Information Sciences 370–371 (2016) 288–302

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Improved data visualisation through multiple dissimilarity modelling

Iain Rice

*Aston University, UK*

ARTICLE INFO

*Article history:*
Received 9 May 2016
Revised 22 July 2016
Accepted 28 July 2016
Available online 4 August 2016

*Keywords:*
Dissimilarity
Visualisation
Multidimensional scaling
Euclidean

ABSTRACT

Popular dimension reduction and visualisation algorithms rely on the assumption that input dissimilarities are typically Euclidean, for instance Metric Multidimensional Scaling, t-distributed Stochastic Neighbour Embedding and the Gaussian Process Latent Variable Model. It is well known that this assumption does not hold for most datasets and often high-dimensional data sits upon a manifold of unknown global geometry. We present a method for improving the manifold charting process, coupled with Elastic MDS, such that we no longer assume that the manifold is Euclidean, or of any particular structure. We draw on the benefits of different dissimilarity measures allowing for the relative responsibilities, under a linear combination, to drive the visualisation process.



## 1. Introduction

In order to create humanly interpretable visualisations, popular algorithms which chart high dimensional data assume some global or local geometric structure. Methods such as the Sammon map [25], Stochastic Neighbour Embedding (SNE) [12] and its' derivatives, the Gaussian Process Latent Variable Model [13], the Generative Topographic Map (GTM) [5], Metric Multidimensional Scaling (MDS) and Curvilinear Component Analysis (CCA) [7] assume global Euclidean structure. The work of [27,29,30] removes the metric constraint in MDS and CCA using the Bregman divergence, however the Euclidean dissimilarity over inputs is retained. Typically observation manifolds are better characterised by Riemannian manifolds than the restrictive Euclidean spaces [2]. These manifolds can be assumed to have locally Euclidean structure, as in Locally Linear Embedding [24], Laplacian Eigenmaps [4], Riemannian Manifold Learning [18] and methods using geodesic distances based upon local Euclidean structure such as Isomap [31], the Geodesic Nonlinear Map [17] and Curvilinear Distance Analysis [16].

These methods relying on a local Euclidean structure, assuming smooth continuity between local charts. This assumption can lead to a poor approximation of the manifold chart when observations are complex non-Euclidean structures or have a fractal dimensionality. Further to this the local estimates of a chart in high dimensional space can be unreliable, particularly when data is sparse. These methods are also sensitive to the choice of the neighbourhood size parameter, potentially leading to false neighbourhoods in the mapping of data.

In particular we can consider the case of GTM which assumes the observations are distributed according to an isotropic Gaussian, akin to a hypershpere, which in high dimensions contains nearly all of its' mass in a thin shell sitting upon the surface (see [15] for an overview of this). This particular case highlights the need for a more thorough approach to characterising reliable dissimilarities over observations in the generation of visualisations.

*E-mail address:* i.rice@aston.ac.uk

http://dx.doi.org/10.1016/j.ins.2016.07.073

The combination of different dissimilarity operators has recieved much interest in recent years, particularly in the area of multiple kernel learning. The use of these multiple kernels allows for measures inducing diverse topologies upon observations to warp high dimensional data, often with complex structure, in order to achieve better regression and classification performance, outlined in [6,10], as well as in the field of manifold learning [1]. The tasks of regression and classification are supervised by nature in that true targets exist. In this paper we consider the unsupervised case where the targets are learned by minimisation of a mapping cost function, such that the data structure is preserved. One such case where a non-Euclidean dissimilarity measure was used to characterise observations is described in [20] where class label information was used as a descriptor to motivate the visualisation process. We present an extension to this work where we no longer rely on a single dissimilarity measure. Similar work was performed in [19] relying on linear discriminant analysis to form visualisations, however the learned mapping is linear and the optimisation method does not generalise to nonlinear mappings as in this paper.

This paper details a method for incorporating several dissimilarity measures into a visualisation framework. By learning a mixture representation for the observation space the learned visualisation is more interpretable and better spans the latent dimensions than the Euclidean counterpart. It was found in [15] that visualisation mappings which learn the position of latent points through optimisation of a non-convex cost function (such as MDS, SNE and CCA) perform better than those whose cost functions are convex (such as PCA, LLE and Isomap). SNE, CCA and their variants require tuning of the mapping parameters such as the perplexity and neighbourhood parameters in CCA and Culrvilinear Distance Analysis. The more traditional approaches of the Sammon map and MDS require no such tuning. Despite constructing a topographic mapping, the local focus of the Sammon mapping is improved by using Elastic MDS as shown in [27]. Further local focus can be achieved with other variants of MDS, however the optimisation of the resulting cost functions requires stochastic gradient descent and poor local minima are likely in the optimisation procedure. We therefore focus on the specific case of Elastic MDS to show the potential improvements in this paper, though the approach does generalise to the other methods mentioned. Five standard datasets are analysed with the proposed mixture dissimilarity approach qualitatively compared to the standard MDS method. Following [33] quantitative quality measures typically used for visualisations are not appropriate so we rely on a visual comparison to show improvement.

## 2. The learning task

The method described in this paper is born from the desire for robust dimension reduction, however we focus on the particular case of visualisation where the mapped data is 2-dimensional. Elastic MDS [22] embeds a dataset, $X$, into a reduced dimensional space, $Y \in \mathbb{R}^V$. In the experiments of this paper we fix $V = 2$. $X$ can be nonvectorial as all that is required for the projection to $Y$ is a matrix of pairwise relative dissimilarities, $D_x(i, j)$, between observations $X_i$ and $X_j$. The latent points corresponding to each observation, denoted $\boldsymbol{y}_i$, are learned through gradient descent of the cost function:

$$E = \sum_{i,j<i} \frac{(D_x(i, j) - D_y(i, j))^2}{(D_x(i, j))^2}, \tag{1}$$

where $D_y(i, j)$ denotes the dissimilarity between visualised points $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, taken as standard to be the Euclidean distance. The term Elastic MDS is due to the quadratic factor in the denominator of Eq. (1), as opposed to that of the Sammon map. This has the effect of forcing the mapping process to focus more on local than on global distance preservation. We naturally desire to observe a physically motivated clustering of observations in a visualisation without imposing this in the mapping procedure.

In order to remove the restrictive assumption that $D_x$ consists of Euclidean distances only we treat the measure as a weighted combination of multiple separate dissimilarity measures:

$$D_x = \sum_{l} \alpha_l D_x^l$$

where $\alpha_l$ is the weight corresponding to the $l$-th dissimilarity measure. To construct a more flexible input dissimilarity matrix we utilise 15 dissimilarity measures for the experiments in this paper. Our approach is flexible such that additional measures can be included and removed if desired. Table 1 taken from [23] lists 14 of the dissimilarities and the final measure is the geodesic distance given by Dijkstra's algorithm [8] as in Isomap and other related algorithms. In this paper we restrict our discussion and experiments to the case where observations are vectorial. However, the approach of learning a mixture of input dissimilarities is generic and a trivial change of the measures of Table 1 to other measures allows for analysis of other data structures. Typical non-vectorial examples include probability distributions, binary data, images, graphs and time series for instance. Dissimilarity measures specific to these applications are detailed in [23].

For the experiments described in this paper the weighted Euclidean (measure 2) parameters were fixed to be the inverse of the sample mean vector and $p$ in the Minkowski distance (measure 5) was fixed to be 1.2 to induce a metric between the city block and Euclidean measures. The weighting matrix, $C$, in the Mahalanobis distance (measure 6) is the data sample covariance matrix. These choices of parameters allow the measures to be invariant to the absolute scaling of the data in each of the observed dimensions such that a single dimension does not disproportionately affect the overall dissimilarity. The geodesic distance (measure 15) is performed over the Euclidean distance as is typical in the literature.