



Attribute Selection Using Contranominal Scales

Dominik Dürschnabel^{1,2} , Maren Koyda^{1,2} , and Gerd Stumme^{1,2} 

¹ Knowledge and Data Engineering Group, University of Kassel, Kassel, Germany
{duerrschnabel,koyda,stumme}@cs.uni-kassel.de

² Interdisciplinary Research Center for Information System Design,
University of Kassel, Kassel, Germany

Abstract. Formal Concept Analysis (FCA) allows to analyze binary data by deriving concepts and ordering them in lattices. One of the main goals of FCA is to enable humans to comprehend the information that is encapsulated in the data; however, the large size of concept lattices is a limiting factor for the feasibility of understanding the underlying structural properties. The size of such a lattice depends on the number of subcontexts in the corresponding formal context that are isomorphic to a contranominal scale of high dimension. In this work, we propose the algorithm **ContraFinder** that enables the computation of all contranominal scales of a given formal context. Leveraging this algorithm, we introduce δ -**adjusting**, a novel approach in order to decrease the number of contranominal scales in a formal context by the selection of an appropriate attribute subset. We demonstrate that δ -**adjusting** a context reduces the size of the hereby emerging sub-semilattice and that the implication set is restricted to meaningful implications. This is evaluated with respect to its associated knowledge by means of a classification task. Hence, our proposed technique strongly improves understandability while preserving important conceptual structures.

Keywords: Formal Concept Analysis · Contranominal scales · Concept lattices · Attribute selection · Feature selection · Implications

1 Introduction

One of the main objectives of Formal Concept Analysis (FCA) is to present data in a comprehensible way. For this, the data is clustered into concepts which are then ordered in a lattice structure. Relationships between the features are represented as implications. However, the complexity of the corresponding concept lattice can increase exponentially in the size of the input data. Beyond that, the size of the implication set is also exponential in the worst case, even when it is restricted to a minimal base. As humans tend to comprehend connections in smaller chunks of data, the understandability is decreased by this exponential nature even in medium sized datasets. That is why reducing large and complex data to meaningful substructures by eliminating redundant information

enhances the application of Formal Concept Analysis. Nested line diagrams [26] and drawing algorithms [8] can improve the readability of concept lattices by optimizing their presentation. However, neither of them compresses the size of the datasets and thus grasping relationships in large concept lattices remains hard. Therefore, our research question is: **How can one reduce the lattice size as much as possible by reducing the data as little as possible?** There are different ways of reducing the data. In this paper, we focus on the removal of attributes. The size of the concept lattice is heavily influenced by the number of its Boolean suborders. A lattice contains such an k -dimensional Boolean suborder if and only if the corresponding formal context contains an k -dimensional contranominal scale [1, 16]. Thus, to reduce the size of the concept lattice it is reasonable to eliminate those. However, deciding on the largest contranominal scale of a formal context is an \mathcal{NP} -complete problem. Therefore, choosing sensible substructures of formal contexts which can be augmented in order to reduce the number of large contranominal scales is a challenging task.

In this work, we propose the algorithm **ContraFinder** that is more efficient than prior approaches in computing all contranominal scales in real world datasets. This enables us to present our novel approach δ -**adjusting** which focuses on the selection of an appropriate attribute subset of a formal context. To this end, we measure the influence of each attribute with respect to the number of contranominal scales. Hereby, a sub-semilattice is computed that preserves the meet-operation. This provides the advantage to not only maintain all implications between the selected attributes but also does not produce false implications and thus retains underlying structure. We conduct experiments to demonstrate that the subcontexts that arise by δ -**adjusting** decrease the size of the concept lattice and the implication set while preserving underlying knowledge. We evaluate the remaining knowledge by training a classification task. This results in a more understandable depiction of the encapsulated data for the human mind.

Due to space constraints, this work only briefly sketches proofs. A version containing all proofs is released on arxiv.org¹.

2 Foundations

We start this section by recalling notions from FCA [10]. A *formal context* is a triple $\mathbb{K} := (G, M, I)$, consisting of an *object set* G , an *attribute set* M and a binary *incidence relation* $I \subseteq G \times M$. In this work, G and M are assumed to be finite. The *complementary formal context* is given by $\mathbb{K}^C := (G, M, (G \times M) \setminus I)$. The maps $\cdot': \mathcal{P}(G) \rightarrow \mathcal{P}(M)$, $A \mapsto A' := \{m \in M \mid \forall g \in A: (g, m) \in I\}$ and $\cdot': \mathcal{P}(M) \rightarrow \mathcal{P}(G)$, $B \mapsto B' := \{g \in G \mid \forall m \in B: (g, m) \in I\}$ are called *derivations*. A pair $c = (A, B)$ with $A \subseteq G$ and $B \subseteq M$ such that $A' = B$ and $B' = A$ is called a *formal concept* of the context (G, M, I) . The set of all formal concepts of \mathbb{K} is denoted by $\mathfrak{B}(\mathbb{K})$. The pair consisting of $\mathfrak{B}(\mathbb{K})$ and the order $\leq \subset (\mathfrak{B}(\mathbb{K}) \times \mathfrak{B}(\mathbb{K}))$ with $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$ defines the

¹ <https://arxiv.org/abs/2106.10978>.

concept lattice $\mathfrak{B}(\mathbb{K})$. In every lattice and thus every concept lattice each subset U has a unique infimum and supremum which are denoted by $\bigwedge U$ and $\bigvee U$. The *contranominal scale of dimension k* is $\mathbb{N}_k^c := (\{1, 2, \dots, k\}, \{1, 2, \dots, k\}, \neq)$. Its concept lattice is the *Boolean lattices of dimension k* and consists of 2^k concepts. Let $\mathbb{K} = (G, M, I)$. We call an attribute m *clarifiable* if there is an attribute $n \neq m$ with $n' = m'$. In addition we call it *reducible* if there is a set $X \subseteq M$ with $m \notin X$ and $m' = X'$. Otherwise, we call m *irreducible*. \mathbb{K} is called *attribute clarified (attribute reduced)* if it does not contain clarifiable (reducible) attributes. The definitions for the object set are analogous. If \mathbb{K} is attribute clarified and object clarified (attribute reduced and object reduced), we say \mathbb{K} is *clarified (reduced)*. This contexts are unique up to isomorphisms. Their concept lattices are isomorphic to $\mathfrak{B}(\mathbb{K})$. A *subcontext* $\mathbb{S} = (H, N, J)$ of $\mathbb{K} = (G, M, I)$ is a formal context with $H \subseteq G$, $N \subseteq M$ and $J = I \cap (H \times N)$. We denote this by $\mathbb{S} \leq \mathbb{K}$ and use the notion $\mathbb{K}[H, N] := (H, N, I \cap (H \times N))$. If $\mathbb{S} \leq \mathbb{K}$ with $\mathbb{S} \cong \mathbb{N}_k^c$ we call \mathbb{S} a *contranominal scale in \mathbb{K}* . For a (concept) lattice (L, \leq) and a subset $S \subseteq L$, $(S, \leq_{S \times S})$ is called *suborder* of (L, \leq) . A suborder S of a lattice is called a *sub-meet-semilattice* if $(a, b \in S \Rightarrow (a \wedge b) \in S)$ holds. In a formal context $\mathbb{K} = (G, M, I)$ with $X, Y \subseteq M$ define an *implication* as $X \rightarrow Y$ with *premise* X and *conclusion* Y . An implication is *valid in \mathbb{K}* if $X' \subset Y'$. In this case, we call $X \rightarrow Y$ an *implication of \mathbb{K}* . The set of all implications of a formal context \mathbb{K} is denoted by $Imp(\mathbb{K})$. A minimal set $\mathcal{L}(\mathbb{K}) \leq Imp(\mathbb{K})$ defines an *implication base* if every implication of \mathbb{K} follows from $\mathcal{L}(\mathbb{K})$ by composition. An implication base of minimal size is called *canonical base* of \mathbb{K} and is denoted by $\mathcal{C}(\mathbb{K})$.

Now recall some notions from graph theory. A *graph* is a pair (V, E) with a set of *vertices* V and a set of *edges* $E \subset \binom{V}{2}$. Two vertices u, v are called *adjacent* if $\{u, v\} \in E$. The adjacent vertices of a vertex are called its *neighbors*. In this work graphs are undirected and have no multiple edges or loops. A graph with two sets S and T with $S \cup T = V$ and $S \cap T = \emptyset$ such that there is no edge with both vertices in S or both vertices in T is called *bipartite* and denoted by (S, T, E) . A *matching* in a graph is a subset of the edges such that no two edges share a vertex. It is called *induced* if no two edges share vertices with some edge not in the matching. For a formal context (G, M, I) the *associated bipartite graph* is the graph where S and T correspond to G and M and the set of edges to I .

3 Related Work

In the field of Formal Concept Analysis numerous approaches deal with simplifying the structure of large datasets. Large research interest was dedicated to altering the incidence relation together with the objects and attributes in order to achieve smaller contexts. A procedure based on a random projection is introduced in [18]. Dias and Vierira [5] investigate the replacement of similar objects by a single representative. They evaluate this strategy by measuring the appearance of false implications on the new object set. In the attribute case a similar approach is explored by Kuitche et al. [17]. Similar to our method, many common prior approaches are based on the selection of subcontexts. For example,

Hanika et al. [12] rate attributes based on the distribution of the objects in the concepts and select a small relevant subset of them. A different approach is to select a subset of concepts from the concept lattice. While it is possible to sample concepts randomly [2], the selection of concepts by using measures is well investigated. To this end, a structural approach is given in [7] through dismantling where a sublattice is chosen by the iterative elimination of all doubly irreducible concepts. Kuznetsov [20] proposes a stability measure for formal concepts based on the sizes of the concepts. The support measure is used by Stumme et al. [25] to generate iceberg lattices. Our approach follows up on this, as we also preserve sub-semilattices of the original concept lattice. However, we are not restricted to the selection of iceberg lattices. Compared to many other approaches we do not alter the incidence or the objects and thus do not introduce false implications.

4 Computing Contranominal Scales

In this section, we examine the complexity of computing all contranominals and provide the recursive backtracking algorithm **ContraFinder** to solve this task.

4.1 Computing Contranominals Is Hard

The problem of computing contranominal scales is closely related to the problem of computing cliques in graphs and induced maximum matchings in bipartite graphs.

The relationship between the induced matching problem and the contranominal scale problem follows directly from their respective definitions.

Lemma 1. *Let (S, T, E) be a bipartite graph, $\mathbb{K} := (S, T, (S \times T) \setminus E)$ a formal context and $H \subset S, N \subset T$. The edges between H and N are an induced matching of size k in (S, T, E) iff $\mathbb{K}[H, N]$ is a contranominal scale of dimension k .*

The lemma follows directly from the definition of induced matchings and contranominal scales. To investigate the connection between the clique problem and the contranominal scale problem, define the conflict graph as follows:

Definition 1. *Let $\mathbb{K} := (G, M, I)$ be a formal context. Define the conflict graph of \mathbb{K} as the graph $cg(\mathbb{K}) := (V, E)$ with the vertex set $V = (G \times M) \setminus I$ and the edge set $E = \{(g, m), (h, n)\} \in \binom{V}{2} \mid (g, n) \in I, (h, m) \in I\}$.*

The relationship between the cliques in the conflict graph and the contranominal scales in the formal context is given through the following lemma.

Lemma 2. *Let $\mathbb{K} = (G, M, I)$ be a formal context, $cg(\mathbb{K})$ its conflict graph and $H \subset G, N \subset M$. Then $\mathbb{K}[H, N]$ is a contranominal scale of dimension k iff $(H \times N) \setminus I$ is a clique of size k in $cg(\mathbb{K})$.*

The lemma follows from the definition of the conflict graph. Furthermore, all three problems are in the same computational class as the clique problem is *NP*-complete [15] and Lozin [21] shows the similar result for the induced matching problem in the bipartite case. Thus, Lemma 1 provides the following:

Proposition 1. *Deciding the **CONTRANOMINAL PROBLEM** is *NP*-complete.*

4.2 Baseline Algorithms

Building on Lemma 2 the set of all contranominal scales can be computed using algorithms for iterating all cliques in the conflict graph. The set of all cliques then corresponds to the set of all contranominal scales in the formal context. An algorithm to iterate all cliques in a graph is proposed by Bron and Kerbosch [3].

An alternative approach is to use branch and search algorithms such as [27]. Those exploit the fact that for each maximum matching and each vertex there is either an adjacent edge to this vertex in the matching or each of its neighboring vertices has an adjacent edge in the matching. Branching on these vertices the size of the graph is iteratively decreased. Note, that this idea, in contrast to our approach described below, does not exploit bipartiteness of the graph.

4.3 ContraFinder: An Algorithm to Compute Contranominal Scales

In this section we introduce the recursive backtracking algorithm **ContraFinder** to compute all contranominal scales. Due to Proposition 1, it has exponential runtime, thus two speedup techniques are proposed in the subsequent section.

The main idea behind **ContraFinder** is the following. In each recursion step a set of tuples corresponding to an attribute set is investigated:

Definition 2. Let $\mathbb{K} = (G, M, I)$ be a formal context and $N \subseteq M$. Define $C(N) := \{(g, m) \notin I \mid g \in G, m \in N \text{ and } \forall x \in N \setminus \{m\} : (g, x) \in I\}$ as the set of characterizing tuples of N . We call N the generator of $C(N)$.

The characterizing tuples encodes all contranominal scales for this attributes:

Lemma 3. Let $\mathbb{K} = (G, M, I)$, $N \subseteq M$ and $H(m) := \{g \in G \mid (g, m) \in C(N)\}$. Then $\mathbb{K}[O, N]$ is a contranominal scale iff O contains exactly one element of each $H(m)$ with $m \in N$.

The proof follows from the fact, that the non-incident pairs of each contranominal scale are represented by the combinations of characterizing tuples with different attributes. Lemma 3 implies that such contranominal scales can exist only if no $H(m)$ is empty and $|N| = |O|$. Both this sets can be reconstructed from a set of characterizing tuples corresponding to N . This is done in `unpack_contranominals` in Algorithm 1. Therefore, N does not have to be memorized in **ContraFinder**. The algorithm exploits the fact that for each set of characterizing tuples $C(N)$ the attributes N can be ordered and iterated in lexicographical order, similar to `NextClosure` [10, sec. 2.1].

Definition 3. Let (M, \leq) be a linearly ordered set. The lexicographical order on $\mathcal{P}(M)$ is a linear order. Let $A = a_1, \dots, a_n$ and $B = b_1, \dots, b_m$ with $a_i < a_{i+1}$ and $b_i < b_{i+1}$. $A < B$ in case $n < m$ if $(a_1, \dots, a_n) = (b_1, \dots, b_n)$ and in case $n = m$ if $\exists i : \forall j \leq i : a_j = b_j$ and $a_i < b_i$.

Algorithm 1. ContraFinder**Input:** Formal Context $\mathbb{K} = (G, M, I)$ **Output:** Set of all Contranominal Scales

```

def compute_contranominal_scales( $G, M, I$ ):
    characterizing_tuples( $\emptyset, M, \emptyset, I$ )

def characterizing_tuples( $C_N, \tilde{M}, F, I$ ):
    for  $m$  in  $\tilde{M}$  in lexicographical order:
         $\tilde{M} = \tilde{M} \setminus \{m\}$ 
         $cand\_C_N = \{(g, n) \in C_N \mid (g, m) \in I\}$ 
         $cand\_m = \{(g, m) \mid (g, m) \notin I, g \notin F, \nexists n : (g, n) \in C_N\}$ 
        if  $|\{g \mid (g, n) \in C_N\}| = |\{g \mid (g, n) \in cand\_C_N\}|$  and  $|cand\_m| > 0$ :
            unpack_contrinominals( $cand\_C_N \cup cand\_m$ )
             $C_{N_{new}} = cand\_C_N \cup cand\_m$ 
             $F_{new} = F \cup \{g \in G \mid (g, m) \notin I\}$ 
            characterizing_tuples( $C_{N_{new}}, \tilde{M}, F_{new}, I$ )

def unpack_contrinominals( $C_N$ ):
     $N = \{m \mid (g, m) \in C_N\}$ 
    for  $O$  in  $\{\{g_{m_1}, \dots, g_{m_{|N|}}\} \mid m_i \in N, g_{m_i} \in \{g \in G \mid (g, m_i) \in C_N\}\}$ 
        report ( $O, N$ ) as contranominal scale

```

Similar to **Titanic**, our algorithm utilises the following anti-monotonic property. Each contranominal scale of dimension k has a contranominal scale of dimension $k - 1$ as subcontext. Thus, only attribute combinations N have to be considered if $\forall N' \subset N : C(N') \neq \emptyset$. The algorithm removes in each recursion step the attributes in \tilde{M} in lexicographical order to guarantee that all attribute combinations of the formal context with contranominal scales are investigated.

In each step the set of forbidden objects F increases, since each contranominal scale contains exactly one non-incidence in each contained object.

Theorem 1. *The algorithm reports every contranominal scale exactly once.*

To proof this theorem, one has to show that the lexicographical order and the anti-monotonic property are respected. **ContraFinder**, combined with Lemma 1, can also be used to compute all maximum induced matchings in bipartite graphs.

4.4 Speedup Techniques

Clarifying and Reducing. In the following, we consider clarified and reduced formal contexts with regards to reconstructing the contranominal scales in the original context from the contranominal scales of the augmented one. This allows to use clarifying and reducing as a speedup technique.

In the clarified context, each pair of objects or attributes is merged if equality of their derivations holds. To deduce the original formal context from the clari-

fied one the previously merged attributes and objects can be duplicated. Thus, contranominal scales containing merged objects or attributes are duplicated.

Now, we demonstrate how to reconstruct the contranominal scales from attribute reduced contexts. Thereby, for each eliminated attribute m we have to memorize the irreducible attribute set that has the same derivation as m .

Definition 4. Let $\mathbb{K} = (G, M, I)$ be a formal context and $R(\mathbb{K})$ the set of all attributes that are reducible in \mathbb{K} . Define the map $\omega: R(\mathbb{K}) \rightarrow \mathcal{P}(M \setminus R(\mathbb{K}))$ with $x \mapsto (N \subset M \setminus (R(\mathbb{K}) \cup \{x\}))$ such that $N' = x'$ and N of greatest cardinality. For a fixed object set $H \subseteq G$, let $\omega_H: R(\mathbb{K}) \rightarrow \mathcal{P}(M \setminus R(\mathbb{K}))$ be the map with $x \mapsto \{y \mid y \in \omega(x), \forall h \in H : (h, x) \notin I \Rightarrow (h, y) \notin I\}$.

Note, that the map ω is well defined as the uniqueness follows directly from the maximality of N . The following lemma provides a way to reconstruct the contranominal scales in the original context from the ones in the reduced one.

Lemma 4. Let $\mathbb{K} = (G, M, I)$ be a formal context with \mathbb{K}_r its attribute-reduced subcontext and \mathcal{K} the set containing all contranominal scales of \mathbb{K}_r . Then the set $\tilde{\mathcal{K}} = \{\mathbb{K}[H, \tilde{N}] \mid \mathbb{K}[H, N = \{n_1, \dots, n_l\}] \in \mathcal{K}, \tilde{N} = \{\tilde{n}_i \mid n_i = \tilde{n}_i \vee n_i \in \omega_H(\tilde{n}_i)\}\}$ contains exactly all contranominal scales of \mathbb{K} .

This follows from the definition of reducibility. Thus, to reconstruct contranominal scales, for each $x \in R(\mathbb{K})$ all $y \in \omega(x)$ are considered. $U \cup x$ is a candidate for the attribute set of a contranominal scale in \mathbb{K} , if there is a $U \subset M \setminus \omega(x)$ with $U \cup y$ attribute set of a contranominal scale \mathbb{S}_y for all y . This candidate forms the contranominal scale $\mathbb{K}[H, U \cup x]$, if and only if all contranominal scales \mathbb{S}_y share the same object set H . The object reducible case can be done dually.

Knowledge-Cores. The notion of (p, q) -cores is introduced to FCA by Hanika and Hirth in [11]. Thereby, dense subcontexts are defined as follows:

Definition 5 (Hanika and Hirth [11]). Let $\mathbb{K} = (G, M, I)$ and $\mathbb{S} = \mathbb{K}[H, N]$ be formal contexts. \mathbb{S} is called a (p, q) -core of \mathbb{K} for $p, q \in \mathbb{N}$, if $\forall g \in H : |g'| \geq p$ and $\forall m \in N : |m'| \geq q$ and \mathbb{S} is maximal under this condition.

Every formal context with fixed p and q has a unique (p, q) -core. Computing knowledge cores provides a way to reduce the number of attributes and objects in a formal context without removing large contranominal scales.

Lemma 5. Let \mathbb{K} be a formal context, $k \in \mathbb{N}$, and $\mathbb{S} \leq \mathbb{K}$ its $(k-1, k-1)$ -core. Then for every contranominal scale $\mathbb{C} \leq \mathbb{K}$ of dimension k it holds $\mathbb{C} \leq \mathbb{S}$.

The lemma follows from the maximality of (p, q) -cores. Thus, to compute all contranominal scales of dimension at least k it is possible to compute them in the $(k-1, k-1)$ -core. Note that in this case however, smaller contranominal scales might get eliminated. Therefore, if the goal is to compute contranominal scales of smaller sizes the $(k-1, k-1)$ -cores should not be computed.

5 Attribute Selection

In this section we propose δ -**adjusting**, a method to select attributes based on measuring their influence for contranominal scales as follows:

Definition 6. Let $\mathbb{K} = (G, M, I)$ be a formal context and $k \in \mathbb{N}$. Call $N \subset M$ k -cubic if $\exists H \subset G$ with $\mathbb{K}[H, N]$ being a contranominal scale of dimension k and $\# \tilde{N} \supseteq N$ such that \tilde{N} is $(k + 1)$ -cubic. Define the contranominal-influence of $m \in M$ in \mathbb{K} as $\zeta(m) := \sum_{k=1}^{\infty} (|\{N \subset M \mid m \in N, N \text{ is } k\text{-cubic}\}| \cdot \frac{2^k}{k})$.

Subcontexts that are k -cubic are directly influencing the concept lattice, as those dominates the structure as the following shows.

Proposition 2. An attribute set is k -cubic, iff the sub-meet-semilattice that is generated by its attribute concepts is a Boolean lattice of dimension k that has no Boolean superlattice in the original concept lattice.

The contranominal influence thus measures the impact of an attribute on the lattice structure. In this, only the maximal contranominal scales are considered since the smaller non maximal-ones have no additional structural impact. As each contranominal scale of dimension k corresponds to 2^k concepts, we scale the number of attribute combinations with this factor. To distribute the impact of a contranominal scale evenly over all involved attributes, the measure is scaled by $\frac{1}{k}$. With this measure we now define the notions of δ -**adjusting**.

Definition 7. Let $\mathbb{K} = (G, M, I)$ be a formal context and $\delta \in [0, 1]$. Let $N \subset M$ minimal such that $\frac{|N|}{|M|} \geq \delta$, $\zeta(n) < \zeta(m)$ for all $n \in N, m \in M \setminus N$. We call $\mathbb{A}_\delta(\mathbb{K}) := \mathbb{K}[G, N]$ the δ -adjusted subcontext of \mathbb{K} and $\mathfrak{B}(\mathbb{A}_\delta(\mathbb{K}))$ the δ -adjusted sublattice of $\mathfrak{B}(\mathbb{K})$.

Note, that δ -**adjusting** always results in unique contexts. Moreover, every δ -adjusted sublattice is a sub-meet-semilattice of the original one [10, Prop 31]. For every context $\mathbb{K} = (G, M, I)$ it holds that $\mathbb{A}_1 = \mathbb{K}$ and $\mathbb{A}_0 = \mathbb{K}[G, \emptyset]$. A context from a medical diagnosis dataset with measured contranominal influence and computed $\frac{1}{2}$ -adjusted subcontext can be retraced in Fig. 1.

It is important to observe that for a context \mathbb{K} and its reduced context \mathbb{K}_r , a different attribute set can remain if they are δ -adjusted, as can be seen in Fig. 2. Therefore, the resulting concept lattices for \mathbb{K} and \mathbb{K}_r can differ. To preserve structural integrity between δ -adjusted formal contexts and their concept lattices we thus recommend to only consider clarified and reduced formal contexts. In the rest of this work, these steps are therefore performed prior to δ -**adjusting**. Note, that since no attributes are generated no new contranominal scales can arise by δ -**adjusting**. Furthermore, removing attributes can not turn another attribute from irreducible to reducible. On the other hand however, objects can become reducible as can be seen again in Fig. 2. While 6 is irreducible in the original context, it is reducible in $\mathbb{A}_{\frac{3}{5}}(\mathbb{K})$.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	Attribute Name	2	3	4	ζ
111			×	×	×			×	×	×					×	a: Lumbar pain y	1	22	6	84.7
119	×				×	×	×	×					×	×	×	b: Bladder inflammation y	1	29	0	79.3
31		×	×	×	×					×	×	×			×	c: Burning n	1	31	9	120.7
32	×		×		×			×	×	×						d: Lumbar pain n	2	19	0	54.7
17		×		×	×	×			×	×	×		×			e: Nausea n	0	16	3	54.7
27		×	×	×	×			×	×	×			×			f: Burning y	1	31	0	84.7
105	×	×				×	×				×	×	×			g: Temp. $\in [40.0, 42.0]$	2	24	5	88.0
58		×		×	×	×				×	×		×	×		h: Micturition pains n	1	18	5	70.0
65	×			×	×			×					×	×	×	i: Temp. $\in [35.0, 37.5]$	3	16	0	48.7
103	×		×				×					×	×		×	j: Pelvis nephritis n	1	19	1	56.7
56		×	×	×	×					×	×		×	×	×	k: Micturition pains y	1	33	0	90.0
98	×	×	×				×				×	×	×			l: Pelvis nephritis y	3	17	0	51.3
43		×	×	×	×			×		×			×	×		m: Urine pushing y	0	21	7	84.0
50	×		×				×		×					×	×	n: Temp. $\in [37.5, 40.0]$	2	23	3	77.3
															×	o: Bladder inflammation n	1	26	1	75.3

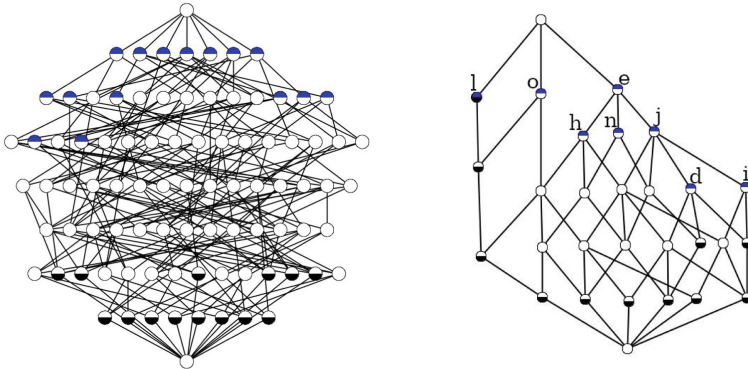


Fig. 1. Top: reduced and clarified medical diagnosis dataset [4]. The $\frac{1}{2}$ -adjusted sub-context is highlighted. The objects are patient numbers. The attributes are described in the figure together with the count of k-cubic subcontexts and their contranominal influence ζ . Bottom: lattice of the original (left) and the $\frac{1}{2}$ -adjusted (right) dataset. (Color figure online)

5.1 Properties of Implications

In this section we investigate δ -adjusting with respect to the influence on implications. Let $\mathbb{K} = (G, M, I)$ be a formal context, $m \in M$ and $X \rightarrow Y$ an implication in \mathbb{K} . If m is part of the implication; i.e., $m \in X$ or $m \in Y$, this implication vanishes. Therefore the removal of m in an implication $X \rightarrow Y$ of some implication base $\mathcal{C}(\mathbb{K})$ is of interest. If m is neither part of a premise nor a conclusion of an implication $X \rightarrow Y \in \mathcal{C}(\mathbb{K})$ its removal has no impact on this implication base. In case $m \in Y$, its elimination changes all implications $X \rightarrow Y$ to $X \rightarrow Y \setminus \{m\}$. Note that, even though all implications can still be deduced

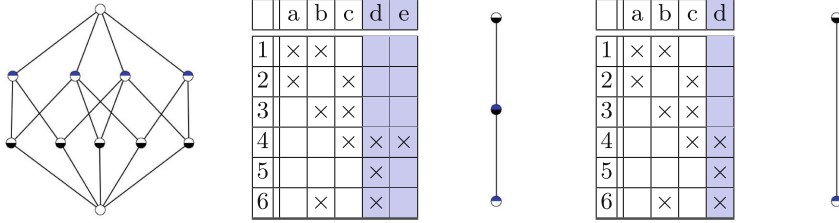


Fig. 2. A concept lattice together with two of its contexts \mathbb{K} and \mathbb{K}_r , whereby \mathbb{K}_r is attribute reduced while \mathbb{K} contains the reducible element e . In both contexts the $\frac{3}{5}$ -adjusted subcontext is highlighted. Their lattices (right to each context) differ. (Color figure online)

from $\mathcal{C}' = \{X \rightarrow Y : X \rightarrow Y \cup \{m\} \in \mathcal{C}(\mathbb{K})\}$ this set is not necessarily minimal and in this case is not a base. Especially if $\{m\} = Y$ the resulting $X \rightarrow \emptyset$ is never part of an implication base. In case $m \in X$, every $Z \rightarrow X$ in the base is changed to $Z \rightarrow X \setminus \{m\} \cup Y$ while $X \rightarrow Y$ is removed. Similarly to the conclusion case, the resulting set of implications can be used to deduce all implications but is not necessarily an implication base. Moreover, no new implications can emerge from the removal of attributes, as the following shows.

Lemma 6. *Let $\mathbb{K} = (G, M, I)$ be a formal context, $N \subset M$ and $X, Y \subseteq N$ with $X \rightarrow Y$ a non-valid implication in \mathbb{K} . Then $X \rightarrow Y$ is also non-valid in $\mathbb{K}[G, N]$.*

The lemma follows from the fact that if $X' \subset Y'$ in \mathbb{K} , then $X' \subset Y'$ in a subcontext of \mathbb{K} with all objects. Thus, the relationship between the implications of a subcontext with all objects and the original context is as follows:

Corollary 1. *Let $\mathbb{K} = (G, M, I)$ be a formal context, $\mathbb{S} = \mathbb{K}[G, N]$ and $N \subset M$. Then $\text{Imp}(\mathbb{S}) \subseteq \text{Imp}(\mathbb{K})$.*

This influences the size of the base of a δ -adjusted subcontext as follows:

Lemma 7. *Let $\mathbb{K} = (G, M, I)$ a formal context, and $\mathbb{S} = \mathbb{K}[G, N]$ and $N \subset M$. Then $|\mathcal{C}(\mathbb{S})| \leq |\mathcal{C}(\mathbb{K})|$.*

To prove this lemma, one can construct an implication set of size at most $|\mathcal{C}(\mathbb{K})|$ that generates all implications. Revisiting the context in Fig. 1 together with its $\frac{1}{2}$ -adjusted subcontext the selection of nearly 50% of the attributes (8 out of 15) results in a sub-meet-semilattice containing only 33% of the concepts (29 out of 88). Moreover, the implication base of the original context includes 40 implications. After the alteration its size is decreased to 11 implications.

6 Evaluation and Discussion

In this section we evaluate the algorithm **ContraFinder** and the process of δ -adjusting using real-world datasets.

6.1 Datasets

Table 1 provides descriptive properties of the datasets used in this work. The zoo [6, 22] and mushroom [6, 23] datasets are classical examples often used in FCA based research such as the TITANIC algorithm. The Wikipedia [19] dataset depicts the edit relation between authors and articles while the Wiki44k dataset is a dense part of the Wikidata knowledge graph. The original Wiki44k dataset was taken from [14], in this work we conduct our experiments on an adapted version by [13]. Finally, the Students dataset [24] depicts grades of students together with properties such as parental level of education. All experiments are conducted on the reduced and clarified versions of the contexts. For reproducibility the adjusted versions of all datasets are published in [9].

Table 1. Datasets used for the evaluation of **ContraFinder** and δ -adjusting.

	Zoo	Students	Wikipedia	Wiki44k	Mushroom
Objects	101	1000	11273	45021	8124
Attributes	43	32	102	101	119
Density	0.40	0.28	0.015	0.045	0.19
Number of concepts	4579	17603	14171	21923	238710
Mean objects per concept	18.48	16.73	20.06	109.47	91.89
Mean attributes per concept	7.32	5.97	5.88	7.013	16.69
Size of canonical base	401	2826	4575	7040	2323

6.2 Runtime of ContraFinder

ContraFinder is a recursive backtracking algorithm that iterates over all attribute sets containing contranominal scales. Thus, the worst case runtime is given by $O(n^k)$ where n is the number of attributes of the formal context and k the maximum dimension of a contranominal scale in it. The Branch-And-Search algorithm from [27] has a runtime of $O(1.3752^n)$ where n is the sum of attributes and objects. Finally the Bron-Kerbosch algorithm has a worst-case runtime of $O(3^{n/3})$ with n being the number of non-incident object-attribute pairs.

To compare the practical runtime of the algorithms we test them on the previously introduced real world datasets. We report the runtimes in Table 2, together with the dimension of the largest contranominal scale and the total number of contranominal scales. Note, that for larger datasets we are not able to compute the number of all contranominal scales using Bron-Kerbosch (from Students) and the Branch-And-Search algorithm (Mushroom) below 24h due to their exponential nature and thus stopped the computations. All experiments are conducted on an Intel Core i5-8250U processor with 16 GB of RAM.

Table 2. Experimental runtimes of the different algorithms on all datasets.

	Zoo	Students	Wikipedia	Wiki44k	Mushroom
ContraFinder	2.43	7.36	17.15	35.65	1961.0
Bron Kerbosch searching cliques	138.70	>86400	>86400	>86400	>86400
Branch and search algorithm	14.40	12005.82	1532.17	16783.58	>86400
Dim. of max. contranominal scale	7	8	9	11	10
Number of contranominal scales	$4.1 \cdot 10^7$	$7.8 \cdot 10^9$	$9.9 \cdot 10^8$	$2.0 \cdot 10^{14}$	$1.2 \cdot 10^{19}$

Table 3. Evaluation of k -adjusted contexts. The standard deviation is given in parenthesis. “Acc of DT” is the abbreviation for “Accuracy of the Decision Tree”.

		Zoo	Students	Wikipedia	Wiki44k	Mushroom
$ \mathfrak{B}(\mathbb{K}) $	$\frac{1}{2}$ -adjusted	90	312	65	323	426
	Sampling	496 (205)	1036 (327)	833 (517)	1397 (627)	8563 (4532)
	Hanika et al.	95	341	67	254	561
$ \mathcal{C}(\mathbb{K}) $	$\frac{1}{2}$ -adjusted	98	105	626	1003	339
	Sampling	95 (17)	156 (35)	758 (101)	1360 (135)	574 (93)
	Hanika et al.	100	105	553	1091	490
Acc of DT	$\frac{1}{2}$ -adjusted	0.88 (0.08)	0.88 (0.06)	0.99 (0.01)	0.98 (0.03)	0.98 (0.02)
	Sampling	0.89 (0.15)	0.81 (0.15)	0.9 (0.14)	0.95 (0.06)	0.92 (0.13)
	Hanika et al.	0.88 (0.09)	0.89 (0.06)	0.99 (0.01)	0.98 (0.16)	0.97 (0.03)

6.3 Structural Effects of δ -Adjusting

We measure the number of formal concepts generated by the formal context as well as the size of the canonical base. To demonstrate the effects of δ -adjusting we focus on $\delta = \frac{1}{2}$. Our two baselines are selecting the same number of attributes using random sampling and choosing the attributes of highest relative relevance as described in [12]. It can be observed, that in all three cases the number of concepts heavily decrease. However, this effect is considerably stronger for $\frac{1}{2}$ -adjusting and the approach of Hanika et al. compared to sampling. Hereby, $\frac{1}{2}$ -adjusting yields smaller concept lattices on four datasets. A similar effect can be observed for the sizes of the canonical bases where our method yields three times in the smallest cardinality.

6.4 Knowledge in the δ -Adjusted Context

To measure the degree of encapsulated knowledge in δ -adjusted formal contexts we conduct the following experiment using once again sampling and the relative relevant attributes of Hanika et al. as baselines. In order to measure if the remaining subcontexts still encapsulates knowledge we train a decision tree classifier on them predicting an attribute that is removed beforehand. This attribute is sampled randomly in each step. To prevent a random outlier from distorting the result we repeat this same experiment 1000 times for each context and method and report the mean value as well as the standard-deviation in Table 3. The experiment is conducted using a 0.5-split on the train and test

data. For all five datasets, the results of the decision tree on the $\frac{1}{2}$ -adjusted context are consistently high, however $\frac{1}{2}$ -adjusting and the Hanika et al. approach outperform the sampling approach. Both this methods achieve the highest score on four contexts, in two of this cases the highest result is shared. The single highest score of sampling is just slightly above the other two approaches.

6.5 Discussion

The theoretical runtime of **ContraFinder** is polynomial in the dimension of the maximum contranominal. Therefore, compared to the baseline algorithms it performs better, the smaller the maximum contranominal scale in a dataset. Furthermore, the runtime of Bron-Kerbosch is worse, the sparser a formal context, as the number of pairs that are non-incident increases and thus more vertices have to be iterated. Finally, the Branch-And-Search algorithm is best in the case that the dimension of the maximum contranominal scale is not bounded. To evaluate, how this theoretical properties translate to real world data, we compute the set of all contranominal scales with the three algorithms on the previously described datasets. Only **ContraFinder** can compute the set of all contranominal scales on the larger datasets on our hardware under 24h. The runtime of **ContraFinder** is thus superior to the other two on real-world datasets.

To evaluate the impact on the understandability of the δ -adjusted formal contexts, we conduct the experiments measuring the sizes of the concept lattices and the canonical bases. All three evaluated methods heavily decrease the size of the concept lattice as well as the canonical base. Compared to the random sampling $\frac{1}{2}$ -adjusting and the method of Hanika et al. influence the size of this structural components much stronger. Among those two, $\frac{1}{2}$ -adjusting seems to slightly outperform the method of Hanika et al. and is thus more suited to select attributes from a large dataset in order to be analyzed by a human.

To evaluate to what extent knowledge in the formal context of reduced size is encapsulated we conduct the experiment with the decision trees. This experiment demonstrates that the selected formal subcontext can be used in order to deduce relationships of the remaining attributes in the context. While meaningful implications are preserved and the implication set is downsized, $\frac{1}{2}$ -adjusted lattices seem to be suitable to preserve large amounts of data from the original dataset. Similar good results can be achieved with the method of Hanika et al.; however, our algorithm combines this with producing smaller concept lattices and canonical bases and is thus more suitable for the task to prepare data for a human analyst by reducing sizes of structural constructs.

We conclude from these experiments that δ -adjusting is a solution to the problem to make information more feasible for manual analysis while retaining important parts of the data. In particular, if large formal contexts are investigated this method provides a way to extract relevant subcontexts.

7 Conclusion

In this work, we proposed the algorithm **ContraFinder** in order to enable the computation of the set of all contranominal scales in a formal context. Using this, we defined the contranominal-influence of an attribute. This measure allows us to select a subset of attributes in order to reduce a formal context to its δ -adjusted subcontext. The size of its lattice is significantly reduced compared to the original lattice and thus enables researchers to analyze and understand much larger datasets using Formal Concept Analysis. Furthermore, the size of the canonical base, which can be used in order to derive relationships of the remaining attributes shrinks significantly. Still, remaining data can be used to deduce relationships between attributes, as our classification experiment shows. This approach therefore identifies subcontexts whose sub-meet-semilattice is a restriction of the original lattice of a formal context to a small meaningful part.

Further work in this area could leverage **ContraFinder** in order to compute the contranominal-relevance of attributes more efficiently to handle even larger datasets. Moreover, a similar measure for objects could be introduced. However, one should keep in mind that hereby false implications can arise.

References

1. Albano, A., Chornomaz, B.: Why concept lattices are large - extremal theory for the number of minimal generators and formal concepts. In: 12th International Conference on Concept Lattices and Their Applications (CLA 2016). CEUR Workshop Proceedings, vol. 1466, pp. 73–86. CEUR-WS.org (2015)
2. Boley, M., Gärtner, T., Grosskreutz, H.: Formal concept sampling for counting and threshold-free local pattern mining. In: SIAM International Conference on Data Mining (SDM 2010), pp. 177–188. SIAM (2010)
3. Bron, C., Kerbosch, J.: Finding all cliques of an undirected graph (algorithm 457). *Commun. ACM* **16**(9), 575–576 (1973)
4. Czerniak, J., Zarzycki, H.: Application of rough sets in the presumptive diagnosis of urinary system diseases. In: Soldek, J., Drobiazgiwicz, L. (eds.) *Artificial Intelligence and Security in Computing Systems*. The Springer International Series in Engineering and Computer Science, vol. 752, pp. 41–51. Springer, Boston (2002). https://doi.org/10.1007/978-1-4419-9226-0_5
5. Dias, S., Vieira, N.: Reducing the size of concept lattices: the JBOS approach. In: 7th International Conference on Concept Lattices and Their Applications (CLA 2010). CEUR Workshop Proceedings, vol. 672, pp. 80–91. CEUR-WS.org (2010)
6. Dua, D., Graff, C.: UCI machine learning repository (2017). <http://archive.ics.uci.edu/ml>
7. Duffus, D., Rival, I.: Crowns in dismantlable partially ordered sets. In: 5th Hungarian Combinatorial Colloquium, vol. I, pp. 271–292 (1978)
8. Dürschnabel, D., Hanika, T., Stumme, G.: Drawing order diagrams through two-dimension extension. CoRR [arXiv:1906.06208](https://arxiv.org/abs/1906.06208) (2019)
9. Dürschnabel, D., Koyda, M., Stumme, G.: Attribute selection using contranominal scales [dataset], April 2021. <https://doi.org/10.5281/zenodo.4945088>
10. Ganter, B., Wille, R.: *Formal Concept Analysis - Mathematical Foundations*. Springer, Heidelberg (1999). <https://doi.org/10.1007/978-3-642-59830-2>

11. Hanika, T., Hirth, J.: Knowledge cores in large formal contexts. CoRR [arXiv:2002.11776](https://arxiv.org/abs/2002.11776) (2020)
12. Hanika, T., Koyda, M., Stumme, G.: Relevant attributes in formal contexts. In: Endres, D., Alam, M., Šotropa, D. (eds.) ICCS 2019. LNCS (LNAI), vol. 11530, pp. 102–116. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23182-8_8
13. Hanika, T., Marx, M., Stumme, G.: Discovering implicational knowledge in wiki-data. In: Cristea, D., Le Ber, F., Sertkaya, B. (eds.) ICFCA 2019. LNCS (LNAI), vol. 11511, pp. 315–323. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21462-3_21
14. Ho, V.T., Stepanova, D., Gad-Elrab, M.H., Kharlamov, E., Weikum, G.: Rule learning from knowledge graphs guided by embedding models. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11136, pp. 72–90. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00671-6_5
15. Karp, R.: Reducibility among combinatorial problems. In: Proceedings of a Symposium on the Complexity of Computer Computations. The IBM Research Symposia Series, pp. 85–103. Plenum Press, New York (1972)
16. Koyda, M., Stumme, G.: Boolean substructures in formal concept analysis. CoRR [arXiv:2104.07159](https://arxiv.org/abs/2104.07159) (2021)
17. Kuitché, R., Temgoua, R., Kwuida, L.: A similarity measure to generalize attributes. In: 14th International Conference on Concept Lattices and Their Applications (CLA 2018). CEUR Workshop Proceedings, vol. 2123, pp. 141–152. CEUR-WS.org (2018)
18. Kumar, C.: Knowledge discovery in data using formal concept analysis and random projections. *Int. J. Appl. Math. Comput. Sci.* **21**(4), 745–756 (2011)
19. Kunegis, J.: KONECT: the Koblenz network collection. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1343–1350 (2013)
20. Kuznetsov, S.: Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. *Nauchno-Tekhnicheskaya Informatiya, Seriya 2* **24**, 21–29 (1990)
21. Lozin, V.: On maximum induced matchings in bipartite graphs. *Inf. Process. Lett.* **81**(1), 7–11 (2002)
22. Rowley, D.: PC/BEAGLE. *Expert. Syst.* **7**(1), 58–62 (1990)
23. Schlimmer, J.: *Mushroom Records Drawn from the Audubon Society Field Guide to North American Mushrooms*. GH Lincoff (Pres), New York (1981)
24. Seshapanu, J.: Students performance in exams, November 2018. <https://www.kaggle.com/spscientist/students-performance-in-exams>
25. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with titanic. *Data Knowl. Eng.* **42**(2), 189–222 (2002)
26. Wille, R.: Lattices in data analysis: how to draw them with a computer. In: Rival, I. (ed.) *Algorithms and Order*. NATO ASI Series (Series C: Mathematical and Physical Sciences), vol. 255, pp. 33–58. Springer, Dordrecht (1989). https://doi.org/10.1007/978-94-009-2639-4_2
27. Xiao, M., Tan, H.: Exact algorithms for maximum induced matching. *Inf. Comput.* **256**, 196–211 (2017)