NTT に基づく応用モデルの検討
-一対比較モデルと多次元モデルを中心に-

宇佐美慧
(東京大学大学院教育学研究科・日本学術振興会)

1　序論

ニューラルテスト理論（Neural Test Theory : NTT : Shojima, 2007, 2008a, 2008b）は，自己組織化マップ（Self-organizing Map : SOM, Kohonen, 1995）のメカニズムを利用した，主にテストデータを分析する為の潜在ランク理論である．NTT モデルにおいては，各項目の各カテゴリに対する選択確率を表す項目参照プロファイル（Item Category Reference Profiles : ICRP) が潜在ランクごとに推定され，同時に回答者は離散的な潜在ランク上に配置される．宇佐美 (2009a) では，NTT において，潜在ランクおよび ICRP の推定精度，項目サンプリングを超えた潜在ランクの推定値の一貫性，分析モデルの構築・改良などの方法論的な観点，および実用に際し潜在ランクに基づく評価が学習者に与える影響に関する，教育心理学・教育社会学的な観点から検討すべき課題を指摘した．今後はさらに，方法論的問題においては NTT モデルそのものに関する，内的な意味での理論的構築だけではなく，例えば NTT モデルと従来の潜在変数モデリング（Latent Variable Modeling : LVM）との理論上・実用上の双方における異同について記述していくような外的な視点も重要である．本発表では，上記の方法論的な観点からの問題において焦点を当て，NTT に基づく応用モデルの検討を，特に多次元モデルと一対比較モデルを中心に行い，さらにこれらのモデルに関して外的な視点も交えた理論的考察を試みる．

2　NTT と一対比較モデル

一対比較データは，対提示された刺激の比較判断の結果から得られるデータである (Thurstone, 1927). 一般に作業が簡便であり，評定尺度法に比べてより精緻で，刺激の差異に対して豊富な情報量が得られるという特色を持っており（e.g.,Maydeu-Olivares & Böckenholt, 2005 ; 宇佐美, 2009b)，行動科学系の分野においては今日でも広く用いられている．(e.g., Causeur & Husson, 2005 ; Hatzinger & Mazanec, 2007). また，Usami(2009) でも述べられているように，比較判断に伴う諸要因 (e.g., 順序効果) を考慮する方法 (e.g., Davidson & Beaver, 1977 ; 宇佐美, 2009b) や，刺激の尺度値の信頼性を考慮した方法 (e.g., Dittrich, Hatzinger, Katzenbeisser, 1998) など，多くのモデル拡張が行われている．本節では，刺激と個人の潜在特性値が同様の一次元上の心理学的連続体上に布置しており，判断確率が刺激と個人の潜在特性値の双方の関数となっている場合を考える．これは，例えば味覚検査における食物（刺激）の好みの比較判断において，各回答者の味覚には個人差があるという場面が挙げられる．

2.1　定式化

刺激 i と $j\,(1,\ldots i,j\ldots I)$ における個人 $h(1,\ldots h,\ldots H)$ の比較判断データ$\boldsymbol{u_h}$および全体のデータ$\boldsymbol{U}$を

$$\boldsymbol{u_h}=(u_{h1},u_{h2},\ldots,u_{hK})\qquad \boldsymbol{U}=(\boldsymbol{u_1}',\boldsymbol{u_2}',\ldots,\boldsymbol{u_H}')' \tag{1}$$

と表す．ただし $K={}_IC_2$ であり，K はとりうる刺激対の総数を表す．また，ここでは簡単のため，各 h に対して全ての刺激対が 1 度ずつ提示され，かつ反応が 2 値（例えば Yes or No に基づく形式）である場合を想定する．これらの制約を超えたより一般的な表現への拡張は容易に行うことができる．この場合，NTT のアルゴリズムに基づけば，各個人 h は任意に設定された，一次元上に順序的に配列されている Q 個の，各々サイズが $K\times 1$ の潜在ランク $\boldsymbol{v_q}(1,\ldots q,\ldots Q;\boldsymbol{v_q}=(v_{q1},\ldots v_{qk},\ldots v_{qK});\boldsymbol{V}=(\boldsymbol{v_1}',\ldots\boldsymbol{v_q}',\ldots\boldsymbol{v_Q}'))$ の中で，ある最適化基準をもっともよく満たすランクに割り振られ，その都度$\boldsymbol{V}$の要素を更新するということになる．

また, 比較判断の個人差を想定した一対比較判断データは，回答者の潜在特性値に近くに布置している方の刺激が高い確率で選択されることが想定される．例えば Andrich(1988) における展開データ (unfolding data) を分析する項目反応モデルにおいては，個人 h が刺激 i と j をそれぞれ選択する確率を，刺激と個人の潜在特

性値を各々 b_i, b_j, θ_h とあらわしたときに,

$$P_{hi} = \frac{\exp(-(\theta_h - b_i)^2)}{1 + \exp(-(\theta_h - b_i)^2)} \quad P_{hj} = \frac{\exp(-(\theta_h - b_j)^2)}{1 + \exp(-(\theta_h - b_j)^2)} \tag{2}$$

のような個人と刺激の潜在特性値の布置の距離の関数で表現している．すなわち，b は刺激の理想点 (ideal point) を意味する．ここでこのモデルをもとに刺激 i と j の一対比較判断を考えた場合に，各刺激への評価に対する反応の独立性を仮定すれば，たとえば刺激 i が選択される確率 P_{ij} は $P_{ij} = P_{hi}/(P_{hi} + P_{hj})$ と考えられ，これは一般に θ_h についての単調増加もしくは単調減少関数となる．このことにより，$\boldsymbol{V}$ の列 $\boldsymbol{V_k}$ の要素は，潜在ランクがいま一次元上に順序的に配置されているので，潜在ランクの変化に対して単調増加もしくは単調減少の関係を想定することができる．したがって，一対比較判断における任意の刺激対についての判断確率の単調性が，一般の累積型の項目反応データと同様に仮定できるために，Shojima(2007) や宇佐美 (2009a) で挙げられているような，従来のアルゴリズムを拡張した方法で推定できる．すなわち，

STEP1 分析前の準備

(1) $\boldsymbol{V}$ の初期値の設定.

まず，何らかの方法 (例えば 1 式を用いてパラメタを直接推定する方法や，個人差を仮定しない場合でまず分析する方法など) を用いて I 個の刺激を一次元上に，その潜在特性値の小さい順に配置する．すなわち I 番目の刺激が最も高い潜在特性値であるとさしあたり仮定する．ここで潜在特性値が高いと仮定された方の刺激についての判断確率を常に想定することで，潜在ランクの番号について単調増加する性質を満たすことができる．そして，例えば以下のように，選択確率が単調増加する性質を満たすような，$\boldsymbol{V}$の初期値を設定する．

$$v_{qk} = \frac{q}{Q+1} \tag{3}$$

(2) 潜在ランクの事前分布の設定.

STEP3 以降でベイズ基準を利用する場合に必要な設定であり，具体的には事前確率 π_q を定める．Shojima(2008a) でも指摘されているように，SOM の性質上，最小二乗基準や最尤基準を用いた場合に，両端の潜在ランク ($q = 1, Q$) の度数に偏りが生じることから，事前確率の設定はその影響を緩和する効果がある．事前確率の設定については，離散一様分布においてその両端の潜在ランクに対応する事前確率を調整する方法や，三角分布を利用する方法が考えられる (宇佐美, 2009a).

STEP2 $\boldsymbol{U}$ の行成分のソート.

これは，STEP3 以降の学習において各行（回答者）について行われる学習の順番をランダムにする為である．

STEP3 潜在ランクへの割り当て.

ある h に対して，最適化基準である距離関数 d を最小にするランク $w(1, \ldots w, \ldots Q)$ を選ぶ．d については，最小二乗基準 (LS)・最尤基準 (ML)・ベイズ基準 (MAP) があり，それぞれ以下の最適化基準を最小にする w が選ばれる．なお，欠損値を含んだより一般的な場合については Shojima(2007)，宇佐美 (2009a) を参照のこと．

$$d_{LS} = \sum_{k=1}^{K}(u_{hk} - v_{qk})^2 \qquad d_{ML} = -\sum_{k=1}^{K} u_{hk}\log v_{qk} - \sum_{k=1}^{K}(1-u_{hk})\log(1-v_{qk}) \qquad d_{MAP} = d_{ML} - \log\pi_q \qquad w = argmin\, d \tag{4}$$

STEP4 $\boldsymbol{V}$ の要素を更新.

更新の方法は必ずしも一義的でないが，本発表では以下のように行う．

$$\boldsymbol{v_q*} = \boldsymbol{v_q} + \boldsymbol{h_{qw}} \bullet (\boldsymbol{u_h} - \boldsymbol{v_q}) \tag{5}$$

ただし $\bullet$ はアダマール積であり，

$$h_{qw} = \frac{\alpha_t Q}{H}\exp(-\frac{(q-w)^2}{2\sigma_t^2}) \quad \alpha_t = \frac{(T-t)\alpha_1 + (t-1)\alpha_T}{T-1} \quad \sigma_t = \frac{(T-t)\sigma_1 + (t-1)\sigma_T}{T-1} \tag{6}$$

である．ここで，$\alpha_1, \alpha_T, \sigma_1, \sigma_T$ は，収束までの速さを調節するための，事前に設定する定数である．
また，T は以下に述べる全体の反復回数を意味し，$t(1, \ldots t, \ldots T)$ である．そしてさらに，以下の単調制約を満たすように $\boldsymbol{v_q*}$ の要素 v_{qk*} を v_{qk**} に更新する．更新における場合分けについての根拠は，STEP1(1) の場合と同様である．

$$v_{(q+1)k**} = \begin{cases} v_{(q+1)k*}, & (v_{(q+1)k*} > v_{qk*}) \\ v_{qk*}, & (v_{(q+1)k*} \le v_{qk*}) \end{cases} \tag{7}$$

STEP5 STEP3-4 を全ての h に対して反復.

STEP6 STEP2-5 を T 回反復.

2.2　比較判断の個人差を無視する場合

同一の心理学的連続体上における個人の布置を想定しない，すなわち判断の個人差を単純な誤差として捉える場合（例えば食物の味覚検査において個人差を考慮しない場合），ないしはスポーツデータなどに見られる，そもそも個人が判断する過程を経ないデータの場合，Q=1 の仮定のもとで分析することに等しい．したがってこれらの場合は，複数の潜在ランクを想定する NTT の本来の目的とは外れたものになる．

3　NTT と多次元モデル

本節では前節とは異なり，対比較ではなく，一般の学力テストデータのように，個人と 1 つの刺激 (項目) の潜在特性値の関数で反応が規定される場合を考える．このようなデータを分析することが当初の NTT の目的であり，そのため既に理論的検討がなされているが (e.g., Shojima, 2007)，多次元的な潜在ランクを想定する場合の検討はなされていない．そこで以下では，個人が L 次元上の潜在ランクに所属している状況を仮定する．

3.1　定式化

前節の一対比較データの場合に沿って，以下のアルゴリズムで計算される．ただし前節における $\boldsymbol{V}$ はここでサイズ $Q \times I$ であり，$\boldsymbol{V} = (\boldsymbol{v_1}, \ldots \boldsymbol{v_l}, \ldots \boldsymbol{v_L})'$，$Q = \sum_{l=1}^{L} Q_l$ と表わされる．I は項目数で，Q_l は第 l 次元における潜在ランク数を意味する．さらに $\boldsymbol{v_l}$ はサイズ $Q_l \times I$ の行列であり，その要素 v_{lq_li} には第 l 次元の $q(1, \ldots q_l, \ldots Q_l)$ 番目の潜在ランクにおける項目 i への正答の寄与を表す数値が入っている．すなわち，反応確率は各次元に所属する潜在ランクに対応する要素の和で決定される．そのため，$\boldsymbol{V}$ のサイズの大きさや，更新の方法，および要素間の制約が若干異なるが，前節と同様のアルゴリズムを応用して推定することができる．すなわち，

STEP1 分析前の準備.
(1) $\boldsymbol{V}$ の初期値の設定．$v_{lq_li} = q_l/(L(Q_l + 1))$ とするが，次元の識別性のために $v_{l11} \leq v_{(l+1)11}$ となるように調整する．他にも，$\boldsymbol{U}$ についての因子分析の結果を利用して適宜ウェイトをつける方法も考えられる．
(2) 潜在ランクの事前分布の設定．　第 l 次元における q_l 番目の潜在ランクの事前確率 π_{lq_l} の設定の方法については前節と同様である．

STEP2 $\boldsymbol{U}$ の行成分のソート.

STEP3 最適なランクの抽出.
ある h に対して，最適化基準である以下の距離関数 d を最小にする，q_l についてのランクの組み合わせ $\boldsymbol{W}(w_1, \ldots w_l, \ldots w_L)$ を求める．

$$d_{LS} = \sum_{i=1}^{I}(u_{hi} - \sum_{l=1}^{L} v_{lq_li})^2 \quad d_{ML} = -\sum_{i=1}^{I} u_{hi} \log(\sum_{l=1}^{L} v_{lq_li}) - \sum_{i=1}^{I}(1-u_{hi}) \log(1-\sum_{l=1}^{L} v_{lq_li}) \quad d_{MAP} = d_{ML} - \sum_{l=1}^{L} \log \pi_{lq_l} \quad \boldsymbol{W} = argmin\, d \tag{8}$$

STEP4 各 l 次元において，$\boldsymbol{V}$ の要素を更新.

$$\boldsymbol{v_{lq_l*}} = \boldsymbol{v_{lq_l}} + \boldsymbol{h_{qw_l}} \bullet (\boldsymbol{u_h} - \sum_{l=1}^{L} \boldsymbol{v_{lq_l}}) \tag{9}$$

ここで，$\boldsymbol{h_{qw_l}} = \boldsymbol{z_l} \bullet \boldsymbol{h_{qw}}$ であり，$\boldsymbol{z_l}(= z_{l1}, \ldots z_{li}, \ldots z_{lI})$ は各次元において，収束までの速さを調節するための定数であり，以下のように定義される

$$z_{li} = \begin{cases} 1 & (l = argmin(u_{hi} - v_{lq_li})) \\ 0 & (l \neq argmin(u_{hi} - v_{lq_li})) \end{cases} \tag{10}$$

すなわち，u_{hi} を最もよく説明した次元のみを更新するように調整することとなる．次に，以下の単調制約を満たすように $\boldsymbol{V}^*$ の要素を更新する．

$$v_{lq*i} \leq v_{l(q+1)*i} \qquad v_{l11} \leq v_{(l+1)11} \tag{11}$$

なお，$\sum_{l=1}^{L} v_{lq_l i} < 1$ の条件を満たさない q_lの 組み合わせがあった場合，要素の修正を行う.

STEP5 STEP3-4 を全ての h に対して反復.

STEP6 STEP2-5 を T 回反復.

3.2　モデルの特徴と応用

Hunter (2004) や Causeur & Husson (2005) では，一対比較データにおいて伝統的に用いられてきている Bradley-Terry モデル (Bradley & Terry, 1952) を多次元に拡張し，例えば推移律が満たされない場合を考慮したモデリングを行っている．前節と前々節の議論をもとにすれば，これらと同様のアルゴリズムに基づく多次元の一対比較モデルの構築は比較的容易であると思われる．さらには多値型データの場合への応用などについても，同様に検討できるであろう．

それでは，SOM のアルゴリズムに基づく，本発表で検討した一対比較モデルと多次元モデルは，理論上・実用上の双方においてどのような特徴があるだろうか．上述のように，最適化基準において最尤基準やベイズ基準が用いられるが，反復計算および更新においてはノンパラメトリックな方法に基づいて離散的な潜在クラス $\boldsymbol{V}$ の要素と各回答者の所属ランクが推定される．また，一対比較モデルは IRT や SEM などの LVM の枠組みで表現できることも踏まえれば (e.g., Ozaki & Toyoda, 2006 ; Maydeu-Olivares & Böckenholt, 2005)，多次元モデルも含め，今回検討したモデルはノンパラメトリックな発想のもとでの，SOM のアルゴリズムを利用した潜在クラスモデルとして位置づけられると思われる．これは，従来の方法論 (例えば混合分布モデルやノンパラメトリック IRT) との類似点を示唆することでもあり，同時に SOM のアルゴリズムを用いていることから，推定アルゴリズムの構築や結果の解釈，および関連する指標についての独自性をも示唆する．とりわけ，ベイズ基準を用いた推定においてはその事前分布 π_{lq_l} の設定による潜在ランクの度数分布の調整が可能であるという点に，また本発表のようにモデルの拡張を行った場合においても比較的一貫したアルゴリズムの構築が可能であるという点に柔軟性（と同時に一部に任意性）があると思われる．これらの点についての考察はまだ不十分であり，ここでの考察も大雑把な指摘であるに過ぎないが，冒頭で述べた NTT の外的な意味での理論的発展のためには検討すべき論点と言えよう．

なお，本発表で検討したアルゴリズムに基づく，データの分析例については当日報告予定である．シミュレーションデータの作成においては，一対比較モデルの場合においては (2) 式を，多次元モデルの場合では，多次元のロジスティックモデル

$$P_{hi} = \frac{\exp(\sum_{l=1}^{L} a_{il}\theta_{hl} + d_i)}{1 + \exp(\sum_{l=1}^{L} a_{il}\theta_{hl} + d_i)} \tag{12}$$

を利用した．ここで a_{il}, θ_{hl} は第 l 次元における項目の識別力母数及び能力母数であり，d_i は困難度に関連する母数である．

4　参考文献

Andrich, D. (1988). The application of an unfolding model of the PIRT Type to the measurement of attitude. *Applied Psychological Measurement, 12*, 33-51.
Bradley, R.A., & Terry, M.E. (1952). The rank analysis of incomplete block designs. : The mathod of paired comparisons. *Biometrika*, 39, 324-345.
Causeur, D; Husson, F. (2005). A 2-dimensional extension of the Bradley-Terry model for paired comparisons. *Journal of Statistical Planning and Inference, 135*(2), 245-259.
Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Applied Statistics, 47*, 511-525.
Hatzinger, R., & Mazanec, J.A. (2007). Measuring the part worth of the mode of transport in a trip package: An extended Bradley-Terry model for paired-comparison conjoint data. *Journal of Business Research. 60*, 1290-1302.
Hunter, D.R. (2004). MM algorithms for generalized Bradley Terry models. *Annals of Statistics, 32*, 386-408
Kohonen, T. (1995). *Self-organizing maps*. Springer.
Maydeu-Olivares, A., & Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods, 10*(3), 285-304.
Ozaki, K., & Toyoda, H. (2006). A paired comparison IRT model using 3-VALUE judgement : estimation of item difficulty parameters prior to the administration of the test. *Behaviormetrika, 33*(2), 131-147.
Shojima, K. (2007). The graded neural test model: A neural test model for ordered polytomous data. *DNC Research Note*, RN07-03.
Shojima, K. (2008a). Neural test theory: A latent rank theory for analyzing test data. *DNC Research Note*, RN08-01.
Shojima, K. (2008b) Structural neurofield mapping: Latent rank model for multivariate data. The 36th Annual Meeting of the Behaviormetric Society of Japan, 179-180.
Thurstone, L., (1927). A law of comparative jugdement. *Psychogical Review, 34*, 273-286.
宇佐美慧 (2009a). ニューラルテスト理論の応用可能性 －方法論的課題の考察と多値型モデルの適用例－ 日本テスト学会誌, *5*(1), 65-80.
宇佐美慧 (2009b). 比較判断に影響する複数の要因を考慮した一対比較データ分析法. 心理学研究, 536-541.
Usami, S. (2009). Bayesian generalized Bradley-Terry model using RJMCMC method. Proceeding of the 36th Annual Meeting of the Behaviormetric Society of Japan. Oita University.

宇佐美慧 (usami_s@p.u-tokyo.ac.jp)