The Institute for Perception

For more technical reports from The Institute for Perception or for information about our short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (2001) 4(3)2,3

Probabilistic Multidimensional Scaling

Daniel M. Ennis

Background: Solutions to many market research, product development and quality assurance problems require the use of various types of product maps[1,2,3]. Sometimes it is of interest to know whether products differ without regard to liking and preference. For instance, product maps are used to guide quality assurance, cost reduction formulations, and 'me-too' product development. It is natural to think of products as points and when two products are less similar than two other products that they are further apart in some product space. In this report we will show how this idea has significant limitations, and that a more meaningful interpretation of similarity data can be made when products are treated as distributions.

Scenario: An unusually high number of consumer complaints concerning product consistency have been received regarding your orange juice brand. An investigation of this problem leads to the conclusion that five of your ten plants are largely responsible for the offending product. In order to understand differences among the products produced at your ten plants, you conduct a consumer study of paired product similarities. One hundred consumers evaluate ten variants (one from each plant) of your current orange juice brand. Table 1 presents the proportion of times (out of 100) that pairs were declared to be the same. Notice that although pairs of identical product are on the diagonal, the diagonal elements do not appear to be identical. Based on the data in Table 1, your goal is to construct a map of the products from the ten plants.

Table 1. Number of times (out of 100) that pairs of products were declared to be the same. Plant numbers are given in the first row and the first column.

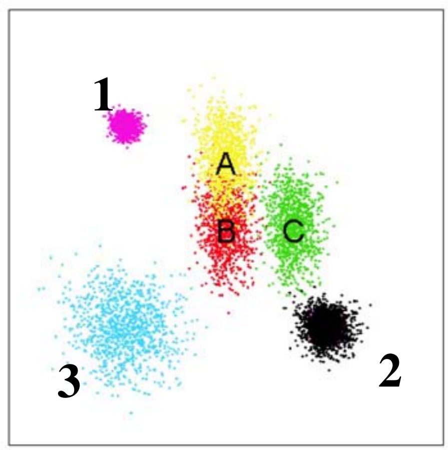
Plant Number	1	2	3	4	5	6	7	8	9	10
1	92	80	87	74	63	86	64	65	88	85
2		86	83	76	77	87	77	72	86	83
3			94	68	68	89	71	62	90	80
4				77	68	75	65	71	75	80
5					75	75	74	68	72	71
6						90	76	70	90	85
7							73	65	73	69
8								68	68	72
9									91	85
10										86

Scaling Multiple Attributes: The first significant contributions to the scaling of similarities occurred in the late 1950's. In this form of scaling, each product was represented as a discrete point, rather than a distribution. We will refer to this type of multidimensional scaling (MDS) as deterministic. Probabilistic MDS treats each product as a distribution and the scaling units are perceptual standard deviations. This approach can be viewed as a multivariate extension of Thurstonian scaling. Deterministic MDS suffers from serious flaws that are absent in its probabilistic counterpart.

When we use deterministic MDS, we assume that each entry in Table 1 corresponds to a unique distance in a perceptual space and as the similarity increases, the distance decreases. Specifically, this model requires that pairs of identical products correspond to identical similarities. In practice, pairs of identical products do not produce identical 'same' response probabilities, as illustrated in Table 1. Hence, the deterministic MDS model is handicapped.

Figure 1 shows products as distributions. Notice that the distributions labelled 1, 2 and 3 all differ in variance. If an individual is presented with pairs of samples from a single one of these distributions and asked to decide if the pairs are the same or different, it is very clear that the object with the smaller distribution (1) will be declared 'same' more often than the other two. A probabilistic account of the product information provides an intuitive understanding of why pairs of identical products do not produce identical 'same' probabilities.

Figure 1. Multivariate distributions corresponding to products that differ in variance.



A second idea from deterministic MDS is that the similarity measure is monotonically related to inter-product distance. This assumption implies that when a pair of products are less similar than a second pair, the first pair are further apart than the second pair. This assumption is also violated in practice. Probabilistic MDS has no problem with violations of this assumption, however. The distance between the means of A and B is equal to the distance between the means of B and C, but their similarity measures are not the same. A is more similar to B than B is to C. A deterministic MDS model would push B and C further apart to account for their greater dissimilarity. In other words, this type of model confounds variance with distance since it only has distance to work with. Most consumer products are variable either because they contain natural ingredients or because manufacturing precision is not absolute. In addition, consumer perception of products varies from time to time. In view of these facts about products and people, we need to consider variance when mapping consumer product perceptions.

How Same-Different Judgments Arise: Multivariate probabilistic models for same-different judgments have been published[4,5]. One of these models has a very simple form and we will used it to interpret Table 1. In this model we assume that the probability of a 'same' response is a decreasing function of distance between momentary values. When the distance is zero, the probability is 1. However, unlike a deterministic model, this function only applies to momentary values from the product distributions. Products with large variances will yield lower 'same' probabilities than those with smaller variances, explaining differences in the 'same' probability between pairs of identical products. In this model, the 'same' probability is not monotonically related to inter-product distances as it also depends on the variance and the relative orientation of products to one another.

Mapping Table 1: Figure 2 shows a fit to Table 1 as a map constructed from the probabilistic similarity model. Products are shown with confidence limit ellipses and it can be seen that as the first dimension intensity increases, so does its variance. Variances of the products on the second dimension remain constant. Differences in total variance for the products explain differences seen in the diagonal of Table 1. For instance, the variance for product 3 is smaller than product 8 and the self-similarity values of 0.94 and 0.68, respectively, are consistent with this. Once the map of similarities has been constructed, descriptive information is used to explain the dimensions identified. The product means on the first dimension correspond mainly to perceived pulpiness while sweetness primarily explains the second dimension. The fact that as products become pulpier the variance increases suggests that control of pulpiness is less easily achieved than sweetness. The problem appears to be that the five plants, 2, 4, 5, 7 and 8, produce excessively pulpy product and that this product is more inconsistent because of the difficulty in maintaining high pulp levels from one container to another. A deterministic map of Table 1, Figure 3, suggests that products from the plants producing pulpier product differ more among themselves relative to products from the remaining plants as shown in Figure 2. It can be seen in this figure how products with greater variability are pushed away from each other to accommodate the variance effect using distance. This model cannot explain the lower self-similarity of the pulpy products and provides no insight into the variability of product produced at specific plants.

Figure 2. Probabilistic multidimensional scaling of same-different judgments for orange juice products from ten plants. Numbers correspond to the plants.

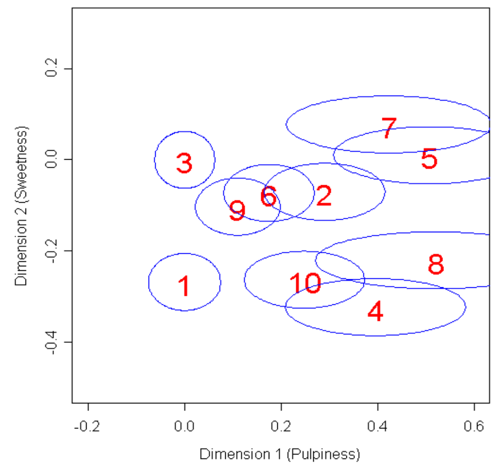
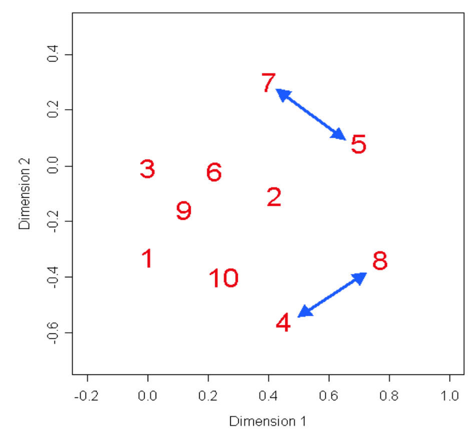


Figure 3. Deterministic multidimensional scaling of same-different judgments for orange juice products from ten plants.



Conclusion: Probabilistic MDS provides compelling solutions to problems in which product and perceptual variability arise. They also provide interesting diagnostic information for researchers interested in the dimensionality of product differences and the variances of products. Insights from these analyses provide useful tools to guide product quality management.

References:

1. Ennis, D.M. and Bi. J. (1998). Drivers of Liking. *IFPress*. **1**(1),2-3.
2. Ennis, D.M. (1999). Multivariate preference mapping. *IFPress*, **2**(2),2-3.
3. Ennis, D.M. (2001) Drivers of Liking® for multiple segments. *IFPress*, **4**(1), 2-3.
4. Ennis, D.M., Palen, J. and Mullen, K. 1988. A multidimensional stochastic theory of similarity, *Journal of Mathematical Psychology*, **32**, 449-465.
5. Ennis, D.M. and Johnson, N.L. 1993. Thurstone-Shepard similarity models as special cases of moment-generating functions. *Journal of Mathematical Psychology*, **37**, 104-110.