Modeling Diversified Preference for Flowers

Hirohisa Kishino
University of Tokyo
Yayoi 1-1-1, Bunkyo-ku,
Tokyo 113-8657, Japan
kishino@wheat.ab.a.u-tokyo.ac.jp

Sho-Ken Tan
University of Tokyo
Yayoi 1-1-1, Bunkyo-ku,
Tokyo 113-8657, Japan
tan@wheat.ab.a.u-tokyo.ac.jp

Yasushi Takano
University of Tokyo
Yayoi 1-1-1, Bunkyo-ku,
Tokyo 113-8657, Japan
takano@wheat.ab.a.u-tokyo.ac.jp

1. Introduction

With the maturity of the consumers' life, its importance is growingly recognized to measure and characterize their diversified preference of consumers. Here, we extend Thurstone's classical model (1927) expressing the discrimination of two stimuli and develop a unified maximum likelihood procedure. It can analyze various types of preference test and effective when fairly a large number of objects are examined. In such cases, the preference test becomes inevitably incomplete.

As a case study, we carried out experiment to examine the preference of $N = 191$ students (subjects) for $H = 20$ varieties (objects) of violet. The experiment consisted of paired comparisons by sixty slides, four-fold comparisons by twenty slides, and choice-of-three-out-of-eight comparisons by twenty slides. Pretest showed that these experiments require six seconds, eight seconds and thirty seconds per slide respectively.

2. Likelihood of generalized Thurstone's model and a hierarchical model

Let $(n_{s1}, \ldots, n_{sK})$ be the set of the varieties out of the H varieties to be compared, which are extracted for the sth slide of K-fold comparisons ($s = 1, \ldots, S$). The dataset consists of Z_{sk}^{n} ($n = 1, \ldots, N$, $s = 1, \ldots, S$, $k = 1, \ldots, K$) which takes value of 1 if the nth subject selects the kth variety in the sth slide, and value of 0 otherwise. The generalized Thurstone's model for K-fold comparisons consisting of S slides is described as

$$L = \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{k=1}^{K} Z_{sk}^{n} \log P\left(X_{v_{sk}} = \max_{k'} (X_{v_{sk'}}) \right). \qquad (1)$$

Here, X_h follow independent normal distributions with means m_h and variance 1 with $\sum_{h=1}^{H} m_h = 0$.

The estimated scores from the four-fold comparison agreed well with those from paired comparison (Figure 1). The estimated scores from paired comparisons have smaller variance for intermediate varieties, whereas those from the four-fold comparisons have smaller variances for preferred varieties. Given a fixed number of slides, the variance of estimates from the paired comparisons became explosively large when a large number of varieties are compared. On the other hand, as far as the preferred varieties are concerned, the variance from four-fold comparison did not depend much on the number of varieties to be compared.

3. Predicted selection probability: correlations and a hierarchical model

Predicted selection probabilities of varieties are obtained from the estimated scores. However, it is important to take account of the effect of assortment when multiple varieties are selected. Correspondence analysis applied to a slide of choice-of-three-out-of-eight comparisons implies correlation structure between the varieties (Figure 2). The likelihood of this experiment is obtained by slight modification of equation 1, considering the probability that the selected varieties have the three largest scores among those on the slide. The model taking account of correlations gave much different predicted selection probabilities from those from the model assuming independence among the varieties. Under the independence model, the selection probabilities of variety 4 and variety 5 increased in parallel with the increased number of selected varieties. On the other hand, under the model taking account of correlations, the selection probability of variety 4 caught up with that of variety 5 with the increased number of selected varieties (Figure 3).

It is also important to take account of diversity of preference for better prediction of the selection probability. Based on the first stage analysis, we propose a hierarchical model with the following prior distribution for the preference of individuals.

$$P\left(m_1^{(1)},\ldots,m_1^{(H)},\ldots,m_N^{(1)},\ldots,m_N^{(H)}\right)=\prod_{n=1}^{N}\prod_{h=1}^{H}\left(\frac{1}{\sqrt{2\pi g^{(h)}s_n^2}}e^{-\frac{\left(m_n^{(h)}-s_n\bar{m}^{(h)}\right)^2}{2g^{(h)}s_n^2}}\right).$$

The three types of hyper-parameters, s_n^2 $(n=1,\ldots,N)$, $\bar{m}^{(h)}$ and $g^{(h)}$ $(h=1,\ldots,H)$ measure the self-inconsistency of individuals, the population preference to varieties and the diversity of preference respectively. Posterior distributions of the parameters and hyper-parameters were obtained by Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). The most and the third preferred varieties had large diversity, whereas the second preferred variety had a small diversity (Figure 4). Difference in diversity turned over the ranks in selection probability between the second and the third preferred varieties.

REFERENCE

Hasting, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57: 97-109.

Metropolis, N., N.A. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. (1953) Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21: 1087-1092.

Thurstone, L.L. (1927) A law of comparative judgement. Psychological Review, 34: 273-286.

RÉSUMÉ

Nous developped un procédé de maximum de vraisemblance du modèle de Thurstone généralisé pour analyser de divers types d'essai de préférence. Pour décrire la préférence diversifiée, nous developped un modèle hiérarchique.

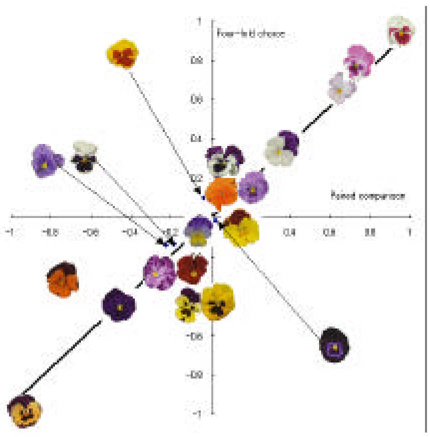

Figure 1 Estimated scores from paired and four-fold comparisons

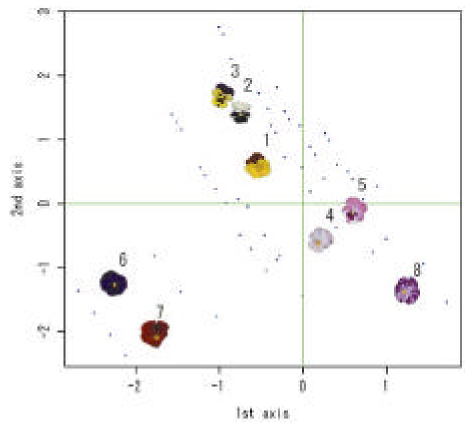

Figure 2 Correspondence analysis applied to one slide of choice-of-three-out-of-eight comparisons

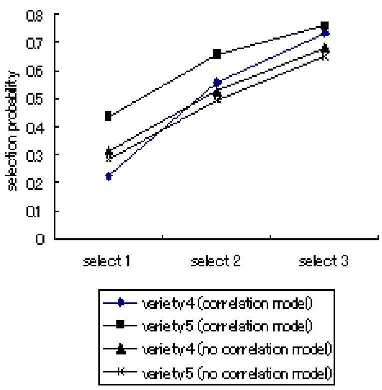

Figure 3 Predicted selection probabilities: the effect of assortment

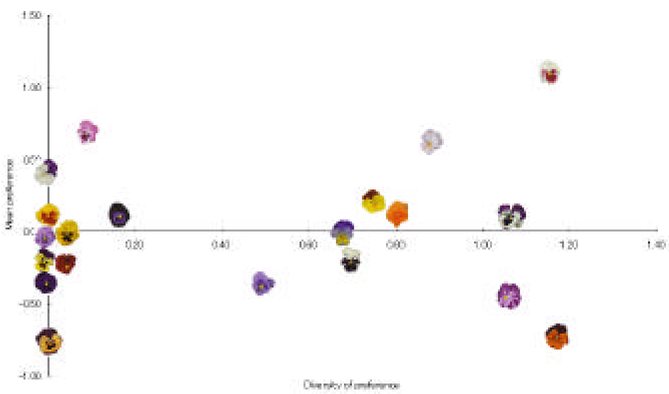

Figure 4 Mean and diversity of preference estimated from a hierarchical model