PERGAMON

Engineering Applications of Artificial Intelligence 11 (1998) 659–667

Engineering Applications of
ARTIFICIAL
INTELLIGENCE

Contributed Paper

# Visualization and the understanding of multidimensional data

Yoh-Han Pao [a,*], Zhuo Meng [b]

[a] *Case Western Reserve University, Cleveland, OH 44106, USA*
[b] *AI WARE, Inc., Beachwood, OH 44122, USA*

Received 1 December 1997; accepted 1 June 1998

**Abstract**

This paper is concerned with the task of making sense out of a body of a large number of multivariate objects. The matter of 'making sense' or 'understanding' is described in terms of knowing three types of relationships and being able to keep such knowledge in mind. The paper argues that both matters, obtaining the knowledge, and keeping it in mind, available for use, would be facilitated by mapping the *N*-dimensional space to 2-dimensional space or 3-dimensional space, in a topologically meaningful manner. A 'ratio-conserving' mapping procedure is described and compared with some other previously proposed methods. A body of 5-dimensional semiconductor materials composition and properties data is used for illustration purposes. Three types of 'meaning' are discussed and it is shown that 'ratio-conserving' mapping does provide a way of obtaining and retaining a visualization of the meaning of large bodies of complex multivariate data. 

*Keywords:* Visualization; Complexity reduction; Reduced-dimension mapping; Neural networks; Materials data mining

## 1. Introduction

This paper is concerned with the task of trying to make sense out of a large body of multivariate data, or equivalently to understand some aspect of the meaning contained in that body of data. It is argued that there are three aspects to the understanding of multidimensional data. The first task is to obtain knowledge of the distribution of the *N*-dimensional data points in the data space, a metric space. Is the distribution uniform, or in the form of clumps or irregular structures or in one form in some region and in another in another region? Obtaining knowledge of such matters is part of the task of 'understanding' the data.

The first component of knowledge of the data is, therefore, knowledge of the interrelationships between the data points in data space or input space. In contrast to that, the second component of understanding consists of knowing whether there are functional relationships between the vector field of input space and the space of property values. Do nearby points in data space correspond to property values which are also nearly alike? Are there smooth trends, are there regions of large inconsistencies, or perhaps even chaotic regions? In principle, all that information is available in the body of data being considered, but it is usually very difficult to extract that information in any useful form, or in any form which can be readily 'understood' or 'visualized' by humans. In practice, massive data-base procedures are used to deal with such matters in a piece-meal manner, as required, and when required. The difficulty is rooted in the multivariate nature of the data.

The third part of understanding consists of the formation of categories. This corresponds to the discernment of the formation of 'clumps' in property space. It is useful to determine how input data points relate to categories. Points quite far apart in input descriptor space might fall into the same category in property space, or objects seemingly similar in input space might be more conveniently allocated to different categories, because of the objectives of the task in question. This third component of understanding addresses the expected and the unexpected—the inconsistencies. It is at the heart of the formation of conjectures, of any theory formation and of the making of discoveries.

* Corresponding author.

0952-1976/98/$ - see front matter 
PII: S0952-1976(98)00031-1

This is also because the formation of categories is a subjective matter. The understanding of the structure and characteristics of categories of multivariate objects is therefore a complex matter.

It is the opinion of the present authors that the impediment to the 'understanding' of large bodies of multivariate data lies not so much in the difficulty of computing various distributions and relationships as in the lack of convenient means of holding such information in convenient and intuitively understandable form, available for use. It is that means for visualization that is the focus of this paper.

## 2. Related work

This present paper is concerned with dimension reduction. In other words, it is proposed that a mathematical procedure can be used to map a high dimensional description of a set of objects into a simpler representation of that same data in a space of much smaller dimensions, 2D representations being of particular interest. There is evidence to indicate that such reduced dimension depictions can be very useful in helping humans obtain overall intuitive understanding of bodies of complex data. Further elaborate complex quantitative manipulations of bodies of data can then be planned more effectively. This matter is discussed in the context of particular dimension reduction approach, the 'ratio-conserving' map.

The thrust of this discussion is to describe how the benefits of a robust 'topologically-correct' dimension-reduction procedure can indeed be used to support visualization and understanding. There are also other suggestions on how dimension reduction might be carried out. In addition to this ratio-conserving approach, six other dimension reduction methods are known to the present authors, these being use of the Karhunen–Loeve transform (Fukunaga and Koontz 1970), the feature map approach of Kohonen (1982, 1995), the generative topographic mapping (GTM) method (Bishop et al., 1998), the autoassociative mapping approach (Kramer, 1991; Pao, 1996), the nonlinear variance conserving approach (NLVC) (Pao and Shen 1997), and the equalized orthogonal mapping (EOM) (Meng, 1998).

In pattern recognition research, the Karhunen–Loeve (K–L) transform is a procedure for finding features which are linear combinations of the original features. Those new features whose values do not vary much from one pattern to another, over the entire ensemble of patterns, do not provide discriminating power and may be dropped from further consideration in so far as classification is concerned, for example. The K–L method can be very effective, depending on the nature of the data. If a linear transform suffices to identify a few new features with large discriminatory power then use of the K–L transform dimension-reduction method would be appropriate for classification purposes. The feature map or self-organizing map (SOM) and the GTM approaches are both grid point methods. Using the scalar product as a measure of similarity, similar data vectors are collected onto nearby points on a 2-dimensional grid.

The autoassociative mapping approach makes use of some characteristics of the conventional multilayer feedforward neural network. The net is trained to yield an $N$-dimensional output vector identical to the input vector for all the data vectors available, as nearly so as possible. The essential point, however, is that an internal layer of the network be severely restricted in the number of nodes available to it, perhaps only two in number. Since all the knowledge necessary for reconstruction of the $N$-dimensional vectors passes through the two internal layer nodes, it is clear that a reduced-dimension representation would have been found which contains all, or nearly all, the intrinsic information in the data patterns, sufficient for reconstruction purposes. This is true provided the training error is sufficiently small; that is if the autoassociative reconstruction is indeed achievable. In practice, the training errors are often large and the mappings so obtained are sensitive to the initial values of the net parameters.

The nonlinear variance conserving mapping is indeed an extension of the K–L transform. The idea is that a nonlinear transform be implemented with a multilayer feedforward neural net with the stipulation that the input vectors are of $N$-dimensions, the original data, and the output representation be of 2D. The network is trained on the criterion that the total output variance be a certain fraction of the input variance. The network learns a net which does the dimension reduction and spreads the variance over the two outputs and over the entire set of data vectors so that the requisite variance value is attained. The method works well.

This approach to dimension reduction was applied to a body of semiconductor data listed in Table 1, provided by Dr A. G. Jackson of The US Air Force Wright Laboratory, Wright–Patterson Air Force Base, OH 45433-6523. In that table, each semiconductor compound is described in terms of 5 supposedly independent variables, namely the electronic band gap for excitation to the conducting states, two crystal unit cell parameters, the atomic weight of the compound, and the 'radius' of the anion.

The EOM arranges to make optimum use of the representation resources available. In other words, the covariance matrix of the two outputs is specified to be diagonal with equal-valued elements. That method produces results which are similar to that of NLVC but with increased efficiency of use of the reduced-dimension space. The ratio-conserving approach is somewhat different from the others.

Table 1
Characteristics and properties of some semiconductor compounds

| $n$ | Compound | gap | $a$ | $c$ | $w$ | $r$ | $\rho$ |
|---|---|---|---|---|---|---|---|
| 1 | ZnS | 3.9 | 3.823 | 6.261 | 97.434 | 53 | 3.536 |
| 2 | AlN | 6.2 | 3.11 | 4.98 | 40.99 | 25 | 3.255 |
| 3 | ZnO | 3.3 | 3.251 | 5.209 | 81.369 | 22 | 5.651 |
| 4 | $AgGaS_2$ | 2.638 | 5.751 | 10.238 | 241.718 | 53 | 4.66 |
| 5 | $CuGaS_2$ | 2.43 | 5.351 | 10.47 | 197.388 | 53 | 4.332 |
| 6 | $LiIO_3$ | 4 | 5.481 | 5.171 | 181.836 | 22 | 4.502 |
| 7 | Se | 1.7 | 4.361 | 4.954 | 78.96 | 66 | 4.819 |
| 8 | GaS | 2.5 | 3.586 | 15.496 | 101.784 | 53 | 3.86 |
| 9 | SiC | 6 | 4.359 | 4.359 | 40.09 | 29 | 3.191 |
| 10 | $SiO_2$ | 8.4 | 4.9134 | 5.4052 | 60.078 | 22 | 2.65 |
| 11 | Te | 0.33 | 4.457 | 5.939 | 236.55 | 82 | 6.25 |
| 12 | AgI | 2.8 | 6.473 | 6.473 | 234.77 | 126 | 6 |
| 13 | CuCl | 3.17 | 5.405 | 5.405 | 98.993 | 77 | 4.137 |
| 14 | CuI | 2.95 | 6.042 | 6.042 | 190.44 | 96 | 5.667 |
| 15 | InSb | 0.23 | 6.479 | 6.479 | 236.55 | 89 | 5.777 |
| 16 | $AgGaSe_2$ | 1.8 | 5.981 | 10.865 | 335.51 | 66 | 5.759 |
| 17 | $AgInSe_2$ | 1.2 | 6.099 | 11.691 | 286.798 | 66 | 5.808 |
| 18 | InAs | 0.36 | 6.268 | 6.479 | 189.79 | 71 | 5.72 |
| 19 | $CdGeAs_2$ | 0.57 | 5.943 | 11.217 | 334.97 | 71 | 5.6 |
| 20 | GaSb | 0.72 | 6.095 | 6.095 | 191.47 | 89 | 5.615 |
| 21 | InSe | 1.25 | 4.002 | 24.946 | 193.76 | 66 | 5.55 |
| 22 | InP | 1.35 | 5.868 | 5.868 | 145.77 | 59 | 4.798 |
| 23 | $Ag_3AsS_3$ | 2 | 10.8 | 8.69 | 494.792 | 53 | 5.6 |
| 24 | GaAs | 1.4 | 5.653 | 5.653 | 144.71 | 71 | 5.316 |
| 25 | $CuGaSe_2$ | 1.7 | 5.606 | 11.006 | 242.468 | 66 | 4.73 |
| 26 | GaSe | 2.021 | 3.747 | 23.91 | 148.68 | 66 | 5.03 |
| 27 | $CuInS_2$ | 1.53 | 5.489 | 11.101 | 242.468 | 53 | 4.73 |
| 28 | HgS | 2.1 | 4.145 | 9.496 | 232.654 | 53 | 7.101 |
| 29 | $\beta$-SiC | 2.26 | 4.359 | 4.359 | 40.09 | 29 | 3.191 |
| 30 | GaP | 2.3 | 5.45 | 5.45 | 100.69 | 59 | 4.135 |
| 31 | ZnTe | 2.3 | 6.101 | 6.101 | 192.97 | 82 | 5.924 |
| 32 | ZnSe | 2.7 | 5.667 | 5.668 | 144.33 | 66 | 5.318 |
| 33 | CuBr | 2.91 | 5.69 | 5.69 | 143.449 | 82 | 4.72 |
| 34 | $CdGeP_2$ | 2.91 | 5.74 | 10.776 | 246.93 | 59 | 4.549 |
| 35 | $ZnSiAs_2$ | 1.74 | 5.606 | 10.88 | 243.43 | 71 | 4.7 |
| 36 | $ZnGeP_2$ | 2.05 | 5.463 | 10.731 | 199.9 | 59 | 4.105 |
| 37 | $CdGa_2S_4$ | 3.05 | 5.568 | 10.04 | 380.096 | 53 | 3.97 |
| 38 | $LiNbO_3$ | 4 | 5.148 | 13.86 | 147.842 | 22 | 4.64 |

## 3. The ratio-conserving approach

The ratio-conserving mapping is also implemented with a multilayer feedforward net such as the net shown in Fig. 1, with two outputs in the case of mapping onto 2D space.

It is interesting to note that the net is trained with no target patterns. In other words, there is no target output vector associated with any single input data vector. Instead, for a set of data points in input $N$-dimensional space, the Euclidean distance between any two data points $p$ and $p'$ is calculated in the full dimensional space and in the reduced-dimension output space, and the requirement is that the value of the ratio of those two distances be the same or nearly the same, as much as possible, for all pairs of input data points. The objective function $E$ is the variance of that ratio for all two point combinations of the data points and the requirement is that the values of the network parameters be varied until the objective function attains a minimum value.

The net can be trained in backpropagation manner in the usual gradient descent manner (Pao, 1989). An expression for computing the variance in the values of the ratio and formulae for learning the network weights are as follows.

Distance definitions:

$$\|o_p - o_{p'}\| = \sqrt{\sum_{k=1}^{K}(o_{kp} - o_{kp'})^2}, \qquad \|x_p - x_{p'}\| = \sqrt{\sum_{i=1}^{I}(x_{ip} - x_{ip'})^2}. \tag{1}$$

Proposed error function:

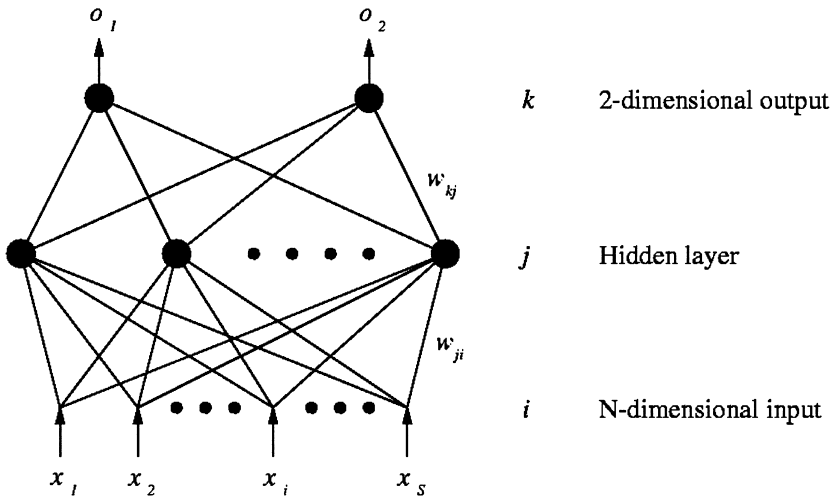


Fig. 1. Illustration of a net for learning ratio-conserving mapping.

$$E=\frac{2}{P(P-1)}\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\frac{\|o_p-o_{p'}\|^2}{\|x_p-x_{p'}\|^2}-\left[\frac{2}{P(P-1)}\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\frac{\|o_p-o_{p'}\|}{\|x_p-x_{p'}\|}\right]^2. \quad (2)$$

Formula for iterative adjustment of weights between output and hidden layers:

$$\Delta w_{kj}=-\eta\frac{\partial E}{\partial w_{kj}}=-\eta\frac{4}{P(P-1)}\times\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\left(\frac{\|o_p-o_{p'}\|}{\|x_p-x_{p'}\|^2}\frac{\partial}{\partial w_{kj}}\|o_p-o_{p'}\|\right)-\frac{8}{P^2(P-1)^2}\left(\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\frac{\|o_p-o_{p'}\|}{\|x_p-x_{p'}\|}\right)\times\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\left(\frac{1}{\|x_p-x_{p'}\|}\frac{\partial}{\partial w_{kj}}\|o_p-o_{p'}\|\right), \quad (3)$$

where

$$\frac{\partial}{\partial w_{jk}}\|o_p-o_{'}\|=\frac{(o_{kp}-o_{kp'})[o_{kp}(1-o_{kp})o_{jp}-o_{kp'}(1-o_{kp'})o_{jp'}]}{\|o_p-o_{p'}\|}. \quad (4)$$

Formula for iterative adjustment of weights between hidden and input layers:

$$\Delta w_{ji}=-\eta\frac{\partial E}{\partial w_{ji}}=-\eta\frac{4}{P(P-1)}\times\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\left(\frac{\|o_p-o_{p'}\|}{\|x_p-x_{p'}\|^2}\frac{\partial}{\partial w_{ji}}\|o_p-o_{p'}\|\right)-\frac{8}{P^2(P-1)^2}\left(\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\frac{\|o_p-o_{p'}\|}{x_p-x_{p'}\|}\right)\times\sum_{p=2}^{P}\sum_{p'=1}^{p-1}\left(\frac{1}{\|x_p-x_{p'}\|}\frac{\partial}{\partial w_{ji}}\|o_p-o_{p'}\|\right), \quad (5)$$

where

$$\frac{\partial}{\partial w_{ji}}\|o_p-o_{p'}\|=\frac{1}{\|o_p-o_{p'}}\times\left\{\sum_{k=1}^{K}(o_{kp}-o_{kp'})[o_{kp}(1-o_{kp})w_{kj}o_{jp}(1-o_{jp})x_{ip}-o_{kp'}(1-o_{kp'})w_{kj}o_{jp'}(1-o_{jp'}(1-o_{jp'})x_{ip'}]\right\}. \quad (6)$$

The expression $w_{kj}$ is for updating the output layer weights and the expression for $w_{ji}$ is for updating the hidden layer nodes. These expressions are given in full detail and in a manner that they can be verified in a straight forward manner, using all the details presented above.

The results of many such learning trials indicate that the reduced-dimension maps so obtained are not unique. This is not surprising because it would be surprising if all the points on a three dimensional figure such a helix, for example, could be mapped faithfully and uniquely on a 2D figure. Clearly, many mappings are possible, each corresponding to a 'perspective'. But is was interesting to note that the relative positions of the various data points tended to be about the same, from training run to training run. And so it seems that such a map might yield a means for understanding

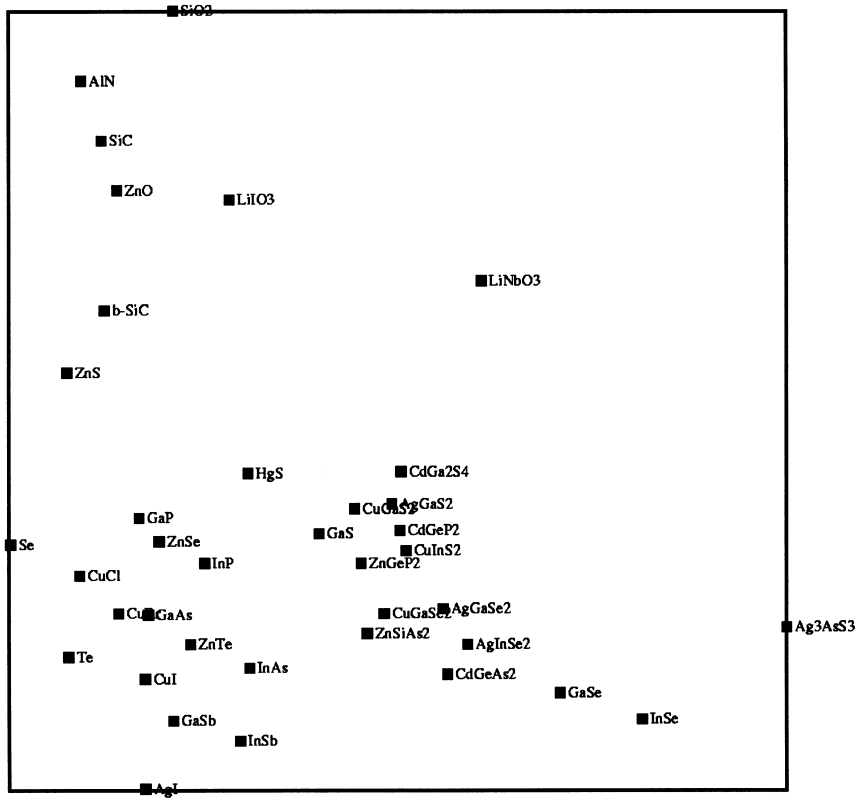


Fig. 2. 2-D display of semiconductor patterns obtained using ratio-conserving approach.

high-dimensional pattern data, in terms of the types of understanding described previously.

For demonstration purposes, a value of the density of the compound is associated with each entry of five values of independent variables. In this discussion, there is no guarantee of the accuracy of the values provided. This body of data is being used as provided for the purposes of demonstrating use of ratio-conserving mapping, and for interpreting the results of that mapping.

The ratio-conserving method has been described briefly in a previous manuscript dealing with discovery of materials (Meng, 1998).

## 4. Visualizing and understanding data distribution

It might be noted that even though the compounds listed in Table 1 are described in terms of five features and even though there are only 38 such compounds in the list, that body of data is nevertheless already sufficiently complex. For example, it is difficult to visualize how the data points are distributed in the 5-dimensional input space of input variables. Are the data points distributed randomly, or in the form of groups? How are the data points positioned relative to each other, are there outliers, and so on?

A 2-dimensional plot of the 38 data points is shown in Fig. 2, the 2D plot having been obtained through application of a ratio-conserving mapping from 5D to 2D space. The 2D plot enables a human to have an idea of how the data points are distributed relative to each other. In other words, such a visualization enables the observer to 'understand' the distribution of the input data. That understanding would be valuable if the mapping is correct in a meaningful way. This is the question that is of some considerable interest to the present authors and to this present discussion.

The input data were also subjected to K-means clustering in the conventional pattern analysis manner (Pao et al., 1977). Some clusters were found to have a significantly larger population than others and so a hierarchical tree of clusters was built to obtain smaller clusters within clusters but with all leaf clusters having about the same number of members in them, as much as possible. This was done to illustrate that some understanding of data distribution can be obtained through use of conventional, well-established methods. The identities of the members of the various clusters are listed in Fig. 3. This is helpful but very little can be kept in mind about distances between clusters. It is difficult to reason about that body of data in any detailed manner because it is difficult to maintain a high-dimensional image of how the various data points are distributed relative to each other. Cluster member-

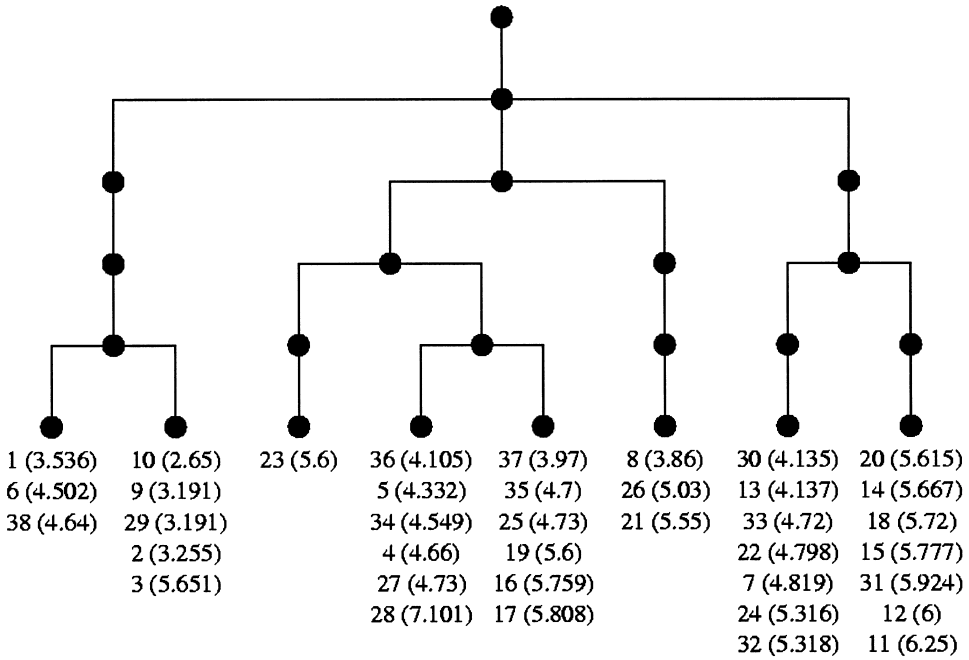


Fig. 3. Hierarchical cluster structure of semiconductor data (compound number and density).

ship is simply not sufficient for visualization purposes. On the other hand, the 2D display shown in Fig. 2 provides a detailed picture of inter-pattern distances, in accordance with some metric, valid on a large scale as well as over short distances.

In one check of the validity of the ratio-conserving approach, the boundaries of the various clusters were sketched in by hand and these are displayed in Fig. 4. It is encouraging to see that all data points which are in the same cluster are also singly connected in the 2D depiction. In further detailed work, variances in the transform ratio are calculated and regions of high or low variances can be investigated.

## 5. Visualizing and understanding functional relationships

As mentioned previously, a second step in the understanding of data consists of having holistic ideas of how the input descriptors relate to the properties of the data objects. Conventionally, for a collection of objects, each object is described in terms of the values of a number of descriptors. These are the input variables and the objects are points in input space. Often, the descriptors are called the input variables or the independent variables, even though the latter characterization may not always be true. In addition, various other attributes may be attached to each of those objects. These are the properties of the objects and the nature of a property can be quite varied. There may be functional relationships between the input variables and the properties, the output or dependent variables. Often it would be helpful to be able to have a general idea of how the properties vary with location in input data space. This is difficult to do when dealing with high-dimensional data.

For example, even if only one property is considered at a time, it is still difficult to have a holistic overall idea of the locations in data space which have interesting values in property space. The present discussion suggests that this overall view of functional relationships would be much more easily attained if the relationship is only from a 2D space to a scalar property.

In the present case, the density of a compound might be taken to be the property of interest. The density values might be coded in gray level manner or with other aspects of pseudo-color and displayed in 2D manner as shown in Fig. 5. A neural network can then be used to learn a quantitative functional relationship between the 2D coordinate values and the property value. That combination of reduced-dimension mapping and learning of a functional relationship, implied through knowledge of instances of the relationship, provides an overall visualization and a detailed knowledge of the multivariate data. This constitutes a component of what might be called the second aspect of understanding data.

## 6. Concluding remarks

A third aspect of the understanding of multivariate data relates to distributions in property space, to the

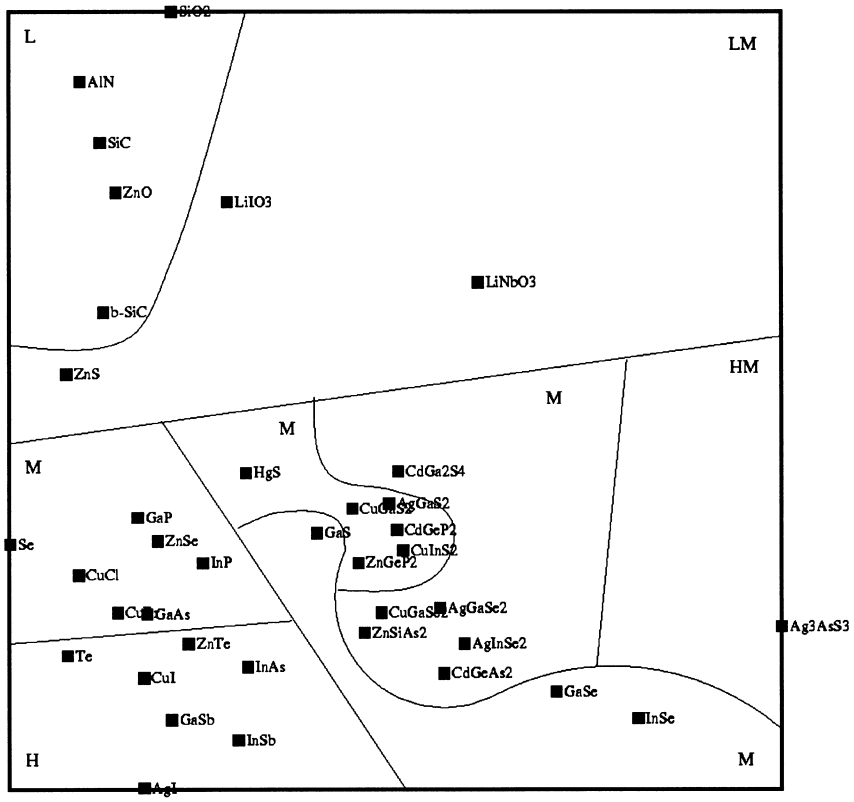


Fig. 4. Illustration of agreement between clustering and reduced-dimemsion display

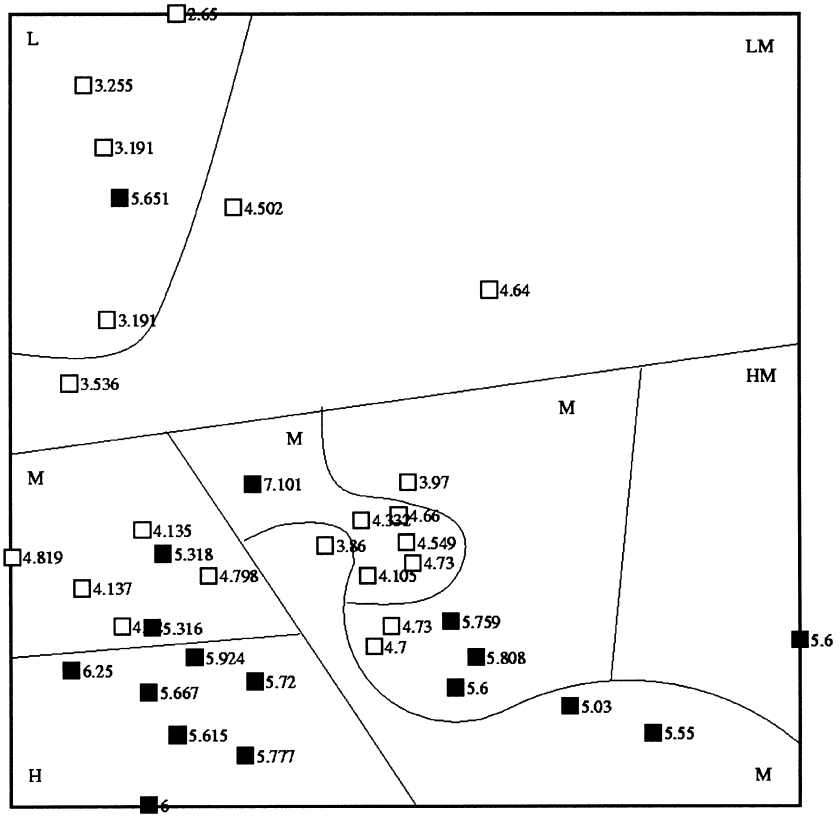


Fig. 5. Use of reduced-dimension plot for obtaining an overview of variation of density.

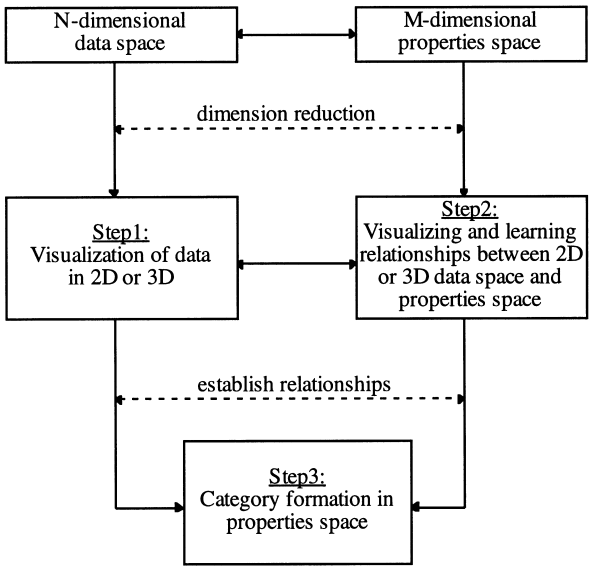


Fig. 6. Visualization and steps in understanding of multidimensional data.

formation of 'categories', and to the visualization and understanding of relationships between clusters in one space and categories in another space, both multidimensional spaces. This is being addressed in the work in progress.

But briefly, the thrust of the present discussion has been that understanding of a body of data includes (a) knowing where the data are, in the sense of knowing the locations and distributions of those locations in metric data space, and (b) knowing how these locations relate to associated properties. The argument is that it is difficult to obtain and hold on to stable mental images of these distributions and relationships if the data points are of high dimension. And that the visualization task becomes feasible if data are of two-dimensions or perhaps of three dimensions at the most. This is summarized in Fig. 6.

In all of this, the dimension-reduction procedure plays an important role. The ratio-conserving map seems to provide a useful and valid method. It helps to realize that such dimension-reduction mappings cannot be completely information-conserving. Something will be lost. The only question is what should be conserved and what can be sacrificed. Of the six methods mentioned briefly previously, the K–L method has the advantage of yielding unique results but is not sufficiently appropriate in general. In particular, the method is not effective when the eigenvalues of the covariance matrix are nearly of the same order of magnitude. The nonlinear variance conserving (NLVC) method seems to be a useful extension of the K–L method especially when used in the equalized orthogonal (EOM) form. The grid methods may not provide sufficient support for the understanding of functional relationships and do not seem to provide a means for displaying the distribution of patterns within groupings.

The ratio-conserving method suffers from one disadvantage, namely that the computational complexity varies with the square of the number of data points. One way of dealing with this difficulty is to train the net used for learning the mapping using cluster centers only. This would alleviate that difficulty of having the computational burden vary with the second power of $P$, the number of patterns, or data points.

## References

Bishop, C.M., Svensen, M., Williams, C.K.I. 1995. GTM: the generative topographic mapping. Neural Comput. In print.

Fukunaga, K., Koontz, W.L.G., 1970. Application of the Karhunen–Loeve expansion to feature selection and ordering. IEEE Trans. Comp. 19, 311–318.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybernetics 43, 59–69.

Kohonen, T., 1995. Self-organization Maps. Springer, New York.

Kramer, M., 1991. Nonlinear principal component analysis using auto-associative neural networks. AIChE 37, 233–243.

Meng, Z. 1988. Visualization and self-organization of multidimensional data through equalized orthogonal mapping. PhD dissertation, Electrical Engineering. Case Western Reserve University, Cleveland, Ohio.

Pao, Y-H., 1996. Dimension reduction, feature extraction and interpretation of data with network computing. IJPRAI 10, 521–535.

Pao, Y-H., Shen, C.Y., 1997. Visualization of pattern data through learning of nonlinear variance-constrained dimension-reduction mapping. Pattern Recog. 30, 1705–1717.

Pao, Y-H., 1989. Adaptive Pattern Recognition and Neural Networks. Addison–Wesley, Reading, MA.

Pao, Y-H., Meng, Z., LeClair, S.R., Igelnik, B., 1997. Materials discovery via topologically-correct display of reduced-dimension data. J. Alloys Comp.

### Authors' Biographies

**Yoh-Han Pao** is the Emeritus George S. Dively Distinguished Professor at Case Western Reserve University, Cleveland, OH, with appointments in Electrical Engineering and in Computer Science. In his career he has served as the Chairman of Case Western's Electrical Engineering Department (1969–1977); as the Director of the Division of Electrical, Computer, and Systems Engineering of the National Science Foundation (1978–1980); and as founding Director of the Center for Automation and Intelligent Systems Research (1984–1989) at Case Western Reserve. He has served as NATO Senior Science Fellow (1972–1973) and as visiting Professor at MIT's AI Laboratory, Cambridge, MA (1980). He has carried out research and lectured at Edinburgh University, the Turing Institute, Tsinghua University, the Chinese Academy of Sciences and other institutions. His industrial career include a total of fourteen years at the duPont Company and at Bell Laboratories. His research interests are adaptive pattern recognition, neural networks, computational intelligence, and signal and image processing. He is the Founding Editor of the Academic Press Quantum Electronics Series and is the author of many technical publications including the Addison-Wesley book on Adaptive Pattern Recognition and Neural Networks (1989).

He is a Fellow of IEEE and of the Optical Society of America. He is on the Editorial board of several major technical journals. He is also co-founder and past President of AI WARE, Inc., a software systems company, currently a Division of Computer Associates International.

**Zhuo Meng** received his B.S. degree in Radio Electronics from Peking University, China in 1991 and obtained his M.S. and Ph.D. degrees in Electrical Engineering and Applied Physics from Case Western Reserve University in 1996 and 1998 respectively. He also holds a second master's degree in Physics from The Ohio State University. He is now with the AI WARE Division of Computer Associates International Inc. His research interests include neural net computing, data visualization and other computational intelligence applications.