Scale Coarsening as Feature Selection

Bernhard Ganter and Sergei O. Kuznetsov

Institut für Algebra
Dresden University of Technology
D-01062 Dresden, Germany
Bernhard.Ganter@tu-dresden.de
Department of Applied Mathematics
Higher School of Economics
Kirpichnaya ul. 33/5
105679 Moscow, Russia
skuznetsov@hse.ru

Abstract. We propose a unifying FCA-based framework for some questions in data analysis and data mining, combining ideas from Rough Set Theory, JSM-reasoning, and feature selection in machine learning. Unlike the standard rough set model the indiscernibility relation in our paper is based on a quasi-order, not necessarily an equivalence relation. Feature selection, though algorithmically difficult in general, appears to be easier in many cases of scaled many-valued contexts, because the difficulties can at least partially be projected to the scale contexts. We propose a heuristic algorithm for this.

1 Introduction

A paper recycling company gets vast amounts of material delivered for recycling every day. The first step in their process is to separate the waste from the recyclable part. This is done automatically: A machine performs certain optical measurements on every single piece and then decides which fraction it goes to. We are interested in the rules by which these decisions are made.

The situation is typical for applications of *Machine Learning* [14], and most likely the decision rules were obtained from a *training data* set, using a method of *supervised learning*. Machine Learning offers powerful algorithms, in particular when the data is numerical in nature. Here we concentrate on the more general case of *qualitative data*, and formalise the learning scenario as follows: We are given a formal context (G, M, I) [8], describing the "observations" or "measurements", together with a set $G_+ \subseteq G$, comprising the objects of interest, also called the *positive examples*. Objects from the complement $G_- = G \setminus G_+$ are called negative examples. The task then is to give a characterisation of G_+ in terms of (G, M, I) (a similar problem may be stated for G_-). The nicest case, of course, is that membership in G_+ is equivalent to some attribute combination, i.e., that G_+ is a concept extent of (G, M, I). But even if that is not the case, often a classification is desired. The second best choice then is to find attribute

R. Medina and S. Obiedkov (Eds.): ICFCA 2008, LNAI 4933, pp. 217–228, 2008.

combinations (*"classifiers"*) that are *sufficient* for membership in G_+. And ideally there should be enough such classifiers to cover all elements of G_+. This motivates our first definition:

Definition 1. *Let* (G, M, I) *be a formal context. A set*

$$G_+ \subseteq G$$

is called ***grounded*** *or, equivalently,* ***definable****, iff*

$$G_+ = \bigcup \{P' \mid P \subseteq M, P' \subseteq G_+\}.$$

The word "grounded" is used in JSM-theory of inductive reasoning [2,3,4], and it is defined there in a slightly different manner: The sets P in the above definition are required to be **(positive) hypotheses** for G_+, that is, *concept intents* P with $P' \subseteq G_+$ (see FCA formalization [5] of [4]). But it is easy to see that this causes no additional difficulty, simply replacing each P by its closure P''.

The word "definable" comes from Rough Set Theory [15,16,18,19], where it is defined in terms of an **indiscernibility relation**, usually an equivalence relation. Our approach generalises this. The role of the indiscernibility will be taken by the *object quasi-order* of the formal context and will not necessarily be symmetric. This is unfolded in Theorem 1 below. The relation between FCA, JSM-reasoning and Rough Set Theory was first studied in [17], but for specific hypotheses from [10].

2 Definability and the Object Quasi-order

The **object quasi-order** of a formal context (G, M, I) is defined by

$$g \leq h :\iff g' \supseteq h' \qquad (g, h \in G).$$

It is indeed reflexive and transitive, but not necessarily anti-symmetric. This makes it a quasi-order (called a *preorder* by some authors). The notion of an **order ideal** is the same as for ordered sets: a subset $S \subseteq G$ such that $h \in S$ and $g \leq h$ always implies $g \in S$. The quasi-order ideals are precisely the extents of the formal context $(G, G, \not\geq)$, as in the case of an ordered set.

Theorem 1. *The definable object sets of* (G, M, I) *are precisely the quasi-order ideals of the object quasi-order. Each subset* $G_+ \subseteq G$ *contains a largest definable set* $\underline{R}(G_+)$*, and has a smallest definable set containing it, denoted* $\overline{R}(G_+)$*.*

Proof. Let $G_+ \subseteq G$ be some subset and let $g \in G$. There exists some $P \subseteq M$ with

$$g \in P' \subseteq G_+$$

iff $g'' \subseteq G_+$. But since

$$g'' = \{h \mid h' \supseteq g'\} = \{h \mid h \leq g\},$$

this is equivalent to G_+ being a quasi-order ideal. Since the family of quasi-order ideals is closed under set union and intersection, the rest of the proposition is immediate.

Several algorithmic questions about the object quasi-order will arise in the sequel. We mention here three of them:

Feature Sets: Given a definable set $G_+ \subseteq G$, which subsets $F \subseteq M$ suffice to make G_+ definable? In other words, what are the subsets $F \subseteq M$ for which $G_+ = \bigcup\{P' \mid P \subseteq F, P' \subseteq G_+\}$?

Global Reducts: Which subsets of the attribute set of a formal context (G, M, I) induce the same definability? In other words, what are the subsets $F \subseteq M$ for which it is true that

$$g \leq h \iff g' \cap F \supseteq h' \cap F \qquad \text{for all } g, h \in G?$$

Separability: Given subsets $L, U \subseteq G$ such that $u \leq l$ holds for no $u \in U$ and no $l \in L$. Which sets also separate L from U, i.e., for which $E \subseteq M$ is it true that

$$u' \cap E \supseteq l' \cap E$$

holds for no $u \in U$ and no $l \in L$?

In what follows we will show that these problems are algorithmically difficult if we require the respective subsets of attributes to be minimal.

The **lower** and **upper approximation operators**, as the operators $\underline{R}(\cdot)$ and $\overline{R}(\cdot)$ occurring in the theorem are called in Rough Set Theory, are given as follows:

$$\underline{R}(G_+) = \bigcup\{P' \mid P \subseteq M, P' \subseteq G_+\}$$
$$\overline{R}(G_+) = \bigcup\{g'' \mid g \in G_+\}.$$

The operators can also be given in terms of the JSM-method. In [6,7] we have introduced the notion of a **hopeless example**, by which we meant a positive example $g \in G_+$ which cannot be classified because there is some object $h \notin G_+$ having all attributes of g. That is $g \in G_+$ is hopeless iff there is some $h \notin G_+$ such that $h \leq g$. In that language then

$$\underline{R}(G_+) = \{g \in G_+ \mid g \text{ is not hopeless}\}$$
$$\overline{R}(G_+) = \underline{R}(G_+) \cup \{h \mid h \leq g \text{ for some hopeless } g \in G_+\}.$$

The next proposition is now immediate.

Proposition 1. *The following conditions are equivalent:*

1. *G_+ is definable,*
2. *$\overline{R}(G_+) = G_+$,*
3. *$\underline{R}(G_+) = G_+$.*

Example. We illustrate our definitions by means of an artificial example. Consider the following context (G, M, I) where positive examples are fruits. This information is given by the target attribute "fruit", which does not belong to the set of attributes M.

	color	firm	smooth	form	fruit
apple	yellow	no	yes	round	+
grapefruit	yellow	no	no	round	+
kiwi	green	no	no	oval	+
plum	blue	no	yes	oval	+
toy cube	green	yes	yes	cubic	–
egg	white	yes	yes	oval	–
tennis ball	white	no	no	round	–

Consider a natural scaling of the context

	white	yellow	green	blue	firm	nonfirm	smooth	nonsmooth	round	nonround	fruit
apple		×				×	×		×		+
grapefruit		×				×		×	×		+
kiwi			×			×		×		×	+
plum				×		×	×			×	+
toy cube			×		×		×			×	-
egg	×				×		×			×	-
tennis ball	×					×		×	×		-

- A minimal feature set for G_+ is the set $\{\text{yellow}, \text{nonfirm}, \text{nonround}\}$.
- A minimal feature set for G_- is $\{\text{white}, \text{firm}\}$.
- A minimal global reduct is, e.g., the set
  $\{\text{white}, \text{yellow}, \text{green}, \text{smooth}, \text{nonsmooth}, \text{round}, \text{nonround}\}$.
- The set $\{\text{white}, \text{firm}\}$ is a minimal set separating G_- from G_+.
- The set of positive examples is definable, since
  $\overline{R}(G_+) = \underline{R}(G_+) = G_+ = \{\textit{apple, grapefruit, kiwi, plum}\}$.
- Consider another positive example *orange*, which is orange, nonfirm, nonsmooth and round. Under the scaling chosen,
  $\textit{orange}' = \{\text{nonfirm}, \text{nonsmooth}, \text{round}\}$.
  Thus this example is hopeless for the scaling, since $\textit{orange}' \subseteq \textit{tennis ball}'$.
- For the extended data set we have
  $\overline{R}(G_+) = \{\textit{apple, grapefruit, kiwi, plum, tennis ball, orange}\}$.
  $\underline{R}(G_+) = \{\textit{apple, grapefruit, kiwi, plum}\}$ and
  Thus including *orange* in the set G_+ of positive examples makes G_+ undefinable (for the given scaling).

3 Feature Sets

Not all attributes in the attribute set M may be necessary for the classification, often a subset may suffice. Such subsets are called *feature sets*. In Rough Set

Theory, minimal feature sets are called *reducts*. The process of thinning out the attributes to obtain a feature set is called *feature selection* [13]. In relation to FCA-based hypotheses this was studied in [6,1]. To relate these issues to the Rough Set Theory, we introduce for arbitrary subsets $N \subseteq M$ the **relative approximation operators**:

$$\underline{R}_N(G_+) = \bigcup\{P' \mid P \subseteq N, P' \subseteq G_+\}$$
$$\overline{R}_N(G_+) = \bigcup\{(g' \cap N)' \mid g \in G_+\}.$$

So the relative approximation operators are simply the approximations operators for the shortened formal context $(G, N, I \cap G \times N)$. They therefore share the properties of approximation operators. We say that G_+ is *definable in terms of* N, shortly N**-definable** or N**-grounded** iff

$$\underline{R}_N(G_+) = G_+ = \overline{R}_N(G_+),$$

where again each of the two equalities implies the other.

If G_+ has only one element, we omit the set brackets and write $\overline{R}_N(g)$ instead of $\overline{R}_N(\{g\})$.

Proposition 2. $\overline{R}_A(g) \cap \overline{R}_B(g) = \overline{R}_{A\cup B}(g)$.

Proof. $\overline{R}_A(g) \cap \overline{R}_B(g) = (g' \cap A)' \cap (g' \cap B)' = ((g' \cap A) \cup (g' \cap B))' = (g' \cap (A \cup B))'$.

There are two different ways to formally define the notion of a feature set. In the *global* view, we look for sets inducing the same definable sets as M does. We call $F \subseteq M$ a **global feature set** if for *all* subsets $S \subseteq G$ it holds that

$$\underline{R}_F(S) = \underline{R}(S) \quad \text{and } \overline{R}_F(S) = \overline{R}(S),$$

which is equivalent to the condition that

$$S \text{ is } F\text{-definable iff } S \text{ is definable.}$$

Finding global feature sets is equivalent to the global reduct problem mentioned above. Its complexity will be treated in the next section.

Our focus here is more on finding feature sets for a given target set G_+ of positive examples. So we are interested in finding, for a fixed given definable set $G_+ \subseteq G$ sets $F \subseteq M$ such that

$$\underline{R}_F(G_+) = G_+ = \overline{R}_F(G_+).$$

Such a set will be called a **feature set for** G_+. Note that we do not restrict ourselves to *minimal* such sets. But finding small ones is indeed intractable, as it is for reducts in the case of Rough Sets:

Proposition 3. *The problem of finding small feature sets, given by*

INSTANCE: *A formal context* (G, M, I)*, a definable set* $G_+ \subseteq G$*, and a natural number* k*.*

QUESTION: *Is there a feature set for G_+ of size $\leq k$, i.e., a subset $F \subseteq M$ such that $\underline{R}_F(G_+) = G_+$ and $|F| \leq k$?*

is $\mathcal{NP}$–complete.

Proof. The problem is in $\mathcal{NP}$, because for testing if a given $F \subseteq M$ is a feature set we only need to check if $\bigcup\{(g' \cap F)' \mid g \in G_+\} = G_+$. This can clearly be done in polynomial time.

To show that the problem is $\mathcal{NP}$-hard, we reduce it to a problem well known to be $\mathcal{NP}$-complete: Finding transversals of a family of sets:

INSTANCE: A set M, a family $S_t, t \in T$ of nonempty proper subsets of M (here T is some index set), and an integer k.
QUESTION: Is there a subset $F \subseteq M$, $|F| \leq k$, such that $F \cap S_t \neq \emptyset$ for all $t \in T$?

Given an instance of the transversal problem, we can construct a formal context (G, M, I) by letting $G := T \cup \{g_0\}$, $t' := S_t$ for $t \neq g_0$ and $g_0' := \emptyset$. Moreover, we set $G_+ := T$. It is easy to check that $F \subseteq M$ is a feature set for G_+ iff it is a transversal for $\{S_t \mid t \in T\}$.

Our approach to finding feature sets for G_+ is an indirect one. Rather than building such sets bottom-up, we assume that we are already given one, say F, where $F = M$ is a possible choice. Then we try thinning F, using the following strategy: We consider some subset of F which is not a feature set for G_+ and investigate which elements of F must be added to extend that subset to a feature set for G_+. There will be no unique answer to this question. Our goal is to describe all possible solutions.

More formally, let F be a feature set for G_+, so that

$$\underline{R}_F(G_+) = G_+ = \overline{R}_F(G_+).$$

Fix some subset $N \subseteq F$ which is not a feature set, so that

$$\underline{R}_N(G_+) \subsetneq G_+ \subsetneq \overline{R}_N(G_+).$$

Then both the **lower boundary**

$$L := G_+ \setminus \underline{R}_N(G_+)$$

and the **upper boundary**

$$U := \overline{R}_N(G_+) \setminus G_+$$

are nonempty sets. The lower boundary consists of those elements $g \in G_+$ which are not in the extent of any hypothesis $H \subseteq N$ with $H' \subseteq G_+$.

Theorem 2. *Let $N, E \subseteq M$ and let L, U denote the lower and upper boundary with respect to N. Then $N \cup E$ is a feature set for G_+ iff for all $g \in L$ it holds that*

$$\overline{R}_N(g) \cap \overline{R}_E(g) \subseteq G_+.$$

A sufficient condition is

$$\overline{R}_E(L) \cap U = \emptyset.$$

Proof. $N \cup E$ is a feature set for G_+ iff each object $g \in G_+$ is implied by some $P \subseteq N \cup E$ with $P' \subseteq G_+$. For objects in $\underline{R}_N(G_+)$ this is clear anyway, so it suffices to consider objects from the lower boundary $L = G_+ \setminus \underline{R}_N(G_+)$. For every such object $g \in L$ we must have that

$$\overline{R}_{N \cup E}(g) \subseteq G_+.$$

By Proposition 2, this is equivalent to

$$\overline{R}_N(g) \cap \overline{R}_E(g) \subseteq G_+ \quad \text{for all } g \in L.$$

Since $\overline{R}_N(g) \subseteq \overline{R}_N(G_+)$ holds anyway, it suffices that

$$\overline{R}_E(g) \cap U = \emptyset$$

holds for all $g \in L$. But because of $\overline{R}_E(L) = \bigcup_{g \in L} \overline{R}_E(g)$ this can be summarised to

$$\overline{R}_E(L) \cap U = \emptyset.$$

4 Global Reducts and Separators

Finding minimal global reducts may be hard, which is expressed by the following

Proposition 4. *The following problem is $\mathcal{NP}$-complete*[1]*:*

INSTANCE: *A formal context (G, M, I) and a natural number k.*
QUESTION: *Is there a subset $F \subseteq M$, $|F| \leq k$, such that*

$$g \leq h \iff g' \cap F \supseteq h' \cap F \qquad \textit{for all } g, h \in G?$$

Proof. We reduce "3-dimensional matching", a well-known $\mathcal{NP}$-complete problem [9], to our problem. It requires to decide, for given disjoint sets X, Y, and Z of equal cardinality k and a set $T \subseteq X \times Y \times Z$, if T contains a *matching*, that is, a subset $T' \subseteq T$ such that $|T'| = k$ and no two elements of T' agree in any coordinate. Such a matching can of course only exist if the coordinates of T cover the sets X, Y, and Z, respectively, so this can be assumed as additional precondition.

Given such an instance T for some $k > 1$, we can construct a formal context having a global reduct of size $\leq k$ if and only if the instance contains a matching. The construction is as follows. Let

$$G_0 := \{(w, 0) \mid w \in X \cup Y \cup Z\}, \text{ and}$$
$$G_1 := \{(w, 1) \mid w \in X \cup Y \cup Z\}.$$

We investigate the formal context $(G, T \,\dot\cup\, \{m_X, m_Y, m_Z\}, I)$ with $G := G_0 \cup G_1$, where the incidence is defined as follows:

[1] See the acknowledgements in Section 7 below.

$m'_X := X \times \{0,1\}$, $m'_Y := Y \times \{0,1\}$, $m'_Z := Z \times \{0,1\}$,
and, for each $t =: (x,y,z) \in T$,
$t' := G_0 \setminus \{(x,0),(y,0),(z,0)\} \cup \{(x,1),(y,1),(z,1)\}$.

When is $g \leq h$ in this formal context? Recall that objects are pairs (w,i), where $w \in X \cup Y \cup Z$ and $i \in \{0,1\}$. An analysis of the possible cases shows that $(w_1,i_1) \leq (w_2,i_2)$ holds if and only if w_1 and w_2 are from the same set (that is, $\{w_1,w_2\}$ is a subset of either X or Y or Z), $w_1 \neq w_2$, $i_1 = 0$ and $i_2 = 1$. Actually, this order is obtained exactly from those subsets of T containing triples such that each element of $X \cup Y \cup Z$ occurs at least once as a component. Such a subset has cardinality $\leq k$ if and only if it is a matching. Therefore the existence of a 3-dimensional matching is reduced to the problem of finding a global reduct with $\leq k$ attributes.

A similar result holds for the problem of finding a minimal separator, i.e., a minimal set of attributes separating a set of objects from another one, as stated by the following

Proposition 5. *The following minimal separator problem is $\mathcal{NP}$-complete:*

INSTANCE: *A formal context (G,M,I), two sets of objects $L,U \subseteq G$ such that $u \leq l$ holds for no $l \in L$, $u \in U$, and a natural number k.*
QUESTION: *Is there a subset $F \subseteq M$, $|F| \leq k$ such that*

$$u' \cap F \supseteq l' \cap F \qquad \textit{holds for no } u \in U, l \in L?$$

Proof. We reduce the minimal transversal problem

INSTANCE: A set M, a family $S_t, t \in T$ of nonempty proper subsets of M (here T is some index set), and an integer k.
QUESTION: Is there a subset $F \subseteq M$, $|F| \leq k$, such that $F \cap S_t \neq \emptyset$ for all $t \in T$?

Given an instance of the transversal problem, we can construct a formal context (G,M,I) by letting $G := T \cup \{g_0\}$, $t' := M \setminus S_t$ for $t \neq g_0$ and $g'_0 := M$. Let $L = \{g_0\}$, $U = T$. It is easy to check that $F \subseteq M$ separates $L = \{g_0\}$ from $U = T$ iff F is a transversal for $\{S_t \mid t \in T\}$. The reduction is completed, its polynomiality, as well as the membership of the minimal separator problem in $\mathcal{NP}$ are obvious.

5 Scale Coarsening

Theorem 2 was tailored for applications to *scaled many-valued contexts*. For understanding this article it is not required to recall the precise definitions (which can be found in [8]). It suffices to understand that these are formal contexts (G,M,I) for which the attribute set M can be subdivided into subsets $M_s, s \in S$, such that each such M_s comes from a standardised formal context

$\mathbb{S}_s := (G_s, M_s, I_s)$, a "scale". Some scales are used frequently because of their interpretation and their particularly simple structure, like "nominal", "ordinal" or "interordinal" scales. For these, the algorithmic problems mentioned above are easy to solve.

The heuristic procedure that we suggest for feature selection in scaled many valued contexts builds on this. Feature selection will result in coarser scales, because some scale attributes will not be used. We propose the following strategy:

- Start with some feature set F, for example $F := M$.
- Then pick a scale, one after another, and
  1. remove the set M_s of scale attributes from the feature set.
  2. The result $N := F \setminus M_s$ may fail to be a feature set. In that case, use Theorem 2 to find an appropriate set $E \subseteq F \setminus N$ such that $N \cup E$ is a feature set.
  3. Replace F by $N \cup E$, and continue.

Note that choosing E can be done in two ways, according to Theorem 2. Either we use the equivalence stated in the first part of the theorem, which gives the precise results. Or we use the sufficient condition given in the second part. Note that this amounts to solving the separation problem stated above, but only of the formal context $(G, E, I \cap (G \times E))$.

This is, as already said, a heuristic procedure. Its result depends on the sequence in which the scales are handled, and even if the set E is chosen minimal in each step, we do not claim that the result is a minimal reduct. This heuristic can be useful for data with very large sets of attributes like those described in [12], where standard context reduction [8] is difficult because it is hard even to keep the context in the memory.

However, we expect that the method leads to reasonably small feature sets in a reasonable computing time, since the application of Theorem 2 avoids exhaustive search in testing whether a subset of attributes is good (but not in finding the minimal reduct itself) by projecting the problem to standardised scales.

But more importantly, the method is flexible enough to include other criteria into the search for good feature sets. Small size is not always the most desirable property, and other aspects may be more important. The next section gives an example of this.

6 To Avoid Overfitting

Recall the example that was mentioned in the introduction, where paper samples were to be classified based on the spectra of the light spectra they emit. We are actually working on such a data set (it is too large to be discussed here in detail). There the spectra are given with such a precision that virtually *every* subset of the training data set is grounded, simply because no two of the spectra coincide precisely in every decimal digit. Thus the condition of definability,

$$G_+ = \bigcup\{P' \mid P \subseteq M, P' \subseteq G_+\},$$

is satisfied because for each $g \in G_+$ we get as a classifying attribute set $P := g'$, with $P' = \{g\}$.

However, such a classification will probably be useless when the classification rules learnt from the training set are to be applied to other data. Then, since the positive examples in the training set have been described so precisely, their descriptions will most likely not fit new examples outside the training set. This effect is called **overfitting**. There are many suggestions how this can be avoided.

In the original version of the JSM-method, for example, it is required that only rules are used for classification that apply to at least two positive examples. A set $G_+ \subseteq G$ is called **sufficiently grounded** if for each g in G_+ there exists some $h \in G_+$ such that

$$\{g, h\}'' \subseteq G_+.$$

This is the case if and only if

$$G_+ = \bigcup \{P' \mid P \subseteq M, P = P'', P' \subseteq G_+, |P'| \geq 2\}.$$

Note that the requirement $P = P''$ can be omitted here.

If the set of positive training examples is sufficiently grounded, it is possible to allowing only attribute sets $P \subseteq M$ as classifiers whose **support** $|P'|$ is at least 2. It is reasonable that this restriction lowers the effect of overfitting, because an attribute combination that applies to at least two different objects is more likely to apply to other objects as well. This approach can, of course, be varied by replacing 2 by other thresholds and so on. We are not going into such details here. Instead, we shall study the following problem: Call $F \subseteq M$ a **strong feature set** for G_+ if

$$G_+ = \bigcup \{P' \mid P \subseteq F, P' \subseteq G_+, |P'| \geq 2\}.$$

Clearly G_+ is sufficiently grounded if and only if there is a strong feature set for G_+. However, even if G_+ is sufficiently grounded, not every feature set for G_+ must be strong. The question to investigate therefore is: How can the feature selection procedure described in Section 5 be modified to obtain strong feature sets? Unfortunately, the necessary modification of Theorem 2 is not very elegant:

Proposition 6. *Suppose that $F \subseteq M$ is a strong feature set for G_+ and that $N, E \subseteq F$. Then $N \cup E$ is a strong feature set for G_+ iff for each $g \in G_+$ there is some h such that*

$$(\{g, h\}' \cap N)' \cap (\{g, h\}' \cap E)' \subseteq G_+.$$

This is rather obvious. Not so obvious, but not a hopeless task, is how this can be made efficient in an algorithm. We pose this as a problem.

7 Conclusion

We considered a framework for selecting important subsets of attributes (or attribute values) in FCA-based knowledge discovery. This framework uses the ideas of reducts, upper and lower approximations of the Rough Set Theory, at the same time generalizing the latter by allowing for a quasi-order (not necessarily equivalence) indiscernibility relation. We showed that choosing smallest representations (global reducts, feature sets) is intractable (NP-complete) in general settings, and propose a heuristic based on coarsening the set of attribute values.

Acknowledgements

We thank Graham Brightwell (LSE, London) and William T. Trotter (Georgia Tech, Atlanta) for giving us most useful hints for proving NP-completeness of the minimal global reducts problem.

References

1. Ferré, S., Ridoux, O.: The Use of Associative Concepts in the Incremental Building of a Logical Context. In: Priss, U., Corbett, D.R., Angelova, G. (eds.) ICCS 2002. LNCS (LNAI), vol. 2393, pp. 299–313. Springer, Heidelberg (2002)
2. Finn, V.K.: On Machine-Oriented Formalization of Plausible Reasoning in the Style of F. Backon–J. S. Mill (in Russian). Semiotika Informatika 20, 35–101 (1983)
3. Finn, V.K.: Plausible Reasoning in Systems of JSM Type (in Russian). Itogi Nauki i Tekhniki, Seriya Informatika 15, 54–101 (1991)
4. Finn, V.K.: Synthesis of cognitive procedures and the problem of induction (in Russian). Nauchno-Tekhnicheskaya Informatsiya 2(1-2), 8–44 (1999)
5. Ganter, B., Kuznetsov, S.O.: Formalizing Hypotheses with Concepts. In: Ganter, B., Mineau, G.W. (eds.) ICCS 2000. LNCS, vol. 1867, pp. 342–356. Springer, Heidelberg (2000)
6. Ganter, B., Kuznetsov, S.O.: Pattern Structures and Their Projections. In: Delugach, H.S., Stumme, G. (eds.) ICCS 2001. LNCS (LNAI), vol. 2120, pp. 129–142. Springer, Heidelberg (2001)
7. Ganter, B., Kuznetsov, S.O.: Hypotheses and Version Spaces. In: Ganter, B., de Moor, A., Lex, W. (eds.) ICCS 2003. LNCS, vol. 2746, pp. 83–95. Springer, Heidelberg (2003)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
9. Garey, M.R., Johnson, D.S.: Computers and Intractability. A Guide to the Theory of NP-Completeness. W.H. Freeman, New York (1979)
10. Grigoriev, P.A., Kuznetsov, S.O., Obiedkov, S.A., Yevtushenko, S.A.: On a Version of Mill's Method of Difference. In: Proc. ECAI 2002 Workshop on Concept Lattices in Data Mining, Lyon, pp. 26–31 (2002)
11. Kuznetsov, S.O.: JSM-method as a Machine Learning System. Itogi Nauki i Tekhniki, ser. Informatika 15, 17–54 (1991)
12. Kuznetsov, S.O., Samokhin, M.V.: Learning Closed Sets of Labeled Graphs for Chemical Applications. In: Kramer, S., Pfahringer, B. (eds.) ILP 2005. LNCS (LNAI), vol. 3625, pp. 190–208. Springer, Heidelberg (2005)
13. Liu, H., Motoda, H.: Feature Selection for Knowledge Discovery and Data Mining. In: The Springer International Series in Engineering and Computer Science, vol. 454, Springer, New York (1998)

14. Mitchell, T.: Machine Learning, The McGraw-Hill Companies (1997)
15. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: Probabilistic versus deterministic approach. International Journal of Man-Machine Studies 29, 81–95 (1988)
16. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishing, Dordrecht (1991)
17. Wolski, M.: Galois connections and data analysis. Fundamenta Informatica 60, 401–415 (2004)
18. Ziarko, W.: Rough sets as a methodology for data mining. In: Rough Sets in Knowledge Discovery 1: Methodology and Applications, pp. 554–576. Physica-Verlag, Heidelberg (1998)
19. Ziarko, W., Shan, N.: Discovering attribute relationships, dependencies and rules by using rough sets. In: Proc. 28th Annual Hawaii International Conference on System Sciences (HICSS 1995), pp. 293–299 (1995)