Brock University

Department of Computer Science

Simplifying contextual structures

Ivo Düntsch and Günther Gediga*

Technical Report # CS-15-01
January 2015

Brock University
Department of Computer Science
St. Catharines
Ontario Canada L2S 3A1

Universität Münster, Institut IV
Department of Psychology
Münster
Germany

www.cosc.brocku.ca

Simplifying contextual structures

Ivo Düntsch⋆ ⋆⋆ and Günther Gediga⋆

[1] Brock University, St. Catharines, Ontario, Canada, L2S 3A1,
duentsch@brocku.ca
[2] Department of Psychology, Institut IV, Universität Münster, Fliednerstr. 21, Münster, Germany,
gediga@uni-muenster.de

Abstract. We present a method to reduce a formal context while retaining much its information content. Although simple, our ICRA approach offers an effective way to reduce the complexity of concept lattices and / or knowledge spaces by changing only little information in comparison to a competing model which uses fuzzy K-Means clustering.

1 Introduction

A very simple data structure is a triple $\mathfrak{C} = \langle U, V, R \rangle$ where R is a binary relation between elements of U and elements of V which we shall call a *formal context* [17]. From this, various data representations can be constructed, one of the more popular ones being the *concept lattice* obtained from $\mathfrak{C}$ introduced by Wille [17]. With each concept a line diagram can be associated which depicts the concept lattice in a consolidated way. For lack of space we shall not describe his further; For details we invite the reader to consult, for example, [18] or [6].

As a context $\mathfrak{C}$ grows large, the construction of the concept lattice is costly and it is difficult to interpret the structure and its associated line diagram. Therefore, various techniques have been proposed to reduce a formal context $\mathfrak{C} = \langle U, V, R \rangle$ or its associated concept lattice such as the stability indices of [13], the reduction using fuzzy K-Means clustering (FKM) [12], or reduction based on objects similarity [2]. Cheung & Vogel [1] propose a way to obtain a quotient – like concept lattice by identifying rows of a context and then considering the resulting concept lattice. However, this approach was shown to be flawed [11].

All these techniques can be subsumed under one of two strategies:

1. Omit attributes (or objects), or
2. Merge attributes (or objects) which are similar according to some criterion.

Both types change the adjacency matrix of R. However, reducing the matrix does not guarantee that the associated concept lattice will be reduced as well, see Example 3 of [11]. In this paper we propose a simple algorithm to reduce the concept which does not increase the size of the associated concept lattice.

⋆ The ordering of authors is alphabetical and equal authorship is implied.

⋆⋆ The author gratefully acknowledges support by the Natural Sciences and Engineering Research Council of Canada.

2 Notation and definitions

Throughout we suppose that $U = \{p_1, \dots, p_n\}$ is a finite set of objects (such as problems) and $V = \{s_1, \dots, s_k\}$ is a finite set of attributes (such as skills). $R \subseteq U \times V$ is a binary relation between elements of U and elements of$V$. For each $p \in U$ we set $R(u) \stackrel{\text{df}}{=} \{s \in V : pRs\}$, and $\mathscr{R} \stackrel{\text{df}}{=} \{R(u) : u \in U\}$. The identity relation on U is denoted by $1'_U$. The relational converse of R is denoted by $R^{\smile}$, and $-R$ is the complement of R in $U \times V$. The set $\mathscr{R}$ is partially ordered by $\subseteq$. The *adjacency matrix of* R has rows labeled by the elements of U, and columns labeled with the elements of V. An entry $\langle u, v\rangle$ is 1 if and only if $u_i R s_j$, otherwise, the entry in this cell is left empty.

A formal context $\langle U, V, R\rangle$ gives rise to several modal–style set operators:

$$\langle R\rangle(X) = \{b \in V : (\exists a \in X) aRb\} = \{b \in B : R^{\smile}(b) \cap X \neq \emptyset\}, \qquad \text{(Possibility)} \tag{2.1}$$

$$[R](X) = \{b \in V : (\forall a \in U)[aRb \Rightarrow a \in X]\} = \{b \in B : R^{\smile}(b) \subseteq X\}, \qquad \text{(Necessity)} \tag{2.2}$$

$$[[R]](X) = \{b \in V : (\forall a \in U))[a \in X \Rightarrow aRb]\} = \{b \in B : X \subseteq R^{\smile}(b)\} \qquad \text{(Sufficiency).} \tag{2.3}$$

It is well known that for all $X, X' \subseteq U$,

$$\langle R\rangle(X \cup X') = \langle R\rangle(X) \cup \langle R\rangle(X'), \tag{2.4}$$

$$[R](X \cap X') = [R](X) \cap [R](X'), \tag{2.5}$$

$$[[R]](X \cup X') = [[R]](X) \cap [[R]](X'). \tag{2.6}$$

The mappings $\langle R\rangle$ and $[[R]]$ are, respectively, the existential (disjunctive) and universal (conjunctive) extension of the assignment $x \mapsto R(x)$ to subsets of U, since it follows immediately from the definitions that for all $x \in U, X \subseteq U$,

$$\langle R\rangle(\{x\}) = [[R]](\{x\}) = R(x), \tag{2.7}$$

$$\langle R\rangle(X) = \bigcup_{x \in X} R(x), \tag{2.8}$$

$$[[R]](X) = \bigcap_{x \in X} R(x). \tag{2.9}$$

The operators $[[R]]$ and $[R]$, as well as $\langle R\rangle$, are related since, clearly,

$$[[R]](X) = [-R](U \setminus X), \tag{2.10}$$

$$\langle R\rangle(X) = V \setminus [[-R]](X). \tag{2.11}$$

For unexplained notation and concepts in lattice theory we refer the reader to [8].

3 Data models based on modal operators

Suppose we have a formal context $\mathfrak{C} = \langle U, V, R\rangle$ which we regard as "raw data". The image sets $R(x)$ are our basic constructs.

As a first approach to a data model (a structural representation of data) based on $\langle U,V,R\rangle$, we define a quasiorder $\preceq$ on U by setting $x \preceq y$ if and only if $R(x) \subseteq R(y)$. We also define the *incomparability relation* by

$$x\#y \stackrel{\text{df}}{\iff} (x \not\preceq y) \text{ and } (y \not\preceq x). \tag{3.1}$$

From this starting point, several more involved data models can be developed. One of the better known models are those based on the sufficiency operators $[[R]]$ ("intent") and $[[R^{\smile}]]$ ("extent"): For each $X \subseteq U$, $[[R]](X)$ is the set of all attributes, common to all elements of X, and for $Y \subseteq V$, $[[R^{\smile}]](Y)$ is the set of all objects which possess all attributes in Y. A pair $\langle [[R^{\smile}]][[R]](X), [[R]](X)\rangle$ is called a *formal concept*. The set of all formal concepts can be made into a lattice which can be drawn as a consolidated line diagram [17] as in Figure 1 [3]. Each node of the diagram represents a formal concept, and for each object x, $R(x)$ is the set of

R	v_1	v_2	v_3	v_4
u_1	1	1	1	
u_2		1	1	1
u_3		1		
u_4			1	
u_5		1	1	

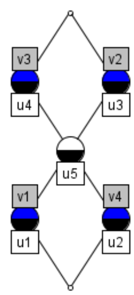


Fig. 1: A context and its line diagram

all attributes above the node labelled x (we interpret "above" and "below" as reflexive relations). In the line diagram of R, $x \preceq y$ if and only if x and y label the same node or the node labelled by y is below the node labelled by x.

In our problem/skill interpretation, $[[R]](X)$ is the set of all skills the possession of which is required by all problems in X. Such conjunctive problem assignment is an assumption e.g. of the "Deterministic Inputs, Noisy And" gate (DINA) model [14,9,10] and the rule space model [16].

A data model which in some sense complements concept lattices are the knowledge spaces introduced in [4]. These are set systems closed under union and can be related to the modal operator $\langle R\rangle$ which is called the *span operator* in [3]. It was shown in [7] that the models arising from $[[R]]$ and $\langle R\rangle$ have the same expressive power and are useful in situations different from those where conjunctive assignments are employed.

Taking $\{R(x) : x \in U\}$ as a starting point, the set of spans and the set of intent go into different directions: It follows from (2.7) and (2.9) that $\mathscr{K}_R \stackrel{\text{df}}{=} \{\langle R\rangle(X) : X \subseteq U\}$ is the $\cup$ – semilattice generated by $\{R(x) : x \in U\}$, and $\mathscr{I}_R \stackrel{\text{df}}{=} \{[[R]](X) : X \subseteq U\}$ is the $\cap$ – semilattice generated by $\{R(x) : x \in U\}$. For $X \subseteq U$, $[[R]]$ is the set

[3] The diagrams were drawn by the ConExp package [19]

of all attributes lying above all objects in X, and $\langle R\rangle(\{x\})$ is the set of all attributes not upwards reachable from object x in the line diagram of $-R$.

4 Reducing the complexity

The simplest way to change the adjacency matrix is to change one bit at a time, according to a given criterion. The question arises which criterion we shall use. If $\preceq$ is a linear quasi order – i.e. if any two objects of U are comparable – then $\mathscr{K}_R$ and $\mathscr{I}_R$ coincide and are equal to $\langle \mathscr{K}_R, \subseteq\rangle$ (possibly with added $\emptyset$ or V); nothing is gained by going from the simple model $\langle |C, \preceq\rangle$ to one of the more involved ones. At the other extreme, if no two different elements of U are comparable with respect to #, then the representations obtained from $\mathfrak{C}$ very strongly depend on the modal operator used and may widely differ. Consider the simple relation depicted in Figure 2. There, $\mathscr{I}_R$ consists of the singletons $\{v_i\}$ and the empty set, while $\mathscr{K}_R$ is the set of all nonempty subsets of V. If we consider the complement of $-R$, then situation is reversed, see Figure 3. Therefore, if

R	v_1	v_2	v_3	v_4
u_1	1			
u_2		1		
u_3			1	
u_4				1

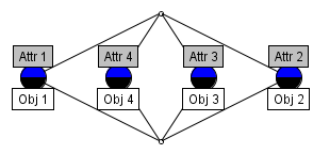


Fig. 2: $\# = U^2 \setminus 1'_U$, 1st example

-R	v_1	v_2	v_3	v_4
u_1		1	1	1
u_2	1		1	1
u_3	1	1		1
u_4	1	1	1	

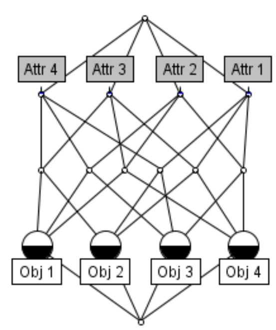


Fig. 3: $\# = U^2 \setminus 1'_U$, 2nd example

the incomparability relation is large, choosing one operator over the other may not provide a meaningful interpretation, and it may not be the wisest choice at the outset to prefer one over the other. Keeping in mind the problem/skill situation, we suggest the *relative incomparability* of objects as a measure of context complexity which we aim to reduce: If $\mathfrak{C} = \langle U, V, R \rangle$ is a formal context and $u \in U$, then we let

$$\texttt{incomp}(u) \stackrel{\text{df}}{=} \{v \in U : u\#v\}, \quad \texttt{incomp}(\mathfrak{C}) \stackrel{\text{df}}{=} \frac{|\{\langle u, v \rangle : u\#v\}|}{n^2 - n},$$

where $n = |U|$. Now, $\texttt{incomp}(\mathfrak{C}) = 0$ if and only if $\preceq$ is a linear quasiorder, and $\texttt{incomp}(\mathfrak{C}) = 1$ if no two different elements are $\preceq$ – comparable. The measure of success is the reduction of $\texttt{incomp}(\mathfrak{C})$ relative to the number of bit changes.

Our *InComparablity Reduction Analysis* algorithm (ICRA)[4] is based on a simple steepest descent method: We consider objects u for which $|\texttt{incomp}(u)|$ is maximal and then invert a bit – i.e. an entry in the adjacency matrix of the relation under consideration – for which the drop of the number of overall incomparable pairs is maximal. This will increase the comparability of objects with respect to $\preceq$ or, equivalently, of sets $R(x)$ without increasing the number of intents, respectively, knowledge states. Indeed, in most cases we have looked at, the complexity of the concept lattice was significantly reduced. If after inverting one bit so that the resulting relation is R' and $x \preceq_{R'} y$ then there will be a path from y to x in the line diagram of R' as well so that the new representation is closer to the data as represented by R.

The stop criterion is a predetermined relative value of incomparable pairs, i.e. a value for $\texttt{incomp}(\mathfrak{C})$, where $\mathfrak{C}$ is the current context, or no more reduction is possible. As a rule of thumb we suggest to require that 50% of pairs with different components should be comparable (*Median InComparablity Reduction Analysis*). An overview of the pseudocode the ICRA algorithm is shown in Figure 4.

```
noEntry := FALSE.
pout:= p                                                          ▷ Initialize stop criterion to 0 ≤ p ≤ 1.
Unmark all object-attribute-pairs.                                 ▷ No pair changed yet.
repeat
    Find the set OBJ of objects belonging to unmarked pairs for which incomp(u) is maximal.
    if incomp(C) ≤ pout then                                       ▷ Goal reached
        NoEntry := TRUE
    else
        Using OBJ find the object-attribute-pairs, which maximally reduce the incomparability,
        when inverting one bit of the matrix under consideration.
        if no reduction is achieved for any of these then
            NoEntry := TRUE
        else
            Invert the entry of one of the maximal object-attribute-pairs and use the new relation.
            Mark the chosen object-attribute-pair.
            Replace C with the revised context.
        end if
    end if
until NoEntry = TRUE.
```

Fig. 4: Pseudocode of the algorithm

[4] The algorithm is implemented in R [15] and the source code is available at http://roughsets.net/FCred.R.

5 Experiments

Even though our procedure is simple, it compares well with other reduction measures. As a case in point we shall consider the reduction using fuzzy K-Means clustering (FKM) proposed in [12]. This method is based on partitioning a set of vectors into k fuzzy clusters, specifying to what degree a vector belongs to the cluster centre. Owing to lack of space we cannot explain their method in detail and refer the reader to [12]. The context $\mathfrak{C}$ of their first example relates documents with keywords and it is shown in Figure 5 along with its context lattice. The relative incomparability of $\mathfrak{C}$ is 94%.

	D1	D2	D3	D4	D5	D6	D7
Baby		1		1	1		1
Child		1	1				
Guide						1	1
Health				1			
Home		1	1				
Infant	1			1			
Proofing					1	1	
Safety			1	1			
Toddler	1			1			

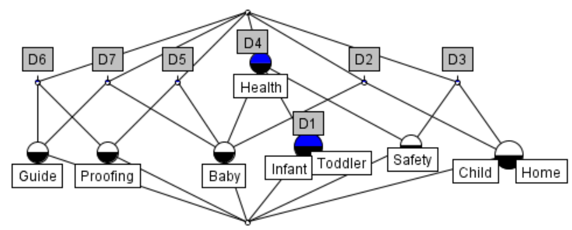


Fig. 5: Example from [12], p 2699

After applying FKM based clustering with $k = 2$, the columns D1 – D2 are identified and the entry $\langle T_i, D1‘ -$ $-D4\rangle$ of the resulting adjacency matrix is $\max\{\langle T_i, D1\rangle, \ldots, \langle T_i, D4\rangle\}$. The reduced context $\mathfrak{C}_1$ and its concept lattice are shown in Figure 6.

	D1–D4	D5	D6	D7
Baby	1	1		1
Child	1			
Guide			1	1
Health	1			
Home	1			
Infant	1			
Proofing		1	1	
Safety	1			
Toddler	1			

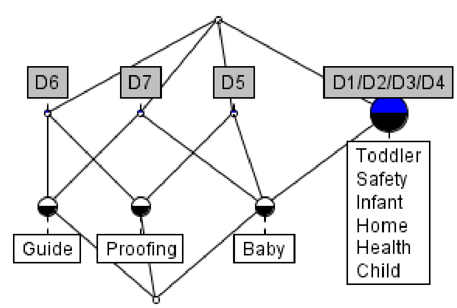
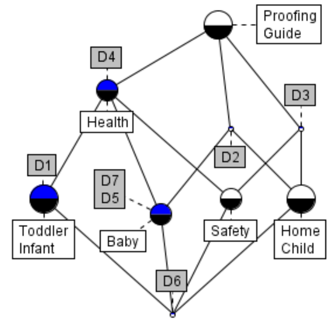


Fig. 6: Example from [12], p 2699, reduced

To achieve the FKM result $\mathfrak{C}_1$ from$\mathfrak{C}$ requires to change 15 bits for a relative incomparability of 49%; this includes the effort to identify columns. In comparison, our algorithm needs only 4 bits for a 50% incomparability, and 9 bits for 0% incomparability. The resulting context along with its line diagram is shown in Figure 7. It has the same number of concepts as the concept lattice obtained from FKM (9), and the same number of edges (14).

5 bits	D1	D2	D3	D4	D5	D6	D7
Baby		1		1	1		1
Child		1	1				
Guide							
Health				1			
Home		1	1				
Infant				1			
Proofing					1	1	
Safety			1	1			
Toddler	1			1			



9 bits	D1	D2	D3	D4	D5	D6	D7
Baby		1	1	1	1		1
Child		1	1	1			
Guide							
Health				1			
Home		1	1	1			
Infant				1			
Proofing					1		
Safety			1	1			
Toddler	1			1			

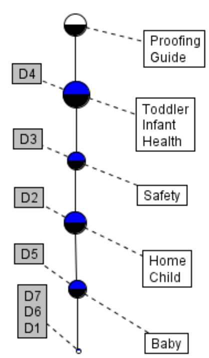


Fig. 7: Reduction of Example 1 from [12] using ICRA

In classification tasks, there is often a trade – off between the (relative) number of correctly classified objects and, for example, the (relative) cost of obtaining the classification or the clarity of a pictorial representation. In some instances, this may be expressed as the amount of errors we are prepared to allow to achieve another aim. A case in point are curves based on receiver operating characteristics (ROC), where the sensitivity (benefit) of a binary classifier is plotted as a function of its FP rate (cost), see [5] for an overview. We can plot the relative incomparability as a function of the number of bits changed to achieve it, see the reducibility graph in Figure 8. If we interpret (in-)comparability as sensitivity and the number of changed bits as cost to retrieve the original data, this can be interpreted as a ROC curve.

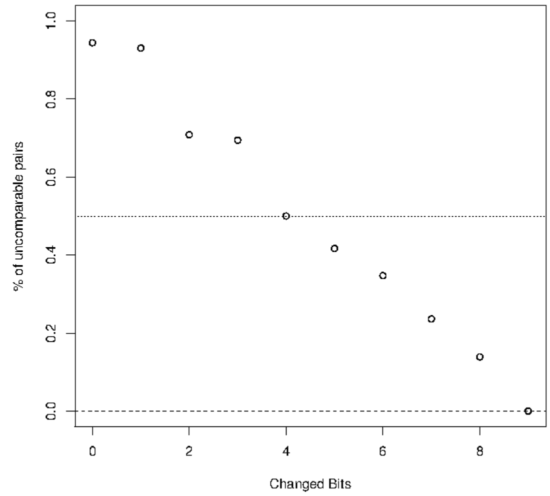


Fig. 8: Reducing relative incomparability with ICRA

The next example for [12] investigates a dataset consisting of various species of bacteria and 16 phenotypic characters, shown in Table 1.

	H2S	MAN	LYS	IND	ORN	CIT	URE	ONP	VPT	INO	LIP	PHE	MAL	ADO	ARA	RHA
ecoli1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
ecoli2	0	1	0	1	1	0	0	1	0	0	0	0	0	0	1	0
ecoli3	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1
styphi1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
styphi2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
styphi3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
kpneu1	0	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1
kpneu2	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1
kpneu3	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1
kpneu4	0	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1
kpneu5	0	1	1	1	0	1	0	1	1	1	0	0	1	1	1	1
pvul1	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0
pvul2	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
pvul3	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
pmor1	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0
pmor2	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
smar	0	1	1	0	1	1	0	1	1	0	1	0	0	0	0	0

Table 1: Bacterial dataset from [12]

For this context $\mathfrak{C}$, the incomparability $\texttt{incomp}(\mathfrak{C})$ turns out to be 81%. $\mathfrak{C}$ is reduced with the FKM method for $k = 5$ and $k = 9$, resulting in contexts $\mathfrak{C}_5$ and $\mathfrak{C}_9$ with $\texttt{incomp}(\mathfrak{C}_5) = 34.5\%$ and $\texttt{incomp}(\mathfrak{C}_9) = 64.7\%$. 40 bits are required to reduce $\mathfrak{C}$ to C_5, and the reduction to $\mathfrak{C}_9$ with 64.7% incomparability needs changing 11 bits. In contrast, our algorithm requires changing 19 bits to achieve an incomparability reduction to 34.6%, and 8 bits for a reduction to 66.1%. Changing 11 bits (as in the FKM reduction with k = 9) results in a reduction to 60.2%. The ICRA reducibility graph is shown in Figure 9.

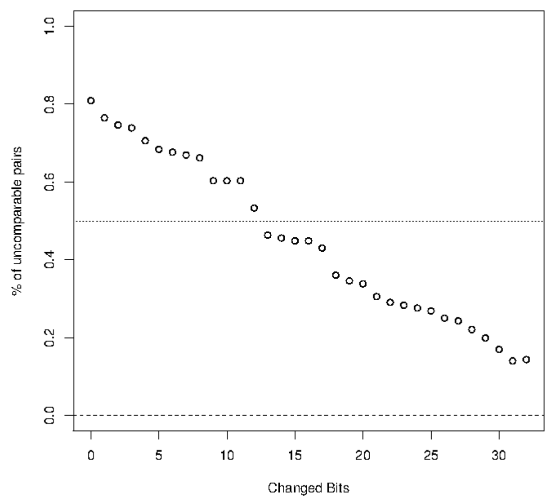


Fig. 9: Reducing relative incomparability of the bacterial dataset with ICRA

6 Conclusion and outlook

We have introduced a simple algorithm ICRA to reduce a formal context, the success criterion of which is a prescribed reduction of incomparable pairs. As a rule of thumb, we propose a relative frequency of incomparable pairs of objects of 50%. This seems a fair compromise between closeness to the data on the one hand, and the additional structure introduced by the chosen model on the other. We have compared the success of our algorithm with several examples of [12] and have found that fewer bits are needed than FKM to obtain similar incomparability ratios. Furthermore, the FKM algorithm requires much more effort and additional model assumptions so that its cost/benefit ratio is much smaller than for the median comparability algorithm. Furthermore, it is not clear which *k* should used for the reduction.

In the available space, only an indication of the impact of the median comparability algorithm could be given. Further work will include investigation of the powers and limitations of the ICRA algorithm using both theoretical and practical analysis. In particular, we shall consider its effects on implication sets and association rules.

References

1. Cheung, K., Vogel, D.: Complexity reduction in lattice-based information retrieval. Information Retrieval 8(2), 285–299 (2005), http://dx.doi.org/10.1007/s10791-005-5663-y

2. Dias, S.M., Vieira, N.J.: Reducing the size of concept lattices: The JBOS approach. Proc. CLA pp. 80–91 (2010)
3. Düntsch, I., Gediga, G.: Approximation operators in qualitative data analysis. In: de Swart, H., Orłowska, E., Schmidt, G., Roubens, M. (eds.) Theory and Application of Relational Structures as Knowledge Instruments, Lecture Notes in Computer Science, vol. 2929, pp. 214–230. Springer–Verlag, Heidelberg (2003)
4. Falmagne, J.C., Koppen, M., Villano, M., Doignon, J.P., Johannesen, J.: Introduction to knowledge spaces: How to build, test and search them. Psychological Review 97 (1990)
5. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters 27, 861–874 (2006)
6. Ganter, B., Wille, R.: Formal concept analysis: Mathematical foundations. Springer–Verlag, Berlin (1999)
7. Gediga, G., Düntsch, I.: Skill set analysis in knowledge structures. British Journal of Mathematical and Statistical Psychology 55, 361–384 (2002), http://www.cosc.brocku.ca/~duentsch/archive/skills2.pdf
8. Grätzer, G.: General Lattice Theory. Birkhäuser, Basel, second edn. (2000)
9. Haertel, E.H.: Using restricted latent class models to map the skill structure of achievement items. Journal of Educational Measurement 26, 301–324 (1989)
10. Junker, B.W., Sijtsma, K.: Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Applied Psychological Measurement 25, 258–272 (2001)
11. Krupka, M.: On complexity reduction of concept lattices: three counterexamples. Information Retrieval 15(2), 151–156 (2012), http://dx.doi.org/10.1007/s10791-011-9175-7
12. Kumar, C.A., Srinivas, S.: Concept lattice reduction using fuzzy K–means clustering. Expert Systems with Applications 37, 2696–2704 (2010)
13. Kuznetsov, S., Obiedkov, S., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In: Priss, U., Polovina, S., Hill, R. (eds.) Conceptual Structures: Knowledge Architectures for Smart Applications, Lecture Notes in Computer Science, vol. 4604, pp. 241–254. Springer Berlin Heidelberg (2007), http://dx.doi.org/10.1007/978-3-540-73681-3_18
14. Macready, G.B., Dayton, C.M.: The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics 2, 99–120 (1977)
15. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2014), http://www.R-project.org/
16. Tatsuoka, K.K..: Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement. Vol 20(4) pp. 345–354 (1983)
17. Wille, R.: Restructuring lattice theory: An approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered sets, NATO Advanced Studies Institute, vol. 83, pp. 445–470. Reidel, Dordrecht (1982)
18. Wolff, K.E.: A first course in formal concept analysis - how to understand line diagrams. In: Faulbaum, F. (ed.) Softstat '93: Advances in Statistical Software 4. pp. 429–438 (1993)
19. Yevtushenko, S.: The Concept Explorer (2000), http://conexp.sourceforge.net/index.html, http://conexp.sourceforge.net/index.html, retrieved Dec 24, 2011