Rectifying Non-Euclidean Similarity Data using Ricci Flow Embedding

Weiping Xu  Edwin R. Hancock  Richard C. Wilson  
Dept. of Computer Science  
University of York  
York, UK  
{elizaxu, erh, wilson}@cs.york.ac.uk

Abstract—Similarity based pattern recognition is concerned with the analysis of patterns that are specified in terms of object dissimilarity or proximity rather than ordinal values. For many types of data and measures, these dissimilarities are not Euclidean. This hinders the use of many machine-learning techniques. In this paper, we provide a means of correcting or rectifying the similarities so that the non-Euclidean artifacts are minimized. We consider the data to be embedded as points on a curved manifold and then evolve the manifold so as to increase its flatness. Our work uses the idea of Ricci flow on the constant curvature Riemannian manifold to modify the Gaussian curvatures on the edges of a graph representing the non-Euclidean data. We demonstrate the utility of our method on the standard “Chicken pieces” dataset and show that we can transform the non-Euclidean distances into Euclidean space.

Keywords—similarity; embedding; Ricci flow;

I. INTRODUCTION

Although individual object feature-vectors provide a convenient way of characterizing an arrangement of objects, dissimilarity representations are sometimes a more natural way of capturing the relationships between objects. In the dissimilarity representation pairwise dissimilarity (or proximity) measure describes the properties of objects in terms of their differences in attributes. Concrete examples of such representations are provided by weighted proximity graphs. As an example of the advantages of such a representation, if a proximity weight matrix is used to represent a distorted arrangement of points [1], then tasks such as point-set alignment can be effected without the need for an explicit image deformation model. These distances, however, are quite commonly non-Euclidean and prevents the use of many geometrically based learning techniques.

Embedding provides a way to apply feature-based classifiers to dissimilarity data and produces a vectorial representation by projecting dissimilarity data into a fixed-dimensional vector space. Multidimensional scaling (MDS) [2] is one of the earliest embedding techniques to find a vectorial representation in a Euclidean space. More recent approaches such as ISOMAP [3] attempt to simultaneously reduce the dimensionality of the embedded space whilst minimizing distortion. This is achieved by inferring a low dimensional manifold on which the data resides.

The common aim of the above embedding methods is to locate a low-dimensional representation. In order to apply non-Euclidean dissimilarity data with traditional geometry based learning techniques, we must attempt to rectify the data so as to minimize the non-Euclidean artifacts. One way is to consider the positive definite subspace of the distances [4]. An alternative route adopted by Pekalska et al. [5] is to add a suitable constant amount to the off-diagonal elements of the dissimilarity matrix. This is equivalent to adding a certain constant to all eigenvalues of the related Gram matrix, and thus compensating for the effect of the negative eigenevalues, while maintaining the same eigenvector structure.

An important question that arises is how to correct the dissimilarity such that the new Euclidean distances are not less discriminating than the non-Euclidean dissimilarity. Our work develops the idea of Ricci flow on a constant curvature Riemannian manifold by evolving the distance measures through updating Gaussian curvatures on the edges of a graph, so that non-Euclidean distance measures can be rectified (i.e. made Euclidean).

II. RICCI FLOW

Our aim is to rectify a given set of similarities so as to make them more Euclidean. Firstly, we consider the objects of interest to be represented by points on a manifold, and the given dissimilarities to be the geodesic distances on the manifold between these points (geodesic distances). For an arbitrary set of non-Euclidean similarities the manifold will be curved. In contrast, a Euclidean space will be flat and the geodesic and Euclidean distances are identical. Our task is then to remove the curvature from the manifold to create a corrected set of Euclidean distances. We achieve this by evolving the manifold using Ricci flow.

Introduced by Richard Hamilton in 1981, the Ricci flow [6] evolves a manifold so that the rate of change of the metric tensor is controlled by the Ricci curvature. Essentially, this is an analogue of a diffusion process for a manifold. The geometric evolution equation is:

$$\frac{dg_{ij}}{dt} = -2R_{ij}$$  (1)
where $y_{ij}$ is the metric tensor of the manifold and $R_{ij}$ is the Ricci curvature.

We model the embedding manifold as consisting of a set of local patches with individual constant Ricci curvatures. These patches can be either elliptic (of positive sectional curvature) or hyperbolic (of negative sectional curvature). It is straightforward to re-express the Ricci flow in terms of the sectional curvature $K$:

$$\frac{dK}{dt} = \begin{cases} -2K^2 & \text{elliptic hypersphere,} \\ 2K^2 & \text{hyperbolic space.} \end{cases}$$

(2)

Under this evolution, the curvature moves towards zero for both types of patch, flattening the manifold. The solution of the differential equation is straightforward. Commencing with the initial conditions $K = K_0$ at time $t = 0$, then at time $t$ we have

$$K = \frac{K_0}{1 \pm 2K_0t}$$

(3)

with the positive sign for the elliptic space.

III. CURVATURE COMPUTATION

We use a Euclidean embedding of the points and use the difference between the geodesic distance $d_G$ on the manifold (from the similarity or dissimilarity matrix) and the Euclidean distance in the embedded space $d_E$ to compute the curvature. We compare experimental results for embeddings obtained with both ISOMAP [3] and the kernel embedding in Section V. Lindman and Caelli [7] give the relationship between the two distances on elliptic, hyperbolic and Euclidean constant curvature manifolds as

$$d_E = \begin{cases} \frac{2}{K^2} \sinh(\frac{1}{2} d_G) & \text{Elliptic,} \\ \frac{2}{|K|^2} \sinh(\frac{1}{2} d_G) & \text{Hyperbolic,} \\ d_G & \text{Euclidean.} \end{cases}$$

One way to estimate Gaussian curvature on the edges of graphs is presented by Xiao in [8]. However, the approximations used only hold for small curvatures. In the data under study here, we find that the curvatures are too large for these approximations to hold. We therefore estimate curvature using Newton’s method. Taking the curvature in an elliptic space as an example, the Newton iteration is

$$K_{n+1}^\frac{1}{2} = K_n^\frac{1}{2} - \frac{K_n^\frac{1}{2} d_E - 2 \sin \frac{K_n^\frac{1}{2} d_G}{d_E - d_G \cos \frac{K_n^\frac{1}{2} d_G}}}{d_G}$$

(4)

Finally, we can compute new geodesic distances for the points based on the updated curvature. We keep the Euclidean distance between the points fixed, while updating the curvature. The updated geodesic distance under the new Gaussian curvature can be represented in terms of the old geodesic distance at the previous iteration. The update equation for the geodesic distance is Equation 5.

IV. THE ALGORITHM

Our aim in this paper is to rectify a given set of non-Euclidean dissimilarity data so as to make them more Euclidean. One way to gauge the degree to which a pairwise distance matrix contains non-Euclidean artefacts is to analyse the properties of its centered Gram matrix. For an $N \times N$ symmetric pairwise dissimilarity matrix $D$ with the pairwise distance as elements, the centered Gram matrix $G = -\frac{1}{2} J D^2 J$, where $D^2$ is element-wise squaring of elements in $D$, $J = I - \frac{1}{N} 1^T 1$ is the centered matrix and 1 is the all-ones vector of length $N$. The degree to which the distance matrix departs from being Euclidean can be measured by using the relative mass of negative eigenvalues or “negative eigenfraction” $J_{eigS} = \sum_{\lambda_i \leq 0} |\lambda_i| / \sum_{i=1}^N |\lambda_i|$ [9].

Given a set $X = \{x_1, \ldots, x_N\}$ of $N$ objects and a dissimilarity measure $d$, a dissimilarity representation is an $N \times N$ matrix $D_G$ with the elements $d_G(u, v)$ representing the pairwise geodesic distance between objects $x_u$ and $x_v$. The following implementation steps show how to rectify the distance matrix from being non-Euclidean to Euclidean.

Begin with a pairwise distance matrix $D_G(X, X)$,

1) Embed the objects in a Euclidean space using either ISOMAP or the kernel embedding to obtain Euclidean distances $d_E$.
2) From geodesic distance $d_G$ and Euclidean distance $d_E$, find the constant curvature space with curvature $K_{old}$ for a pair of objects using equation 4.
3) Update the Gaussian curvature with a small time step from Equation 3.
4) Obtain the new geodesic distance $d_G^{new}$ from the previously available geodesic distance together with the curvatures under fixed Euclidean distance based on Equation 5.
5) Get the new distance matrix $D_G$ composed of new geodesic distances between objects, and repeat from step 1 until $D_G$ is Euclidean, that is, there are no negative eigenvalues from its centered Gram matrix.

V. EXPERIMENTS

The “Chicken pieces” dataset [9], [4] is used for experimentation. The chicken pieces data contains 446 binary image in five classes: breast (96 examples), back (76 examples), thigh and back (61 examples), wing (117 examples) and drumstick (96 examples). The data exists in terms of a set of non-Euclidean dissimilarity matrices generated using different parameter settings $L$ and $C$. Our experimental results are from the data with $C = 120$ and $L = \{5, 10, 15, 20, 25, 30\}$. The originally asymmetric dissimilarities are made symmetric by averaging [4].

We can measure the degree of Euclidean behavior of a distance matrix [9] using the “negative eigenfraction. This measure is zero for Euclidean distances and increases as the
Equation 5

\[
d_{G_{n+1}} = \begin{cases} 
\frac{2}{K_{n+1}^{\frac{1}{2}}} \arcsin \left( \frac{K_{n+1}^{\frac{1}{2}}}{K_{n}^{\frac{1}{2}}} \sin \left( \frac{K_{n}^{\frac{1}{2}}}{2} d_{G_n} \right) \right) & \text{elliptic hyperspace} \\
\frac{2}{|K_{n+1}|^{\frac{1}{2}}} \arcsinh \left( \frac{|K_{n+1}|^{\frac{1}{2}}}{|K_{n}|^{\frac{1}{2}}} \sinh \left( \frac{|K_{n}|^{\frac{1}{2}}}{2} d_{G_n} \right) \right) & \text{hyperbolic space}
\end{cases}
\]

Figure 1. \( J_{\text{eigS}} \)

Distance becomes increasingly non-Euclidean. The negative eigenfraction of the chickenpieces data with \( L = 5.0 \), \( C = 120 \) as the manifold evolves is shown in Figure 1. As the curvatures are updated the negative eigenfraction decreases, indicating that the dissimilarity measure becomes increasingly Euclidean. Figure 2 shows the Gaussian curvature for the edge with the largest initial curvature, the edge with median curvature and the edge with minimum curvature. Each of the curvatures move towards zero, indicating that the evolution process transforms the hyperbolic space (negative Gaussian curvature) to a Euclidean space (zero Gaussian curvature). Figure 3 shows the curvatures as a function of distances obtained using the kernel embedding. The figure shows that the Kernel embedding preserves the global distances. The larger the distances, the smaller the curvatures.

Finally, some classification experiments were performed to explore whether the evolution preserves the class structure in the data. The classification results were obtained with the 1-NN classifier and 10-fold cross validation. We compare these results with those obtained using some alternative Euclidean correction procedures. The methods explored were a) projecting onto the positive subspace, b) projecting onto the associated Euclidean space [4], and c) using the original distances. Our classification performance results are shown in Figure 4. Each of the embedding methods distort the data to some extent. However, the Ricci flow with ISOMAP gives the smallest degradation. By comparison, the results obtained from Ricci flow and kernel embedding are poor. We believe that this is because the curvature increases much more rapidly for shorter distance in the kernel embedding (Figure 3). Thus, the change in distance induced by Ricci flow is rapid and potentially unstable.

VI. CONCLUSION

In this paper, we have presented a method for evolving a non-Euclidean dissimilarity measure into a Euclidean one. Our distance rectification method evolves the distance measure by updating Gaussian curvatures on the edges of the graph embedded on a manifold, based on the Ricci flow on a constant curvature Riemannian manifold. Applying our method to the chicken pieces dataset, we demonstrate that the distance measures can be transformed into a Euclidean space, but with some loss of discriminating power. Using the ISOMAP embedding gives good performance, but the performance is poorer when the kernel embedding is used. The loss of information may be caused by the effect of the Ricci evolution process as it is applied independently on each edge and ignores the local structure of the manifold. Hence, one way to improve our current work is to maintain local structure during the Ricci flow smoothing process.

ACKNOWLEDGEMENTS

This work was supported by the EU FET project SIMBAD (213250). Edwin Hancock was supported by a Royal Society Wolfson Research Merit Award.
Figure 3. (a) and (b) are initial edge curvatures for the Kernel and ISOMAP embeddings. (c) and (d) are edge curvatures after Ricci Flow for the Kernel and ISOMAP embeddings.

Figure 4. Error rate from 1NN

REFERENCES


3327