# Formalizing Hypotheses with Concepts

Bernhard Ganter and Sergei O. Kuznetsov

[1] Technische Universität Dresden
ganter@math.tu-dresden.de
[2] Technische Universität Dresden
kuznetso@math.tu-dresden.de

**Abstract.** The frequently used strategy of forming hypotheses, based on observations, to generate predictions can be formalized in several ways. We study an elaborated approach that has some conceptual flavour and translate the basic ideas in the language of Formal Concept Analysis, blazing the trail for applications to other conceptual structures as well. We investigate the relation to pseudo intents, discuss algorithmic questions, and give an example.

## Motivation

Imagine that on your way to work there is a junction where you frequently run into a traffic jam. You would like to know in advance whether you are likely to get in the hold up or not, because then you could take a detour to avoid a delay. But you known of no single obvious cause for the congestion. You know that certain constellations are very likely to cause such a problem, and you also know of situations where it is rather safe. Probably you will develop your personal hypotheses, based on your positive and negative experiences, to predict traffic jams at that junction.

A similarly structured problem, concerning pharmacological applications, is known as the Structure-Activity Relationship (SAR) Problem [14], where structural properties of chemical compounds (such as particular subgraphs of their molecular graphs) are used as predictors for certain biological activities (like mutagenicity, sedativity, or such).

The logical framework of such problems has been extensively studied. One of the broader approaches is called the *JSM-method* (after John Stuart Mill, an English philosopher of the 19th century, who was one of the first to systematically consider schemes of inductive reasoning). It was proposed by V. Finn [3] in 1983 and has been developed ever since by his group at the Moscow VINITI Institute. From the standpoint of Artificial Intelligence, that method can be considered as a formal system of inductive plausible reasoning based on examples and knowledge about the domain (see Finn [4]) or as a method of Machine Learning that employs examples and counterexamples of a goal attribute (compare [10]).

The original formalization of the JSM-method used by Finn was the language of first-order predicate calculus with two sorts of variables, six truth-value types

B. Ganter and G.W. Mineau (Eds.): ICCS 2000, LNAI 1867, pp. 342–356, 2000.

and "quanitification over tuples with variable length" [1]. However, the underlying combinatorial structure of the JSM-method suggests analogies to another method of knowledge processing, that of the Formal Concept Analysis (FCA) [6]. Some attempts to connect these approaches were already made in [11], [12]. In this paper we propose a partial translation from the language of the JSM-theory into that of FCA.[1]

# 1 Hypotheses

We consider a finite set $M$ of "structural attributes", a set $G$ of objects (or observations) and a relation $I \subseteq G \times M$, such that $(g, m) \in I$ if and only if object $g$ has the attribute $m$. Such a triple $\mathbb{K} = (G, M, I)$ is called a **formal context**. It is the basic data type of Formal Concept Analysis. Using the **derivation operators**, defined for $A \subseteq G$, $B \subseteq M$ by

$$A' := \{m \in M \mid gIm \text{ for all } g \in A\},$$
$$B' := \{g \in G \mid gIm \text{ for all } m \in B\},$$

we can define a **formal concept** (of the context $\mathbb{K}$) to be a pair $(A, B)$ satisfying $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$. $A$ is called the **extent** and $B$ is called the **intent** of the concept $(A, B)$. These concepts, ordered by

$$(A_1, B_1) \geq (A_2, B_2) \iff A_1 \supseteq A_2$$

form a complete lattice, called **the concept lattice** of $\mathbb{K} := (G, M, I)$. Double application of the derivation operators, i.e., $(B')'$ or $(A')'$ are abbreviated as $B''$ or $A''$, repsectively. It can easily be shown that $''$ is a closure operator, i.e., it is monotone, idempotent, and extensive. Further on, this operator will be referred to as **closure** and sets $A \subseteq G$, $B \subseteq M$ such that $A'' = A$, $B'' = B$ as **closed**.

In what follows we shall need the notion of implication between attribute sets [6]. An **implication** $A \to B$ between a pair of subsets of the attribute set $M$ **holds** for a given formal context $\mathbb{K} := (G, M, I)$ if every object that has all the attributes from $A$ also has all attributes from $B$. This is equivalent to $A' \subseteq B'$.

In addition to the structural attributes of $M$ we consider a **goal attribute** $w \notin M$. This divides the set $G$ of all objects into three subsets: The set $G_+$ of those objects that are known to have the property $w$ (these are the *positive examples*), the set $G_-$ of those objects of which it is known that they do not have $w$ (the *negative examples*) and the set $G_\tau$ of *undetermined examples*, i.e., of those objects, of which it is unknown if they have property $w$ or not. This gives three subcontexts of $\mathbb{K} = (G, M, I)$:

$$\mathbb{K}_+ := (G_+, M, I_+), \mathbb{K}_- := (G_-, M, I_-), \text{ and } \mathbb{K}_\tau := (G_\tau, M, I_\tau),$$

where for $\varepsilon \in \{+, -, \tau\}$ we have $I_\epsilon := I \cap G_\varepsilon \times M$.

[1] In terms of the JSM-theory, we consider only "counterexample forbidding hypotheses" and the "atomistic case," where there is a single goal attribute

Intents are, as defined above, attribute combinations shared by some of the observed objects. In order to form hypotheses about structural causes of the goal attribute $w$, we are interested in sets of structural attributes that are shared by some positive, but by no negative examples. Thus, a **positive hypothesis**, or a (+)-hypothesis, for $w$ is an intent of $\mathbb{K}_+$ that is not contained in the intent $g'$ of any negative example $g \in G_-$. A **negative hypothesis**, or a (-)-hypothesis, is defined accordingly. If $g$ is an object such that its intent[2]

$$g' := \{m \in M \mid (g,m) \in I\}$$

contains a positive or a negative hypothesis then we say, for short, that $g$ contains that hypothesis.

We intend to use these hypotheses as predictors for the undetermined examples, predicting that an object $g \in G_\tau$ has the goal attribute if it contains a positive hypothesis and does not have $w$ if it contains a negative one. These cases may be not exclusive. It may happen that an object contains both a positive and a negative hypothesis. An object that contains a positive, but no negative hypothesis will be *classified positively.* Negative classifications are defined similarly. If $g'$ contains hypotheses of both kinds, or if $g'$ contains no hypothesis at all, then the classification is contradictory or undetermined, respectively. The underlying principle can be formulated in the line of J.S.Mill as "common effects are brought about by common causes." We may restrict to *minimal* (w.r.t. inclusion $\subseteq$) hypotheses, positive as well as negative, since an object obviously contains a positive hypothesis if and only if it contains a minimal positive hypothesis, etc.

Of course there is, without further assumptions, little hope that these predictions will be correct. Realistically, we have no reliable information on the unknown set $G_w$ of those objects in $G$ which do have $w$, except for $G_+ \subseteq G_w$ and $G_- \cap G_w = \emptyset$. We have not even excluded the possibility that the goal does not depend at all on the structural attributes. It might happen that there are positive and negative examples with exactly the same intents. All we can guarantee is the following, trivial fact:

**Proposition 1.** *A positive example $g \in G_+$ contains a positive hypothesis if and only if $g'$ is not contained in the intent of any negative example.*

Nevertheless it seems natural to operate with hypotheses, at least in situations where we expect that the goal attribute somehow depends on the structural attributes. An analogue of the following assumption on the (unknown) set $G_w$ is called the **Spinoza-axiom** in [4]:

The goal attribute $w$ is **properly implied** in $(G, M, I)$ if there are sets $\mathcal{P}, \mathcal{N} \subseteq \mathcal{P}(M)$[3] such that

1. $g \in G_w \iff P \subseteq g'$ for some $P \in \mathcal{P}$,
2. $g \notin G_w \iff N \subseteq g'$ for some $N \in \mathcal{N}$.

[2] For brevity sake we write $g'$ and $g''$ instead of $\{g\}'$ and $\{g\}''$, respectively.
[3] $\mathcal{P}(M)$ is the power set of $M$

Thus, the goal attribute is properly implied if there is a set $\mathcal{P}$ of *positive attribute combinations* that force the goal attribute $w$, and a set $\mathcal{N}$ of *negative attribute combinations* that exclude $w$, and if, moreover, these combinations would suffice to classify all objects in $G$.

However, this condition does not require that these positive or negative attribute combinations have been observed as such. Let us therefore assume that each $P$ and each $N$ is supported by some example, i.e., that

3. for each $P \in \mathcal{P}$ there is some $g \in G_+$ such that $P \subseteq g'$,
4. for each $N \in \mathcal{N}$ there is some $g \in G_-$ such that $N \subseteq g'$.

It seems that this condition is rather strong. It does however not suffice to make the hypotheses work, as the following example shows:

| | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ | $+,-,\tau$ | $w$ |
|---|---|---|---|---|---|---|---|---|
| $g_1$ | × | | × | | | | $+$ | yes |
| $g_2$ | | × | × | | | | $+$ | yes |
| $g_3$ | | | | × | × | | $-$ | no |
| $g_4$ | | | | × | | × | $-$ | no |
| $g_5$ | | × | × | × | | | $\tau$ | yes, but undetermined |
| $g_6$ | | | × | × | × | | $\tau$ | no, but undetermined |

In this formal context, we have $\mathcal{P} := \{\{a, c\}, \{b, c\}\}$, $\mathcal{N} := \{\{d, e\}, \{d, f\}\}$. Each set in $\mathcal{P}$ is a positive hypothesis, each set in $\mathcal{N}$ is a negative hypothesis. But $\{c\}$ is also a positive hypothesis and $\{d\}$ is a negative one, which has the effect that each undetermined object contains both a positive and a negative hypothesis.

Note that in this example each $g \in G_w$ contains a positive hypothesis and every $g \notin G_w$ contains a negative one.

Another condition would be that whatever is not causal for the goal attribute will be observed to be non-causal. This may be formalized as follows:

5. If for a set $S \subseteq G_+$ there is no $P \in \mathcal{P}$ with $P \subseteq S'$, then there is some $h \in G_-$ such that $S' \subseteq h'$,
6. if for a set $T \subseteq G_-$ there is no $N \in \mathcal{N}$ with $N \subseteq T'$, then there is some $h \in G_+$ such that $T' \subseteq h'$.

**Proposition 2.** *If conditions 1), 2), 5), and 6) are satisfied, then every positive hypothesis contains some $P \in \mathcal{P}$ and every negative hypothesis contains some $N \in \mathcal{N}$.*

Then, in particular, any object containing a positive hypothesis must be in $G_w$ and those containing a negative hypothesis must be in the complement of $G_w$. The converse is not necessarily true: it may happen that the observed hypotheses are too large and that, therefore, not every $g \in G_w$ contains a positive hypothesis in its intent. We may then try to improve our hypotheses by enlarging $G_+$ and $G_-$ by those elements from $G_\tau$ that contain a positive or a negative hypothesis, respectively. Conditions 1), 2), 5), and 6) will remain satisfied, but the new configuration may lead to new hypotheses which are smaller then the ones before.

This process may be iterated. For such an iteration process suggested by Finn there is no guarantee that the conditions 1) to 6) are fulfilled. The dynamics of such an iteration process will not be studied here.[4]

## 2 Hypotheses and Implications

We have introduced (positive) hypotheses as attribute combinations that imply the goal attribute $w$, but with an additional "conceptual" condition: only intents of formal concepts are considered as hypotheses. We explain shortly why this is reasonable.

Suppose that $C$ is some attribute combination occurring only in positive examples, so that we may expect that $C$ implies $w$. But suppose that there are further attributes, say $d_1, \ldots, d_n$, that also "come with $C$" in the sense that each observed object having $C$ also has the set of attributes $\{d_1, d_2, \ldots, d_n\}$. Then there are two typical strategies to generate a hypothesis:

> The *courageous* strategy would argue: "Since we have observed that $w$ occurs whenever $C$ does, we shall predict $w$ when we observe $C$."
> The *cautious* strategy would say: "We have observed that in case of $C$ we also have $d_1, \ldots, d_n$ and $w$, so we shall predict $w$ if we meet the same conditions, i.e., $C \cup \{d_1, \ldots, d_n\}$. If you think that the set of attributes $\{d_1, \ldots, d_n\}$ is superfluous, give an example in which, given $C$, the set of conditions $\{d_1, \ldots, d_n\}$ does not hold, but $w$ does."

The JSM-approach follows the cautious strategy. It is assumed that hypotheses are closed under implications to other structural attributes.

However, the subsets closed under implications are precisely the intents of $\mathbb{K}$. Therefore, it is sensible to allow only intents (of concepts of $\mathbb{K}$) as hypotheses.

There is a powerful tool for studying implications in formal contexts: the notion of a **pseudo intent**. The definition is recursive[5]. A **pseudo intent** of a formal context $(G, M, I)$ is a set $P \subseteq M$ satisfying

1. $P \neq P''$, and
2. for every pseudo intent $Q \subseteq P$, $Q \neq P$, we have $Q'' \subseteq P$.

Pseudointents can be used to give an irredundant representation of the implicational theory of a formal context, see [6] for details. We show that they are closely related to the minimal hypotheses:

Let $\mathbb{K}$, $\mathbb{K}_+$ and $\mathbb{K}_-$ be as above and let

$$\mathbb{K}_\pm := (G_+ \cup G_-, M \cup \{w\}, I_+ \cup I_- \cup G_+ \times \{w\})$$

be the context of the positive and the negative examples, extended by the goal attribute $w$ in the natural way. The derivation and closure operators of this

[4] For details, see [1], [3], [4].

[5] ...and a little confusing, because there seems to be no base case. But note that the empty set automatically fulfills the second condition, since it contains no proper subsets at all.

context will be denoted by superscripts $^{\pm}$ and $^{\pm\pm}$, respectively. In particular, for a given set $P \subseteq M$ of attributes we write $P^{\pm\pm}$ to denote the smallest intent of $\mathbb{K}_{\pm}$ that contains $P$. Derivation and closure operators for contexts $\mathbb{K}_{+}$ and $\mathbb{K}_{-}$ will be denoted by superscripts $^{+}$, $^{-}$ and $^{++}$, $^{--}$, respectively.

**Proposition 3.** *Let $H$ be some minimal positive hypothesis for $w$ (so in particular $w \notin H$). Then*

1. *if $P \subseteq H$ is a pseudo intent with $w \in P^{\pm\pm}$ then $P^{\pm\pm} = H \cup \{w\}$,*
2. *there is some pseudo intent $P$ of $\mathbb{K}_{\pm}$ such that $P \subseteq H$ and $P^{\pm\pm} = H \cup \{w\}$.*
3. *If $P$ is a pseudo intent such that $w \notin P$ but $w \in P^{\pm\pm}$, then $P^{\pm\pm} \setminus \{w\}$ is a positive hypothesis*[6].

*Proof.* Suppose that $H$ contains a pseudo intent $P$ with $w \in P^{\pm\pm}$. Then, since $H$ is an intent of $\mathbb{K}_{+}$, we get $P^{++} \subseteq H^{++} = H$, thus $P^{++}$ is a hypothesis for $w$. The minimality of $H$ implies $H = P^{++}$, which proves 1).

Since $H^{\pm\pm} = H \cup \{w\}$, there must be some pseudo intent $P \subseteq H$ with $w \in P^{\pm\pm}$. This gives 2).

The closure of $P$ is $P^{\pm\pm}$ and thus $P^{\pm\pm} \setminus \{w\} = P^{++}$ is closed in $\mathbb{K}_{+}$ and is contained in no negative example, since otherwise $w \notin P^{\pm\pm}$. Therefore this set is a hypothesis.

In what follows, a pseudo intent $P$ such that $w \notin P$ and $w \in P^{\pm\pm}$ will be called *($w$-)generative.*

## 3 An Example

To illustrate the notion of a minimal hypothesis and the relation of hypotheses with pseudointent-based implications we consider an expert analysis of 17 winter wheel chains (that improve behavior of a car on winter roads). Information was given by the table below from the ADAC Magazine (1999, no. 11).

The values of the **system** attribute give the type of a chain system: SK - rope chain (Seilkette), SRK - steel ring chain (Stahlringkette), SMS - quick mounting chain (Schnellmontage-System).

The **mount** attribute takes the values F and F or R to denote that a chain of particular type can be mounted either only on the front wheels or both on the front and rear wheels. The values of **price** are given in DM, the values of **con** give the average expert asessment of the conveniency of a particular type of chain; the values of **snow** give average expert asessments of the maneuverability of a car, with a particular kind of chain, on snow; **ice** means the same for ice; the values of **dur** give average expert asessments of the durability of a particular kind of chain; the values of **grade** give average expert asessments of the general quality of a particular chain type. Smaller values of attributes **con**, **snow**, **ice**, **dur**, and **grade** correspond to better asessments of the corresponding chain properties.

[6] not necessarily a minimal one

| type | system | mount | price | con | snow | ice | dur | grade |
|---|---|---|---|---|---|---|---|---|
| 1 | SK | F | 206 | 1.9 | 1.4 | 1.8 | 2.7 | 1.8 |
| 2 | SRK | F or R | 520 | 2.1 | 0.8 | 3.8 | 2.3 | 1.9 |
| 3 | SK | F | 160 | 1.7 | 1.9 | 1.6 | 3.7 | 2.1 |
| 4 | SK | F | 213 | 1.7 | 2.0 | 2.4 | 3.4 | 2.1 |
| 5 | SMS | F or R | 598 | 1.6 | 2.4 | 2.7 | 2.8 | 2.2 |
| 6 | SK | F | 109 | 2.0 | 1.9 | 2.4 | 3.7 | 2.3 |
| 7 | SRK | F or R | 325 | 2.0 | 2.1 | 3.2 | 2.8 | 2.3 |
| 8 | SMS | F or R | 498 | 1.5 | 3.3 | 3.5 | 2.0 | 2.4 |
| 9 | SRK | F or R | 396 | 2.8 | 2.1 | 3.1 | 2.5 | 2.6 |
| 10 | SRK | F or R | 325 | 2.2 | 2.2 | 4.6 | 3.2 | 2.6 |
| 11 | SRK | F or R | 389 | 2.0 | 2.2 | 3.3 | 4.3 | 2.6 |
| 12 | SRK | F | 298 | 2.5 | 2.3 | 3.3 | 2.8 | 2.6 |
| 13 | SK | F | 149 | 1.9 | 2.5 | 4.0 | 3.8 | 2.6 |
| 14 | SMS | F or R | 684 | 1.7 | 3.3 | 4.4 | 2.2 | 2.6 |
| 15 | SK | F | 99 | 2.8 | 2.2 | 2.5 | 4.0 | 2.7 |
| 16 | SK | F | 140 | 2.6 | 2.3 | 3.3 | 3.4 | 2.7 |
| 17 | SK | F | 215 | 2.3 | 3.8 | 4.8 | 2.3 | 3.1 |

Here the values of the **type** attribute substitute tradenames.

As goal attributes we considered the **grade** (obtained as an average expert asessment of quality) and the **price**. Since the information was given in numerical values, we had to scale it before obtaining hypotheses and pseudointents. Scalings in both cases were similar, except for the goal attributes.

### 3.1 Grade

First, we made an assumption that the values of **grade** less or equal to 2.1 testify to the high quality of an item and the values of **grade** greater or equal to 2.6 testify to the low quality of an item. Thus, items 1-4 were treated as positive and items 9-17 were treated as negative examples, respectively. The items with numbers 5-8 were neglected as those with ambiguous medium-value grades. Thus, the positive and negative contexts w.r.t. the goal attribute **grade** are given as follows (the values of the **grade** attribute are given in brackets to indicate that this is the goal attribute and its actual values are insignificant within a context, either positive or negative):

**Positive context**

| type | system | mount | price | con | snow | ice | dur | (grade) |
|---|---|---|---|---|---|---|---|---|
| 1 | SK | F | 206 | 1.9 | 1.4 | 1.8 | 2.7 | (1.8) |
| 2 | SRK | F or R | 520 | 2.1 | 0.8 | 3.8 | 2.3 | (1.9) |
| 3 | SK | F | 160 | 1.7 | 1.9 | 1.6 | 3.7 | (2.1) |
| 4 | SK | F | 213 | 1.7 | 2.0 | 2.4 | 3.4 | (2.1) |

**Negative context**

| type | system | mount | price | con | snow | ice | dur | (grade) |
|---|---|---|---|---|---|---|---|---|
| 9 | SRK | F or R | 396 | 2.8 | 2.1 | 3.1 | 2.5 | (2.6) |
| 10 | SRK | F or R | 325 | 2.2 | 2.2 | 4.6 | 3.2 | (2.6) |
| 11 | SRK | F or R | 389 | 2.0 | 2.2 | 3.3 | 4.3 | (2.6) |
| 12 | SRK | F | 298 | 2.5 | 2.3 | 3.3 | 2.8 | (2.6) |
| 13 | SK | F | 149 | 1.9 | 2.5 | 4.0 | 3.8 | (2.6) |
| 14 | SMS | F or R | 684 | 1.7 | 3.3 | 4.4 | 2.2 | (2.6) |
| 15 | SK | F | 99 | 2.8 | 2.2 | 2.5 | 4.0 | (2.7) |
| 16 | SK | F | 140 | 2.6 | 2.3 | 3.3 | 3.4 | (2.7) |
| 17 | SK | F | 215 | 2.3 | 3.8 | 4.8 | 2.3 | (3.1) |

The present example is not yet fully compatible with our definitions from Section 1. To obtain a formal context, we apply a conceptual scaling (see [6] for details) replacing a given many-valued attribute by one-valued ones. We only list the scale attributes. The first two scales are nominal, the other are ordinal.

| | | | | |
|---|---|---|---|---|
| **system** | SK | SRK | SMS | |
| **mount** | F | F or R | | |
| **price** | $\leq 160$ | $\leq 215$ | $\leq 520$ | $> 520$ |
| **con** | $\leq 2.1$ | $\leq 2.5$ | $> 2.5$ | |
| **snow** | $\leq 2.0$ | $> 2.0$ | | |
| **ice** | $\leq 2.4$ | $\leq 3.0$ | $\leq 4.0$ | $> 4.0$ |
| **dur** | $\leq 3$ | $\leq 3.7$ | $> 3.7$ | |

The table is read as follows. Original many-valued attributes are listed in the first column. Each many-valued attribute staying in the beginning of the row is replaced by several Boolean attributes that stay in other row positions. For example, the many-valued attribute **system** is replaced by attributes SK, SRK, and SMS, so that each object gets exactly one of them. The attribute **mount** is scaled similarly. This type of scaling is called *nominal* in [6]. The many-valued attribute **price** is replaced by four Boolean attributes $\leq 160$, $\leq 215$, $\leq 520$, and $> 520$. In contrast to the nominal attributes **system** and **mount**, the objects that have the attribute $\leq 160$, have also attributes $\leq 215$ and $\leq 520$ (in what follows, for brevity sake, we do not write this explicitly in the descriptions of object intents); the objects that have the attribute $\leq 215$ have also the attribute $\leq 520$. The other numerical attributes are scaled similarly. This type of scaling is called *ordinal* in [6].

The following positive and negative minimal hypotheses were obtained (they are unique, since the intersections of all positive and of all negative example intents are nonempty):

- the minimal positive hypothesis: {**con** $\leq$ 2.1, **snow** $\leq$ 2.0, **ice** $\leq$ 4, **dur** $\leq$ 3.7};
- the minimal negative hypothesis: {**snow** $>$ 2.0}.

The generative (w.r.t. the **grade** attribute) pseudointents of $K_\pm$ are $\{\textbf{snow} \leq 2.0\}$ and $\{\textbf{con} \leq 2.1, \textbf{ice} \leq 4, \textbf{dur} \leq 3.7\}$.

The first one $\{\textbf{snow} \leq 2.0\}$ is obviously a pseudointent of $\mathbb{K}_\pm$, since it is a nonclosed one-element subset of $M$. The second one is a pseudointent of $\mathbb{K}_\pm$, since it is not closed w.r.t. $\pm$ and all its subsets are closed.

If we introduce the "antigoal attribute" as a new attribute (in this case it corresponds to the "low quality of an item"), then the corresponding generative pseudointent will be $\{\textbf{snow} > 2.0\}$.

One can see that the positive minimal hypothesis is the closure (in the positive context with examples 1-4 and all attributes except for the general grade) of the both generative pseudointents (compare with Proposition 3). The implications

$$\{\textbf{snow} \leq 2.0\} \ \rightarrow \ \{\text{good quality}\},$$
$$\{\textbf{con} \ \leq 2.1, \ \textbf{ice} \ \leq 4, \ \textbf{dur} \ \leq 3.7\} \ \rightarrow \ \{\text{good quality}\},$$

which correspond to the generative pseudointents, coincide here with the minimal conditions for good quality. They can be used, e.g., by a producer who wants to attain the best sales at the lowest cost. For example, it suffices for a chain to make a car behave good on snow to make customer consider it as a good one. The implication

$$\{\textbf{con} \leq 2.1, \ \textbf{snow} \leq 2.0, \ \textbf{ice} \leq 4, \ \textbf{dur} \leq 3.7\} \ \rightarrow \ \{\text{good quality}\},$$

which corresponds to the minimal positive hypothesis, informs one about the whole bunch of attributes relative to the notion "good chain." This can be interpreted as a viewpoint of a customer who wants to know what is really a good chain. A good chain should be convenient, behave excellently on snow and at least satisfactorily on ice, and its life time should not be very small.

Both viewpoints are justified at their own and the consideration of both provides one with multifacetous understanding of the situation under study.

Note that for the above consideration of causes of good quality it is also reasonable to consider nonminimal hypotheses. For example, the (+)-hypothesis

$$\{\text{SK, F}, \ \textbf{price} \leq 215, \ \textbf{con} \leq 2.1, \ \textbf{snow} \leq 2.0, \ \textbf{ice} \leq 4.0, \ \textbf{dur} \leq 3.7\}$$

describes a class of relatively cheap chains, which have the same system and same mounting possibilities with good behavior on snow and satisfactory behavior on ice, that have good asessment of quality. One can also indicate (-)-hypotheses

$$\{\text{SRK, F or R}, \ \textbf{price} \leq 520, \ \textbf{snow} > 2.0\};$$
$$\{\text{SK, F}, \ \textbf{price} \leq 215, \ \textbf{snow} > 2.0\},$$

which describe different classes of chains with only low quality. These classes use different systems, have different mounting possibilities and their prices range within different intervals.

Nonminimal hypotheses can give information about taxonomy of positive and negative examples, which may be useful for the understanding of real causes of the goal attribute or its absence. If a customer identifies the quality of a good with some construction and utilization specificity, then these hypotheses can allow him to form simple consumer heuristics like the following one: "the quality of SRK chains with front or rear mounting is not high, but the quality of SK chains with front mounting can be different depending on other parameters."

### 3.2 Price

In the case of this goal attribute we took the items with costs less or equal to DM 215 to be cheap (positive examples, items 1, 3, 4, 6, 13, 15, 16, 17) and the items with costs greater or equal to DM 430 to be expensive (negative examples, items 2, 5, 8, 14). The items with numbers 7, 9, 10, 11, 12 were neglected as those with ambiguous medium-value prices. The **grade** attribute is scaled to obtain the following three ordinal attributes **grade** $\leq 2.1$, **grade** $\leq 2.5$, **grade** $> 2.5$. Other attributes were scaled in the same way as in the case of the **grade** goal attribute.

The following positive and negative minimal hypotheses were obtained (they are unique, since the intersections of all positive and of all negative example intents are nonempty):

- the minimal positive hypothesis: {SK, F}
- the minimal negative hypothesis: {$F$ or R, **con** $\leq 2.1$, **dur** $\leq 3.0$}.

There are two generative (w.r.t. the "cheap" antigoal attribute) pseudointents: {SK} and {F}.

If we introduce the "antigoal attribute" as a new attribute (in this case it corresponds to the "expensiveness"), then the corresponding three generative pseudointents are {F or R}, {**con** $\leq 2.1$}, and {**dur** $\leq 3.0$}.

As in the case with the **grade** goal attribute, each pseudointent generative w.r.t. the goal attribute **price** gives small sufficient conditions for the occurrence of the goal attribute, but minimal hypotheses give more detailed description of what does a "cheap chain" mean and what does an "expensive chain" mean.

The consideration of the hypotheses and pseudointent-based implications shows that the quality of chains (asessed by experts) is fairly independent of their price.

Among nonminimal hypotheses that may be of interest here we can indicate the negative hypothesis

$$\{\text{SMS, F or R, } \mathbf{con} \ \leq 2.1, \ \mathbf{snow} \ \geq 2.0, \ \mathbf{dur} \ \leq 3.0\},$$

which describes a class of expensive chains with a certain system, certain mounting possibilities, convenient, but with bad behavior on snow.

## 4 Algorithmic Problems

As it was shown in [5], the set of all formal concepts of a formal context can be generated in time $O(|G|^2 \cdot |M| \cdot |\underline{\mathfrak{B}}(K)|)$, where $|\underline{\mathfrak{B}}(K)|)$ is the size of the concept lattice of the context $K$, by an algorithm with polynomial delay. Recall that an algorithm for listing a family of combinatorial structures is said to have *delay d* [9] if and only if it satisfies the following conditions whenever it is run with any input of length $p$:

1. It executes at most $d(p)$ computation steps before either outputting the first structure or halting.

2. After any output it executes at most $d(p)$ machine instructions before either outputting the next structure or halting. An algorithm whose delay is bounded from above by a polynomial in the length of the input is called a *polynomial delay* algorithm[9].

In [7] the algorithm **Next Concept** from [5] was extended for the construction of the concept lattice for a given context (also with polynomial time delay). Algorithms for generating hypotheses can be found in [13]. However, to generate the set of all hypotheses, one can also adapt the **Next Concept** algorithm, running it in the bottom-up order (from least extents to least intents) in the following way. To obtain all hypotheses one should repeatedly call the following procedure, where $\mathcal{H}_+$ denotes the current set of positive hypotheses.

**Next Hypothesis**

```
A: = a hypothesis from H+
FOUND: = false;
g: = LAST_IN(G+);
while not (FOUND or PREVIOUS_IN(g, G+) = g) do
begin
        if g ∉ A then
        begin
                A := A ∪ {g};
                FOUND:= min(A++\A) > g and HYP(A+);
        end;
         A := A\{g};
         g:=PREVIOUS_IN(g, G+);
 end;
 next_hyp:=FOUND;
 H+: = H+ ∪ {A+};
```

The function HYP$(X)$ tests whether the positive intent $X$ is a hypothesis, i.e., the condition $\forall f \in G_-\ X \not\subseteq f^-$.

For $X \subseteq G_+$ the function $\min(X)$ returns the smallest (w.r.t. the ordering in $G_+$) element of $X$.

The function PREVIOUS_IN$(g, G_+)$ returns $g$ if $g$ is the first element of $G_+$, and returns the greatest element of $G_+$ smaller than $g$, otherwise.

The function LAST_IN$(G_+)$ returns the greatest element of $G_+$ (i.e., the element with the greatest number).

When the algorithm that computes the set of all hypotheses by calling **Next Hypothesis** generates a new positive intent, it needs to test whether this intent was not generated before and whether it is not contained in a negative example (line 8). Execution of line 8 takes $(|G_+| \cdot |M| + |G_-| \cdot |M|)$ time. If the positive intent is not contained in any negative example, then the process of concept generation goes further, otherwise, the intent is not a hypothesis and the algorithm backtracks. Thus, the test is executed for no more than $O(|G_+| \cdot |\mathcal{H}_+|)$ positive concepts ($\mathcal{H}_+$ is the set of all positive hypotheses) and the resulting time complexity is $O(|G_+| \cdot |G_-| \cdot |M| \cdot |\mathcal{H}_+|)$. The algorithm for finding all hypotheses has delay $O((|G_+|^2 + |G_-|) \cdot |M|)$, i.e., is a polynomial delay algorithm.

To find all minimal hypotheses, one needs to replace the line 8 with the following line 8* (and make some minor changes in the procedure that calls **Next Hypothesis**):

8*. FOUND:= $\min(A^{++} \backslash A) > g$ and $\forall f \in G_- (A^+ \not\subseteq f^-$ and MINHYP$(A^+))$;

The function MINHYP$(X)$ tests whether the hypothesis $X$ is a minimal hypothesis: first, all sets $(X \cap g^+)$, $g \in G_+ \backslash X$ maximal by inclusion are generated, then the condition $(X \cap g^+) \not\subseteq g_-^-$ is tested for each generated set and each $g_- \in G_-$.

The MINHYP test requires additional $O(|G_+|^2|M| \cdot |G_-|)$ operations for at most $|\mathcal{H}_+|$ hypothesis intents, so the resulting batch algorithm that constructs the set of all minimal hypothesis has $O(|G_+|^2|M| \cdot |G_-| \cdot |\mathcal{H}_+|)$ time complexity. The algorithm that computes all minimal hypotheses by calling **Next Hypothesis** with line 8* is not polynomial-delay. It is not clear whether it satisfies a weaker notion of efficiency, namely has cumulative polynomial delay. Recall that an algorithm listing a set of objects is said to have a *cumulative delay* $d$ [8] if it is the case that at any point of time in any execution of the algorithm with any input of length $p$ the total number of instructions that have been executed is at most $d(p)$ plus the product of $d(p)$ and the number of structures that have been output so far. The cumulative polynomial delay means that $d(p)$ is a polynomial of $p$.

Below, we give an incremental algorithm **Next Example** that modifies the list $\mathcal{H}_+^m$ of all minimal positive hypotheses if the example context $K$ is updated with a new object $g_n$.

• If $g_n$ has the goal attribute $w$, then all implications $H \to w$ of the old context remain true in the new context, but it may happen that some of $H \in \mathcal{H}_+^m$ are no longer minimal hypotheses. These must be replaced by $H \cap g_n^+$, which is clearly a minimal hypothesis.

• If $g_n$ does not have $w$, then all implications $H \to w$ of the old context with $H \subseteq g_n^-$ become false. If $H \in \mathcal{H}_+^m$ and $H \subseteq g_n^-$, then $H$ is no longer a positive hypothesis (it is called a *falsified* hypothesis), and it is deleted from $\mathcal{H}_+^m$. There may be new minimal hypotheses, which must be of the form $(H \cup \{m\})^{++}$, where $m \notin g_n^-$. These new minimal hypotheses are "most general specializations" of the old ones relative to the new negative example. We systematically generate the sets $(H \cup \{m\})^{++}$ and check if they are minimal hypotheses.

Therefore, our procedure is this:
1. For every $H \in \mathcal{H}_+^m$, $H \subseteq g_n^-$ we delete $H$ from $\mathcal{H}_+^m$.
2. For every $H \in \mathcal{H}_+^m$, $H \subseteq g_n^-$ and every $m \in M \setminus g_n^-$ we construct $F := (H \cup \{m\})^{++}$. The set $\mathcal{F}$ (of tentatively new minimal hypotheses) is updated with $F$.
3. Each element of $\mathcal{F}$ that is a superset of another element of $\mathcal{F}$ is deleted from $\mathcal{F}$.
4. The family $\mathcal{H}_+^m$ is updated with sets from $\mathcal{F}$. Finally, each element of $\mathcal{H}_+^m$ that is a supersets of $\mathcal{H}_+^m$ is deleted from $\mathcal{H}_+^m$.

Below we present a more formal pseudocode description of the algorithm. Here for two families of sets $\mathcal{X}$ and $\mathcal{Y}$ the function MERGE$(\mathcal{X}, \mathcal{Y})$ takes the union of $\mathcal{X}$ and $\mathcal{Y}$, discards every element set of it that is a superset of another set from the union (only one set is left for a pair of equal sets), and returns the resulting family of sets. Obviously, MERGE$(\mathcal{X}, \mathcal{X})$ selects all sets from $\mathcal{X}$ minimal with respect to the set-theoretic inclusion $\subseteq$.

The function NEXT_IN$(\mathcal{H}_+^m, X,$ Cond) returns "true" if there is an element of $\mathcal{H}_+^m$ that is lectically greater than $X$ and satisfies condition Cond (in this case $X$ takes the value of this element) and returns "false" otherwise. The function FIRST_IN$(\mathcal{H}_+^m, X,$ Cond) returns "true" if there is an element of $\mathcal{H}_+^m$ that satisfies condition Cond (in this case $X$ takes the value of the lectically smallest such element) and returns "false" otherwise.

**Next Example**

0 $\mathcal{F} := \emptyset$,
1 <u>if</u> $w \in g_n^{\pm}$ <u>then</u>
2     <u>if</u> FIRST_IN$(\mathcal{H}_+^m, H, \not\subseteq g_n^+)$ <u>then</u>
3         <u>repeat</u>
4             <u>if</u> for all $h \notin w^{\pm}$    $g_n^+ \cap H \not\subseteq h^+$ <u>then</u>
5                 $\mathcal{H}_+^m := \mathcal{H}_+^m \setminus \{H\} \cup \{H \cap g_n^+\}$
6         <u>until</u> not NEXT_IN$(\mathcal{H}_+^m, H, \subseteq g_n^+)$
7     <u>else</u>
8 <u>if</u> FIRST_IN$(\mathcal{H}_+^m, H, \subseteq g_n^-)$ <u>then</u>
9     <u>repeat</u>
10             $\mathcal{H}_+^m := \mathcal{H}_+^m \setminus \{H\}$
11             for all $m \notin g_n^-$ <u>do</u> $\mathcal{F} := \mathcal{F} \cup \{(H \cup \{m\})^{++}\}$
12     <u>until</u> not NEXT_IN$(\mathcal{H}_+^m, H, \subseteq g_n^-)$
13   <u>if</u> $\mathcal{F} \neq \emptyset$ <u>then</u>
14     <u>begin</u>
15         $\mathcal{F} := MERGE(\mathcal{F}, \mathcal{F})$
16         $\mathcal{H}_+^m := MERGE(\mathcal{H}_+^m, \mathcal{F})$
17     <u>end</u>

**Complexity.** If $g_n$ is a new positive example (i.e., $w \in g_n^{\pm}$), then the algorithm terminates in time $O(|\mathcal{H}_+^m| \cdot |G_-| \cdot |M|)$, since it intersects each old minimal hypothesis with $g_n^+$ and tests the inclusion of the result in the intent of each negative example (lines 1-6).

If $g_n$ is a new negative example (i.e., $w \notin g_n^{\pm}$), then in the worst case it will take much more time to terminate. The lines 9-12 are executed in the worst case for each element of $\mathcal{H}_+^m$ and for each $m \in g_n^-$ and, therefore, require $O(|\mathcal{H}_+^m| \cdot |M| \cdot |G_+| \cdot |M|)$ time, since line 11, the most time-consuming operation of the cycle 9-12, is done in $O(|G_+| \cdot |M|)$ time,

Since $\mathcal{F}$ has at most $|\mathcal{H}_+^m| \cdot |M|$ elements, the lines 15 and 16 can be executed in $O(|\mathcal{H}_+^m|^2 |M|^2)$ time each. Note that line 15 is not necessary, it can reduce the execution time in practice, but does not affect the upper bound of the worst-case complexity. Thus, the complexity of executing lines 14-17 is also $O(|\mathcal{H}_+^m|^2 |M|^2)$ and the total time complexity of the algorithm is $O(|\mathcal{H}_+^m|^2 |M|^2 + |\mathcal{H}_+^m| \cdot |M|^2 |G|)$.

Note that the worst-time complexity of the algorithm is quadratic in the size of the set of all minimal hypotheses and not of the set of all hypotheses. Therefore, when the number of all minimal hypotheses is much less than the number of all hypotheses, the incremental algorithm can operate faster than the batch algorithm, whose time complexity is linear in the number of all hypotheses.

## 5 Conclusion

We considered the JSM-method of generating hypotheses from positive and negative examples in terms of Formal Concept Analysis. We presented conditions that ensure correct behavior of hypotheses relative to examples. We showed how hypotheses and minimal hypotheses are related to pseudointent-based implications. We proposed a batch algorithm for computing all hypotheses and/or all minimal hypotheses in time linear in the number of all hypotheses. We also proposed an incremental algorithm for computing all minimal hypotheses, which runs in time quadratic in the number of minimal hypotheses. The example concerning consumer properties of wheel chains illustrated the relationship between hypotheses and implications.

### Acknowledgments

The second author was supported by the Alexander von Humboldt Foundation.

## References

1. O. M. Anshakov, V. K. Finn, and D. P. Skvortsov, On axiomatization of many-valued logics associated with the formalization of plausible reasonings, *Studia Logica*, **25**, No. 4, 23-47 (1989).
2. V. Duquenne and J.-L. Guigues, Familles minimales d'implications informatives résultant d'un tableau de données binaires, *Math. Sci. Hum.*, **24**, No. 95, 5-18 (1986).
3. V. K. Finn, On Machine-Oriented Formalization of Plausible Reasoning in the Style of F. Backon–J. S. Mill, *Semiotika i Informatika*, No. 20, 35-101, 1983 [in Russian].
4. V. K. Finn, Plausible Reasoning in Systems of JSM Type, *Itogi Nauki i Tekhniki, Seriya Informatika*, **15**, 54-101, 1991 [in Russian].

5. B. Ganter, Algorithmen zur Formalen Begriffsanalyse, in: *Beiträge zur Begriffsanalyse*, (B. Ganter, R. Wille, and K. E. Wolff, eds.), B. I. Wissenschaftsverlag, Mannheim, 241-254, 1987.
6. B. Ganter and R. Wille, Formal Concept Analysis. Mathematical Foundations, Berlin, Springer, 1999.
7. B. Ganter and S. Kuznetsov, Stepwise Construction of the Dedekind-McNeille Completion, Proc. 6th Int. Conference on Conceptual Structures, Montpellier (France), 1998, Lecture Notes in Artificial Intelligence, no. 1453, 295-302.
8. L.A. Goldberg, Efficient Algorithms for Listing Combinatorial Structures, Cambridge University Press, 1993.
9. D.S. Johnson, M.Yannakakis, and C.H. Papadimitriou, On Generating all Maximal Independent Sets, *Information Processing Letters*, **27**, 1988, 119-123.
10. S. O. Kuznetsov, JSM-method as a Machine Learning System, *Itogi Nauki i Tekhniki, ser. Informatika*, no. 15, 17-54, 1991 [in Russian].
11. S.O. Kuznetsov, Mathematical Aspects of Concept Analysis, *Journal of Mathematical Science, Series Contemporary Mathematics and Its Applications*, **18**, 1654-1698, 1996.
12. S.O. Kuznetsov and V.K. Finn, On a model of learning and classification based on similarity operation, *Obozrenie Prikladnoi i Promyshlennoi Matematiki* **3**, no. 1, 66-90, 1996 [in Russian].
13. S.O. Kuznetsov, Learning of Simple Conceptual Graphs from Positive and Negative Examples, in Proc. *Principles of Data Mining and Knowledge Discovery, Third European Conference, PKDD'99*, J. Zytkow, J. Rauch, Eds., Lecture Notes in Artificial Intelligence, **1704**, pp. 384-392, 1999.
14. H. van de Waterbeemd, Ed., *Structure-Property Correlations in Drug Research*, Landes, Austin, 1996.