On Stability of a Formal Concept

Sergei O. Kuznetsov *

Abstract: *In this paper we define and analyze stability of a formal concept. A stability index is based on the idea of a dependency in a dataset that can be reconstructed from different parts of the dataset. This idea, underlying various methods of estimating scientific hypotheses, is used here for estimating concept-based hypotheses. Combinatorial properties of stability indices, algorithmic comlpexity of their computation, as well as their dynamics with arrival of new examples are studied.*
Keywords: formal concepts, hypotheses, stability, computation.

1 Introduction

Assume that we are interested in scientific hypotheses about causes of a natural phenomenon. We make some observations which are to certain extent "random." A good hypothesis about a cause of the phenomenon should be independent of this randomness, and thus, to some extent, be independent of each particular piece of data. This sort of independence we call stability.

The idea of stability has been used to assess plausibility of hypotheses of different kinds. For example, this idea is implicitly used in the construction of extrapolation polynominals. Let $x_1, \ldots, x_n$ be points of the space R^n and a polynominal whose graph contains these points is to be constructed. Generally, it is possible to construct a polynomial of degree not greater than n that satisfies these conditions. However, if it is possible to construct a polynominal P from a certain subset of points $\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \{x_1, ..., x_n\}$ in such a way that all the points $x_1, \ldots, x_n$ lie on the curve that corresponds to the polynomial and the degree of the polynominal will be of degree not higher than k. Thus, P as a hypothesis for regularity in data given by $\{x_{i_1}, \ldots, x_{i_k}\} \subseteq \{x_1, ..., x_n\}$ is simpler and, therefore, more reliable than the hypothesis for which the polynominal can be constructed only from the whole set $\{x_1, \ldots, x_n\}$.

The idea of stability underlies certain nonparametric statistical methods. In particular, in the jackknife method the variances of arbitrary statistics (i.e., sample functions) are estimated in the following way [2]. From an initial sample of size n all possible subsamples of size $n-1$ are composed. For the i-th subsample we calculate the value S_i, of the statistic S that we wish to examine. Taking the average S_* of S_i values, we then calculate the mean of the squares of deviations S_i from S_*. The result (within insignificant arithmetic transformation) gives the estimate of the variance of the statistic S according to the jackknife method. This method can give good estimates of the variance [2]. Modifications of this method that make use of all subsamples of size $n-2$, $n-3$, etc, are also possible. However, they require a considerable amount of computation. The bootstrap method evaluates the variance of a statistic in a similar way, with the difference that new samples, also of size n, are generated from the initial sample by n-fold selection with replacement (each element of the initial sample can appear from zero to n times in the new sample).

Ideas similar to that of stability also underly research in probabilistic logics, which dates back to the work of R. Carnap on inductive logic [1]. The confirmation of a statement is evaluated by the number of universes where the statement is derivable. In this article we shall consider a realization of the idea of stability of hypotheses based on similarity of object descriptions. Hypotheses of this kind were introduced in [3] and were redefined in terms of Formal Concept Analysis in [9], [6].

2 Main Definitions

2.1 Concepts and Hypotheses

First, we recall some basic notions from Formal Concept Analysis (FCA) [14], [5]. Let G and M be sets, called the set of objects and attributes, respectively, and let I be a relation $I \subseteq G \times M$: for $g \in G$, $m \in M$,

*Institut für Algebra, Technische Universität Dresden, Germany, All-Russia Institute for Scientific and Technical Information (VINITI), Moscow, Russia. Mail: serge@viniti.ru

gIm holds iff the object g has the attribute m. The triple $K = (G, M, I)$ is called a *(formal) context.* If $A \subseteq G$, $B \subseteq M$ are arbitrary subsets, then the *Galois connection* is given by the following *derivation operators*:

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

The pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$ is called a *(formal) concept (of the context K)* with *extent* A and *intent* B (in this case we have also $A'' = A$ and $B'' = B$). For $B, D \subseteq M$ the *implication* $B \to D$ holds if $B' \subseteq D'$. Implications between subsets of objects are defined similarly.

In [9] and [6] we presented a learning model from [4] in terms of FCA. This model complies with the common paradigm of learning from positive and negative examples (see, e.g., [11]): given positive and negative examples of a *target attribute*, construct a generalization of descriptions of the positive examples that would not "cover" any negative example.

Assume that $w \notin M$ is a *target* attribute. *Positive examples* (or $(+)$-examples) are objects that are known to have the property w and *negative examples* (or $(-)$-examples) are objects that are known not to have this property. *Undetermined examples* (or (τ)-examples) are those that are neither known to have the property nor known not to have the property. The results of learning are supposed to be used for the classification of undetermined examples (in other words, for the forecast of property w for these examples).

In terms of formal concept analysis, this situation can be described by three subcontexts: a positive context $K_+ = (G_+, M, I_+)$, a negative context $K_- = (G_-, M, I_-)$, and an undetermined one $K_\tau = (G_\tau, M, I_\tau)$. The derivation operators in these contexts are denoted by $(\cdot)^+$, $(\cdot)^-$, and $(\cdot)^\tau$, respectively. A *positive hypothesis* by V.K. Finn [3], [4] (called "counterexample forbidding hypothesis" there) can be defined in terms of FCA as a nonempty intent $H \subseteq M$ of K_+ such that $|H^+| \geq 2$ and H is not contained in the intent of any negative example $g \in G_-$, i.e., for all $g \in G_-$ $H \not\subseteq g^-$. In other words, a hypothesis is an intersection of positive example intents (i.e., their "similarity") that is contained in no intent of a negative example (further, for simplicity sake we say "is (not) contained in example"). Hypotheses can be used for classification of undetermined examples from G_τ: an undetermined example is classified positively if its intent contains a positive hypothesis and does not contain a negative one. Similar for negative classification. Since we do not consider classifications here, for details we refer the reader to [9], [6].

2.2 Stability Indices: Definition and the Main Properties

Definition 1. For a context $K = (G, M, I)$ and a concept $C = (A, B)$

$$\langle C\rangle_j = \{Y \subset A \mid |Y| = j,\ Y' = B\}, \langle C\rangle_\Sigma = \bigcup_{j=2}^{n-1} \langle C\rangle_j,$$
$$\gamma_j(C) = |\langle C\rangle_j|,\ \gamma_\Sigma(C) = |\langle C\rangle_\Sigma|,\ n = |A|.$$

The summation limits are justified by the fact that the empty set and all one-element subsets of A cannot be extents of hypotheses by definition. Usually, when it is clear what concept is meant, we omit the argument C and simply write γ_j or γ_Σ. Now *stability indices* are defined as follows:

(1) *stability index* $J_j(C)$ of the jth level $(2 \leq j \leq n-1)$:

$$J_j(C) = \frac{\gamma_j(C)}{\binom{n}{j}},$$

(2) *integral stability index* $J_\Sigma(C)$:

$$J_\Sigma(C) = \frac{\gamma_\Sigma(C)}{2^n - n - 2},$$

Inverses of stability indices (i.e., values $1 - J_j$ and $1 - J_\Sigma$, respectively) can be considered as measures of "dispersion" of a concept-based hypothesis. Using the analogy with nonparametric statistics, the inverse of a stability index is related to the derivation operator $(\cdot)'$ in the same way as the sample variance calculated by the jackknife method is related to the sample average.

Note another aspect of stability of hypotheses defined above as positive intents not contained in negative examples. A positive hypothesis H with $H' = \{g_1, \ldots, g_n\} \subseteq G_+$ is the intersection of example intents (i.e., "descriptions" of examples): $g_1' \cap \ldots \cap g_n' = H$). High stability of a hypothesis means that subsets of $g_1', \ldots, g_n'$ that distinguish $g_1', \ldots, g_n'$ from each other have little in common. If stability of H is small, it is more likely that the hypothetical cause of the target attribute resides not in H, but in some $H_1, \ldots, H_q \supset H$, each of which is a "common description of some subsets of $\{g_1, \ldots, g_n\}$. This justification of stability holds for hypotheses defined by arbitrary derivation operators (or, equivalently, by semilattices of "descriptions"), such as lattices of graph sets [10], [7].

Stability indices are related to implications as follows:

Proposition 1. *Let $K = (G, M, I)$ be a context and $C = (A, B)$ be one of its concepts, $A \subseteq G$, $B \subseteq M$. Then $\langle C \rangle_\Sigma = \{D \subset A \mid |D| > 1, D \to A\}$*

Proof. By definition of $\langle C \rangle_\Sigma$, one has $\langle C \rangle_\Sigma = \{D \subset A \mid D' = B\} = \{D \subset A \mid D' = A'\}$. If $D' = A'$, then by definition of implication, one has $D \to A$. In the other direction, if $D \to A$, then by definition of implication, we have $D' \subseteq A'$. By the condition $D \subset A$ and antimonotonicity of the derivation operator $(\cdot)'$ one has $D' \supseteq A'$. Therefore, $D' = A'$ and $\langle C \rangle_\Sigma = \{D \subset A \mid D \to A\}$. $\diamond$

Corollary. *Let $K = (G, M, I)$ be a formal context, $C = (A, B)$ be a formal concept of it, and $|A| = n$. Then*

$$J_\Sigma(C) = J_2(C) = \ldots = J_{n-1}(C) = 0 \text{ iff } D \to A \text{ for no } D \subset A.$$
$$J_\Sigma(C) = J_2(C) = \ldots = J_{n-1}(C) = 1 \text{ iff } D \to A \text{ for all } D \subset A.$$

A property of stability indices given in the following proposition is related to a property of monotone Boolean functions: the relative number of units of a monotone Boolean function in the $(j+1)$th layer of the Boolean hypercube is greater than that in the jth layer. For a fixed hypothesis $H = (A, B)$, the corresponding monotone function is

$$f(Y) = \begin{cases} 1, & \text{if} \quad Y \subseteq A \quad \text{and} \quad Y' = B; \\ 0, & \text{if} \quad Y \subseteq A \quad \text{and} \quad Y' \neq B. \end{cases}$$

Proposition 2. *Let $K = (G, M, I)$ be a context and $C = (A, B)$ be a concept of K, then $J_2 \leq \ldots \leq J_{|A|-1}$.*

Proof. We consider families $\langle C \rangle_j$, $\langle C \rangle_{j+1}$, and the bipartite graph induced by the layers j and $j+1$ of the line (Hasse) diagram of the Boolean lattice $\mathbf{2}^A$. In this graph, each of $\binom{n}{j+1}$ vertices of the layer $(j+1)$ is connected with $(j+1)$ vertices of the j-th layer; each of $\binom{n}{j}$ vertices of the j-th layer is connected with $n - j$ vertices of the $(j+1)$-th layer. In this graph we isolate the vertices that correspond to the families $\langle C \rangle_j$ and $\langle C \rangle_{j+1}$. Since any superset of size $j+1$ of sets from $\langle C \rangle_j$ is a set from $\langle C \rangle_{j+1}$, the number of edges in the graph that connect vertices corresponding to sets from $\langle C \rangle_j$, with the vertices corresponding to sets from $\langle C \rangle_{j+1}$, is $\gamma_j(n-j)$. On the other hand, generally, not each subset of size j of a set from $\langle C \rangle_{j+1}$ is a set from $\langle C \rangle_j$ and hence, the number of edges is not greater than $\gamma_{j+1}(j+1)$. Thus, $\gamma_j(n-j) \leq \gamma_{j+1}(j+1)$. Since $\frac{\binom{n}{j+1}}{\binom{n}{j}} = \frac{n-j}{j+1}$, we have $\frac{\gamma_j}{\gamma_{j+1}} \leq \frac{j+1}{n-j} = \frac{\binom{n}{j}}{\binom{n}{j+1}}$ and $J_j = \frac{\gamma_j}{\binom{n}{j}} \leq \frac{\gamma_{j+1}}{\binom{n}{j+1}} = J_{j+1}$. $\diamond$

The notion of stability can be captured differently, by means of other indices. Here, we give some alternative definitions of stability indices of a concept $C = (A, B)$ with $|A| = n$:

- "Middle layer": J_j for $j = [n/2]$.
- "Average stability": J_j averaged over $j \in \{2, \ldots, n-1\}$.
- "Minimal covering": $J_s = J_j$, where $j = \min\limits_{2 \leq i \leq n-1} (J_i \neq 0)$;
- "Maximal anticovering": $J_s = J_j$, where $j = \max\limits_{2 \leq i \leq n-1} (J_i \neq 1)$.

These indices will not be considered in this paper, because the first two are similar to the integral stability, whereas the last two seem to be "biased", like level-wise indices.

3 An Example

Consider a many-valued context with results of an expert analysis of 17 winter wheel chains. This context originated from information given by a table in the ADAC Magazin (1999, no. 11). Hypotheses and implications of this context were considered in [6].

The initial dataset is given in Table 1. The attributes are given in bold face and are described as follows.

The values of the **type** attribute substitute trade-names of the chains. 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 assessment of the conveniency of a particular type of chain; the values of **snow** give average expert assessments 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 assessments of the durability of a particular kind of chain; the values of **grade** give average expert assessments of the general quality of a particular chain type. Smaller values of attributes **con**, **snow**, **ice**, **dur**, and **grade** correspond to better assessments of the corresponding chain properties.

Here, we consider **grade** (obtained as an average expert assessment of quality) as the target attribute. We make 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. So, the positive context w.r.t. the target attribute **grade** is given in Table 2, where the values of the **grade** attribute are given in brackets to indicate that this is the target attribute and its actual values are insignificant within the positive context. Negative examples can be read from Table 1.

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

Table 1

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)

Table 2

We scale the original table (i.e., convert many-valued attributes to binary ones) in the way shown in Table 3.

system	SK	SRK	SMS	
mount	F	F or R		
price	≤ 160	≤ 215	≤ 500	> 500
con	≤ 2.1	≤ 2.5	> 2.5	
snow	≤ 2.0	> 2.0		
ice	≤ 2.4	≤ 3.0	≤ 4.0	> 4.0
dur	≤ 3.0	≤ 3.7	> 3.7	

Table 3

Table 3 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 scaled *nominally* [5]: it is replaced by Boolean attributes SK, SRK, and SMS, so that each object has exactly one of them. The attribute **mount** is also scaled nominally. The many-valued attribute **price** is scaled *ordinally* [5]: it is replaced by four Boolean attributes ≤ 160, ≤ 215, ≤ 500, and > 500. In contrast to the nominal attributes **system** and **mount**, the objects that have the attribute ≤ 160, have also attributes ≤ 215 and ≤ 500 (in what follows, for brevity sake, we do not write this explicitly in the descriptions of object intents); the objects that have the attribute ≤ 215 have also the attribute ≤ 500. The other numerical attributes are also scaled ordinally.

We have the following (unique) minimal positive hypothesis:

$$\{\textbf{con} \ \leq 2.1, \textbf{snow} \ \leq 2.0, \textbf{ice} \ \leq 4, \textbf{dur} \ \leq 3.7\}.$$

The corresponding extent is $\{1, 2, 3, 4\}$.

Other positive hypotheses are

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

with extent $\{1, 3, 4\}$ and

$$\textbf{con} \leq 2.1,\ \textbf{snow} \leq 2.0,\ \textbf{ice} \leq 4.0,\ \textbf{dur} \leq 3.0\}$$

with extent $\{1, 2\}$, and hypotheses corresponding to intents of positive examples, (i.e., with extents $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$).

Thus we have only two hypotheses with extents > 2, namely that with extent $\{1, 2, 3, 4\}$ (the minimal hypothesis) and that with extent $\{1, 3, 4\}$.

Consider stability indices of the minimal hypothesis. Since

$$\{1, 2, 3, 4\}' = \{1, 2, 3\}' = \{1, 2, 4\}' = \{2, 3, 4\}' = \{2, 3\}' = \{2, 4\}',$$

and there is no other subset of $\{1, 2, 3, 4\}$ giving the same intent, we have

$\gamma_3(\{1, 2, 3, 4\}, \{1, 2, 3, 4\}') = |\{\{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 4\}\}| = 3,$

$J_3 = \frac{3}{\binom{4}{3}} = \frac{3}{4};$

$\gamma_2(\{1, 2, 3, 4\}, \{1, 2, 3, 4\}') = |\{\{2, 3\}, \{2, 4\}| = 2, J_2 = \frac{2}{\binom{4}{2}} = \frac{1}{3};$

$\gamma_\Sigma(\{1, 2, 3, 4\}) = |\{\{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 4\}, \{2, 3\}, \{2, 4\}\}|,$

$J_\Sigma(\{1, 2, 3, 4\}, \{1, 2, 3, 4\}') = \frac{5}{\binom{4}{3} + \binom{4}{2}} = \frac{1}{2}.$

For the hypothesis with extent $\{1, 3, 4\}$ we have

$\gamma_2(\{1, 3, 4\}, \{1, 3, 4\}') = \gamma_\Sigma(\{1, 3, 4\}, \{1, 3, 4\}') =$

$|\{\{1, 3\}, \{1, 4\}, \{3, 4\}| = 3,$

$J_2(\{1, 3, 4\}, \{1, 3, 4\}') = J_\Sigma(\{1, 3, 4\}, \{1, 3, 4\}') = \frac{3}{\binom{3}{2}} = 1.$

Thus, the (minimal) hypothesis with extent $\{1, 2, 3, 4\}$ has integral stability index smaller than that of the hypothesis with extent $\{1, 3, 4\}$. This illustrates the interpretation of stability indices: hypothesis with extent $\{1, 2, 3, 4\}$ could not have been obtained without example 2, whereas the hypothesis with extent $\{1, 3, 4\}$ can be obtained when any single example is missing. The set $\{1, 2, 3, 4\}$ has a subset of similar examples, whose similarity is more important as a cause of the target attribute than the similarity of the whole set.

4 Stability Dynamics with the Growth of the Set of Examples

In this section we study the possible behavior of stability indices when a dataset is updated with new examples. We assume that intent of a new example is a subset of the "old" attribute set M. When it does not lead to confusions, the derivation operator in the new context is also denoted by $(\cdot)'$. For any positive hypothesis H a new negative example with intent containing H "falsifies" H (with arrival of such an example, H is no more a hypothesis by definition), other new negative examples do not affect stability of H at all. Therefore, to study behavior of stability indices of a positive hypothesis with the growth of the dataset, we can consider only the positive context $K_+ = (G_+, M, I_+)$. Since negative context is ignored, for simplicity sake, we omit subscript "+" and speak of a context $K = (G, M, I)$. A new positive example g can affect stability of hypothesis H, which corresponds to intent B of a concept $C = (A, B)$ if the intent of g is a subset of B. In this case we say that g *confirms* H (or confirms B and C).

Upon arrival of k new examples stability indices of a concept $C = (A, B)$ are denoted by J_j^k, $j \in \{2, \ldots, |A| + k - 1\}$ and J_Σ^k, respectively. For the sake of convenience, we will also set $|A| = n$, $J_n = 1$, $J_j = 0 \quad \text{for} \quad j \in Z \setminus \{2, \ldots, n\}$.

Theorem 3. *Let $K = (G, M, I)$ be a context and $C = (A, B)$ be a concept of this context. Let the set G be updated with k new objects, then the stability indices of the concept C lie within the following bounds:*

$$\underline{J}_j^k \leq J_j^k \leq \overline{J}_j^k,$$
$$\underline{J}_\Sigma^k \leq J_\Sigma^k \leq \overline{J}_\Sigma^k,$$

where

$$\underline{J}_j^k = \frac{1}{\binom{n+k}{j}} \left(\gamma_j + \binom{k}{1} \gamma_{j-1} + \ldots + \binom{k}{k-1} \gamma_{j-k+1} + \gamma_{j-k} \right),$$

$$\overline{J}_j^k = \frac{1}{\binom{n+k}{j}} \left(\gamma_j + \binom{n}{j-1} + \ldots + \binom{n+k-1}{j-1} \right),$$

$$\underline{J}_\Sigma^k = \frac{2^k \cdot \gamma_\Sigma + 2^k - 1}{2^{n+k} - (n+k+2)},$$

$$\overline{J}_\Sigma^k = \frac{\gamma_\Sigma + 2^n (n^k - 1) - k}{2^{n+k} - (n+k+2)}.$$

Proof. Though we consider changing contexts, we shall use notation $'$ to denote the operation of taking intents of examples, both new and old. This slight misuse of notation will not result in any trouble, since the set of attributes M does not change. We set $A = \{g_1, \ldots, g_n\}$.

(1) The lower bounds for indices J_j^k. The set of new examples will be denoted by $\{\underline{g}_1, \ldots, \underline{g}_k\}$. For a given initial value of the index J_j, the value J_j^k is minimal if any subset of the set of k new confirming examples yields B in the intersection of only those sets of examples from A, which themselves give B in the intersection. In more rigorous terms, $J_j^k = \frac{\gamma_j^k}{\binom{n+k}{j}}$ takes the minimal value if for any $p, q \quad 1 \leq p < k, \quad 1 < q < n, \quad \underline{g}'_{i_1} \cap \ldots \sqcap \underline{g}'_{i_p} \cap (g'_{t_1} \cap \ldots \cap g'_{t_q}) = B$ if and only if $g'_{t_1} \cap \ldots \cap g'_{t_q} = B$. Here, $\{\underline{g}_{i_1}, \ldots, \underline{g}_{i_p}\} \subseteq \{\underline{g}_1, \ldots, \underline{g}_k\}$, $\{g_{t_1}, \ldots, g_{t_q}\} \subseteq \{g_1, \ldots, g_n\} = A$.

Let us consider the terms that make up the value of γ_j^k. The first term, i.e., γ_j corresponds to the cardinality j of subsets of the set of initial examples $\{g_1, \ldots, g_n\}$. The other terms appear as $\gamma_{j-s}\binom{k}{s}$. They are obtained by virtue of the fact that an arbitrary set $\{g_{i_1}, \ldots, g_{i_{j-s}}\}$ from $\langle C \rangle_{j-s}$ can be supplemented with any new confirming examples $\underline{g}_{r_1}, \ldots, \underline{g}_{r_s}$ to obtain a set $\{\underline{g}_{i_1}, \ldots, \underline{g}_{i_{j-s}}, \underline{g}_{r_1}, \ldots, g_{r_s}\}$, which also give B in the intersection of all its intents $g'_{i_1} \cap \ldots \cap g'_{i_{j-s}} \cap \underline{g}'_{r_1} \cap \ldots \cap \underline{g}'_{r,s} = B$. Therefore, $\gamma_j^k = \sum_{s=0}^{k} \gamma_{j-s} \cdot \binom{k}{s}$ and the lower estimate for J_j^k has been proved.

(2) The lower bound for the index J_Σ^k, i.e., $\underline{J}_\Sigma^k$. When a new confirming example is received, each value of γ_i^1, $\;2 \le i \le n-1$ becomes not less than the value indicated in 1). The entire sequence of values of γ_i^1 is $\gamma_2, \;\; \gamma_2+\gamma_3, \ldots, \gamma_{n-2}+\gamma_{n-1}, \;\; \gamma_{n-1}+1$. Their sum is $\gamma_\Sigma^1 = \gamma_2+(\gamma_2+\gamma_3)+\ldots+(\gamma_{n-2}+\gamma_{n-1})+(\gamma_{n-1}+1) = 2\gamma_\Sigma+1$. Therefore, after receiving k new confirming examples, the value of γ_Σ is not less than $2^k \cdot \gamma_\Sigma + 2^k - 1$ (i.e., $\gamma_\Sigma^k \le 2^k \cdot \gamma_\Sigma + 2^k - 1$) and

$$J_\Sigma^k \ge \frac{2^k \cdot \gamma_\Sigma + 2^k - 1}{2^{n+k} - (n+k+2)}.$$

(3) The upper bounds for the indices J_j^k, i.e., $\overline{J}_j^k$, for $j \in \{2, \ldots, n+k-1\}$ are obtained from the analysis of the sequence of new examples $\overline{g}_1, \ldots, \overline{g}_k$ that confirm the hypothesis H and are of the following form: $\overline{g}_i$ in the intersection with any previous example confirming H yields B. In this case, the intersection of any set of new examples will give B. A sequence of this type will give the upper bound of J_j^k, since we cannot affect "old" intersections (obtained by intersecting old examples) and all "new" intersections will coincide with B. More precisely, $\overline{g}_i \cap g = B$ for $g \in \{g_1, \ldots, g_n, \overline{g}_1, \ldots, \overline{g}_{i-1}\}$. In this case, for $k = 1, \gamma_j^1$ is the sum of the number of old examples γ_j and the new examples formed from $j-1$ old examples and a single new example: $\gamma_j^1 = \gamma_j + \binom{n}{j-1}$. Suppose that for $k = t, \quad \gamma_j^t = \gamma_j + \binom{n}{j-1} + \ldots + \binom{n+t-1}{j-1}$. In this case, for $k = t+1$, γ_j^{t+1} is the sum of the number of old examples γ_j^t and new examples formed from $j-1$ old examples and a single new one:

$$\gamma_j^{t+1} = \gamma_j^t + \binom{n+t}{j-1} = \gamma_j + \binom{n}{j-1} + \ldots$$

$$\ldots + \binom{n+t-1}{j-1} + \binom{n+t}{j-1}.$$

(4) The upper bounds for the index J_Σ, i.e., $\overline{J}_\Sigma^k$. The initial values of γ_i were $\gamma_2, \ldots, \gamma_{s-1}, + \quad \gamma_s, \gamma_{s+1}, \ldots, \gamma_{n-2}, \gamma_{n-1}$; after the arrival of g^1, these values are expressed by $\gamma_2 + \binom{n}{1}, \ldots, \gamma_s + \binom{n}{s-1}, \ldots, \gamma_{n-1} + \binom{n}{n-2}, \gamma_n + \binom{n}{n-1}$, respectively, where $\gamma_n = 1$. In this case,

$$\gamma_\Sigma^1 = \gamma_2 + \ldots + \gamma_{n-1} + \binom{n}{1} + \ldots$$

$$\ldots + \binom{n}{s-1} + \ldots + \binom{n}{n-1} + 1 = \gamma_\Sigma + 2^n - 1.$$

Hence $\gamma_\Sigma^k = \gamma_\Sigma + 2^{n+k-1} + \ldots + 2^n - k = \gamma_\Sigma + 2^n(2^k - 1) - k$ and

$$\overline{J}_\Sigma^k = \frac{\gamma_\Sigma + 2^n(2^k - 1) - k}{2^{n+k} - (n+k+2)}.$$

◇

Theorem 3 allows us to consider the limits of the upper and lower bounds of stability indices with the growth of k. The upper bound of stability indices behave uniformly: they increase monotonically, and, in the limit, tend to 1. The lower bounds of the level indices behave differently: for the indices of the upper levels, they approach 1 in the limit; for the lower levels, they tend to 0. The behaviour in the limit of the lower bounds of the middle level indices remains unclear. The limit of the lower bound of the integral stability index is strictly greater than zero and smaller than one. Since $\lim_{k\to\infty} \underline{J}_\Sigma^k(k) = \frac{\gamma_\Sigma+1}{2^n} > 0$, with arrival of new examples, it is more likely that the value of J_Σ will increase, because there is practically "no room" for its further decrease. This observation suggests the existence of "soft dependence" between J_Σ and the size of extent: under natural assumptions about distributions of attributes with example intents it is likely that a concept with large extent has stability greater than that of a concept with small extent.

Hence, it is natural to compare stability indices with support of association rules in data mining, which have the form $X \to Y$, $X, Y \subseteq M$ (see, e.g., [12]). Support is defined as $supp(X \to Y) = \frac{|(X \cup Y)'|}{|G|}$, i.e., as the relative number of examples whose description includes both the antecedent and the consequent of the association rule. The first difference between stability indices and support is that the former are computed relative not to the set of all examples, but to all possible subsets of a concept extent. The second difference is that stability indices measure support of an *exactly* specified subset of M (i.e., the intent), not of any superset of it.

5 Algorithmic Complexity of Computing Stability Indices

Unfortunately, exact computation of stability indices in general case is intractable, as established by the following

Theorem 4. *Let $K = (G, M, I)$ be a context and $C = (A, B)$ be one of its concepts. Then the problem of determining the stability index $J_\Sigma(C)$ as well as the problem of determining stability index J_j for arbitrary $2 \leq j \leq |A|$ is #P-complete.*

Proof. We introduce the following auxiliary problems.
(1) The problem of the number of vertex coverings (NVC).
INPUT. Graph $\Gamma = (V, E)$.
OUTPUT. The number of vertex coverings, i.e., $\#\{V' \subseteq V |$ if $(u, v) \in E$, then $u \in V'$ or $v \in V'\}$.
(2) The problem of the number of implicants (NI).
INPUT. Monotone 2-CNF, i.e., the formula $F = C_1 \wedge \ldots \wedge C_r$, where

$$C_i = (x_{i_1} \vee x_{i_2}), \quad x_{i_s} \in X = \{x_1, \ldots, x_n\}.$$

OUTPUT. The number of implicants, i.e., $\#\{Y | Y \subseteq X, \wedge_{x_j \in Y} x_j \to F\}$.
(3) Problem of the number of subfamilies with fixed intersection (NSFI).
INPUT. A finite set $\mathcal{U}$ and $\mathcal{X} \subset 2^{\mathcal{U}}$, a family of different sets $\mathcal{X} = \{X_1, \ldots, X_k\}$, where $X_1 \cap \ldots \cap X_k = h$.
OUTPUT. The number of subfamilies $\mathcal{X}'$ of the family $\mathcal{X}$ such that the intersection of all members of the subfamily $\mathcal{X}$ is h, i.e.,

$$\{\mathcal{X}' \subseteq \mathcal{X} | \mathcal{X}' = \{X_{i_1}, \ldots, X_{i_s}\} \quad \text{and} \quad X_{i_1} \cap \ldots \cap X_{i_s} = h\}.$$

Lemma 5. *The NI problem is #P-complete.*

Proof. We will reduce to the NI problem the following one: "the number of 0-1 n-tuples that satisfy 2-CNF $F = C_1 \wedge C_2 \wedge \ldots \wedge C_s$, where $C_i = (y_{i_1} \vee y_{i_2})$ and $y_{i_j} \in X$". The $\#P$-completeness of this problem has been proved in [13]. Suppose that $|X| = n$ and $A = (a_1, \ldots, a_n)$ is a 0-1 tuple that satisfies F. Let $\{j_1, \ldots, j_k\}$ be serial numbers of single components of the 0-1 tuple A. We form a conjunction $Y_j = y_{j_1} \wedge \ldots \wedge y_{j_k}$, where $y_{j_i} \in X$. Obviously, Y_j is the implicant of 2-CNF F. Conversely, each implicant $Y_m = y_{m_1} \wedge \ldots \wedge y_{m_s}$ of the formula F has a corresponding 0-1 tuple $A^m = (a_1^m, \ldots, a_n^m)$, where positions $m_1, \ldots, m_s$ are one-positions and the remaining positions are zero-positions; the tuple A^m satisfies F. The reducibility is accomplished. Since the reducibility is polynomial in s and n, the lemma is proved. $\diamondsuit$

Lemma 6. *The NVC problem is #P-complete.*

Proof. The membership of the NVC problem in the class $\#P$ is obvious, since every solution to the corresponding decision problem ("is there a vertex covering") is tested in polynomial time. We will prove the $\#P$-completeness of the problem by reducing to it $\#P$-complete problem NI. As in [?], from an arbitrary 2-CNF F we construct a graph $\Gamma = (V, E)$, where $V = \{u_1, \ldots, u_n\}$ (each u_i corresponds to a variable x_i from F) and $E = \{(u_i, u_j) | (x_i \vee x_j)$ is included in the conjunction$\}$. Any vertex covering of Γ corresponds to the implicant of F, and conversely; any implicant of F corresponds to a certain vertex covering of G. The reducibility is implemented within a time linear with respect to the size of F. $\diamondsuit$

Lemma 7. *The NSFI problem is #P-complete.*

Proof. We will reduce by Turing the NVC problem to a special case of the NSFI problem (when $h = \emptyset$). Suppose that we have an arbitrary graph $\Gamma = (V, E)$ with no isolated vertices (which does not impair the generality), where $V = \{a_1, \ldots, a_n\}$, $E \subseteq V \times V$. For a vertex v by $N(v)$ we denote the set of edges from E that are incident to the vertex v; $\overline{N}_E(v) = E \setminus N(v)$. For $v_i, v_j \in V$, we have $N(v_i) = N(v_j) \neq \emptyset$ if and only if $N(v_i) = N(v_j) = \{(v_i, v_j)\}$, i.e., v_i and v_j are not connected to any other vertices from $E\{v_i, v_j\}$. In this case the edge (v_i, v_j) will be called isolated. We will calculate the NVC of the graph as follows: Suppose that there are k isolated edges in the graph Γ. The vertex covering of these edges can be executed in 3^k ways (three ways per each edge: it can be covered by either of the two vertices or by both). If the remaining edges of the graph Γ can be covered in d various ways, then all the edges of the graph have $d \cdot 3^k$ coverings. We have only to determine the number of vertex coverings of non-isolated edges of the graph (we will denote this set by E_1). Suppose that edges from E_1 cover vertices from a set $V_1 \subseteq V$. Now, for any two different vertices $v_i, v_j \in V$, we have $N(v_i) \neq N(v_j)$. By definition, the vertices $v_1, \ldots, v_r \in V_1$ form a covering of E if and only if $N(v_1) \cup \ldots \cup N(v_r) = E_1$ or, by de Morgan's law, $\overline{N}_{E_1}(v_1) \cap \ldots \cap \overline{N}_{E_1}(v_r) = \emptyset$. Thus, the set $\{v_1, \ldots, v_r\}$ forms the vertex covering of the graph $\Gamma = (V_1, E_1)$ if and only if the intersection of all sets of the family $N_{E_1}(v_1), \ldots, N_{E_1}(v_r)$ is empty. We have thus reduced the NVC problem for $\Gamma = (V, E)$ to NSFI problem with $\mathcal{X} = \{\overline{N}_{E'}(a_1), \ldots, \overline{N}_{E_1}(a_n)\}$. The reducibility is polynomial because the size of the set $\mathcal{X}$ (i.e., the total number of edges in $\mathcal{X}$) is not greater than $n\left(\binom{n}{2} - (n-1)\right)$, i.e., $O(n^3)$. $\Diamond$

Theorem 4 is a simple corollary of Lemma 7. Under conditions where the hypothesis H is not contradictory and the set of all examples that give rise to it is $\mathcal{X}$, NSFI is $\gamma_\Sigma + 1$.

Corollary. *The problem of determining $J_j(C)$ of a given concept $C = (A, B)$ for an arbitrary j: $2 \leq j \leq n-1$ is #P-complete.*

Having these intractability results, we describe an algorithm for computing stability, which can be considered optimal in the sense that its time complexity is linear in $\langle C \rangle_\Sigma$ and polynomial in the size of the context. To describe an algorithm for computing stability index $J_\Sigma(A, A')$ of a concept (A, A') we need to introduce the following notions and functions. Let elements of G be in one-to-one correspondence with natural numbers from 1 to $|G|$. Then each subset of $A \subseteq G$ is represented by an (increasingly ordered) tuple of numbers of its elements. Let $\leq$ be the following order on tuples of this kind: $X \leq Y$ iff $min(X \setminus Y) \in X$)subsets of G. Let $<$ be the corresponding strict order and $\prec$ be a precedence relation (for $X, Y \subseteq A$ $X \prec Y$ iff $X \leq Y$ and there is no $Z \subseteq A$: $X < Z < Y$). The functions $max(X)$ and $min(X)$ return the maximal and minimal element of the set X, respectively. The function *next* is defined as follows: if $X = max(A)$, then $next(X) := 0$, otherwise $next(X) := Y$, where $X \prec Y$. The variable *count* counts $\gamma_\Sigma(C)$, i.e., the number of elements in $\langle C \rangle_\Sigma$. The function $tail(X)$ is defined as $tail(X) := \{max(X), max(X)+1, \ldots, |G|\} \cap A$.

Algorithm `Stability Count`
Input: a context (G, M, I) and a concept (A, A') of it.
Output: Stability index $\gamma_\Sigma(A, A')$.
Initialization
$X{:=}min(A)$, $count := 0$;
1. **until** $next(X) = 0$
2. **begin**
3. **if** $(A \setminus X)' \neq A'$
4. **then** $X := next(X \cup\ tail(X))$ **else do**
5. **begin**
6. $count{:} = count + 1$
7. $X{:=}next(X)$
8. **end**
9. **end**
10. **return** count

Theorem 8. *Let $K = (G, M, I)$ be a formal context and $C = (A, B)$ be a formal concept of the context K. Then the integral stability index $J_\Sigma(C) = \frac{\gamma_\Sigma(C)}{2^n - n - 2}$ and a stability index $J_j(C) = \frac{\gamma_j(C)}{\binom{n}{j}}$, $1 \leq k \leq n-1$ are computed by* **Stability Count** *in time $O(|G|^3|M|\gamma_\Sigma)$ and $O(|G|^3|M|\gamma_j)$, respectively.*

Theorem 8, together with the $\#P$-completeness of the problems of computing stability indices (Theorem 4) indicates that this algorithm is optimal within a factor polynomial in the input (namely, $O(|M| \cdot |G|^2)$), see [?].

The following result concerning upper and lower bounds of the integral J_Σ stability index provides some means for approximate computation of the indices within $O(|M| \cdot |A|^{k+r})$ time, where k and r are arbitrary preset integers, $k, r < n$.

Theorem 9. *The following inequalities hold for the integral and average stability indices of a concept $C = (A, B)$ (where $n = |A|$, $2 \leq k, r \leq n-1$):*

$$\frac{\gamma_1 + \ldots + \gamma_k}{\binom{n}{2} + \ldots + \binom{n}{k}} \leq J_\Sigma \leq \frac{\gamma_{n-r} + \ldots + \gamma_{n-1}}{\binom{n}{n-r} + \ldots + \binom{n}{n-1}}.$$

Proof. To prove the theorem, one uses Proposition 2: $J_2 \leq \ldots \leq J_{|A|-1}$ and the following

Lemma 10. *For arbitrary sequences (a_i), (b_i) such that $a_i \geq 0$, $b_i > 0$ and $\frac{a_i}{b_i} \leq \frac{a_{i+1}}{b_{i+1}}$, and for an arbitrary $s \geq 2$, we have*

$$\frac{a_1 + \ldots + a_{s-1}}{b_1 + \ldots + b_{s-1}} \leq \frac{a_1 + \ldots + a_s}{b_1 + \ldots + b_s} \leq \frac{a_2 + \ldots + a_s}{b_2 + \ldots + b_s} \leq \frac{a_s}{b_s}.$$

Proof. The proof of the first inequality will be done by induction over s.

For $s = 2$ we have $\frac{a_1}{b_1} \leq \frac{a_2}{b_2}$. In this case, $\frac{a_1}{b_1} \leq \frac{a_1+a_2}{b_1+b_2} \leq \frac{a_2}{b_2}$. Indeed,

$$\frac{a_1 + a_2}{b_1 + b_2} - \frac{a_1}{b_1} = \frac{a_1 b_1 + a_2 b_1 - a_1 b_1 - a_1 b_2}{b_1(b_1 + b_2)} \geq 0.$$

$$\frac{a_1 + a_2}{b_1 + b_2} - \frac{a_2}{b_2} = \frac{a_1 b_2 + a_2 b_2 - a_2 b_1 - a_2 b_2}{b_2(b_1 + b_2)} \leq 0.$$

Suppose that for $s < m$ the statement has been proved and $a_1 + \ldots + a_{m-1} = \alpha$, $b_1 + \ldots + b_{m-1} = \beta$. In that case,

$$\frac{a_1 + \ldots + a_m}{b_1 + \ldots + b_m} - \frac{a_1 + \ldots + a_{m-1}}{b_1 + \ldots + b_{m-1}} = \frac{\alpha + a_m}{\beta + b_m} - \frac{\alpha}{\beta} =$$

$$= \frac{\alpha\beta + a_m\beta - \beta\alpha - b_m\alpha}{\beta(\beta + b_m)} = \frac{a_m\beta - b_m\alpha}{\beta(\beta + b_m)} = \Delta.$$

By the inductive hypothesis we have $\frac{a_{m-1}}{b_{m-1}} \geq \frac{\alpha}{\beta}$ and by the conditions of the lemma we have $\frac{a_m}{b_m} \geq \frac{a_{m-1}}{b_{m-1}}$. Hence $\frac{a_m}{b_m} \geq \frac{a_{m-1}}{b_{m-1}} \geq \frac{\alpha}{\beta}$, and the numerator of Δ is nonnegative. Other two inequalities of the lemma are proved similarly. $\diamond$

Stability indices were applied in the analysis of defects of polymers at "NPO PLASTMASSY" in Moscow [8], [9]. The results of the analysis show that all hypotheses about causes of defects that were accepted by experts had stability indices greater than average values.

6 Conclusion

We proposed a definition of stability of a formal concept, which can be used as a plausibility measure of concept-based hypotheses. Stability measures independence of hypotheses on particular pieces of data that can be random. We showed the interrelation of stability indices, as well as relation of the latter to implications between attributes. The indices were illustrated by an example with a real dataset. The behavior of indices with the increase of the number of examples was studied. Complexity of computing stability indices was studied. On the one hand, the problem of computing integral stability index was proved to be #P-complete. On the other hand, an algorithm that we proposed for the solution of this problem is worst-case optimal modulo a polynomial in the input. We also proposed a polynomial algorithm for computing approximations of the integral index. As further research of stability indices we intend to study their relation to different interestingness and plausibility measures in data mining.

Acknowledgments

The author thanks Oleg Gorbachev and Alexander Kogan for helpful discussions.

This work was supported by the Alexander von Humboldt Foundation and Russian Foundation for Basic Research, project no. 02-01-00537a.

References

[1] R. Carnap, *The Logical Foundations of Probability*, University of Chicago Press, Chicago, 1962.

[2] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, 38, SIAM, Philadelphia, 1982.

[3] V.K. Finn, *On Machine-Oriented Formalization of Plausible Reasoning in the Style of F. Bacon–J. S. Mill*, Semiotika Informatika, **20**, 35-101, 1983 [in Russian].

[4] V.K. Finn, *Plausible Reasoning in Systems of JSM-Type*, Itogi Nauki i Tekhniki, ser. Informatika, **15**, 54-101, 1991 [in Russian].

[5] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Springer, 1999.

[6] B. Ganter and S.O. Kuznetsov, *Formalizing Hypotheses with Concepts*, Proc. 8th Int. Conf. on Conceptual Structures, ICCS'00, G. Mineau and B. Ganter, Eds., Lecture Notes in Artificial Intelligence, **1867**, 342-356, 2000.

[7] B. Ganter and S.O. Kuznetsov, Pattern Structures and Their Projections, *Proc. 9th Int. Conf. on Conceptual Structures, ICCS'01*, G. Stumme and H. Delugach, Eds., Lecture Notes in Artificial Intelligence, **2120** 2001, pp. 129-142.

[8] S.O. Kuznetsov, *Stability as an Estimate of the Degree of Substantiation of Hypotheses Derived on the Basis of Operational Similarity*, Nauchn. Tekh. Inf., Ser. 2 no. 12, 21-29, 1990 [in Russian].

[9] S.O. Kuznetsov, *Mathematical Aspects of Concept Analysis*, J. Math. Sci., **80**, no. 2, 1654-1698, 1996.

[10] S.O. Kuznetsov, Learning of Simple Conceptual Graphs from Positive and Negative Examples. In: J. Zytkow, J. Rauch (eds.), Proc. *Principles of Data Mining and Knowledge Discovery, Third European Conference, PKDD'99*, Lecture Notes in Artificial Intelligence, **1704**, pp. 384-392, 1999.

[11] T. Mitchell, *Machine Learning*, Mc Graw Hill, 1997.

[12] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, *Efficient Minining of Association Rules Using Closed Itemset Lattices*, J. Inf. Systems, **24**, 25-46, 1999.

[13] L.G. Valiant, *The Complexity of Enumeration and Reliability Problems*, SIAM J. Comput., **8**, no. 3, 410-421, 1979.

[14] R. Wille, *Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts*, Ordered Sets, I. Rival, Ed., Reidel, Dordrecht–Boston, 445-470, 1982.