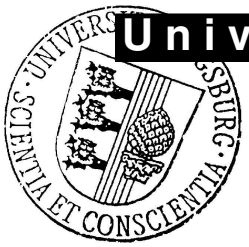

Universität Augsburg

Institut für Mathematik

Anatol Sargin, Ali Ünlü

Inductive Item Tree Analysis: Corrections, Improvements, and Comparisons

Preprint Nr. 24/2008 — 05. Juni 2008

Institut für Mathematik, Universitätsstraße, D-86 135 Augsburg http://www.math.uni-augsburg.de/

Impressum:

Herausgeber:

Institut für Mathematik
Universität Augsburg
86135 Augsburg
http://www.math.uni-augsburg.de/forschung/preprint/

ViSdP:

Anatol Sargin
Institut für Mathematik
Universität Augsburg
86135 Augsburg

Inductive item tree analysis: Corrections, improvements, and comparisons

Anatol Sargin, Ali Ünlü

University of Augsburg, D-86135 Augsburg, Germany

Abstract

There are various methods in knowledge space theory for building knowledge structures or surmise relations from data. Few of them have been thoroughly analyzed, making difficult to decide which of these methods provide good results and when to apply each of the methods.

In this paper, we investigate the method inductive item tree analysis and discuss the advantages and disadvantages of this algorithm. In particular, we introduce some corrections and improvements to it, resulting in two newly proposed algorithms. These algorithms and the original inductive item tree analysis procedure are compared in a simulation study and with empirical data.

Key words: Inductive item tree analysis, knowledge space theory, deriving knowledge structures

1 Introduction

In this paper, we analyze the inductive item tree analysis (IITA) method for building a knowledge structure from data (Schrepp (1999, 2002, 2003, 2007)). A knowledge structure belongs to knowledge space theory (KST) (Albert & Lukas (1999); Doignon & Falmagne (1999)): Assume a set Q of dichotomous items. Mastering an item $j \in Q$ may imply mastering another item $i \in Q$. If no response errors are made, these implications, $j \rightarrow i$, entail that only certain response patterns are possible. Those response patterns are called knowledge states, and the set of all knowledge states is called a knowledge structure.

Email addresses: anatol.sargin@math.uni-augsburg.de/ (Anatol Sargin), ali.uenlue@math.uni-augsburg.de (Ali Ünlü).

URLs: http://stats.math.uni-augsburg.de/mitarbeiter/sargin/ (Anatol Sargin), http://www.math.uni-augsburg.de/∼uenlueal/ (Ali Ünlü).

Implications are assumed to form a quasi order, that is, a reflexive, transitive binary relation, on the item set Q. Quasi orders are referred to as surmise relations in KST, and bijectively correspond to specific knowledge structures (Doignon & Falmagne (1999)).
Applications are, for example, a questionnaire, where people can agree or disagree to a statement, or an aptitude test, where people can solve or fail to solve a question. We use the latter interpretation to illustrate the algorithms.
Implications are latent and not directly observable, due to random response errors. A person who is actually unable to solve an item, but does so, makes a lucky guess. On the other hand, a person makes a careless error, if he fails to solve an item which he is capable of mastering. A probabilistic extension of the knowledge structure model covering random response errors is the basic local independence model in KST (Doignon & Falmagne (1999)).
Random errors in the responses of an examinee are the reason why deriving a knowledge structure from data is difficult. Several data-analytic methods have been proposed, but none of these procedures has proved to give optimal results nor is, at least, better than all the other methods.
In this paper, we review the IITA procedure, show the advantages and disadvantages of the method, and give modifications to correct and improve it. The results gained in our investigations are illustrated using simulated and empirical data.

2 Algorithm of inductive item tree analysis

IITA is a data-analytic method for deriving a surmise relation on an item set. It is similar to item tree analysis, which is another data-analytic method developed by van Leeuwe (1974). In both algorithms, binary relations are generated and a fit measure for every relation is computed in order to find the one that fits the data best.

2.1 Original algorithm

One of the main parts of IITA is the inductive generation of surmise relations (giving the algorithm its name). We introduce some definitions before explaining the algorithm:
For two items i, j, the value $b_{ij} := |\{r \in R | r(i) = 0 \wedge r(j) = 1\}|$ is the number of counterexamples, that is, the number of observed response patterns in the data R contradicting $j \rightarrow i$. Based on these values, binary relations $\sqsubseteq_L$ for $L = 0, ..., m$ are defined, where m is the number of subjects in the dataset:

$$i \sqsubseteq_L j :\Leftrightarrow b_{ij} \leq L.$$

The relation $\sqsubseteq_0$ is transitive, and based on that, all the other transitive relations $\sqsubseteq_L$ are constructed inductively.
Assume $\sqsubseteq_L$ is a transitive relation. Define the set $S_{L+1} := \{(i,j)|b_{ij} \leq L+1 \wedge i \not\sqsubseteq_L j\}$. This set consists of all item pairs that are not already contained in the relation $\sqsubseteq_L$ and have at most $L+1$ counterexamples. From these item pairs those are excluded that cause an intransitivity in $\sqsubseteq_L \cup S_{L+1}$, and the remaining item pairs are referred to as $\tilde{S}_{L+1}$. This process continues iteratively until no intransitivity is caused anymore. The generated relation $\sqsubseteq_{L+1} := \sqsubseteq_L \cup \tilde{S}_{L+1}$ is then transitive by construction. Because $\sqsubseteq_0$ is reflexive, all generated relations are. Hence $\sqsubseteq_L$ is a quasi order for every $L = 0, ..., m$.
Besides the construction of the quasi orders, it is very important to find that quasi order which fits the data best. In IITA, the idea is to estimate the number of counterexamples for each quasi order, and to find the minimum value for the discrepancy between the observed and expected numbers of counterexamples over all competing quasi orders.
Let $p_i := \frac{|\{r \in R | r(i)=1\}|}{m}$ be the relative solution frequency of an item i. A violation of an underlying implication is only possible due to random errors. To compute the expected number of counterexamples, b^*_{ij}, error probabilities are needed. In this algorithm, the error probabilities are assumed to be equal for all items. This single error rate is estimated by

$$\gamma_L := \frac{\sum\{b_{ij}/(p_j m) | i \sqsubseteq_L j \wedge i \neq j\}}{(|\sqsubseteq_L| - n)},$$

where n is the number of items, and $|\sqsubseteq_L| - n$ is the number of non-reflexive item pairs in $\sqsubseteq_L$.
In the next step, under every relation, the algorithm computes the expected number of counterexamples for each item pair. If the relation $\sqsubseteq_L$ provides an implication $j \rightarrow i$, the expected number of counterexamples is computed by $b^*_{ij} = \gamma_L p_j m$. If $(i,j) \notin \sqsubseteq_L$, no dependency between the two items is assumed, and $b^*_{ij} = (1-p_i)p_j m(1-\gamma_L)$. In this formula, $(1-p_i)p_j m$ is the usual probability between two independent items, and the factor $(1-\gamma_L)$ states that no random error occurred. As we discuss later in detail, the main criticism on the algorithm is on the estimates b^*_{ij}.
A measure for the fit of each relation $\sqsubseteq_L$ to the data R is the *diff*-coefficient. It is defined as

$$\mathit{diff}(\sqsubseteq_L, R) := \sum_{i \neq j} \frac{(b_{ij} - b^*_{ij})^2}{n(n-1)}.$$

It gives the sum of the quadratic differences between the observed and expected numbers of counterexamples under the relation $\sqsubseteq_L$. The smaller the *diff*-value the better is the fit of the relation to the data. Therefore, the IITA algorithm looks for the smallest value of the *diff*-coefficient and returns the

corresponding quasi order.

2.2 Problems of the original algorithm

The inductive construction of the quasi orders is stated as one of the main advantages of this algorithm (Schrepp (1999, 2003)). However, the inductive construction can be criticized as folllows. It is possible that two implications would cause together an intransitivity, but not if added separately. Consider on a set of four items $\{a, b, c, d\}$ the implications $a \to b, a \to c, a \to d$ and $c \to d$. Assume that the implications $b \to d$ and $d \to c$ are possible candidates to be added in the next step. Together these implications lead to an intransitivity, and the procedure excludes both implications until $b \to c$ is added. Each of the two implications, $b \to d$ and $d \to c$, could be added separately without violating transitivity. But the procedure does not incorporate this. Moreover, the underlying (correct) quasi order is not necessarily contained in the selection set of all quasi orders. In the simulation study reported in this paper, the underlying quasi order is contained in the selection set 57% of the trials. In the other 43% it is impossible to reveal the underlying quasi order.
The major problem lies in the computation of the *diff*-coefficient. It uses estimates b^*_{ij} of the expected numbers of counterexamples. Two problems arise in the calculation of these estimates. For $(i, j) \notin \sqsubseteq_L$, the estimate is $b^*_{ij} = (1 - p_i)p_j m(1 - \gamma_L)$. But the algorithm does not take two different cases into account, namely $(j, i) \notin \sqsubseteq_L$ and $(j, i) \in \sqsubseteq_L$. In the first case, independence holds, and the correct estimator is $b^*_{ij} = (1 - p_i)p_j m$. This estimator is used in the first version of IITA (Schrepp (1999, 2002)), but is changed in Schrepp (2003). This is the common approach in statistics when independence is present, for instance in the analysis of two-way contingency tables.
In the second case independence can not be assumed, as $j \sqsubseteq_L i$. In Schrepp (2003), this problem is briefly mentioned, but not further pursued or even solved. This, in particular, explains why the original IITA version gives bad results when longer chains of items are present in the underlying quasi order (Schrepp (1999)). The correct estimator for b^*_{ij} is $(p_j - p_i + p_i\gamma_L)m$, instead of $(1 - p_i)p_j m(1 - \gamma_L)$.

3 Corrections and improvements to the algorithm

3.1 Corrected estimation

In this section, we introduce the correct estimators b^*_{ij} for the expected numbers of counterexamples. These are very important for computing the *diff*-

coefficient, which is the fit measure for finding the best quasi order. The correct choice for b^*_{ij} for $(i,j) \notin \sqsubseteq_L$ depends on whether $(j,i) \notin \sqsubseteq_L$ or $(j,i) \in \sqsubseteq_L$.

- If $(i,j) \notin \sqsubseteq_L$ and $(j,i) \notin \sqsubseteq_L$, set $b^*_{ij} = (1-p_i)p_j m$. As stated in Section 2.2, independence holds, and the additional factor $(1-\gamma_L)$ is omitted.

- If $(i,j) \notin \sqsubseteq_L$ and $(j,i) \in \sqsubseteq_L$, set $b^*_{ij} = (p_j - p_i + p_i\gamma_L)m$. This estimator is derived as follows. The observed number of people who solve item i is $p_i m$. Hence the estimated number of people who solve item i and item j is $p_i m - b^*_{ji} = (p_i - p_i\gamma_L)m$. Eventually this gives the estimate $b^*_{ij} = p_j m - (p_i - p_i\gamma_L)m = (p_j - p_i + p_i\gamma_L)m$.
  This estimator not only is mathematically motivated, but is also interpretable. The first term, $p_j m$, gives the number of people solving item j. The second term, $(p_i - p_i\gamma_L)m$, stands for the number of people solving both items, because $p_i m$ is the number of people solving item i, and $p_i\gamma_L m$ represents the number of people solving item i and failing to solve item j.

3.2 Minimizing the fit measure

Next, we discuss minimizing the *diff*-coefficient as a function of the error probability γ_L, for every quasi order $\sqsubseteq_L$. This minimizes the discrepancies between the observed and expected numbers of counterexamples.
The *diff*-coefficient can be decomposed as

$$\begin{aligned} \textit{diff} &= \frac{\sum_{i \neq j}(b_{ij} - b^*_{ij})^2}{n(n-1)} \\ &= \frac{\sum_{i \not\sqsubseteq_L j} b_{ij}^2 - 2b_{ij}(p_j - p_i + p_i\gamma_L)m + (p_j - p_i + p_i\gamma_L)^2 m^2}{n(n-1)} + \frac{\sum_{i \sqsubseteq_L j} b_{ij}^2 - 2b_{ij}p_j\gamma_L m + (p_j\gamma_L)^2 m^2}{n(n-1)}. \end{aligned}$$

Setting equal to zero the derivative of the *diff*-coefficient with respect to γ_L gives

$$\frac{\sum_{i \not\sqsubseteq_L j} -2b_{ij}p_i m + 2p_i p_j m^2 - 2p_i^2 m^2 + 2p_i^2 m^2\gamma_L}{n(n-1)} + \frac{\sum_{i \sqsubseteq_L j} -2b_{ij}p_j m + 2p_j^2 m^2\gamma_L}{n(n-1)} = 0.$$

This is equivalent to

$$\underbrace{\sum_{i \not\sqsubseteq_L j} -2b_{ij}p_i m + 2p_i p_j m^2 - 2p_i^2 m^2}_{=:x_1} + \underbrace{\sum_{i \not\sqsubseteq_L j} 2p_i^2 m^2}_{=:x_3}\gamma_L + \underbrace{\sum_{i \sqsubseteq_L j} -2b_{ij}p_j m}_{=:x_2} + \underbrace{\sum_{i \sqsubseteq_L j} 2p_j^2 m^2}_{=:x_4}\gamma_L = 0.$$

Solving for γ_L results in $\gamma_L = \frac{-(x_1+x_2)}{x_3+x_4}$. Note that this expression always gives a value in $[0, 1]$. This error probability can now be used for an alternative IITA procedure, in which a minimized *diff*-value is computed for every quasi order.

4 Comparisons of the three algorithms

The three algorithms are the original IITA version by Schrepp (2003), and our corrected and minimized corrected IITA versions introduced above. In the following, the performances of these procedures are compared in a simulation study. Simulations were realized using the R statistical computing environment (R Development Core Team (2006); http://www.r-project.org/). The source files are freely available from the authors.

4.1 Settings of the simulation study

Throughout the simulation study nine items are used. The general simulation scheme consists of three parts. First, quasi orders are generated randomly. Second, each of these quasi orders is used for simulating the data. Third, the three algorithms are applied to and compared on that data. More precisely:

(1) All reflexive pairs are always added to the underlying relation R. A constant δ is set, which gives the probability for adding each of the remaining 72 item pairs to the relation. The transitive closure $\sqsubseteq$ of this relation R is computed.
(2) From the set $S = \{s : I \to \{0, 1\} | (i \sqsubseteq j \wedge s(j) = 1) \to s(i) = 1\}$ of all patterns consistent with $\sqsubseteq$ an element is drawn randomly. For this drawn pattern all entries are changed from 1 to 0 or from 0 to 1, with a same prespecified error probability τ.
(3) The three algorithms are applied to the simulated data. They are compared with respect to two criteria: the symmetric differences between the data-analytic solutions of the algorithms and the underlying quasi order, and the number of erroneously detected implications.

This simulation scheme replicates the one described in Schrepp (2003). However, the following important changes are made. Schrepp (2003) draws δ randomly from the entire unit interval. This leads to the following problem. For δ-values greater than (approximately) 0.42, the average number of non-reflexive implications contained in the underlying quasi order already turns out to be (approximately) 70. This can be seen from Figure 1.

[Insert Figure 1 about here]

Figure 1 shows the average number of non-reflexive implications as a function of δ. For δ ranging from 0 to 1, in steps by 0.01, 100 quasi orders are generated, and the corresponding average numbers of non-reflexive implications are calculated. In particular, Figure 1 shows that Schrepp's choice of δ-values mostly results in generating large quasi orders. This definitely does not give a representative sample of the collection of all quasi orders.
To avoid this problem, we pursue the following sampling. The δ-values are drawn from a normal distribution with $\mu = 0.16$ and $\sigma = 0.06$. Values less than 0 or greater than 0.3 are set to 0 or 0.3, respectively. This assures a uniform distribution of average numbers of non-reflexive implications. This can be seen from Figure 2. It shows the average number of non-reflexive implications calculated for 100 generated quasi orders to 500 δ-values drawn according to our sampling.

[Insert Figure 2 about here]

The following settings are made in the simulation study. The error probability τ takes the values $0.03, 0.05, 0.08, 0.10, 0.15$ and 0.20. The sample sizes $50, 100, 200, 400, 800, 1600$ and 6400 are used. For each combination of these settings, 1000 simulations are made. In each of these simulations, an underlying quasi order is generated, a data matrix is simulated, and for each of the three algorithms the data-analytic solution is derived.

4.2 Results of the simulation study

For each of the three algorithms, for every combination of error probability and sample size, two summary statistics are computed. One is the mean of the numbers of elements in the 1000 symmetric differences between the underlying quasi order and the data-analytic solutions; in the sequel referred to as *dist*-value. The other is the mean of the 1000 *diff*-values obtained for the data-analytic solutions. These summary statistics are reported in Table 1; first, second, and third lines refer to the original, corrected, and minimized corrected IITA algorithms, respectively.

[Insert Table 1 about here]

Table 1 shows the following results:

(1) For all combinations of settings, the same ranking for the *diff*-coefficient is obtained. The minimized corrected version gives the smallest average *diff*-value, second comes the corrected version, and the original algorithm has the largest *diff*-value. Hence, the match between the observed and

expected numbers of counterexamples can be ranked accordingly. It is seen from Table 1 that smaller *diff*-values do not necessarily imply smaller *dist*-values.

(2) The average *dist*-values are quite similar (maximum discrepancy of 0.76) for the corrected and minimized corrected algorithms. Moreover, in 24 of the 42 combinations the corrected algorithm performs better, in three they perform identically, and in 15 the performance of the minimized corrected version is better. In particular, the minimized corrected version gives smaller *dist*-values for an error probability of 0.20. On average, however, the corrected algorithm shows a smaller *dist*-value.

(3) For the very small error rates 0.03 and 0.05, the original version gives better results than the corrected and minimized corrected algorithms. It may seem surprising that, though of the incorrect estimators used in the original IITA algorithm, this algorithm gives better results. We suppose that the incorrect estimation, in the case of very small error rates, has a less negative effect for the underlying quasi order than for the other relations. This is an interesting open problem which needs to be investigated in further research. However, for $\tau = 0.08$, the results are approximately the same, and for the higher error rates $0.10, 0.15$ and 0.20, the original version is clearly outperformed. The differences in these cases are substantially larger than the ones obtained when the original version performs better. On average, the corrected and minimized corrected algorithms show smaller *dist*-values than the original algorithm.

(4) With increasing sample size, the improvements obtained for the two new algorithms are greater than the improvements for the original algorithm. For $\tau = 0.10$, for instance, the original algorithm improves from a *dist*-value of 9.01 to 6.87 (difference of 2.14), the corrected algorithm from a value of 9.61 to 4.25 (difference of 5.36), and the minimized corrected version from 9.60 to 4.58 (difference of 5.02).

(5) An interesting observation is as follows. For our two algorithms, for any two error probabilities, the differences between the *dist*-values decrease as the sample size increases. For the original algorithm, these differences range around a constant. For instance, take the error probabilities 0.05 and 0.15. The sequence of differences for the original algorithm is 12.44, 10.91, 11.35, 11.29, 12.95, 12.73 and 13.82. The sequences for the other algorithms are 5.29, 4.43, 2.86, 2.02, 1.84, 1.56 and 1.47 (corrected version), and 5.03, 4.22, 2.69, 1.71, 1.15, 1.19 and 0.82 (minimized corrected version).

From a practical point of view, it may be important to have only few false implications being added to the correct underlying quasi order (cf. Schrepp (2003, 2007)). False implications can lead to incorrect conclusions, and it is inefficient to try to exclude them afterwards. In the following, we compare the three IITA algorithms with respect to the average numbers of erroneously detected implications. This summary statistic is reported in Table 2; first,

second, and third lines refer to the original, corrected, and minimized corrected IITA algorithms, respectively.

[Insert Table 2 about here]

Table 2 shows the following results:

(1) Except for $\tau = 0.10$ and sample sizes 50 and 100, the corrected and minimized corrected IITA algorithms yield smaller average numbers of falsely detected implications. For example, for the error rates 0.15 and 0.20, the original version is clearly outperformed. On average, the corrected and minimized corrected algorithms falsely detect 1.01 and 1.05 implications, respectively, while the original version adds 2.59 false implications.
(2) The results are quite similar (maximum discrepancy of 0.20) for the corrected and minimized corrected algorithms. Moreover, in 25 of the 42 combinations the corrected algorithm performs better, in six they perform identically, and in 11 the performance of the minimized corrected version is better. For smaller sample sizes, the corrected algorithm performs better than the minimized corrected one. For larger sample sizes, there seems to be no noticeable difference.
(3) The results for the corrected and minimized corrected versions improve for increasing sample sizes. The original version, however, jitters between smaller and larger values, with no decreasing trend observable for larger error probabilities. For $\tau = 0.10$, for instance, the sequences of decreasing values for the corrected and minimized corrected versions are 2.30, 1.49, 0.92, 0.73, 0.55, 0.50 and 0.46, and 2.50, 1.56, 0.99, 0.73, 0.55, 0.49 and 0.43, respectively. The sequence for the original version is 1.95, 1.40, 1.73, 0.95, 1.45, 1.22 and 1.81.

An important remark is in order regarding the simulation study in Schrepp (2003). The results reported in this simulation study are much better than the results we have obtained for the original IITA algorithm. There are substantial discrepancies between the average *dist*-values and average numbers of falsely detected implications. For instance, for $\tau = 0.08$ and sample size 200, Schrepp's study gives 1.67 and 0.09, respectively, while our simulation study yields 5.12 and 1.57.
This can be explained by the following flaw in Schrepp's simulation methodology. As mentioned in Section 4.1, his choice of δ-values leads to the problem that mostly large quasi orders are generated. The incorrect estimation scheme of the original IITA algorithm now produces good results for large quasi orders. For a large quasi order $\sqsubseteq$, there are predominantly the cases $i \sqsubseteq j$ and $j \sqsubseteq i$, for which correct estimators are used. For $i \not\sqsubseteq j$ and $j \sqsubseteq i$, or $i \not\sqsubseteq j$ and $j \not\sqsubseteq i$ incorrect estimators are applied, and hence the discrepancies between the observed and expected numbers of counterexamples are large. This implies

that, for an underlying large quasi order, the *diff*-values for small quasi orders of the selection set are large. As a result, the underlying quasi order is more frequently recovered. Moreover, it is not astonishing that in Schrepp (2003) smaller average numbers of falsely detected implications are obtained. For quasi orders containing an average number of 70 non-reflexive implications, there are, on average, two implications left to be added erroneously.

5 Applications to empirical data

In this section, we apply the three IITA algorithms to two empirical datasets. One is the Aphasic dataset, which is also used in Schrepp (2003), and the other is from the Programme for International Student Assessment (PISA; http://www.pisa.oecd.org/).

5.1 IITA analyses of the Aphasic dataset

The Aphasic dataset (Gloning, Lienert & Quatamber (1972)) consists of 162 aphasic patients tested on five tasks. This dataset is used in Schrepp (2003) for comparing the original IITA algorithm to feature pattern analysis and configural frequency analysis. For details on the dataset, the latter two methods, and the obtained results, see Schrepp (2003).
Analyses of the Aphasic dataset using the corrected and minimized corrected IITA algorithms (not reported in detail in this paper) give the same quasi order as obtained for the original algorithm. Interestingly, though the same quasi order is obtained for the three algorithms, the computed *diff*-values are considerably smaller for the corrected (61.54) and minimized corrected (60.93) than for the original (165.98) versions.

5.2 IITA analyses of the PISA dataset

We analyze part of the 2003 PISA data consisting of 340 German students answering eight questions on mathematical literacy (available from the authors). These items are chosen to form a Rasch scale. That is, the dichotomous one-parameter logistic model (Fischer & Molenaar (1995)) fits (goodness-of-fit and item fit) the data very well. Under this model, the following item difficulties are estimated for the eight questions: -2.09, -1.58, -1.23, -0.04, 0.28, 0.66, 1.46 and 2.20. Hence, the items most likely form a chain, which is considered as the underlying quasi order (see Figure 3) in the subsequent analyses.

[Insert Figure 3 about here]

Analyzing the PISA dataset using the original IITA algorithm and the corrected and minimized corrected IITA algorithms gives the quasi orders shown in Figures 4 and 5, respectively.

[Insert Figure 4 about here]

[Insert Figure 5 about here]

The original IITA algorithm yields a *dist*-value of 19, in contrast to the corrected and minimized corrected versions, which give a clearly smaller *dist*-value of 5. Since under all three algorithms no false implications are added, these are the numbers of true implications missed by the algorithms. The corrected and minimized corrected versions outperform the original algorithm.
This can be explained as follows. The underlying quasi order $\sqsubseteq$, which is a chain, consists only of cases $i \sqsubseteq j$ and $j \not\sqsubseteq i$. As mentioned in Section 2.2, for these cases incorrect estimators are used in the original version. This leads to larger discrepancies between the observed and expected numbers of counterexamples, hence to a larger *diff*-value. The corrected and minimized corrected IITA algorithms, however, use correct estimators and therefore detect true implications more properly.

6 Conclusion and further research

The original IITA algorithm is one of the few data-analytic methods in KST. We have made some corrections and improvements to this algorithm. In particular, we have introduced two new versions, the corrected and minimized corrected IITA algorithm. These three algorithms have been compared in a simulation study, and on two empirical datasets. On average, the corrected and minimized corrected versions have performed better than the original one, in terms of both smaller *dist*-values and numbers of erroneously detected implications.
The current simulation study is a starting point for more in-depth analyses of the IITA algorithms, especially of the corrected and minimized corrected versions. Future research may address the effects of deviations from the uniform probability distribution on the set S of all consistent response patterns, and from a single error probability τ (cf. Section 4.1). This would allow investigating whether the present algorithms should be generalized to include different lucky guess and careless error probabilities for every item, and to be applicable to skew population distributions.
Another interesting direction for further research is to modify the *diff*-coefficient. As apparent from the presented simulation study, smaller *diff*-values do not

necessarily imply better reconstructions of underlying quasi orders. It seems that an aggregation (*diff*-coefficient) of local, two-dimensional views of the data (b_{ij}) does not provide acceptable results on the relationships among all items mutually in $|Q|$ dimensions. One could consider developing fit measures incorporating higher-dimensional views of the data.
Work on the generated selection set should definitely be pursued in future research. So far, for the IITA algorithms the quality of the inductively generated set of quasi orders has not been systematically investigated. In our simulation study, on average, the underlying quasi order is contained only 569 (out of 1000) times in the selection set. Since it is computationally intractable to evaluate all possible quasi orders in large-scale applications, better search methods are needed to improve the selection set. A data-analytic method operating on a set of candidate models is only as good as the quality of the selection set is.

References

Albert, D., Lukas, J. (Eds.), 1999. Knowledge Spaces: Theories, Empirical Research, and Applications. Lawrence Erlbaum Associates, Mahwah.

Doignon, J.-P., Falmagne J.-C., 1999. Knowledge Spaces. Springer-Verlag, Berlin.

Fischer, G., Molenaar I. (Eds.), 1995. Rasch Models: Foundations, Recent Developments, and Applications. Springer-Verlag, New York.

Gloning, J., Lienert, G.A., Quatamber, R., 1972. Konfigurationsfrequenzanalyse aphasiespezifischer Testleistungen [Configural frequency analysis of aphasia specific test performances]. Zeitschrift für Klinische Psychologie und Psychotherapie 20, 115-122.

R Development Core Team, 2006. R: A language and environment for statistical computing (ISBN 3-900051-07-0). Vienna, Austria: R Foundation for Statistical Computing.

Schrepp, M., 1999. On the empirical construction of implications between bivalued test items. Mathematical Social Sciences 38, 361-375.

Schrepp, M., 2002. Explorative analysis of empirical data by boolean analysis of questionnaires. Zeitschrift für Psychologie 210, 99-109.

Schrepp, M., 2003. A method for the analysis of hierarchical dependencies between items of a questionnaire. Methods of Psychological Research 19, 43-79.

Schrepp, M., 2007. On the evaluation of fit measures for quasi-orders. Mathematical Social Sciences 53, 196-208.

van Leeuwe, J.F.J., 1974. Item tree analysis. Nederlands Tijdschrift voor de Psychologie 29, 475-484.

7 Tables

Table 1
Average *dist*- and *diff*-values (in parentheses) under original, corrected, and minimized corrected IITA algorithms (first, second, and third lines, respectively)

	Sample size						
	50	100	200	400	800	1600	6400
τ							
0.03	3.71(3.79)	2.89(11.67)	2.14(41.16)	2.29(152.10)	1.87(599.61)	1.69(2438.30)	1.99(36528.70)
	5.44(1.66)	5.18(4.28)	4.54(11.58)	4.86(38.37)	4.30(128.13)	3.90(489.21)	3.81(7414.34)
	5.44(1.58)	5.30(4.04)	4.67(10.76)	5.11(34.92)	4.77(114.65)	4.58(438.49)	4.51(6635.50)
0.05	4.24(3.73)	4.05(11.32)	2.87(38.28)	2.59(138.35)	2.11(517.96)	1.77(2094.56)	1.10(32968.50)
	6.89(1.93)	5.69(5.13)	4.91(14.35)	5.09(44.49)	4.22(152.74)	4.33(587.63)	3.63(8172.90)
	6.90(1.81)	5.67(4.68)	5.02(12.67)	5.40(39.24)	4.91(130.82)	4.89(489.52)	4.39(6892.42)
0.08	7.66(3.96)	5.95(12.18)	5.12(41.09)	4.91(149.30)	4.43(606.56)	4.47(2392.01)	3.69(37444.40)
	8.61(2.24)	6.36(6.05)	5.70(17.35)	5.28(52.26)	4.46(196.09)	4.5(702.46)	3.99(9870.19)
	8.45(2.09)	6.30(5.58)	5.90(15.38)	5.70(45.48)	4.88(163.70)	5.06(582.79)	4.55(8102.32)
0.10	9.01(4.23)	7.87(12.71)	7.89(45.51)	6.14(166.79)	6.67(682.10)	6.37(2808.66)	6.87(44472.91)
	9.61(2.43)	7.65(6.75)	6.37(18.70)	5.37(59.49)	5.26(203.67)	4.35(765.31)	4.25(11373.20)
	9.60(2.31)	7.47(6.21)	6.42(16.84)	5.66(52.69)	5.58(175.46)	4.85(644.15)	4.58(9491.62)
0.15	16.68(4.55)	14.96(14.81)	14.22(58.53)	13.88(221.76)	15.06(935.71)	14.50(3664.33)	14.92(62646.07)
	12.18(2.59)	10.11(7.45)	7.77(21.90)	7.11(67.43)	6.06(250.24)	5.89(877.53)	5.10(14659.14)
	11.93(2.48)	9.89(7.09)	7.71(20.49)	7.11(62.26)	6.06(226.58)	6.08(790.18)	5.21(13027.16)
0.20	23.38(4.53)	25.41(16.69)	24.93(62.96)	24.02(276.97)	23.72(1148.28)	24.65(4699.31)	23.46(76842.70)
	14.81(2.59)	11.40(7.34)	9.81(22.34)	8.00(71.94)	7.96(254.09)	6.79(930.46)	6.79(14769.23)
	14.68(2.52)	11.36(7.12)	9.62(21.53)	7.91(68.79)	7.93(240.95)	6.75(879.06)	6.58(13893.23)

Table 2
Average numbers of erroneously detected implications under original, corrected, and minimized corrected IITA algorithms (first, second, and third lines, respectively)

	Sample size						
	50	100	200	400	800	1600	6400
τ							
0.03	2.69	2.23	1.80	1.91	1.47	1.43	1.65
	1.82	0.92	0.48	0.38	0.23	0.17	0.18
	1.90	0.96	0.48	0.37	0.21	0.16	0.17
0.05	2.30	2.08	1.24	1.28	1.11	0.51	0.23
	2.10	1.14	0.64	0.45	0.30	0.20	0.13
	2.20	1.20	0.66	0.42	0.28	0.19	0.11
0.08	2.69	1.79	1.57	1.37	0.99	1.23	0.98
	2.26	1.47	0.92	0.58	0.42	0.40	0.35
	2.44	1.54	0.95	0.59	0.42	0.40	0.34
0.10	1.95	1.40	1.73	0.95	1.45	1.22	1.81
	2.30	1.49	0.92	0.73	0.55	0.50	0.46
	2.50	1.56	0.99	0.73	0.55	0.49	0.43
0.15	3.02	2.03	2.33	3.13	3.78	4.08	3.84
	2.57	1.72	1.28	1.08	0.97	0.82	0.82
	2.76	1.85	1.36	1.13	1.02	0.82	0.83
0.20	3.46	5.68	5.80	6.38	6.89	8.38	7.00
	2.71	2.08	1.52	1.27	1.09	1.11	0.99
	2.87	2.16	1.57	1.34	1.16	1.17	1.02

8 Figure captions

Fig. 1. Average number of non-reflexive implications as a function of δ. The δ-values range from 0 to 1, in steps by 0.01. For each δ-value, 100 quasi orders are generated, and the corresponding average number of non-reflexive implications is shown.

Fig. 2. Average number of non-reflexive implications calculated for 100 generated quasi orders to 500 δ-values drawn according to our sampling. Points are ordered by average number of non-reflexive implications.

Fig. 3. Rasch scale of the eight assessment items (from bottom to top, items sorted according to increasing difficulty). Assumed to underly the PISA dataset.

Fig. 4. Quasi order obtained for the PISA dataset under the original IITA algorithm.

Fig. 5. Quasi order obtained for the PISA dataset under the corrected and minimized corrected IITA algorithms.

9 Figures

Fig. 1

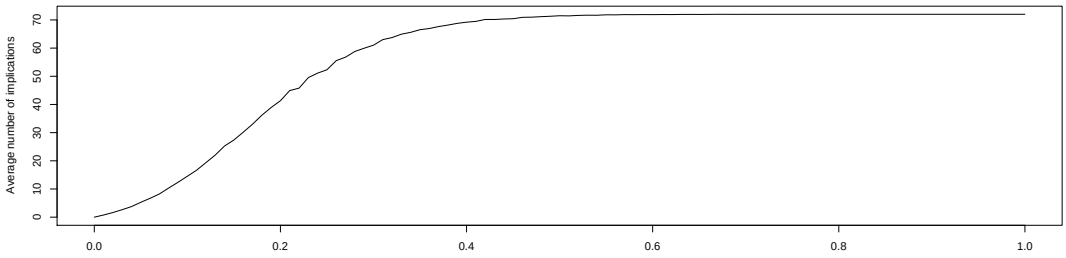

Fig. 2

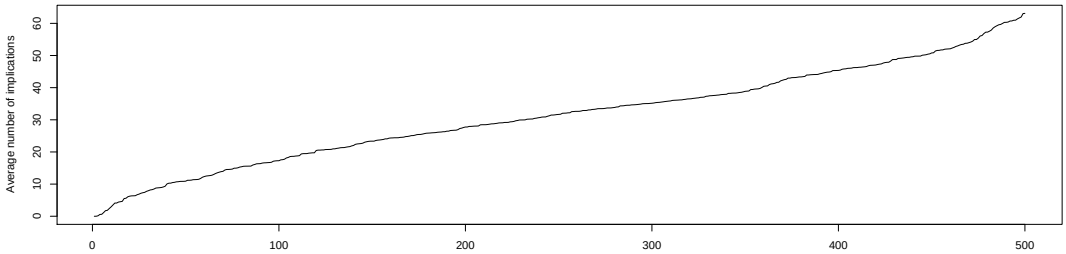
Average number of implications
0
10
20
30
40
50
60
0
100
200
300
400
500

Fig. 3

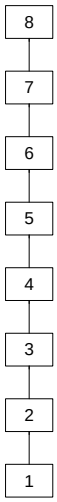

Fig. 4

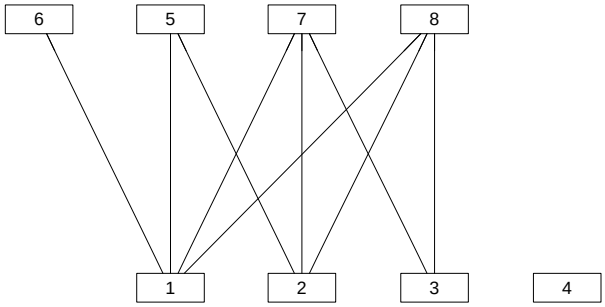

Fig. 5