TPM®

EXTRACTING A KNOWLEDGE STRUCTURE FROM THE DATA BY A MAXIMUM RESIDUALS METHOD

EGIDIO ROBUSTO
LUCA STEFANUTTI
UNIVERSITY OF PADOVA

A major issue in knowledge space theory (KST) is the specification of a knowledge structure for a given set Q of problems. In the literature two distinct approaches can be found: the theory-driven and the data-driven. The data-driven approach includes a number of statistical methods for deriving a knowledge structure from a large enough data set. In the present article a procedure for extracting knowledge structures from data is proposed, which is less restrictive compared to other methods. The procedure constructs a chain of knowledge structures $\mathcal{K}_i$ of increasing size. In each step i an updating rule is applied, which is based on the residuals of the response pattern's frequencies obtained by an application of the basic local independence model. In a simulation study, this updating rule was compared to the one at the basis of the method proposed by Schrepp (1999). The results indicate superiority of the proposed procedure in a number of different conditions. An empirical application to a data set of fraction subtraction problems shows the viability of the method.

Key words: Knowledge Space Theory; Basic Local Independence Model; Data-driven methods; Maximum Residuals Method; Model selection.

Correspondence concerning this article should be addressed to Egidio Robusto, Department of Philosophy, Pedagogy, Sociology and Applied Psychology, University of Padova, Via Venezia 8, 35131 Padova (PD), Italy. Email: egidio.robusto@unipd.it

Given a set Q of problems in a certain field of knowledge, in knowledge space theory (KST) the knowledge state of an individual is the subset $K \subseteq Q$ of all problems in Q that the individual is capable of solving. The collection $\mathcal{K}$ of all knowledge states existing in a population is named *knowledge structure*. A basic assumption in KST is that both the empty set Ø and the full set Q belong to $\mathcal{K}$. A major issue in KST is the specification of a correct knowledge structure for a given set Q of problems. In the literature two distinct approaches can be found: the theory-driven and the data-driven.

The theory-driven approach comprises the so-called expert query procedures (Dowling, 1993; Düntsch & Gediga, 1996; Kambouri, Koppen, Villano, & Falmagne, 1994; Koppen & Doignon, 1990; Stefanutti & Koppen, 2003) and construction procedures based on task analysis and skill maps (Doignon, 1994; Heller, Ünlü, & Albert, 2013; Korossy, 1999; Lukas & Albert, 1993). All these methods are based on theoretical principles that allow to establish whether a given subset $K \subseteq Q$ of problems is a knowledge state or not.

The data-driven approach includes a number of statistical methods for deriving a knowledge structure from a large enough data set. Among the existing procedures we mention the method proposed by Schrepp (1999), the item tree analysis (ITA; Schrepp, 2003) and the inductive ITA (IITA; Sargin & Ünlü, 2009). Unlike Schrepp's method, both ITA and IITA pose spe-

cific restrictions on the type of knowledge structure that can be extracted from the data. These restrictions require that the structure is closed under both union and intersection (thus corresponding to a quasi order on the set Q of problems). On the other hand, Schrepp's procedure is based on strong assumptions that hardly hold in practice. This issue is discussed in the section titled "Extracting knowledge structures from data: Two procedures." Cosyn and Thiéry (2000) developed a data-driven method for refining a knowledge structure initially constructed by querying experts.

In the present article a procedure for extracting knowledge structures from data is proposed, which represents a less restrictive alternative to Schrepp's method. Our approach allows the construction of a chain of knowledge structures $\mathcal{K}_i$ of increasing size. In each step i an updating rule is applied, which is based on the residuals of the response pattern's frequencies obtained by an application of the basic local independence model (BLIM; Doignon & Falmagne, 1999; Falmagne & Doignon,1988).

The article is organized as follows: the following section is an overview of the BLIM; the next one presents the proposed procedure for extracting a knowledge structure from data; the section titled "Simulation studies" illustrates the results of two simulation studies, in which the accuracy of the procedure is assessed; in the section titled "Application to a real data set" the method is applied to a real data set; finally, in the last section we offer our conclusions.

THE BASIC LOCAL INDEPENDENCE MODEL

The BLIM is a probabilistic model for knowledge structures developed by Falmagne and Doignon (1988). Properties and extensions of this model have been investigated in a number of articles (de Chiusole, Stefanutti, Anselmi, & Robusto, 2013; Heller & Wickelmaier, 2013; Robusto, Stefanutti, & Anselmi, 2010; Stefanutti, Anselmi, & Robusto 2011; Stefanutti, Heller, Anselmi, & Robusto 2012; Stefanutti & Robusto, 2009).

Let Q be a nonempty finite set containing n distinct dichotomous items, and $\mathcal{K}$ be a knowledge structure on Q. Both Q and $\mathcal{K}$ are fixed throughout the section. The response pattern of a student randomly sampled from the population is represented by a discrete random variable $\mathbf{R}$ whose realizations are subsets $R \subseteq Q$, containing all items that have been correctly solved. The unknown knowledge state of this student is represented by a discrete random variable $\mathbf{K}$ whose realizations are elements $K \in \mathcal{K}$. The probability of sampling a student whose response pattern is R is denoted by $P(\mathbf{R} = R)$, and the probability that the knowledge state of this student is $K \in \mathcal{K}$ is denoted by $P(\mathbf{K} = K)$. The connection between the observable response patterns and the unobservable knowledge states is given in the BLIM by the following unrestricted latent class model (see, e.g., Goodman, 1973; Haberman, 1979)

$$P(\mathbf{R} = R) = \sum_{K \in \mathcal{K}} P(\mathbf{R} = R | \mathbf{K} = K) P(\mathbf{K} = K) \tag{1}$$

where $P(\mathbf{R} = R|\mathbf{K} = K)$ is the conditional probability that the response pattern of a randomly sampled student is $R \subseteq Q$ given that his knowledge state is $K \in \mathcal{K}$. The BLIM is then characterized by three types of parameters: a parameter π_K specifying the probability of each knowledge state, a careless error parameter β_q and a lucky guess parameter η_q for every item $q \in Q$. The parameter β_q is interpreted as the probability that a student will fail q given that this item is indeed solvable from his knowledge state. The parameter η_q specifies the probability that a student solves q given that this last is not in his knowledge state.

Assuming local independence among the responses, given the knowledge states, the conditional probability of a response pattern R given knowledge state $K \in \mathcal{K}$ takes on the form

$$P(\mathbf{R} = R | \mathbf{K} = K) = \left(\prod_{q \in K \setminus R} \beta_q \right) \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \left(\prod_{q \in R \setminus K} \eta_q \right) \left(\prod_{q \in Q \setminus (K \cup R)} (1 - \eta_q) \right) \tag{2}$$

Equations (1) and (2) are the two basic equations of the BLIM.

EXTRACTING KNOWLEDGE STRUCTURES FROM DATA: TWO PROCEDURES

In this section two procedures are presented for the extraction of a knowledge structure from an observed data set consisting of a large enough number of response patterns. Given some reference population of students, we assume that a "true" knowledge structure $\mathcal{K}$ exists for the finite set Q of items, in such population. We also assume that every knowledge state $K \in \mathcal{K}$ has some fixed probability π_K to be sampled from the population, that is, $\pi : \mathcal{K} \rightarrow \mathbb{R}$ is a probability distribution on the knowledge states. Moreover, for every item $q \in Q$ there are fixed careless error probabilities β_q and lucky guess probabilities η_q. Henceforth, we will refer to the n-tuple $(Q, \mathcal{K}, \pi, \beta, \eta)$ as the *true model* (or, sometimes, the *generating model*). All these quantities, as well as the true knowledge structure $\mathcal{K}$ are supposed to be unknown. The observable data consist of a sample of N distinct response patterns, each of which is the collection R of problems in Q that have been correctly solved by some randomly sampled student. With $F_o(R)$ we denote the observed frequency of response pattern $R \subseteq Q$. The aim is to recover the structure $\mathcal{K}$, as well as the probabilities β, η and π from the observed response patterns, when no prior information is given about $\mathcal{K}$.

Starting from the simplest knowledge structure $\mathcal{K}_0 = \{\emptyset, Q\}$ the proposed recovery procedure generates a chain $\mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_n$ ($n \leq 2^{|Q|} - 2$) of knowledge structures. For each structure $\mathcal{K}_i$ in the chain, the BLIM parameters β, η and π are estimated by maximum likelihood, and the resulting model is tested (by some suitable GOF statistics) against the observed data. Then a criterion is established for terminating the procedure at step n.

Two critical aspects that have to be specified for the procedure described above are the *updating rule* and the *termination criterion*. The updating rule establishes which new state is chosen and added to $\mathcal{K}_i$ in each step i of the procedure, thus yielding $\mathcal{K}_{i+1}$. The termination criterion specifies the step i at which the procedure stops. We say that the recovery procedure behaves *perfectly* if both of the following conditions are met

1. in each step i the new state added to $\mathcal{K}_i$ belongs to $\mathcal{K}$, that is, $\mathcal{K}_{i+1} \setminus \mathcal{K}_i \subseteq \mathcal{K}$
2. the procedure terminates at step $n = |\mathcal{K}| - 2$.

It is clear that if both Conditions (1) and (2) hold then $\mathcal{K}_n = \mathcal{K}$ and thus $\mathcal{K}$ is perfectly recovered by the procedure. However such conditions are only ideal since, irrespectively of the updating rule and termination criterion that one decides to use, there will be some positive probability of violating Condition (1) in every single step i of the procedure. Thus one can only hope to keep such probability as small as possible.

As for the updating rule, in this article we consider and examine two alternative rules. They are both based on the assumption that the collection $\mathcal{K}$ of true knowledge states is a subset of the collection $\mathcal{R}$ of the observed response patterns. This assumption has the clear disadvantage that, if some state does not belong to the collection of the observed response patterns, then it has a null probability of belonging to the recovered structure, with the implication that the true knowledge structure will

not be perfectly recovered. This might happen if either of the following conditions occurs: (a) the sample size is not large enough, compared to the number of knowledge states in $\mathcal{K}$; (b) the probability π_K of some states is too small, and no student in the sample is in knowledge state K; (c) careless error and lucky guess probabilities of the items are too high (although some student is in knowledge state K, due to too high error probabilities, there is no response pattern exactly reproducing that state). In the rest of this section we will assume that the condition $\mathcal{K} \subseteq \mathcal{R}$ holds true.

As stated above, two alternative updating rules have been studied. For $i = 1,2,\ldots,n$, let $\mathcal{N}_i = \mathcal{R} \backslash \mathcal{K}_i$ be the collection of all observed patterns that have not yet been added to $\mathcal{K}_i$. The first updating rule (henceforth called FMax) adds to $\mathcal{K}_i$ the element $R \in \mathcal{N}_i$ having the largest observed frequency $F_o(R)$ (Schrepp, 1999). If more than one element satisfies this condition, then a random choice among them is done.

Since $\mathcal{K} \subseteq \mathcal{R}$, the collection $\mathcal{R}$ can be partitioned into the two subcollections $\mathcal{K}$ and $\overline{\mathcal{K}} = \mathcal{R} \backslash \mathcal{K}$. Then it is easy to see that the FMax rule never violates Condition (1) if and only if

$$\min\{F_o(R) : R \in \mathcal{K}\} > \max\{F_o(R') : R' \in \overline{\mathcal{K}}\}.$$

It is thus clear that the FMax rule fails whenever there is some *non-state*-pattern having a frequency $F_o(R)$ which is higher than that of some *state*-pattern.

The second updating rule is based on a systematic comparison between the expected and observed frequencies of the response patterns. Given any step i of the procedure let $F_e^{<i>}(R)$ denote the expected frequency of response pattern R computed at step i. This frequency is obtained by an application of the BLIM model equations to the values of the parameters $\beta^{<i>}$, $\eta^{<i>}$, $\pi^{<i>}$ estimated at that step. Therefore $F_e^{<i>}$ is the frequency that we would expect to observe for pattern R, in a sample of size N, if $\mathcal{K}_i$ was the correct knowledge structure. After computing the residual $\Delta_i(R) = F_o(R) - F_e^{<i>}(R)$ for each response pattern in $\mathcal{R}$, the pattern for which such residual is maximum (most positive) is added to $\mathcal{K}_i$. For this reason this rule will be called MaxRes henceforth. Again, if two or more patterns happen to have the same maximum residual, only one of them is chosen at random.

Concerning the termination criterion, an ideal condition would be to stop the procedure at the step $i > 0$ in which some suitable measure of the distance of the reconstructed structure $\mathcal{K}_i$ from the true structure $\mathcal{K}_i$ is minimum. In the literature (see, e.g., Falmagne & Doignon, 2011) a distance measure between two knowledge structures $\mathcal{K}$ and $\mathcal{K}'$ on the same set Q of items is defined by

$$d(\mathcal{K},\mathcal{K}') = \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} d_{min}(K,\mathcal{K}') \tag{3}$$

where

$$d_{min}(K,\mathcal{K}') = \min\{|K \,\Delta\, K'| : K' \in \mathcal{K}'\}$$

and

$$K \,\Delta\, K' = (K \backslash K') \cup (K' \backslash K)$$

is the symmetric difference between K and K'. It should be noted that (i) the distance $d(\mathcal{K}, \mathcal{K}')$ is not symmetric, since in general $d(\mathcal{K}, \mathcal{K}') \neq d(\mathcal{K}', \mathcal{K})$; moreover (ii) $d(\mathcal{K}, \mathcal{K}') = 0$ whenever $\mathcal{K}' \subseteq \mathcal{K}$. A mean

$$d^*(\mathcal{K},\mathcal{K}') = \frac{1}{2}(d(\mathcal{K},\mathcal{K}') + d(\mathcal{K}',\mathcal{K}))$$

of the two distances was considered in this article. This distance satisfies the two properties:

1. $d^*(\mathcal{K}, \mathcal{K}') = d^*(\mathcal{K}', \mathcal{K})$ (symmetry),
2. $d^*(\mathcal{K}, \mathcal{K}') = 0$ if and only if $\mathcal{K} = \mathcal{K}'$.

Therefore, an ideal stopping criterion would terminate the reconstruction procedure at the step i in which $d^*(\mathcal{K}_i, \mathcal{K})$ is minimum.

Since the true knowledge structure is unknown, so is the distance $d^*(\mathcal{K}_i, \mathcal{K})$. One possibility is to try to obtain an estimate of d^* from the observed data. The more direct choice would be to use the distance $d^*(\mathcal{K}_i, \mathcal{R})$ as a possible estimate of $d^*(\mathcal{K}_i, \mathcal{K})$. However that would not work. In fact, since $\mathcal{K}_i \subseteq \mathcal{R}$ at any step $i > 0$, the distance $d(\mathcal{R},\mathcal{K}_i)$ would always be zero. The consequence is that $d^*(\mathcal{K}_i, \mathcal{R}) = d(\mathcal{K}_i, \mathcal{R})/2$, which is monotone decreasing and attains zero in the trivial case $\mathcal{K}_i = \mathcal{R}$. For this reason $d^*(\mathcal{K}_i, \mathcal{R})$ cannot be used as an estimate of $d^*(\mathcal{K}_i, \mathcal{K})$. In this article, the behavior of three well-known alternative model selection criteria was studied. The three indexes were the Akaike Information Criterion (AIC; Akaike, 1974), the Bayesian Information Criterion (BIC; Schwarz, 1978), and the corrected AIC (AICc; Burnham & Anderson, 2002).

Simulation Studies

In order to assess the behavior of the FMax and MaxRes procedures, two simulation studies were carried out. In the first of them, the aim was to compare the performances of the two updating rules presented in the above section. In the second study the performance of the three model selection criteria AIC, BIC, and AICc was assessed. In both studies the two procedures were tested in a series of eight simulation conditions, in which the number of items, the number of knowledge states and the maximum value of the error parameters (lucky guesses and careless errors) were manipulated. In each simulation condition, a "true" model was generated, with the number of items and states as specified in Table 1. The lucky guess and careless error parameters of the true model were randomly drawn from a uniform distribution on the open interval $(0,\theta_{max})$, where θ_{max} was 0.2 in odd conditions, and 0.4 in even conditions.

Table 1
Simulation conditions

Condition	# of items	# of states	Max error
1	10	50	0.2
2	10	50	0.4
3	10	250	0.2
4	10	250	0.4
5	10	500	0.2
6	10	500	0.4
7	20	250	0.2
8	20	250	0.4

Simulation Study 1

The FMax and MaxRes procedures were compared in each of the eight simulation conditions, with respect to the capability of reconstructing the true knowledge structure. In each condi-

tion the simulation design was as follows: (i) the expected frequency of each response pattern $R \subseteq Q$ was computed as

$$F(R) = N \sum_{K \in \mathcal{K}} P(R|K)\, \pi_K;$$

where $N = 1000$ is the sample size, $P(R|K)$ is the conditional probability of response pattern R, given knowledge state K, computed through Equation (2), π_K is the probability of state K; (ii) the FMax and MaxRes procedures were applied to the expected frequencies $F(R)$; (iii) for each of the two procedures, and at each step i the distance $d^*(\mathcal{K}_i, \mathcal{K})$ was computed. The results are displayed in Figure 1, where each of the numbered diagrams refers to a different condition. In each diagram the number of states added by the reconstruction procedures is along the x-axis, whereas the distances are measured along the y-axis. The continuous curves correspond to the MaxRes procedure, whereas the dashed curves correspond to the FMax procedure. The dotted vertical lines indicate the number of states in the true knowledge structure.

Overall, it can be noted that the MaxRes procedure attains smaller distances from the true structure than the FMax procedure does. This holds irrespectively of the number of states in the true structure, the size of the error parameters, and the number of items. Moreover, the MaxRes procedure seems to be less sensitive to the size of the error rates. In fact, in moving from a condition with max error 0.2 (the even conditions) to the corresponding one with max error 0.4 (the odds conditions), the distance of FMax increases much more than that of MaxRes, and this holds true irrespectively of the number of items or states. Concerning the minimum distance attained by MaxRes, it is observed that this value is close to zero in all conditions with small error and in almost all conditions with large error. Furthermore, the point of minimum along the x-axis is equal or very close to the number of states in the true knowledge structure. On the basis of the results obtained in this simulation study, the conclusion can be drawn that MaxRes performs better than FMax. For this reason, in the simulation study presented in the next section the analysis is restricted to MaxRes.

Simulation Study 2

In each of the eight conditions of Table 1, the BLIM model was used to generate 100 simulated data sets, each of size 1000. Then the MaxRes procedures were applied to each of the 100 datasets, obtaining a reconstructed knowledge structure for each of them. On the whole the MaxRes procedure was applied to 100 × 8 = 800 data sets. For each of them, in every single iteration i of the MaxRes procedure, the values of the AIC, AICc, and BIC criteria were computed for the BLIM estimated for knowledge structure $\mathcal{K}_i$. In this way, for each of the three criteria, a Conditions × Datasets × Iterations matrix was obtained.

Starting from this matrix, averages across datasets were computed and are displayed in Figure 2. Each of the eight diagrams in the figure corresponds to a different simulation condition. In each diagram the number of states added by the procedure is along the x-axis, whereas the values of AIC, AICc, and BIC are along the y-axis. In each of the diagrams three curves are displayed. The continuous curve represents the average AIC at each iteration of the procedure, the dash-dotted curve represents the AICc, and the dashed curve represents the BIC. The continuous, dash-dotted and dashed vertical lines represent the point of minimum of AIC, AICc, and BIC respectively. Finally, the dotted vertical line represents the number of states in the true knowledge structure.

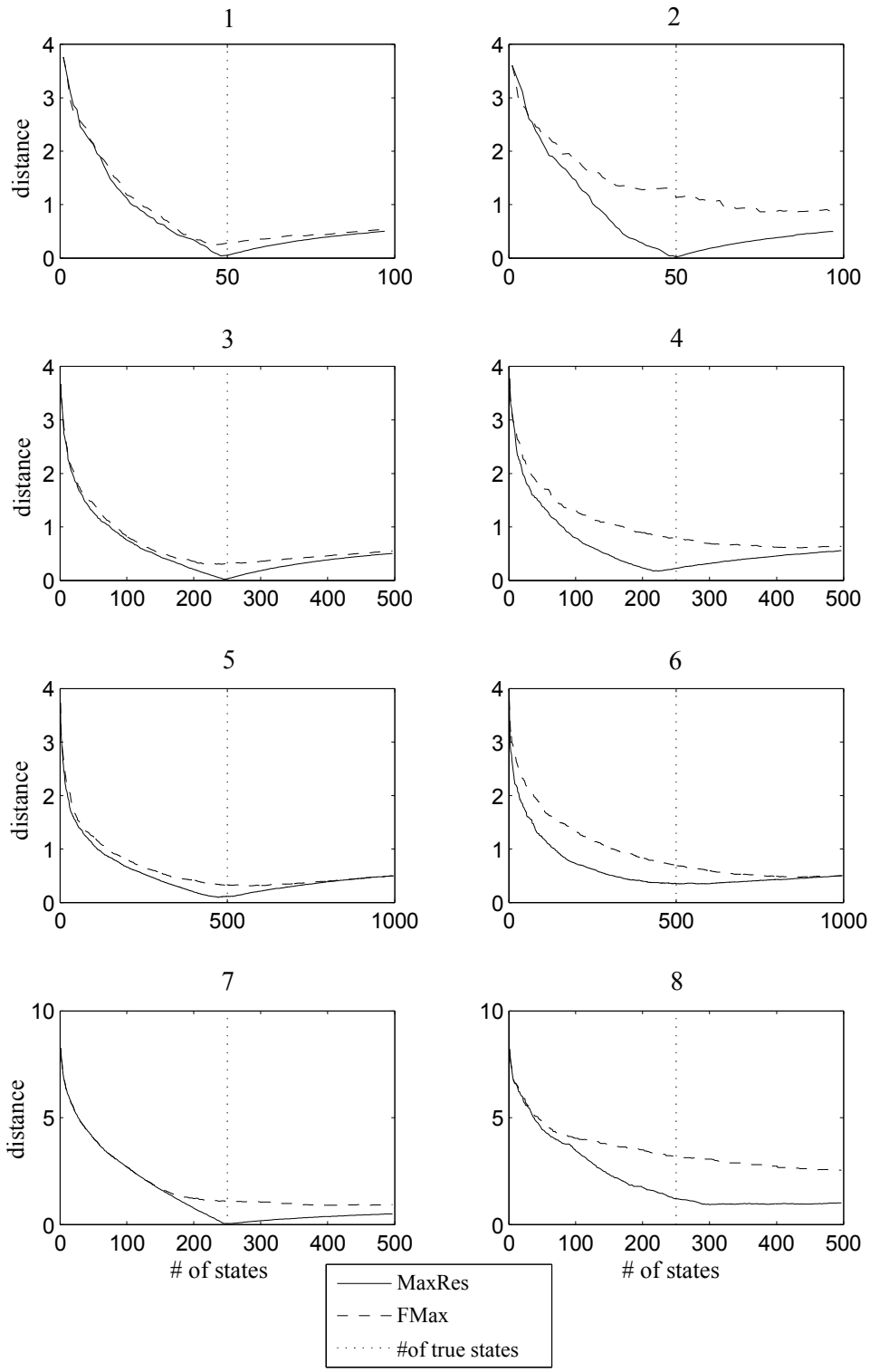

FIGURE 1
Distance from the true structure in every iteration of the MaxRes and FMax procedures in each of the eight simulation conditions. The dotted vertical lines indicate the # of states in the true structure.

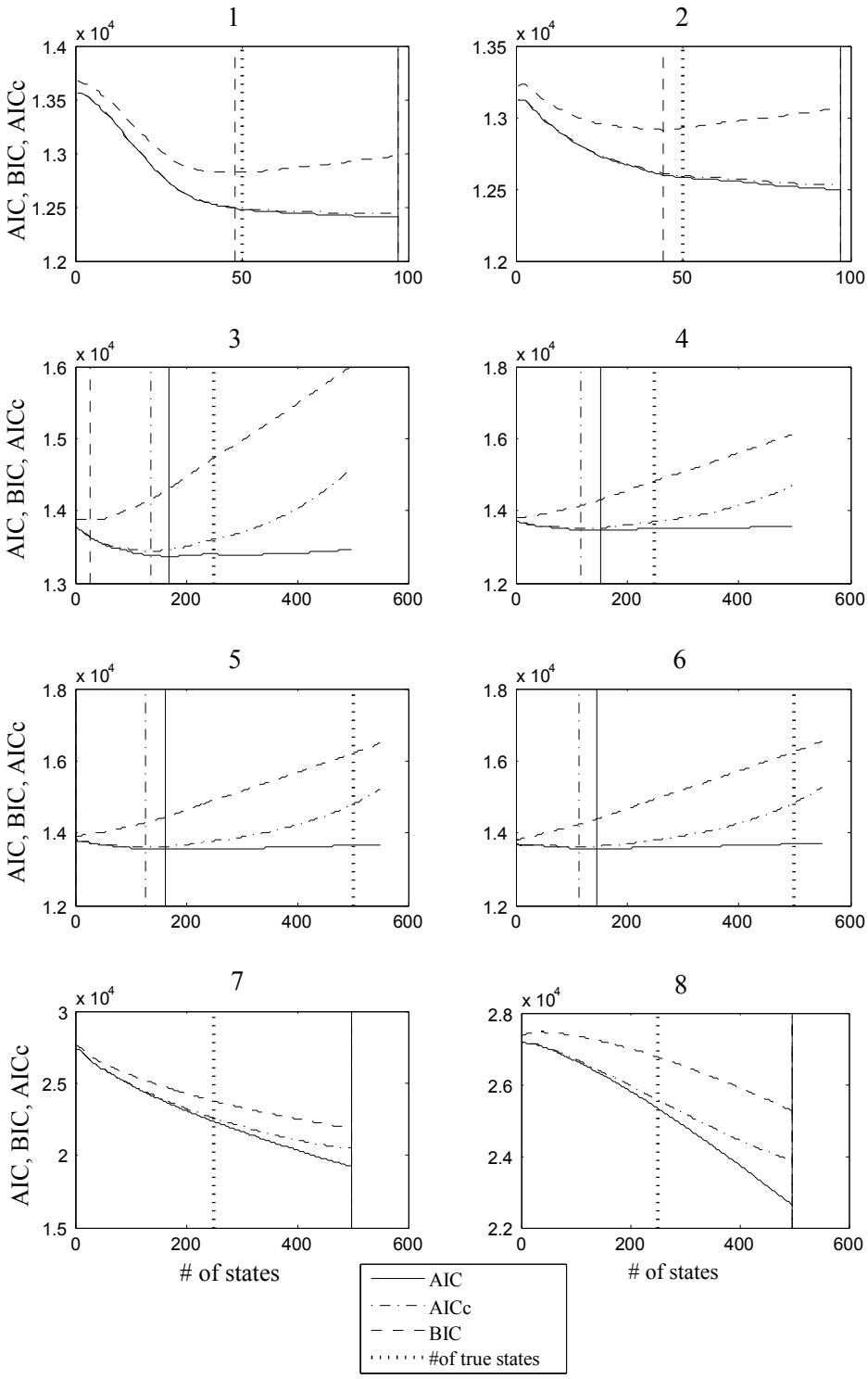

FIGURE 2
AIC, AICc, and BIC values in every iteration of the MaxRes procedure, in each of the eight simulation conditions. The dotted vertical lines indicate the # of states in the true structure. The continuous, dash-dotted and dashed vertical lines indicate the iterations in which AIC, AICc, and BIC, respectively, attain their minimum values.

Recalling that, for each of the three selection criteria, the point of minimum should be used as a stopping criterion of the procedure, in general none of the three criteria behaves satisfactorily. In fact in almost all the conditions a comparison between the point of minimum of each of the three criteria and the true number of states shows that these criteria would terminate the procedure too early or too late. There are two special cases in which the BIC criterion seems to work a little better, namely Conditions 1 and 2, where the number of states is the smallest. However, it is clear that in a real application this number is usually unknown. Therefore, it would be difficult to establish whether the values of this criterion are reliable or not.

A fine grained comparison, across conditions, among the three criteria is possible. In the first place we observe that, with the same number of items and states, the error rate seems not affecting the behavior of these criteria (this appears evident by a comparison of even and odd conditions). In Conditions 1 and 2 (i.e., with 10 items and 50 states) AIC and AICc strongly overestimate the number of knowledge states. In Conditions from 3 to 6 (10 items, 250 to 500 states) all criteria tend to underestimate the true number of knowledge states. In Conditions 7 and 8 (20 items, 500 states) all criteria strongly overestimate the number of states.

Application to a Real Data Set

The MaxRes procedure was applied to a real data set. The data consisted of binary response patterns of 536 students who responded to 20 problems on fraction subtraction (Tatsuoka, 1990). As observed in the simulation study, the MaxRes procedure might produce unreliable results when applied to more than 10 items. For this reason, 10 of the 20 fraction subtraction problems were selected for the analysis (the item identification numbers, in the original Tatsuoka's data set, were: 4, 6, 7, 11, 12, 14, 15, 17, 19, 20). The procedure that led to the selection of these items was as follows. The MaxRes method was repeatedly applied to different subsets of 10 items over the 20. The subset was selected for which the number of states obtained by the MaxRes procedure was closer to 50, that is the number used in Conditions 1 and 2 of simulation Study 2. In fact these two conditions are those in which the BIC criterion selects the model with a number of states close enough to the true number of states. The exact number of knowledge states obtained by the application of the MaxRes procedure to the selected items was 58.

Equation (3) was applied to obtain the two distances $d(\mathcal{K}_T, \mathcal{K}_M)$ and $d(\mathcal{K}_M, \mathcal{K}_T)$ between the knowledge structure $\mathcal{K}_M$, obtained by an application of the MaxRes procedure, and the theoretical knowledge structure $\mathcal{K}_T$ corresponding to the skill map proposed by Mislevy (1996). This skill map is displayed in Table 2. The aim was to establish which of the two structures better predicts the other one. The results were: $d(\mathcal{K}_T, \mathcal{K}_M) = 1.44$ and $d(\mathcal{K}_M, \mathcal{K}_T) = .68$. Recalling that $d(\mathcal{K}, \mathcal{K}')$ equals zero when $\mathcal{K}' \subseteq \mathcal{K}$, we observe that this distance is closer to zero when $\mathcal{K}' = \mathcal{K}_T$. This result says that, by approximating a state in the empirical structure with a state in the theoretical one, on average the error would be 1.44 items. In the opposite case the error would be .68 items.

Noting that the size of the theoretical knowledge structure was $|\mathcal{K}_T| = 19$ (thus much smaller than the empirical one), the following additional statistics were computed:

- number of common states $|\mathcal{K}_T \cap \mathcal{K}_M| = 8$;
- number of states in $\mathcal{K}_T$, not contained in $\mathcal{K}_M$: $|\mathcal{K}_T \backslash \mathcal{K}_M| = 11$;
- number of states in $\mathcal{K}_M$, not contained in $\mathcal{K}_T$: $|\mathcal{K}_M \backslash \mathcal{K}_T| = 50$.

TABLE 2
Skill map used in Mislevy (1996) for the selected items

			Method A				Method B			
Item #	Text	1	2	5	6	7	2	3	4	5
4	$3\frac{1}{2}-2\frac{3}{2}$	✓			✓		✓	✓	✓	
6	$\frac{6}{7}-\frac{4}{7}$	✓								
7	$3-2\frac{1}{5}$	✓		✓	✓		✓	✓	✓	✓
11	$4\frac{1}{3}-2\frac{4}{3}$	✓	✓		✓		✓	✓	✓	
12	$\frac{11}{8}-\frac{1}{8}$	✓	✓				✓			
14	$3\frac{4}{5}-3\frac{2}{5}$	✓			✓			✓		
15	$2-\frac{1}{3}$	✓	✓	✓				✓	✓	✓
17	$7\frac{3}{5}-\frac{4}{5}$	✓	✓		✓			✓	✓	
19	$7-1\frac{4}{3}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
20	$4\frac{1}{3}-1\frac{5}{3}$	✓	✓		✓	✓	✓	✓	✓	

Note. Skills: (1) basic fraction subtraction; (2) simplify/reduce; (3) separate whole number from fraction; (4) borrow one from whole number to fraction; (5) convert whole number to fraction; (6) convert mixed number to fraction; (7) column borrow in subtraction.

The BLIM model was fitted to the data for both the knowledge structure $\mathcal{K}_M$ obtained by the MaxRes procedure, and the theoretical knowledge structure $\mathcal{K}_T$. The aim was to compare the two structures with respect to data fit. Table 3 shows the Chi-square statistic, degrees of freedom, bootstrapped *p*-value, AIC, AICc, BIC for each of the two models. As can be seen, both of them exhibit a pretty good fit. However the model selection criteria are all in favor of the $\mathcal{K}_M$ model. Furthermore, as shown in Table 4, on the average the BLIM has smaller careless error and lucky guess parameters for the empirical structure $\mathcal{K}_M$, compared to the theoretical one.

FINAL REMARKS

In the present article a data-driven procedure was illustrated for extracting a knowledge structure from a large data set. Compared to other existing methods like the one by Schrepp (1999), ITA and IITA methods (Sargin & Ünlü, 2009; Schrepp, 2003), the proposed procedure is less restrictive. In particular, unlike the ITA and IITA, it does not require closure under union or intersection of the extracted structure. Unlike Schrepp's method, it allows careless error and lucky guess probabilities to vary across items, and it does not require a uniform distribution on the knowledge states.

TABLE 3
Model fit comparison in the empirical application. $\mathcal{K}_T$ denotes the theoretical knowledge structure corresponding to the Mislevy's (1996) skill map; $\mathcal{K}_M$ denotes the empirical structure obtained by an application of the MaxRes procedure

Structures	χ^2	*df*	Bootstrapped *p*-value	AIC	AICc	BIC
	1273.0	985	0.15	4566.6	4572.4	4729.2
	1121.1	946	1.00	4325.5	4351.7	4655.4

TABLE 4
Item error parameter estimates for the theoretical and the empirical knowledge structures

	$\mathcal{K}_M$		$\mathcal{K}_T$	
Item #	βq	ηq	βq	ηq
4	.09	.04	.02	.22
6	.02	.15	.05	.04
7	.09	.00	.17	.12
11	.04	.00	.08	.08
12	.00	.09	.05	.09
14	.01	.03	.03	.03
15	.06	.02	.06	.04
17	.07	.03	.12	.04
19	.00	.00	.23	.02
20	.00	.01	.16	.01
Mean	.04	.04	.10	.07

The proposed method is based on the BLIM model. Starting from the simplest knowledge structure $\mathcal{K}_0$ on a given set Q of items, it constructs a chain of knowledge structures $\mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_n$ by adding one state at the time. In each step i of the procedure, the new state is selected by applying the MaxRes updating rule. Among the observed response patterns, the MaxRes rule selects the one for which the Chi-square residual is largest. When the procedure terminates, standard model selection criteria can be applied to select the best knowledge structure.

Simulation Study 1 showed that compared to FMax (i.e., the updating rule of the Schrepp's procedure), the MaxRes updating performs better. In fact for this last rule, the distance of the extracted structure from the "true" structure is systematically smaller, irrespectively of the simulation conditions (number of items, number of states, maximum error rate). Moreover, MaxRes seems to be less sensitive to the size of the error rates.

Simulation Study 2 is a comparison among the most common model selection criteria, applied to our procedure. From it, the conclusion is that none of the three selection criteria behaves satisfactorily. In general, all of them tend to overestimate or underestimate the number of knowledge states. Notwithstanding, BIC criterion works a little better in the conditions with 10 items and a small number of states.

An empirical application showed viability of the proposed procedure, which was applied to a data set on fraction subtraction problems (Tatsuoka, 1990). A subset of 10 of the original 20 items was selected for this application and a knowledge structure containing 58 knowledge states was extracted by the MaxRes procedure, with the BIC criterion. This number of states is close enough to 50, which is the number of states used in those simulation conditions where the BIC works reliably. In other words the procedure was applied in a setting where, as suggested by the simulation studies, the BIC performs adequately. Moreover, for the extracted knowledge structure, the fit of the BLIM was pretty good, with small lucky guess and careless error parameter estimates.

An open question concerning the method presented in this article is about the termination criterion. In fact all considered model selection criteria were not fully adequate. Further research is needed to elaborate some ad hoc criteria, which can better manage the complexity level characterizing the knowledge structure approach.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*, 716-723. doi:10.1109/TAC.1974.1100705

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Cosyn, E., & Thiéry, N. (2000). A practical procedure to build a knowledge structure. *Journal of Mathematical Psychology*, *44*, 383-407. doi:10.1006/JMPS.1998.1252

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2013). Assessing parameter invariance in the blim: Bipartition models. *Psychometrika*, *78*, 710-724. doi:10.1007/SFF336-013-9325-5

Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 111-121). New York, NY: Springer-Verlag.

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge Spaces*. Berlin-Heidelberg, Germany, and New York, NY: Springer-Verlag.

Dowling, C. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology*, *37*, 21-48. doi:10.1006/JMPS.1993.1002

Düntsch, I., & Gediga, G. (1996). On query procedures to build knowledge structures. *Journal of Mathematical Psychology*, *4*, 160-168. doi:10.1006/JMPS.1996.0015

Falmagne, J.-C., & Doignon, J.-P. (1988). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, *41*, 1-23. doi:10.111/j.2044-8317.1988.tb00884.x

Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces: Interdisciplinary applied mathematics*. Berlin-Heidelberg, Germany: Springer-Verlag.

Goodman, L. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, *60*, 179-192. doi:10.1093/biomet/60.1.179

Haberman, S. J. (1979). *Analysis of qualitative data* (Vol. 2). New York, NY: Academic Press.

Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies and knowledge structures. In J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, & X. Hu (Eds.), *Knowledge spaces* (pp. 229-242). Berlin-Heidelberg, Germany: Springer-Verlag.

Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, *42*, 49-56. doi:10.1016/j.endm.2013.05.145

Kambouri, M., Koppen, M., Villano, M., & Falmagne, J.-C. (1994). Knowledge assessment: Tapping human expertise by the query routine. *International Journal of Human Computer Studies*, *40*, 119-151. doi:10.1006/i-jhc.1994.1006

Koppen, M., & Doignon, J.-P. (1990). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, *34*, 311-331. doi:10.1016/0022-2496(90)90035-8

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, applications* (pp. 103-132). Mahwah, NJ: Lawrence Erlbaum Associates.

Lukas, J., & Albert, D. (1993). Knowledge assessment based on skill assignment and psychological task analysis. In G. Strube & K. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 139-160). Amsterdam, Netherlands: North-Holland.

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, *33*, 379-416. doi: 10.1111/j.1745-3984.1996.tb00498.x

Robusto, E., Stefanutti, L., & Anselmi, P. (2010). The gain-loss model: A probabilistic skill multimap model for assessing learning processes. *Journal of Educational Measurement*, *47*, 373-394. doi: 10.1111/j.1745-3984.2010.00119.x

Sargin, A., & Ünlü, A. (2009). Inductive item tree analysis: Corrections, improvements, and comparisons. *Mathematical Social Sciences*, *58*, 376-392. doi:10.1016/j.mathsocsci.2009.06.001

Schrepp, M. (1999). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, *52*, 213-224. doi:10.1348/000711099159071

Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *MPR-online*, *8*, 43-79.

Schwarz, G. E. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*, 461-464.

Stefanutti, L., Anselmi, P., & Robusto, E. (2011). Assessing learning processes with the gain-loss model. *Behavior Research Methods*, *43*, 66-76. doi:10.3758/s.13428-010-036-x

Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behavior Research Methods*, *44*, 1197-1211. doi:10.3758/s13428-012-0187-z

Stefanutti, L., & Koppen, M. (2003). A procedure for the incremental construction of a knowledge space. *Journal of Mathematical Psychology*, *47*, 265-277. doi:10.1016/s0022-2496(02)00022-6

Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83-96. doi:10.1007/s11336-008-9095-7

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453-488). Hillsdale, MI: Lawrence Erlbaum Associates.