Testlet Response Theory - Scoright

Kathrin Gruber

19 01 2010

Überblick

Was ist ein „testlet“

- Ursprünglich die Bezeichnung für eine Sammlung von Aufgaben die gemeinsam vorgegeben werden
  Um die Effizienz eines Tests, in Situationen bei denen Personen einen bestimmten Stimulus verstehen mussten, zu verbessern (Bsp.: Leseverständnisaufgaben - „reading passage“, “information graphs“, „musical passage“, „table of numbers“, ...)
- Einfache Methode um einen Test für „große Stimuli“ in effizienter Weise zu gestalten → Wende als *computerized adaptive testing (CAT)* umsetzbar wurde

Fundamentale Annahme des CAT:

Der *Itemkennwert* bleibt *konstant*, unberücksichtigt des *Kontextes* = „**assumption of item fungibility**“ (auch bekannt als *Kontexteffekte* in Erhebungen („survey designs“))

Def. Kontexteffekt:

Jeder Einfluss, den ein Item dadurch erzeugt, dass es in Beziehung zu einem anderen Item des selben Tests steht.

Die Itemposition: die Position des Items in einem Test (od. versch. Tests) hat einen Einfluss auf die Schätzung der Itemparameter

Cross-information: ein Item trägt versehentlich Information zu einem anderen Item bei

Unausgeglichener Inhalt: betrifft formale, inhaltliche Spezifikationen bzgl. der Aufgaben

Reihenfolge der Itemschwierigkeiten: ordnen der Items in aufsteigendem Schwierigkeitsgrad (lineare, nicht adaptive Tests)

Alternative zum CAT $\rightarrow$ „testlet“

„testlet“ als Lösungsansatz

Def. „testlet“:

Eine Gruppe von Items, die als Einheit entwickelt wurden und gemeinsam vorgegeben werden.

Auch die Items innerhalb eines „testlets“ können

- verzweigt („branched“),
- adaptiv (durch hierarchische Strukturierung) oder
- in aufsteigendem bzw. absteigendem Schwierigkeitsgrad

vorgegeben werden.

Ziel: durch das *Bündeln* von Items wird die Likelihood dieser nachteiligen Effekte reduziert (ohne die Effizienz des adaptiven Tests zu verringern) und die Teststruktur passt besser zu dem Konstrukt über welches man eine Aussage treffen will.

Zusammenfassung

3 Gründe „testlets“ zu benutzen:

1. Um die Bedenken bzgl. des Einzelfallcharakters von alleinstehenden Items zu reduzieren.
2. Um Kontexteffekte in adaptiven Designs zu reduzieren.
3. Um die Effizienz des Testens bei einem verlängerten Stimulus zu verbessern.

Bock's Model (1972)

Ansatz ist stark orientiert an *Rosenbaum's theorem of item bundles*, durch Gebrauch eines IRT-Models für mehrkategorielle Daten um die lokal unabhängigen Itembündel („testlets“) zu verrechnen.

Grundidee: jene Items die ein „testlet“ formen, besitzen (ev.) überhöhte lokale Abhängigkeit $\rightarrow$ das „testlet“ wird als alleinige Einheit betrachtet und mehrkategoriell verrechnet

$$P(T_{jx} = m_j) = \frac{exp(a_{jk}\theta + c_{jk})}{\sum_{k=0}^{mj} exp(a_{jk}\theta + c_{jk})} \quad (1)$$

J ... „testlets“ mit $j = 1, \ldots, J$

m_j ... Items für jedes „testlet“ J

$\{a_k, c_k\}$ ... Itemkategorienparameter für $k = 0, 1, \ldots, m_j$

Zusätzliche Restriktion $\sum_{k=0}^{m_j} a_{jk} = \sum_{k=0}^{m_j} c_{jk} = 0$

Vorteile des Modells

Dieses Vorgehen führt zu:

- interpretierbaren Itemparametern (inkl. Standardfehlern) und
- einem Maß für die Güte der Anpassung

$\rightarrow$ *expected score function* = *testlet information function* welche explizit den Beitrag des „testlets“ zum Gesamttest zeigt.

SCORIGHT - Warum bayesianischer Ansatz?

1. Wenn man mehr Information über das „testlet“ erhalten will
2. Ad-hoc Konstruktion eines „testlets“ innerhalb einer CAT-Vorgabe

ad 2.: Der Itemauswahlalgorithmus wählt den „testlet“-Stimulus und wählt danach ein Item aufgrund dessen Inhalt, dessen psychometrischen Eigenschaften und der bisherigen Antworten der Testperson.

Unterschied frequentistischer vs. bayesianischer Ansatz

Frequentistisch

Bsp.: bei unendlich vielen Versuchen $\longrightarrow$ betrachte Erfolge im Verhältnis zu Gesamtversuchen ohne jegliche Grundannahmen bzgl. der Verteilung zu treffen.

Bayesianisch

Bsp.: Sammeln von Daten, eine Hypothese bzgl. deren Verteilung ist bereits vorhanden. Unter der Vorraussetzung, dass die Verteilungsannahme (*Prior*-Verteilung) korrekt ist, bedingt den Daten $\longrightarrow$ schätze *Posterior*-Verteilung welche genauer ist als die *Prior*-Verteilung (d.h. updaten des Priors).

SCORIGHT benutzt die *empirical Bayes Methode*, d.h. den Parametern des Ausgangsmodels werden *Priors* zugewiesen und diese werden dann (empirisch) anhand der Daten geschätzt.

Ad. Prior:

z.B.: Mittelwert und Varianz der (angenommenen) Verteilung der Schwierigkeitsparameter.

Beachte:

Die Wahl des *Priors* beinhaltet eine gewisse Quelle der Unsicherheit: die Likelihood kann genauso falsch sein wie der gewählte Prior.

„testlet“ response Modelle

In das formale 3PL-Modell wird ein zusätzlicher Parameter $\gamma_{id(j)}$ eingebaut $\rightarrow$ beschreibt die „within - testlet“ Kovariation.

dichotomer Fall: 3PL-Modell

$$P(Y_{ij} = 1) = c_j + (1 - c_j) logit^{-1}(t_{ij}) \quad (2)$$

2PL-Modell als Spezialfall des 3PL-Modells wenn $c_{ij} = 0$

mehrkategorieller Fall: ordinal-response-Modell

$$P(Y_{ij} = r) = \phi(d_r - t_{ij}) - \phi(d_{r-1} - t_{ij}) \quad (3)$$

Y_{ij} ... Antwort von Tp i bzgl. Item j, c_j ... untere Asymptote (Rateparameter; für dichotome Items), d_r, d_{r-1} ... threshold Parameter (für mehrkategorielle Items), $logit^{-1} = log(\frac{x}{1-x})$, ϕ ... CDF der NV, t_{ij} ... latente lineare Score-Prediktor.

„testlet“ response Modelle

Ebenfalls in SCORIGHT implementiert:

- „testlet“ Modelle für dichotome UND mehrkategorielle Items
- „testlet“ Modelle mit Kovariaten

Der „testlet“-Parameter

Um die Abhängigkeit durch das „testlet“ zu modellieren, wird der lineare Prediktor erweitert:

$$t_{ij} = a_j(\theta_i - b_j - \gamma_{id(j)}) \tag{4}$$

a_j ... Anstieg (Krümmung)

b_j ... Itemschwierigkeit

θ_i ... latente Fähigkeit der Tp

$\gamma_{id(j)}$ ... „testletEffekt (Interaktion) von Item j mit Person i welcher im „testlet“ d_j genestet ist.

lt. Definition $\gamma_{id(j)} = 0$ für alle unabhängigen Items

Um alle Inforamtionen über die Tpn, die Items und die „testlets“ zu kombinieren $\rightarrow$ *hierarchischer Bayes-Ansatz*

Der hierarchische Bayes-Ansatz

Die Likelihoodfunktion für das 3PL-Modell:

$$P(Y_{ij} = 1|\lambda_{ij} = (\theta_i, a_j, b_j, c_j)) = \tag{5}$$
$$= \prod_{i=1}^{I} \prod_{j=1}^{N_i} (c_j + (1 - c_j) logit^{-1}(a_j(\theta_i - b_j)))$$

Besitzt die unbekannten Parameter $\lambda = (\lambda_{ij})$ welche unter dem Bayes-Ansatz *Prior*-Verteilungen benötigen $\longrightarrow P(\lambda|\Lambda)$

Priors:

$$\begin{aligned} \theta_i &\sim N(0,1) \\ log(a_j), b_j, logit(c_j) &\sim N_3(\mu = (\mu_a, \mu_b, \mu_c), \Sigma)^3 \\ \gamma_{id(j)} &\sim N(0, \sigma^2_\gamma) \end{aligned} \tag{6}$$

Der hierarchische Bayes-Ansatz

Um die Modellspezifikationen abzuschließen benötigt man noch die Spezifizierung einer *Hyperprior*-Verteilung $\pi(\Lambda)$

Hyperpriors:

$$\begin{aligned} \mu &\sim N_3(0, \nu \times I_3) \\ \Sigma &\sim Inv. - Wishart(S, n_0) \end{aligned} \tag{7}$$

Die marginale *Posterior*-Verteilung liegt nun nicht in geschlossener Form vor

Posterior-Verteilung:

$$p(\lambda|Y) \propto \int p(Y|\lambda)p(\lambda|\Lambda)\pi(\Lambda)d\Lambda \tag{8}$$

Der hierarchische Bayes-Ansatz

Lösung mittels MCMC

$\rightarrow$ Definieren einer Markov-Kette (d.h. jede Iteration des MCMC-Samplers hängt von der vorherigen Iteration ab)

SCORIGHT: Gelman und Rubin Methode (1992)

normalerweise benutzt man 3 - 5 unabhängige MCMC-Ketten

Um die Konvergenz dieser multiplen Ketten zur stationären Verteilung zu prüfen, wird ein F-Test benutzt welcher die *across-chain* Variation gegen die *within-chain* Variation testet. Ist dieses Verhältnis klein $\rightarrow$ Konvergenz!

Beispiel 1

Simulierte Datenmatrix: 500 Personen, 12 Items

Alle 12 Items dichotom

2 Testlets enthalten: Items 1 bis 4 (abhängig), Items 10 bis 12 (abhängig)

Keine Kovariaten enthalten

Beispiel 2

Simulierte Datenmatrix: 500 Personen, 10 Items

Alle 10 Items dichotom

2 Testlets enthalten: Items 1 bis 4, Items 8 bis 10 $\longrightarrow$ hier sind alle 10 Items abhängig

Keine Kovariaten enthalten

Literatur

Wainer, H., Bradlow, E. & Wang, X. (2007). *Testlet Response Theory and its Applications*. Cambridge University Press: New York.

User's Guide:
http : *//www.ets.org/Media/Research/pdf/RR* − 04 − 49.*pdf*