PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **20** (2011), 5–12

PLISKA
STUDIA MATHEMATICA
BULGARICA

MANY FACET, COMMON LATENT TRAIT POLYTOMOUS IRT MODEL AND EM ALGORITHM

Dimitar Atanasov

There are many areas of assessment where the level of performance can not be determined using ordinal, "objective" test. For example, there are aspects of writing language skills which should be subjectively qualified by judges based on some criteria. Usually such data is studied using the Many Facet Rasch Model. To address the problem with Item Response Theory (IRT) approach a Many Facet IRT model for polytomous item response in the case of common latent trait is presented. For estimation of the parameters of the model the EM-algorithm is derived.

1. Introduction

The methods of analysis of a classical test items determine the result of the exam as a interaction between the abilities of the individuals and the properties of the test items. A common problem arise when the the answer of the individual is judged by a referee or rater. Such is the case of essay evaluation, where number or raters rate the essay according to some criteria on a fixed scale. This adds to the process of evaluation additional facets of rater severity and the difficulty of the particular criteria. In a similar way, if an individual can choose different themes for the essay, there is a facet, representing the difficulty of the chosen theme. In order to have a fair assessment of the student ability all these facets should be included in the process of evaluation.

2000 *Mathematics Subject Classification:* 91E45.
Key words: item response theory, many facets, EM algorithm.

From different point of view it is needed to have coordination and agreement between the raters, so their severity should be compared as well the way of application of different criteria.

One of the classical approaches to the problem is the so called Many Facet Rasch Model (Linacre, 1994). This model expands the Rasch-based models for achievement evaluation (Andrich, 1985) and includes the influence of the different facets to the final result.

In this work we present an approach to this problem based on the Item Response Theory (IRT) properties (parameters) of the facets. An EM algorithm for estimation of these parameters is derived.

We assume that the abilities of the students can be estimated using the unidirectional scale. Therefore one can suppose that there is a common latent trait, which is responsible for the performance of the student on the rating criteria.

2. The Model

Let us consider the classical essay scoring, where a number of raters evaluates the essays according to the set of criteria. For any of the criteria, the rater gives a score using a linear non-decreasing scale. The estimation of the parameters of the model is based on the following representation, common in the Many Facet approach (Linacre, 1994)

$$\log\left(\frac{P_{nijk}}{P_{nij(k-1)}}\right) = B_n - D_i - C_j - F_k,$$

where P_{nijk} is the probability an individual n to be rated on criteria i from rater j with rate k. The parameter B_n is the student ability, D_i is the criteria difficulty, C_j is the rater severity and F_k is the level, needed to achieve grade k from grade $k-1$.

The model can be expanded with additional facets as well as with an assumption that the step F_k can differ for different raters.

According to the IRT model (Crocker & Algila, 1986), the probability for correct performance on the given test item, defined by 2-parametric logistic model is

$$P(\theta) = \frac{1}{1 + e^{-aD(\theta-b)}},$$

where θ is the level of the measured abilities of the individual, the parameters of the model b and a represent the difficulty and the level of discrimination of the item respectively. The value $D = 1.7$ is a scaling parameter, giving an approximation to the Normal ogive curve.

To represent the model in the terms of the IRT concept let us consider the following model.

Let θ be the latent ability of the individual. Let the essay of the student be rated on the set of J criteria using a scale from 0 to $K - 1$. A specific ability $\Theta_j, j = 1, \ldots, J$ is related to any of these criteria. For the relation between specific abilities and the generic latent ability of the individual the following factor model holds

$$\Theta_j = \alpha_j \theta + \varepsilon_j,$$

where α_j is a factor loading and ε_j is a $N(0, \sigma^2)$ random error.

Let us note with $r = (r_1, \ldots, r_J)$ the rates of the essay over all criteria. Then for $j = 1, \ldots J$ we have

$$P(r_j \geq k \mid \theta) = P(\Theta_j \geq \delta_k \mid \theta) =$$

$$= P(\alpha_j \theta + \varepsilon_j \geq \delta_k) = P(\varepsilon_j \geq \delta_k - \alpha_j \theta) =$$

$$= 1 - \Phi_\varepsilon(\delta_k - \alpha_j \theta) = \Phi(\frac{\alpha_j \theta - \delta_k}{\sqrt{\sigma^2}}),$$

where δ_k is the is the level of the specific ability θ_j for an examinee to achieve grade of k or higher, Φ_ε is the distribution function of ε and Φ is the distribution function of the standard normal distribution.

Let the level δ_k be a result of influence of h different facets of the assessment procedure. In the case of a equitable assessment these facets should not depend on the assessment scale. Then δ_k can be represented by the factor model

$$\delta_k = \sum_{i=1}^{h} \gamma_i \Delta_i + \Delta^k + \epsilon,$$

where Δ_i are the assessment facets and γ_i are their factor loadings, Δ^k is a specific value, needed for rate k, ϵ is a $N(0, \sigma_\delta^2)$ random error, independent from $\varepsilon_j, j = 1, \ldots, J$. Then

$$P(\alpha_j \theta + \varepsilon_j \geq \delta_k) = P(\varepsilon_j - \epsilon \geq \sum_{i=1}^{h} \gamma_i \Delta_i + \Delta^k - \alpha_j \theta) = \Phi(\alpha_j \theta - \sum_{i=1}^{h} \gamma_i \Delta_i - \Delta^k),$$

for properly normalized α_j, γ_i and Δ^k.

For the probability of to receive grade category k from a person with latent trait θ we have

$$P(r_j = k \mid \theta) = P(\delta_k \leq \theta < \delta_{k+1} \mid \theta) =$$

$$= \Phi(\alpha_j\theta - \sum_{i=1}^{h} \gamma_i \Delta_i - \Delta^k) - \Phi(\alpha_j\theta - \sum_{i=1}^{h} \gamma_i \Delta_i - \Delta^{k+1}).$$

If we use the approximation of the normal ogive curve with a logistic function we have

$$(1) \quad P(r_j = k \mid \theta) = \frac{1}{1 + e^{-D(\alpha_j\theta - \sum_{i=1}^{h} \gamma_i \Delta_i - \Delta^k)}} - \frac{1}{1 + e^{-D(\alpha_j\theta - \sum_{i=1}^{h} \gamma_i \Delta_i - \Delta^{k+1})}}$$

Remark

In the general factor model, used for the relation between the specific abilities and the latent trait of the individual the set of criteria can be replaced by different facet of the assessment if it is inherit in the individual. A model, where all the factor loading $\alpha_j, j = 1, \ldots, J$ are equal, can be used as well. In that case the assessment criteria should be considered as one of the facets in the factor model for δ_k. In similar way an interactions between the latent trait and other facets can be included in the model. Hence all the configuration of facets given in Linacre (1994) can be presented.

3. The EM Algorithm

For the estimation of the parameters of the model an EM algorithm is derived. A similar algorithm in the case of IRT model without influence of different facets is given in Woodruff & Hanson (1997). A complete description of the concept of the EM algorithm is given in McLachlan & Krishnan (2008). The idea is to maximize the complete likelihood function on the set of unknown parameters of the model as well as on the value of the unknown latent trait θ.

Let θ take discrete values $\theta_{[1]}, \ldots, \theta_{[P]}$ with corresponding probabilities $\pi_1, \ldots, \pi_P$. Let $R = \{r_{ij}\}$ be a matrix of the students rates, where r_{ij} is the grade of the individual $i, i = 1, \ldots, N$ on the criteria j, $j = 1, \ldots, J$. Let $H = H_{ij}$, i, $i = 1, \ldots, N$; $j = 1, \ldots, h$ be the matrix, which defines the levels of the assessment facets, where H_{ij} is the level of the facet j, which influences the grade of the individual i. With r^i and H^i we note the i-th row of the matrices R and H respectively.

Then the probability of a given grade r^i over the all population of individual is

$$f(r^i \mid \Delta, \pi, H^i) = \sum_{t=1}^{P} f(r^i \mid \theta_{[t]}, \Delta, \pi, H^i)\pi_t,$$

where Δ is the set of unknown parameters in the model (1) and π is the vector $\pi = (\pi_1, \ldots, \pi_P)$. If the rates of the different criteria are independent we have

$$f(r^i \mid \theta_{[t]}, \Delta, H^i) = \prod_{j=1}^{J} \prod_{k=1}^{K-1} P(r_{ij} = k \mid \theta_{[t]}, \Delta, H^i)^{I(r_{ij}=k)}, \tag{2}$$

where$P(r_{ij} = k \mid \theta_{[t]}, \Delta, H^i)$ is the probability, defined by (1) and $I(\cdot)$ is the indicator function.

Therefore the likelihood function can be represented as

$$L(R \mid \theta, \Delta, \pi, H) = \prod_{i=1}^{N} \sum_{t=1}^{P} \pi_t \left(\prod_{j=1}^{J} \prod_{k=1}^{K-1} P(r_{ij} = k \mid \theta_{[t]}, \Delta, H^i)^{I(r_{ij}=k)} \right).$$

The probability that a randomly chosen individual with a given level of abilities $\theta_{[t]}$ has a set of rates r_i is

$$f(r^i, \theta_{[t]} \mid \Delta, \pi, H) = f(r^i \mid \theta_{[t]}, \Delta, H^i).\pi_t.$$

Then the complete likelihood is

$$L(R, \theta \mid \Delta, \pi, H) = \prod_{i=1}^{N} \prod_{t=1}^{P} \prod_{j=1}^{J} \prod_{k=1}^{K-1} P(r_{ij} = k \mid \theta_{[t]}, \Delta, H^i)^{I(r_{ij}=k)} \pi_t =$$

$$= \prod_{t=1}^{P} \prod_{j=1}^{J} \prod_{k=1}^{K-1} \prod_{i=1}^{N} P(r_{ij} = k \mid \theta_{[t]}, \Delta, H^i)\pi_t =$$

$$\prod_{t=1}^{P} \prod_{j=1}^{J} \prod_{k=1}^{K-1} \prod_{H \in \aleph} P(r_{.j} = k \mid \theta_{[t]}, \Delta, H^i)^{\eta^H_{jtk}} \pi_t^{\nu_t}, \tag{3}$$

where $\aleph$ denotes the set of different values of the observed facets H^i, $P(r_{.j} = k \mid \theta_{[t]}, \Delta, H^i)$ is the probability of the rate k on the criterion j given the value of the other parameters kept fixed. Note with ν_t the number of individual with the value of the latent trait $\theta_{[t]}$ and η^H_{jtk} is the number of individual with latent trait $\theta_{[t]}$, rated on criterion j with grade category k, under the values of the facets in H. The (η^H_{jtk}, ν_t), $t = 1, \ldots, P$, $j = 1, \ldots, J$, $k = 0, \ldots, K-1$, $H \in \aleph$ is a sufficient statistics.

The aim of the algorithm is to find the expected value of the η^H_{jtk} and ν_t given set of values for Δ and π (E-step). The result is substituted in (3) and after that (M-step) the complete likelihood function (3) is maximized on Δ and π. The obtained ML estimators are substituted in the E-step for calculating new values for the expectation of η^H_{jtk} and ν_t and so on. This procedure continues until some optimal criteria is reached, for example if the change of the likelihood function is relatively small.

Let us first find the distribution of the latent trait θ given the set of rates r^i and the parameters of the model Δ and π. The values $\Delta^{(s)}$ and $\pi^{(s)}$ are found on the previous step $(s-1)$ of the algorithm.

$$f(\theta_{[t]} \mid r^i, \Delta^{(s)}, \pi^{(s)}, H^i) = \frac{f(\theta_{[t]}, r^i \mid \Delta^{(s)}, \pi^{(s)}, H^i)}{f(r^i \mid \Delta^{(s)}, \pi^{(s)}, H^i)} =$$

$$= \frac{f(\theta_{[t]}, r^i \mid \Delta^{(s)}, \pi^{(s)}, H^i)}{\sum_{p=1}^{P} f(r^i \mid \theta_{[p]} \Delta^{(s)}, \pi^{(s)}, H^i) P(\theta_{[p]} \mid \Delta^{(s)}, \pi^{(s)}, H^i)} =$$

$$= \frac{f(r^i \mid \theta_{[t]}, \Delta^{(s)}, \pi^{(s)}, H^i) P(\theta_{[t]} \mid \Delta^{(s)}, \pi^{(s)}, H^i)}{\sum_{p=1}^{P} f(r^i \mid \theta_{[p]} \Delta^{(s)}, \pi^{(s)}, H^i) \pi^{(s)}_p} =$$

$$= \frac{f(r^i \mid \theta_{[t]}, \Delta^{(s)}, \pi^{(s)}, H^i) \pi^{(s)}_t}{\sum_{p=1}^{P} f(r^i \mid \theta_{[p]} \Delta^{(s)}, \pi^{(s)}, H^i) \pi^{(s)}_p} =$$

$$= \frac{\prod_{j=1}^{J}\prod_{k=1}^{K-1} P(r_{ij}=k \mid \theta_{[t]}, \Delta, H^i)^{I(r_{ij}=k)}\pi_t^{(s)}}{\sum_{p=1}^{P}\prod_{j=1}^{J}\prod_{k=1}^{K-1} P(r_{ij}=k \mid \theta_{[p]}, \Delta^{(s)}, H^i)^{I(r_{ij}=k)}\pi_p^{(s)}}, \tag{4}$$

where in the last equation (2) is substituted and $P(\theta_{[t]} \mid \Delta^{(s)}, \pi^{(s)}, H^i) = \pi_t^{(s)}$
E-step. Then

$$\hat{\nu}_t = E(\nu_t \mid R, \Delta^{(s)}, \pi^{(s)}, H) = \sum_{i=1}^{N} f(\theta_{[t]} \mid r^i, \Delta^{(s)}, \pi^{(s)}, H^i) =$$

$$= \sum_{i=1}^{N} \frac{\prod_{j=1}^{J}\prod_{k=1}^{K-1} P(r_{ij}=k \mid \theta_{[t]}, \Delta^{(s)}, H^i)^{I(r_{ij}=k)}\pi_t^{(s)}}{\sum_{p=1}^{P}\prod_{j=1}^{J}\prod_{k=1}^{K-1} P(r_{ij}=k \mid \theta_{[p]}, \Delta^{(s)}, H^i)^{I(r_{ij}=k)}\pi_p^{(s)}}$$

$$\hat{\eta}_{jtk}^H = E(\eta_{jtk}^H \mid R, \Delta^{(s)}, \pi^{(s)}, H) = \sum_{i=1}^{N} P(r_{ij}=k \mid \theta_{[t]}, \Delta^{(s)}),$$

where $P(r_{ij}=k \mid \theta_{[t]}, \Delta^{(s)}$ is calculated from (1).

M-step. The values $\hat{\nu}_t$ and $\hat{\eta}_{jtk}^H$ are substituted in the complete likelihood function (3), after that the new values $\Delta^{(s+1)}$ and $\pi^{(s+1)}$ are obtained by its maximization.

Taking the logarithm of (3) we have

$$\log L(R, \theta \mid \Delta, \pi, H) =$$

$$\sum_{t=1}^{P}\sum_{j=1}^{J}\sum_{k=1}^{K-1}\sum_{H\in\aleph} \hat{\eta}_{jtk}^H P(r_{\cdot j}=k \mid \theta_{[t]}, \Delta, H^i) + \sum_{t=1}^{P}\hat{\nu}_t \log(\pi_t) \tag{5}$$

The maximization of (5) is equivalent to the maximization of both terms separately. The second term is the logarithm of the the likelihood function of the multinomial distribution. Therefore its maximum is in the point $\pi_t^{(s+1)} = \dfrac{\hat{\nu}_t}{N}$.

This is the value, used in the next E-step. The value of Δ is calculated by maximizing the first term in the equation (5).

REFERENCES

[1] D. Andrich.. An Elaboration of Guttman Scaling with Rasch Models for Measurement, Sociological Methodology. Vol. **15** 1985, 33–80.

[2] L. Crocker, J. Algila. Introduction to Classical and Modern Test Theory. Warsworth, 1986.

[3] M. Linacre. Many Faceted Rasch Measurement, Univ of Chicago Social Research, 1994.

[4] G. McLachlan, T. Krishnan. The EM Algorithm and Extentions. Second Edition. John Weley, 2008.

[5] D. Woodruff, B. Hanson. Estimation fot Item Response Models using EM Algorithm for Finite Mixtures. Presented at the Annual Meeting of the Psychometric Society, Gatlinburg, Tennessee, 1997.

Dimitar Atanasov
New Bulgarian University
21, Montevideo Str.
Sofia, Bulgaria
e-mail: `datanasov@nbu.bg`