AN ADAPTIVE METHOD FOR THE CONFORMATION PROBLEM

İ. Kuban Altınel, Necati Aras, Kurtuluş Barış Öner
Boğaziçi Üniversitesi, Endüstri Mühendisliği Bölümü, İstanbul
B. J. Oommen
Carleton University, School of Computer Sciences, Ottowa, Canada

Abstract: The conformation problem is the determination of a spatial structure of a group of objects through some relations among those objects. In this work, an adaptive method for the conformation problem (ACON) in which the available data are pairwise distances between objects is proposed. Beside its neural motives, this new method is also based on a physical intuition which is very similar to that used in the force-directed methods of graph drawing. ACON is compared to some multidimensional scaling techniques through two types of experiments: Map reconstruction experiments and dimensionality reduction experiments. Normalized stress and Sammon's stress values are the main measures for the comparisons. For the map reconstruction case, a measure for location error and two strategies based on additional information are also proposed. One of these strategies, namely 1-fixed strategy (1FS), makes use of the information of the coordinates of a single city, while the other strategy, 2-fixed strategy (2FS), takes into account the known coordinates of two cities. Experiments with missing data are also performed for both the map reconstruction and the dimensionality reduction cases. In most of our experiments, the results reveal that ACON outperforms the other methods with respect to stress and location error values. It is also the most robust method: Final configurations are not sensitive to initial configurations.

Keywords*: Conformation Problem, Neural Networks, Force-Directed Methods, Multidimensional Scaling, Map Reconstruction.*

1. Introduction

The conformation of an object is its exact spatial structure at a given time. Then, given n objects and inter-object relations, the conformation problem can be defined as the determination of a spatial structure for those n objects by using inter-object relations. Mainly, the conformation problems in which the available data are the pairwise inter-object distances will be focused and one of such problems, the map reconstruction problem, is handled in this work.

Given the locations of some cities, estimation of the pairwise inter-city road distances is a straightforward problem. Since distance estimation is an important real life problem, it has been well defined and studied extensively. Map reconstruction is just the inverse of the distance estimation problem: Given the pairwise road distances for a set of cities, the locations of cities are estimated in such a way that the distance relations are preserved. So, an abstract map reflecting the hidden structure is searched for in map reconstruction.

Multidimensional scaling (MDS) is a technique used to visualize and analyze similarities or dissimilarities between objects using a lower dimensional representation of their feature vectors, while preserving the hidden structure (Borg and Groenen, 1997). By MDS methods, configurations of objects are constructed and in these configurations, the basic rule is that the objects with high similarity/low dissimilarity are located closer to each other as compared to the objects with low similarity/high dissimilarity. Consideration of the intercity road distances as dissimilarities makes it possible to apply MDS techniques for the map reconstruction problem.

A graph $G(V, \xi)$ is a collection of nodes V and edges ξ where nodes represent some objects and edges represent some pairwise relationships among these objects. Graph drawing is the problem of building a visual representation of a graph which reflects its inherent structure (Battista et al., 1999). The proximity of the map reconstruction problem and the graph drawing problems can be noticed when their definitions are compared.

In this work, an adaptive method for the conformation problem (ACON) is proposed. In the spirit of this new method, some neural network motives exist for sure. There are also neural MDS techniques which are very close to ACON mathematically. In some sense, ACON also resembles force-directed methods in graph drawing.

2. The New Algorithm

As a configuration of n objects in a p-dimensional space is to be constructed, each object is represented by a point in a p-dimensional Euclidean space. ACON performs some operations on these points iteratively. The operations in ACON remind of the ones in vector quantization (VQ) (Linde et al.,

1980) and Kohonen’s Self-Organizing Maps (SOM) (Kohonen, 1995). Throughout these operations, points are manipulated in such a way that inter-point Euclidean distances in the final *p*-dimensional configuration approximate real distances as well as possible.

Beside its similarity to VQ and SOM in terms of operations, ACON has a physical analogy that resembles the one used in force-directed methods in graph drawing (Battista et al., 1999). As calculations of “forces” differ in both, the major difference is in their update mechanisms.

There also exists two neural MDS methods very close to ACON (Van et al., 1997).

In ACON, the initial coordinate values of the representative points are uniformly distributed in a predefined region. The area of this region is scaled according to the real distances. Initialization is simply handled in the following manner: First real distances are scaled with respect to the maximum real distance and then the coordinate values are initialized within the interval $[0,1]$ along all dimensions.

ACON runs by epochs. At each epoch, all object pairs are manipulated iteratively. At each iteration, a pair (i, j) of objects is randomly selected and the real distance between these objects (δ_{ij}) is compared with the Euclidean distance between the points that represent these objects ($d_{ij}(t)$). If the Euclidean distance between the representative points is larger than the real distance between the objects, that is, $d_{ij}(t) > \delta_{ij}$, the points are “far away to comfort”. To compensate for this, the coordinates of the points are moved towards each other along the line connecting them by an amount proportional to the difference between $d_{ij}(t)$ and δ_{ij}. The proportionality factor is a parameter called *step size*, $\mu(e)$ (*e* denotes the number of epochs). Both points are moved towards each other by an amount equal to $\mu(e)(d_{ij}(t)- \delta_{ij})/2$ so the distance between the points decreases by $\mu(e)(d_{ij}(t)- \delta_{ij})$. The update scheme for the inverse case is just in symmetry. That is, if the Euclidean distance between the representative points is smaller than the real distance between the objects ($d_{ij}(t) < \delta_{ij}$), the coordinates of the points are moved away from each other again along the line connecting them by an amount equal to $\mu(e)(\delta_{ij} - d_{ij}(t))/2$, so the distance between the points increases by $\mu(e)(\delta_{ij} - d_{ij}(t))$. The value of $\mu(e)$ is reduced with the number of epochs.

Even though the operations are not directly based on minimizing an error function, ACON minimizes the errror function $E_{ij} = [\delta_{ij} - d_{ij}(t)]^2$ at each iteration and the raw stress, so the normalized stress (Borg and Groenen, 1997), in the overall process.

3. Two New Strategies with Additional Data and Location Error

Configurations generated with respect to distances are subject to translation, rotation, reflection and dilation, and this is not a problematic behavior in most of the cases (Borg and Groenen, 1997). Nevertheless, for some conformation problems, such as the map reconstruction problem, this may be of concern. Not only the distance preservation but also the estimation of exact locations of the cities is important in the map reconstruction.

If, somehow, the location(s) of one or two cities in the original map is (are) available, easy attempts to make the reconstructed map translation- and rotation-free are possible. When the location of a single city in the real map is known, the point representing that city in the reconstructed map is fixed at that location. Such an additional constraint leads to a new translation-free map. We call this strategy the *one-fixed strategy* (1FS). When coordinates of two cities are known, the points representing those cities in the reconstructed map are fixed, just in the same way as in 1FS. This makes the final configuration not only translation-free, but also rotation-free. We name this strategy the *two-fixed strategy* (2FS). As the two strategies are introduced, we call the normal strategy in which no fixation occurs *no-fixed strategy* (NFS).

To measure the deviation between the reproduced map and the original one for the map reconstruction problem, we define a criterion which we name *average location error* (LE): $LE = \sum_{i=1}^{n} e_i / n$, where e_i is the Euclidean distance between the representative points of city *i* in the real map and the reconstructed map. It is important to note that all the city coordinates are required for the calculation of location error. In real-life applications, such information is not available; otherwise map reconstruction would be meaningless. The motivation in using this error measure is to show the value of extra information (the coordinates of at most two cities). It can also be viewed as a measure for testing a method.

4. Experiments and Conclusion

Experiments with 6 methods on 9 data sets are conducted. 7 data sets are for the map reconstruction problem: Australia, Canada, England, France, Germany, Türkiye and the United States. Each set except Türkiye and Germany consists of the pairwise distances of 15 cities. Türkiye data set has 80 cities and Germany data set has 1056 cities. Beside its large size, Germany has another important feature: 7182 of distances are missing. Two dimensional real coordinates of the cities in each data set except Germany are also used for location error calculations. The remaining two data sets, Iris and Wine, are used in the dimensionality reduction experiments. These sets have sizes of 147 and 178, and diemensionality of 4 and 13, respectively.

The 6 methods we compare are ACON, classical MDS (CMDS) (Borg and Groenen, 1997), Sammon's Mapping (SM) (Sammon, 1969), SAMANN (Mao and Jain, 1995), ALSCAL (Takane et al., 1977) and a negative gradient algorithm to minimize the square of normalized stress (NG). For the comparison of performances, the main measures used are the normalized stress and Sammon's stress (Sammon, 1969) values of the configurations obtained by the methods. ACON, SM and NG constitutes a special group in our analysis, since the manner they operate in allows us to experiment with common settings easily. That is why 1FS and 2FS strategies are applied only with ACON, SM and NG in the map recostruction cases. Location error comparison is the heart of these experiments. Average run time is another measure used. The performance of the three methods, ACON, SM and NG with missing data are also compared.

In the map reconstruction experiments, CMDS, SM, NG and ACON are run over the 7 data sets. As NFS, 1FS and 2FS results are considered, ACON provides the smallest stress and location error values for most of the data sets. It is also observed that ACON is the most robust method with respect to the differences in the initial configurations and points fixed in 1FS and 2FS. On the other hand ACON is not the best in terms of average run times when no fixing strategy is applied. With the fixed points in the strategies, the deviations in the stress values of the two methods, SM and NG, increases considerably. Their average run times also exceeds those of ACON. In parallel with our expectations, the location error values become better in all of the methods when these strategies are applied. ACON also achieves the smallest stress and location error values in the map reconstruction experiments with missing data in most of the cases. It is also the most robust method in missing data experiments. In the dimensionality reduction experiments, CMDS, SM, NG, ACON and SAMANN are run over the two data sets. Missing data experimentations are also performed. The results reveal that ACON provides the best stress and location error values and it is the most robust one in general, again.

We have also used the Procrustes Analysis (Borg and Groenen, 1997) to measure the similarities of the configurations generated by different methods. By this analysis, similarities between the configurations obtained by ACON, SM, and NG are detected, as the error values for their pairwise matchings are smaller with respect to the others. The small error values for some of the matching of the configurations obtained by ALSCAL and CMDS can be interpreted as there is similarity between the those configurations.

References

Borg, I. and P. Groenen, *Modern Multidimensional Scaling, Theory and Applications*, Springer Series in Statsitics, Springer-Verlag New York, Berlin, Heidelberg, 1997.

Di Battista, G., P. Eades, R. Tamassia, and I.G. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice-Hall, New Jersey, 1999.

Linde, Y., A. Buzo, and R.M. Gray, “An Algorithm for Vector Quantization”, *IEEE Transactions and Communications*, COM-28, pp. 61-71, 1980.

Kohonen, T., *The Self-Organizing Map*, Springer-Verlag, Berlin, 1995.

Van Wezel, M.C., J.N. Kok, and W.A. Kosters, “Two Neural Network Methods for Multidimensional Scaling”, *Proceedings of the European Symposium on Artificial Neural Networks, (ESANN'97)*, pp. 97-102, 1997.

Sammon, J.W., “A Nonlinear Mapping for Data Structure Analysis”, *IEEE Transactions on Computers*, Vol. C-18, pp. 401-409, 1969.

Mao, J. and A.K. Jain, “Artificial Neural Networks for Feature Extraction and Multivariate Data Projection”, IEEE Transactions on Neural Networks, Vol. 6, No. 2, pp. 296-317, 1995.

Takane, Y., F.W. Young, and J. de Leeuw, “Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features”, *Psychometrika*, Vol. 42, pp. 7-67, 1977.