPA decreases by magnitudes compared to the disparity from using PA. Also, the probabilities of obtaining the correct corresponding observation within the five nearest neighbors from the second set of data points for each point in the first set of data points are radically increased by using NLPA compared to PA. For instance when the t-SNE parameter is 20, the disparity mean value decreases from 0.2913 in the case of using PA to a

mere 0.00066 upon using NLPA. The probability of obtaining the correct corresponding observation within the five nearest neighbors for the same comparison, increases from 60.37 when using PA to 98.45 in case of using NLPA. In conclusion, NLPA outperforms PA by providing much better alignments for the low-dimensional manifolds on the same data set. This indicates that t-SNE low-dimensional manifolds are locally stable which is the main achievement of this study.

Future work will explore extensions of NLPA for aligning low-dimensional manifolds of disparate data sets. Then t-SNE low-dimensional manifolds of disparate data sets will be compared using NLPA to discover the common manifolds of the disparate data sets to be used for Transfer learning.

REFERENCES

- [1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [2] Pedro Latorre Carmona and Reiner Lenz. Performance evaluation of dimensionality reduction techniques for multispectral images. *International Journal of Imaging Systems and Technology*, 17(3):202–217, 2007.
- [3] Michael Thomas Wong et al. A high-dimensional atlas of human t cell diversity reveals tissue-specific trafficking and cytokine signatures. *Immunity*, 45(2):442 – 456, 2016.
- [4] Francisco J García-Fernández, Michel Verleysen, John Aldo Lee, and Ignacio Díaz Blanco. Sensitivity to parameter and data variations in dimensionality reduction techniques. In *ESANN*, 2013.
- [5] Francisco Javier García Fernández, Michel Verleysen, John A Lee, and Ignacio Díaz Blanco. Stability comparison of dimensionality reduction techniques attending to data and parameter variations. In *Eurographics Conference on Visualization (EuroVis)(2013)*. The Eurographics Association, 2013.
- [6] Antonio Gracia, Santiago González, Victor Robles, and Ernestina Menasalvas. A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Information Sciences*, 270:1–27, 2014.
- [7] Hisashi Handa. On the effect of dimensionality reduction by manifold learning for evolutionary learning. *Evolving Systems*, 2(4):235–247, 2011.
- [8] Geoffrey E Hinton and Sam T Roweis. Stochastic neighbor embedding. In *Advances in neural information processing systems*, pages 857–864, 2003.
- [9] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
- [10] Chenping Hou, Changshui Zhang, Yi Wu, and Yuanyuan Jiao. Stable local dimensionality reduction approaches. *Pattern Recognition*, 42(9):2054–2066, 2009.
- [11] Jihan Khoder, Rafic Younes, and Fethi Ben Ouedzou. Stability of dimensionality reduction methods applied on artificial hyperspectral images. In *International Conference on Computer Vision and Graphics*, pages 465–474. Springer, 2012.
- [12] Kyoungok Kim and Jaewook Lee. Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, 47(2):758–768, 2014.
- [13] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.
- [14] John A Lee and Michel Verleysen. Scale-independent quality criteria for dimensionality reduction. *Pattern Recognition Letters*, 31(14):2248–2257, 2010.
- [15] Jiaxi Liang, Shojaeddin Chenouri, and Christopher G Small. A new method for performance analysis in nonlinear dimensionality reduction. *arXiv preprint arXiv:1711.06252*, 2017.
- [16] Pedro E Lopez-de Teruel, Oscar Canovas, and Felix J Garcia. Visualization of clusters for indoor positioning based on t-sne. In *Proceedings of the Indoor Positioning and Indoor Navigation Conference*, 2016.
- [17] Jens Lundström, Wagner O De Morais, Maria Menezes, Celeste Gabrielli, João Bentes, Anita SantAnna, Jonathan Synnott, and Chris Nugent. Halmstad intelligent home-capabilities and opportunities. In *International Conference on IoT Technologies for HealthCare*, pages 9–15. Springer, 2016.
- [18] Jens Lundström, Eric Järpe, and Antanas Verikas. Detecting and exploring deviating behaviour of smart home residents. *Expert Systems with Applications*, 55:429–440, 2016.
- [19] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.
- [20] Ndivhuwo Makondo, Benjamin Rosman, and Osamu Hasegawa. Knowledge transfer for learning robot models via local procrustes analysis. In *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*, pages 1075–1082. IEEE, 2015.
- [21] Deyu Meng, Yee Leung, and Zongben Xu. A new quality assessment criterion for nonlinear dimensionality reduction. *Neurocomputing*, 74(6):941–948, 2011.
- [22] Andreas C Müller and Sarah Guido. Introduction to machine learning with python: a guide for data scientists. 2017.
- [23] Nicola Pezzotti, Boudewijn PF Lelieveldt, Laurens van der Maaten, Thomas Höllt, Elmar Eisemann, and Anna Vilanova. Approximated and user steerable tsne for progressive visual analytics. *IEEE transactions on visualization and computer graphics*, 23(7):1739–1752, 2017.
- [24] Hassan Radvar-Esfahlan and S-A Tahan. Performance study of dimensionality reduction methods for metrology of nonrigid mechanical parts. *International Journal of Metrology and Quality Engineering*, 4(3):193–200, 2013.
- [25] Amy Ross. Procrustes analysis. *Course report, Department of Computer Science and Engineering, University of South Carolina*, 2004.
- [26] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.
- [27] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [28] Mohammed Shukur, T Rani, S Bhavani, G Sastry, and Surampudi Raju. Local and global intrinsic dimensionality estimation for better chemical space representation. *Multi-disciplinary Trends in Artificial Intelligence*, pages 329–338, 2011.
- [29] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.
- [30] Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.
- [31] Lisa Torrey and Jude Shavlik. Transfer learning. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, 1:242, 2009.
- [32] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71, 2009.
- [33] Jarkko Venna et al. *Dimensionality reduction for visual exploration of similarity structures*. Helsinki University of Technology, 2007.
- [34] Chang Wang and Sridhar Mahadevan. Manifold alignment without correspondence. In *IJCAI*, volume 2, page 3, 2009.
- [35] Peng Zhang, Yuanyuan Ren, and Bo Zhang. A new embedding quality assessment method for manifold learning. *Neurocomputing*, 97:251–266, 2012.