DATA ANALYSIS BASED ON THE VISUALIZATION: A SURVEY

[1] Sapna Singh [2]Dr. Bhawna Narain

[1]Ph. D. scholar, [2] Associate professor of MATS school of IT,

[1]Information Technology

[1]MATS university Raipur India

Abstract: Data visual image is an important part of genomic information analysis quite new and promising field in computing that refers to the techniques accustomed communicates information or data by coding it as visual objects contained in graphics. The paper offers an evaluate and systematic comparison of those techniques. The performances of the nonlinear strategies are investigated on artificial and natural duties.

First, we get to compare with data visualization equipment and its associated outcomes. then we are able to look through a few preferred algorithms to do the records visualization. and we additionally explain these outcomes by figuring out weaknesses of modern-day nonlinear strategies**.**

Index Terms - Data visualization, nonlinear techniques

I. INTRODUCTION

The Visualization or visualisation design process typically begins with ill-defined, imprecise design goals. because the layout manner maintains similarly design dreams for developing pics, diagrams, or animations to communicate a message. and constraints are advanced till a final layout that fulfils the layout necessities is met.[1] A primary purpose of information visualization is to communicate facts truly and correctly through statistical images, plots and facts photographs with the emergence of huge and high dimensional facts sets, the task of statistics visualization has come to be an increasing number of crucial in clustering system gaining knowledge of and information mining[2]. Multidimensional analysis is a data analysis process that groups data into two categories: Data dimensions and measurements. A popular way to carry out such an automatic evaluation is via learning a low-dimensional embedding of the statistics. in a low-dimensional embedding, every (excessive-dimensional) object is represented by using a low-dimensional factor in this kind of manner, that nearby factors correspond to comparable objects and that remote points correspond to multiple objects. The low-dimensional embedding can effortlessly be visualized in, to expose a hard and fast of points in an n-dimensional area, a backdrop is drawn at the side of n parallel traces, commonly vertical and in addition spaced. a point in n-dimensional (N>=3) statistics visualization, we have numerous methods to do it like use regular 2/three-dimensional algorithms plus numerous attributes to symbolize facts in other dimensions, use animation and many others. space is represented as a polyline with vertices on the parallel axes. the position of the vertex on the n-th axis corresponds to the n-th coordinate of the factor to are required to determine the position of elements. e.g., a scatter plot or a parallel coordinate plot, or it could be used as the premise for the construction of greater advanced visualizations, together with magnificence-conditional density maps [3]

Various dimension reduction techniques have been introduced based on the aim of preserving specific properties of the original data. A plethora of embedding strategies had been proposed over the last decade with the aid of tenenbaum et al. (2000); roweis and saul (2000); saul et al. (2006); and van der maaten and hinton (2008); van der maaten et al. (2009)   carreira-perpi~n_an (2010); and burges (2010 lawrence (2011); laplacian eigenmaps. non-linear techniques can be widely labelled into two companies: those who provide a mapping (either from the high-dimensional space to the low-dimensional embedding or vice versa), and people that just supply a visualisation because it's dimensional we use a unique color for the factors to indicate the distinction inside the fourth size records. within the context of system mastering. (gadget gaining knowledge of is a subject of computer technology and artificial intelligence that gives computers the potential to study without being sincerely programmed) mapping methods (data mapping is the system of making statistics element mappings among two wonderful facts fashions) may be viewed as a preliminary function extraction step, and then pattern reputation algorithms are carried out. commonly people who just provide a visualization are primarily based on proximity data – that is, distance measurements. because in excessive dimensional spaces, most effective small pair wise distances are dependable, maximum of those techniques handiest try to accurately version such small pair smart distances inside the low-dimensional.[4]

This paper is organized as follows. In section 2, we discuss the related work. Section 3, we introduce different visualization tools., Section 4 concludes this work.

II Related work

We survey work on the different method of graphical representation techniques for visualization of large datasets represented as

graph data. As the velocity of information boom exceeds moore's law at the start of this new century, immoderate statistics is making first rate troubles to humans. but, there are so much ability and relatively useful values hidden in the huge extent of facts. To overcome this problem, different visualization techniques are used. By following some of these techniques, we will reach the conclusion, Some of these Visualization techniques are

(1). t-SNE. (2) SNE (3) PCA, (4) Isomap, (5) LLE, (6) CCA ,(7) Sammon mapping

Stochastic Neighbour Embedding

Generally, dimensional reducing methods convert a high-dimensional data into a less-dimensional data set. Stochastic Neighbour Embedding (SNE) starts by converting the high-dimensional Euclidean distances between data points in conditional probabilities that represent similarities by Hinton and Roweis. The similarity of the data point is to the data point xi is the conditional probability. The similarity of data point x_i is the conditional probability. $P_{j|i}$, x_i will pick x_j as it neighbour if neighbour were picked in properties to their probability desity under a gaussian cantered at x_i . For nearby data points, Centred data points, pj|i are relatively high, whereas for widely separated data points, pj|i will be almost infinitesimal (Gaussian, _i) of the variance for reasonable values. Mathematically, the conditional probability pj|i is given by

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|^2/2\sigma_i^2\right)},$$

(1)

Where σi is that the variance of the Gaussian below this cantered information. the tactic for xi deciding the worth of σi is given later during this section. as a result of we have a tendency to area unit solely inquisitive about modelling try wise similarities, we have a tendency to set the worth of pi|i to zero. For the low-dimensional Counterparts Loloish and yj were encapsulated high-dimensional knowledge points xi and xj, it's doable then calculate the same chance, that we have a tendency to denote by qj|i. The set2 the variance of the Gaussian that's used within the computation of the conditional possibilities. qj|i to $\frac{1}{\sqrt{2}}$, we have a tendency to model the similarity of map purpose yj to map point yi

$$q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|^2\right)}.$$

(2)

again, considering that we are interested in modelling pair smart similarities, we set set qi | i = 0. if the map factors yi and yj are correctly model the similarity among the excessive-dimensional data points xi and xj, the conditional possibilities i and qj can be identical. motored with the aid of this statement, SNE targets to discover a low-dimensional records illustration that minimizes the in shape among i. a herbal measure i models pj that's in this situation same to the go-entropy as much as an additive consistent). SNE minimizes the sum of kullback-leibler divergences usual records points the use of a gradient descent method [5]. The cost function C is given by

$$C = \sum_i KL(P_i||Q_i) \sum_i \sum_j p_{j|i} \log\frac{Pji}{Pji},$$

(3)

during this respect, SNE is inferior to strategies that permit umbel-like optimization Associate in Nursing it'd be helpful to search out an optimization technique that offers sensible results with outré squiring the additional computation time and parameter selections introduced by the simulated hardening.

t-Distributed Stochastic Neighbor Embedding

presenting a brand-new technique referred to as t-SNE that visualizes high dimensional knowledge by giving every information a location in a very 2 or 3-dimensional map. that the question is however will we have a tendency to get a feel, however will we have a tendency to get Associate in Nursing intuition for and the way this high dimensional knowledge area unit organized in knowledge space? if we glance at visualize image info they need developed lots of technique visualization image.

We area unit acquainted with parallel coordination, radial graph layout, tree maps and that they visualize few variable at a time, that the question is the way to visualize terribly high demonical knowledge at a time. however will we have a tendency to take care of terribly sizable amount of variables? and then straightforward plan to resolve this downside. Would be to do and build a map during which, every of high demonical object goes to represent by a going by some extent in such the simplest way that similar objects area unit getting to be represent and also the dissimilar objects area unit getting to be painted by distance points. and now were

going plant in some low dimensional or is also two or three dimensions. If we have a tendency to manage the right construct the map then we will lay out this 2 3dimensional map [6]. Basically we can visualize is as a scatter plot. The distance in low dimensional map reflect the similarities in the original high dimensional data to minimize some objective function that basically measures the discrepancy and so we want to end up with is more something like this, where distance in low dimensional map reflect the original high dimensional data, some objective function that basically measures the discrepancy between the similarity in the original data and the dissimilar in the map this is called a dimension reduction or embedding or multidimensional scaling. which basically finds a liner projection in high dimensional data in such a w

Fig.1 Visualization by t-SNE

Automatic dimensionality reduction is an important "toolkit" operation in machine learning, both as a pre processing step for other algorithms. Let's perceive why you ought to realize t-SNE and also the algorithmic details of t-SNE. t-SNE is Associate in Nursing improvement on the random Neighbor Embedding (SNE) algorithmic rule.

In t-SNE, we have a tendency to use a Student t-distribution with one degree of freedom (which is that the same as a Cauchy distribution) because the heavy-tailed distribution within the low-dimensional map. victimization this distribution. to the current finish, t-SNE defines joint possibilities foreign terrorist organization that live the try wise similarity between objects xi and xj by symmetrizing 2 conditional possibilities.

Similar to SNE we set pii and qii to 0. they refer to this of SNE as symmetric SNE
Let xi tb the ith object in high dimensional space
Let yi tb the ith object in low dimensional space
So, the way they construct the condition probability of the similarity pj is given i, so if you have point j and i is given be this formula.

$$p_{ij} = \frac{\exp(-||y_i - y_j||^2 / 2\sigma_2{}^2)}{\sum_{k \neq i} \exp(-||y_k - y_l||^2 / 2\sigma_2{}^2)}$$

(4)

its look like a kernel of normal distribution. The main thing that from this formula, if you have two data point that are very close to each other xi-xj where xj is going to be small close to ZERO so when we exponent ate a number that is close to ZERO it's going to be around 1 and that for the reason if two points are relatively close to each other then the conditional probability pj given I is going to be large but if two point xi and xj far apart this difference is going to be very large and then we put negative sign and exponent ate a very large negative number it's going to be very close to zero and in that case pj given i is to be going to be infinitesimal[7].

$$q_{ij} = \frac{\exp(-||x_i - x_j||^2 / 2\sigma_2{}^2)}{\sum_{k \neq i} \exp(-||x_k - x_l||^2 / 2\sigma_2{}^2)}$$

(5)

l and then qj given I is going to be the conditional probability similarity in the low dimensional space and its very similar to pj given i except for the fact that authors assume that the variance in the low dimensional space. After calculating this two probabilities there is also an additional assumption that if you have comparing the exact same point i and i is
pi|i= qi|i= 0
after constructing these two conditional probabilities in both high and low dimensional space the objective is to produce the difference between pj|i and qj|i again and when we project high dimensional data to low dimensional it looks as similar to the high dimensional as possible.

$$C= KL(P||Q) \sum_{i} \sum_{j} p_{ij} \log \frac{Pji}{Pji},$$

(6)

As another to minimizing the add of the Kullback-Leibler divergences or KL divergences between the conditional possibilities pj|i and qj|i, it's additionally doable to reduce one Kullback-Leibler divergence between a chance distribution, P, within the high-dimensional houseand a chance distribution, Q, within the low-dimensional space

This is our cost function C and its represented as KL($P||Q$) which is probability distribution of point i in the high dimensional.

we have a tendency to subdivide techniques for spatiality reduction into umbel-like and non-convex techniques. umbel-like techniques optimize Associate in Nursing objective perform that doesn't contain any native optima, whereas non-convex techniques optimize objective functions that do contain native optima. The any subdivisions within the taxonomy area unit mentioned within the review within the following 2 sections..
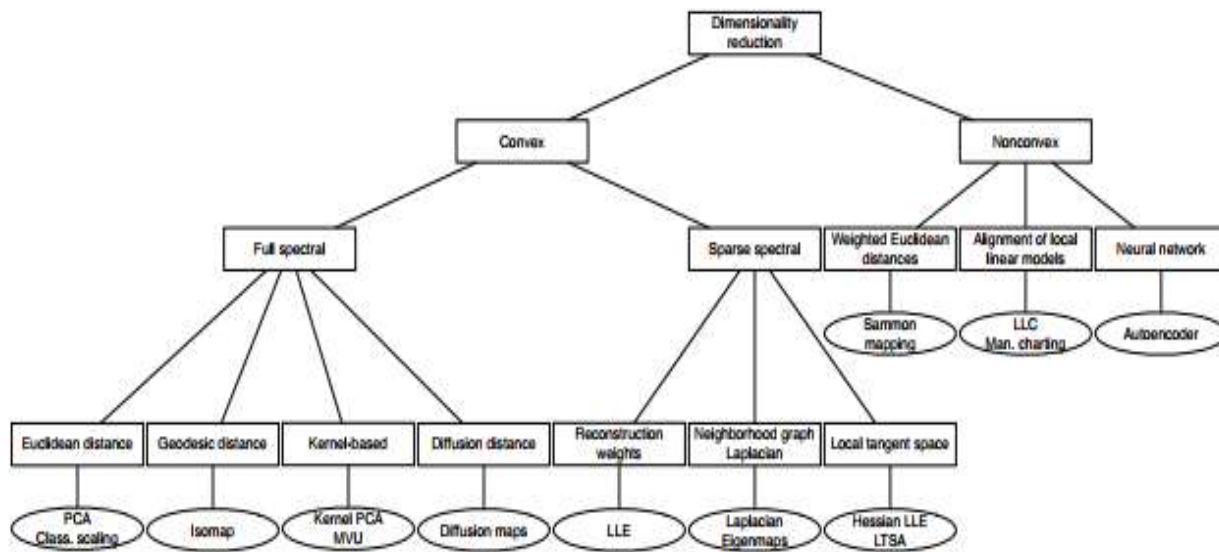

Fig. 2 dimensionality reduction techniques

To evaluate t-SNE, we have a tendency to gift experiments during which t-SNE is compared to seven alternative non-parametric techniques for spatiality reduction., we have a tendency to solely compare t-SNE with: one. PCA (2) Sammon mapping, (3) Isomap, and (4) LLE. (5) CCA,
 (6) SNE, 7) Sammon mapping

visualization image techniques will facilitate researchers explore terribly massive amounts of discretionary, three-dimensional knowledgeand notice attention-grabbing knowledge sets-hot spots, clusters of comparable knowledge, or correlations between totally different dimensions Our approach to those "Data mining" functions retrieval with new data mental image techniques Visualization techniques can help researchers explore very large amounts of arbitrary, multidimensional data and find interesting data sets-hot spots, clusters of similar data, or correlations between different dimensions Our approach to these "Data mining" functions and information retrieval with new data visualization techniques[8]**.**

PCA (principal component analysis)
The central plan of principal element analysis (PCA) is instructional of a knowledge set consisting of an outsized range of reticulate variables whereas retentive the maximum amount as doable of the variation gift within the dataset. this can be achieved by remodel ingto a brand new set of variables, the principal elements (PCs), that area unit unrelated, and that area unit ordered so the primary few area unit most of the variations gift all told the initial variables.[9]. we have a tendency to accomplish by protruding knowledge from the next dimension to a lower dimensional manifold specified the error incurred by reconstructing the info within the higher dimension is decreased.
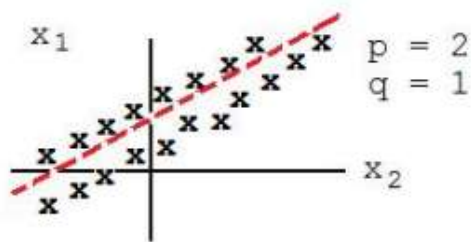
Fig 3 2D data projection

Figure 3: A plot of x's in 2D $\mathbb{R}^p$ space and an example 1D $\mathbb{R}^q$ space (dashed line) to which the data can be projected.

An example of this can be given by Figure three, wherever second knowledge are often projected to the one D house painted by the broken line with fairly tiny. Suppose that we have a tendency to feed a group of three-dimensional points that everyone lie on a two-dimensional plane to PCA. PCA can come 2 vectors that span the plane alongside a 3rd vector that's orthogonal to the plane. the 2 vectors that span the plane are given a positive weight, however the third vector can have a weight of zero, since the info doesn't vary on that direction. PCA can notice a basis for the linear topological space and permit one to disregard the inapplicable options. is termed a manifold. but it doesn't mean that t-SNE is healthier than PCA. t-SNE may be a nice piece of Machine Learning however one will notice several reasons to use PCA rather than it.

- *Stochasticity of final solution*- One gets a nice visualisation and then her colleague gets another visualisation and then they get artistic which looks better and if a difference of $0.03\%0.03\%$ in the $KL(P\|Q)KL(P\|Q)$ divergence is meaningful.
- *Interpretability of mapping*. This relates to the above point but let's assume that a team has agreed in a particular random seed/run. Now the question becomes what this shows. t-SNE tries to map only local / neighbours correctly so our insights from that embedding should be very cautious; global trends are not accurately represented (and that can be potentially a great thing for visualisation). On the other hand, PCA is just a diagonal rotation of our initial covariance matrix and the eigenvectors represent a new axial system in the space spanned by our original data
- *Application to new/ unseen data*. t-SNE is *not* learning a function from the original space to the new (lower) dimensional one and that's a problem. On that matter, t-SNE is a non-parametric learning algorithm so approximating with parametric algorithm is an ill-posed problem. The embedding is learned by directly *moving* the data across the low dimensional space. That means one does not get an *eigenvector* or a similar construct to use in new data. In contrast, using PCA the eigenvectors offer a new axes system what can be directly used to project new data
- *Incomplete data*. Natively tt-SNE does not deal with incomplete data. In fairness, PCA does not deal with them either but numerous extensions of PCA for incomplete data (eg. Probabilistic PCA) are out there and are *almost standard* modelling routines'-SNE currently cannot handle incomplete data.
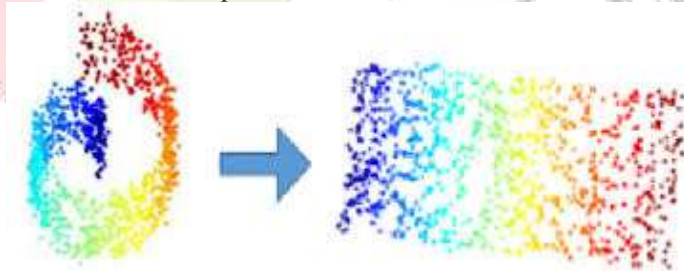

Fig. 4 Data visualization by PCA

Isomap

Isomap (Tenenbaum et al., 2000) is an interesting variant of MDS which again finds a configuration of output coordinates matching a given distance matrix. The difference is that Isomap does not compute pairwise input-space distances as simple Euclidean distances but as geodesic distance*s* along the manifold of the data (technically, along a graph formed by connecting all *k*-nearest neighbors)[10].

What's a reliable, even on these type of manifolds are very small once, very small Euclidean distances, between points and their nearest neighbours. those are still pretty accurate but also on very curve manifold like PCS. So the coming up a new technology that ware basically focusing on preserving this local structure and preserving small pair wise distance in the map and as an example of such a technique is called isomap. Isomap is a nonlinear dimensionality reduction method .it tries to estimate the distance between points in the original space via some geodesic distance tries to estimate the distance and then uses those distance as input into PCA On input the

Isomap algorithm requires the distances dX(i,j) between all pairs i, j from N data points in the high-dimensional input space X, measured using either the standard Euclidean metric, or some domain-specific metric. The algorithm outputs coordinate vectors Yi in a (lower) d-dimensional Euclidean space Y that best represents the intrinsic geometry of the data[11]

- Isomap performs better than PCA and LLE for nonlinear data sets.
- The optimal parameter for modified Isomap, discriminate isometric mapping, can be chosen using our method, in the same manner as for Isomap

Fig. 5 Data visualization by PCA

LLE (LOCALLY LINEAR EMBEDDING)

we attempt to estimate the space over the manifold within the case ,and you'll be able to see that embedding is bit higher. for intense the cluster of once, that is orange cluster here, is fairly well secreted, from the remainder of the info. PCA and MDS area unit secured,
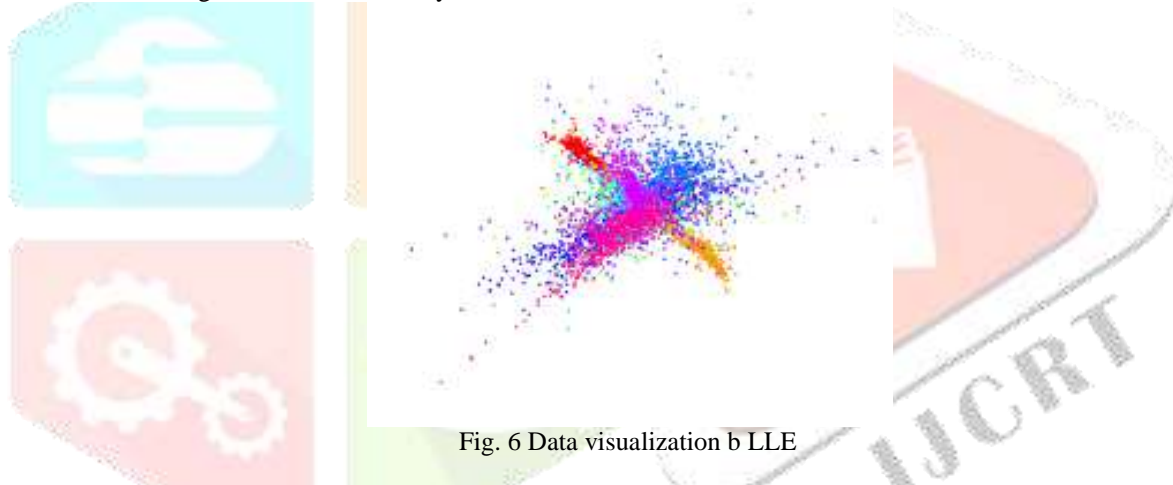
Fig. 6 Data visualization b LLE

To recover truth structure of linear manifolds, Isomap is secured asymptotically to recover truth spatiality and geometric structure of a strictly larger category of nonlinear manifolds. just like the Swiss rol[12] but even here, if we would take out the colours you would not see so much structure. So the disadvantage of isomap is

- Graph discreteness overestimates the geodesic distance
- k must be high to avoid "linear shortcuts" near regions of high surface curvature

It embedding a bit better for intense cluster of once, another method Is called a locally liner imbedding. The LLE is very similar to t-SNE in the sense that is really tries to preserve small pair wise distances. The input of the LLE algorithm consists of – dimensional vectors (X). The output consists of – dimensional vectors (Y) The LLE algorithmic rule has 3 steps. within the commencement, one identifies neighbours of every information . totally different criteria for neighbour choice are often adopted; the only chance is to decide on the -nearest neighbours consistent with the geometrician distance. within the second step, one computes the weights that reconstruct every information best from its neighbours , minimizing subject to the constraints and this can be a typical affected statistical procedure optimization downside, which might be simply answered by finding a linear system of equations. The third step consists in mapping every information to a low-dimensional vector Y , that best preserve high-dimensional neighbourhood pure mathematics painted by the weights[13] . That is, the weights are fixed

Sammon mapping

However as you'll be able to see it's a sort of things wherever what's love to try and do is collapse all points till the origin .this is primarily the origin of the info and so it uses a bunch of outliers to satisfy a really straightforward variance constraint that it has one on this embedding[14]. therefore you embedding with LLE may be a whole bunch of points collapse, and few string flying out ,basically satisfy this variance constraint that. we have a tendency to primarily turn on the concept of LLE and that we tries to come back up with a higher technique that doesn't have this downside attempting to collapse all purpose onto the only purpose Sammon

mapping or Sammon   projection is Associate   in   Nursing algorithmic   rule that   maps a   high-dimensional house to   of lower spatiality by attempting to preserve the structure of inter-point distances in high-dimensional space within the lower-dimension projection

So you embedding with LLE is a whole bunch of points collapse, and few string flying out ,basically satisfy this covariance constraint that. we basically build on the idea of LLE and we tries to come up with a better technique that doesn't have this problem trying to collapse all point onto the single point [15]

Fig. 7 Data visualization by sammon mapping

SM learns Associate in Nursing implicit non-linear projection from the data's original high-dimensional house to, typically, a 2- or third-dimensional projection house. As you'll be able to see SM learns Associate in Nursing approximate isometric from one house to the opposite. However, it's the actual fact that it will use nearly every which way outlined dissimilarities that created SM a really helpful and broadly speaking applicable technique in spatiality reduction, manifold learning, preliminary knowledge analysis and specially mental image of information. sammon mapping the space live between 2 patterns is often the Euclidean metric and error function(of n pattern) to minimize the sammons stress outlined. When the gradient-descent method to go looking for the minimum couldn't reach so, a major range of experiments with totally different random initializations is also necessary implementation becomes inappropriate. SAMANN network for Sammon's projection, nonlinear projection supported Kohonen's Kyrgyzstani monetary unit (NP-SOM), Moreover, the SAMANN network offers the generalization ability of protruding new knowledge, that isn't gift within the original Sammon's projection algorithm [16].

CCA (curvilinear component analysis)

Another mather is curvilineal element Associate in Nursingalysis (CCA) this algorithmic rule was projected as an improvement to the Kohonen self-organizing maps, CCA may be a technique of knowledge illustration that borrows each the ideas of variable data analysis. The aim of CCA is primarily to relinquish a revealing illustration of information in low dimension, getting ready a basis for any bunch and   classification. this   type of illustration helps to   grasp the   structure of the info set and so to   pick out the suitable techniques   for any automatic process.(curvilineal element Analysis: A   Self-Organizing   Neural   Network for nonlinear Mapping of information Sets capital of South Dakota Demartines and Jeanny Heraul). the essential plan behind CCA lies aloof from the PCA. CCA tries toreproduce the information `topology' from the initial house $\mathbb{R}_n$ to the projection house $\mathbb{R}_p$ . In CCA  spirit,  the  word  `topology' suggests that the distances between all pairs of vectors within the database'. Actually, CCA merely tries to finish vectors xp$\mathbb{R}_p$ within the projection house $\mathbb{R}_p$ specified they reproduce the distances measured within the initial house $\mathbb{R}_n$. Here CCA differs from ancient MDS strategies. the concept is to target conserving distances between points that area unit close to one another within the mental image. The cost function .[17]

$$E = \frac{1}{2}\sum_i \sum_{j \neq i}(d(x_i, x_j) - d(y_i, y_j))^2 F\big(d(y_i, y_j), \lambda_y\big).$$

(7)

The term F(d(yi , yj), λy) determines how strongly errors in reproducing the distance between the points i and j contribute to the cost function.
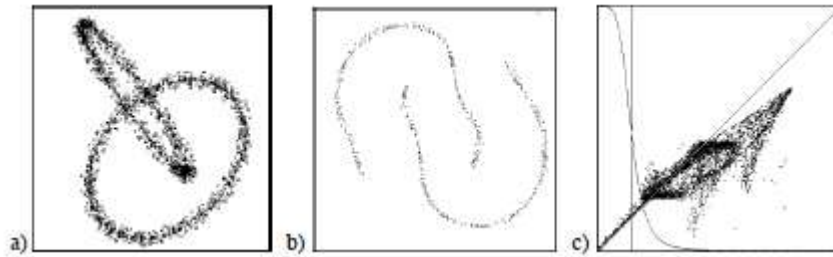
Fig 8. Mapping of a 3D knowledge set of 2 latticelike rings onto a 2-D illustration house. a) input house, b) output house wherever the 2rings area unit tamed order to satisfy at the best the 2-D illustration, c) the dx/dy distribution showing the native projection and also the quality of the flowering
Conclusion

III Conclusion

We will measure the performance of the seven strategies for the visual judgment of the second 3D projection maps. These method are: 1)**t- SNE,** 2**) SNE,** 3) Rubner's **PCA**, 4) **LLE** , 5) **CCA** , 6) **SAMANN,** and 7) **ISOMAP.**

, that retain native structure of knowledge– area unit a lot of seemingly to be influenced by tiny changes in each data and parameter variations. They additionally tend to produce littered visualizations, whereas knowledge points in t-SNE, Isomap and PCA area unit far more scattered.   t-SNE, because   of the   character   of   its   gradient,   tends to   create   tiny clusters within   the embedding.   The projection of graph-based techniques –Isomap, and LLE– might dissent lots reckoning on the info, whereas within the case of SNE and t-SNE –or PCA, that isn't shown–, the projections area unit fairly stable for identical parameter. Isomap has the advantage of speed and algebraically solidity.   CDA depends on a lot of difficult techniques like vector quantisation and random gradient.   Though, once these options area   unit well   parameterized, they   permit CDA to relinquish results that preserves higher some characteristics of the projected knowledge set.

Although we've shown that t-SNE favourably to alternative technologies for knowledge mental image, t-SNE has 3 potential weaknesses: (1) it's unclear however t-SNE performs on general dimensionality-reduction tasks, (2) the comparatively native nature of t-SNE intrinsic spatiality of the info, and (3) t-SNE isn't sure to converge to a world optimum of its value perform Below, we have a tendency to discuss the three weakness in additional detail.

The LDA network additionally implements a linear, however not essentially orthogonal, projection. It tries to search out a plane within the high dimensional house and so skews the plane so as to extend the discriminatory ability of each the projection axes.

IV References

[1]   S. Goguelin, J. M. Flynn, W. P. Essink, and V. Dhokia, "A Data Visualization Dashboard for Exploring the Additive Manufacturing Solution Space," *Procedia CIRP*, vol. 60, pp. 193–198, 2017.

[2]   T. Iwata, N. Houlsby, and Z. Ghahramani, "Active Learning for Interactive Visualization," *Jmlr*, vol. 31, pp. 342–350, 2013.

[3]   M. Oakleaf, "Writing information literacy assessment plans: A guide to best practice," *Commun. Inf. Lit.*, vol. 3, no. 2, pp. 80–90, 2009.

[4]   L. van der Maaten, "Accelerating t-SNE using Tree-Based Algorithms," *J. Mach. Learn. Res.*, vol. 15, pp. 3221–3245, 2014.

[5]   G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," *Adv. Neural Inf. Process. Syst.*, pp. 833–840, 2002.

[6]   M. B. Pouyan and M. Nourani, "Flow-SNE : A New Approach for Flow Cytometry Clustering and Visualization," 2015.

[7]   J. F. Kruiger, P. E. Rauber, R. M. Martins, A. Kerren, S. Kobourov, and A. C. Telea, "Graph Layouts by t-SNE," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 283–294, 2017.

[8]   D. A. Keim and H. Kriegel, "VisDB : Database Exploration Using Multidimensional Visualization."

[9]   I. T. Jolliffe, "Principal Component Analysis, Second Edition."

[10]   "FEATURE SPACE DIMENSIONALITY REDUCTION FOR THE OPTIMIZATION OF VISUALIZATION METHODS Andreea Griparis , Daniela Faur Department of Applied Electronics and Information Engineering , Politehnica University of Bucharest , Romania Mihai Datcu German Aerospace," pp. 1120–1123, 2015.

[11]   I. T. Jolliffe, "Principal Component Analysis and Factor Analysis," pp. 2–3, 1986.

[12]   E. Ettinger, V. de Silva, and J. B. Tenenbaum, "Global (ISOMAP) versus Local (LLE) Methods in Nonlinear Dimensionality Reduction," 2005.

[13]   R. Karbauskaite, O. Kurasova, and G. Dzemyda, "Selection of the Number of Neighbours of Each Data Point in the Locally Linear Embedding," *Inf. Technol. Control*, vol. 36, no. 4, pp. 359–364, 2007.

[14]   M. P. Olga Kouropteva, Oleg Okun, "Selection of the Optimal Parameter Value for the Locally Linear Embedding Algorithm," *1st Int. Conf. Fuzzy Syst.*, no. 1, pp. 359--363, 2002.

[15]   C. J. Wang, H. Fang, and H. Wang, "ESammon : A Computationaly Enhanced Sammon Mapping based on Data Density," 2016.

[16]   J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *IEEE Trans. Neural Networks*, vol. 6, no. 2, pp. 296–317, 1995.

[17]   J. Hérault, C. Jausions-Picaud, and A. Guérin-Dugué, "Curvilinear component analysis for high-dimensional data representation: I. Theoretical aspects and practical use in the presence of noise," vol. 1607, no. March, pp. 625–634, 1999.