Exploring Scale-Measures of Data Sets

Tom Hanika[1,2] and Johannes Hirth[1,2(✉)]

[1] Knowledge and Data Engineering Group, University of Kassel, Kassel, Germany
{tom.hanika,hirth}@cs.uni-kassel.de
[2] Interdisciplinary Research Center for Information System Design,
University of Kassel, Kassel, Germany

Abstract. Measurement is a fundamental building block of numerous scientific models and their creation in data driven science. Due to the high complexity and size of modern data sets, it is necessary to develop understandable and efficient scaling methods. A profound theory for scaling data is scale-measures from formal concept analysis. Recent developments indicate that the set of all scale-measures for a given data set constitutes a lattice. In this work we study the properties of said lattice and propose a novel and efficient scale-measure exploration algorithm, motivating multiple applications for (semi-)automatic scaling.

Keywords: Measurements · Data scaling · Formal concepts · Lattices

1 Introduction

An inevitable step of any data-based knowledge discovery process is *measurement* [14] and the associated (explicit or implicit) *scaling* of the data [17]. The latter is particularly constrained by the underlying mathematical formulation of the data representation, e.g., real-valued vector spaces or weighted graphs, the requirements of the data procedures, e.g., the presence of a distance function, and, more recently, the need for human understanding of the results. Considering the scaling of data as part of the analysis itself, in particular formalizing it and thus making it controllable, is a salient feature of formal concept analysis (FCA) [5]. This field of research has spawned a variety of specialized scaling methods, such as logical scaling [15], and in the form of *scale-measures* links the scaling process with the study of *continuous mappings* between *closure systems*.

Recent results by the authors [11] revealed that the set of all scale-measures for a given data set constitutes a lattice. Furthermore, it was shown that any scale-measure can be expressed in simple propositional terms using disjunction, conjunction and negation. Among other things, the previous results allow a computational transition between different scale-measures, which we may call *scale-measure navigation*, as well as their *interpretability* by humans. Despite these advances, the question of how to identify appropriate and meaningful scale-measures for a given data set with respect to a human data analyst remains

---

Authors are given in alphabetical order. No priority in authorship is implied.

unanswered. In this paper, we propose an answer by adapting the well-known *attribute exploration algorithm* from FCA to present a method for exploring scale measures. Very similar to the original algorithm does *scale-measure exploration* inquire a (human) scaling expert for how to aggregate, separate, omit, or introduce data set features. Our efforts do finally result in a (semi-)automatic scaling framework. Please note, some proofs are outsourced to the journal version [10].

2   Scales and Measurement

Formalizing and understanding the process of *measurement* is, in particular in data science, an ongoing discussion, for which we refer the reader to *Representational Theory of Measurement* [13,18] as well as *Numerical Relational Structure* [14], and *algebraic (measurement) structures* [16, p. 253]. Formal concept analysis (FCA) is well equipped to handle and comprehend data scaling tasks. Within this work we use the standard notation, as introduced by B. Ganter and R. Wille [5]. A fundamental approach to comprehensible scaling, in particular for nominal and ordinal data as studied in this work, is the following.
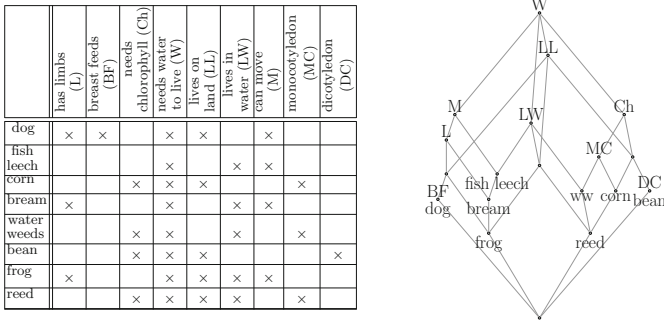

	has limbs (L)	breast feeds (BF)	needs chlorophyll (Ch)	needs water to live (W)	lives on land (LL)	lives in water (LW)	can move (M)	monocotyledon (MC)	dicotyledon (DC)
dog	×	×		×	×		×		
fish				×		×	×		
leech				×		×	×		
corn			×	×	×			×	
bream	×			×		×	×		
water weeds			×	×		×		×	
bean			×	×	×				×
frog	×			×	×	×	×		
reed			×	×	×	×		×	

Fig. 1. This figure shows the *Living Beings and Water* context in the top. Its concept lattice is displayed at the bottom and contains nineteen concepts.

Definition 1 (Scale-Measure (cf. Definition 91, [5])). *Let $\mathbb{K} = (G, M, I)$ and $\mathbb{S} = (G_\mathbb{S}, M_\mathbb{S}, I_\mathbb{S})$ be formal contexts. The map $\sigma : G \to G_\mathbb{S}$ is called an $\mathbb{S}$-measure of $\mathbb{K}$ into the scale $\mathbb{S}$ iff the preimage $\sigma^{-1}(A) := \{g \in G \mid \sigma(g) \in A\}$ of every extent $A \in Ext(\mathbb{S})$ is an extent of $\mathbb{K}$.*

This definition resembles the idea of *continuity between closure spaces* (G_1, c_1) and (G_2, c_2). We say that the map $f : G_1 \to G_2$ is *continuous* if and only if for all $A \in \mathcal{P}(G_2)$ we have $c_1(f^{-1}(A)) \subseteq f^{-1}(c_2(A))$. In the light of the definition above we understand σ as an interpretation of the objects from $\mathbb{K}$ in $\mathbb{S}$. Therefore we view the set $\sigma^{-1}(Ext(\mathbb{S})) := \bigcup_{A \in Ext(\mathbb{S})} \sigma^{-1}(A)$ as the set of extents that is *reflected* by the scale context $\mathbb{S}$.

We present in Fig. 2 the scale-context for some scale-measure and its concept lattice, derived from our *running example* context *Living Beings and Water* $\mathbb{K}_W$, cf. Fig. 1. This scaling is based on the original object set G, however, the attribute set is comprised of nine, partially new, elements, which may reflect species taxons. We observe in this example that the concept lattice of the scale-measure context reflects twelve out of the nineteen concepts from $\mathfrak{B}(\mathbb{K}_W)$.

	W	LW	plants	animals	land plants	water plants	land animal	water animal	mammal
dog	×			×			×		×
fish leech	×	×		×				×	
corn	×		×		×				
bream	×	×		×				×	
water weeds	×	×	×			×			
bean	×		×		×				
frog	×	×		×			×	×	
reed	×	×	×		×	×			

plants := Ch
animals := M
land plants := LL ∧ plant
water plants := LW ∧ plant
land animal := LL ∧ animal
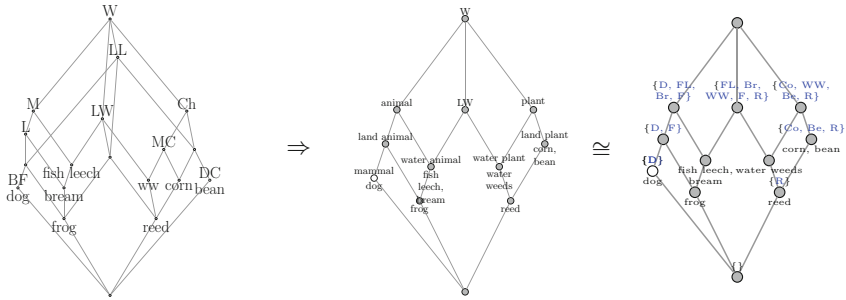water animal := LW ∧ animal
mammal := animal ∧ BF

Fig. 2. A scale context (top), its concept lattice (bottom middle) for which id_G is a scale-measure of the context in Fig. 1. The reflected extents by the scale $\sigma^{-1}(\mathrm{Ext}(\mathbb{S}))$ of the scale-measure are indicated in gray in the contexts concept lattice (bottom left). The concept lattice on the bottom right is the output scale of the scale-measure exploration algorithm and is discussed in Sect. 4. The employed object order is: Be > Co > D > WW > FL > Br > F > R

In our work [11] we derived a *scale-hierarchy* on the set of scale-measures, i.e., $\mathfrak{S}(\mathbb{K}) := \{(\sigma, \mathbb{S}) \mid \sigma \text{ is a } \mathbb{S}-\text{measure of } \mathbb{K}\}$, from a natural order of scales introduced by Ganter and Wille [5, Definition 92]). We say for two scale-measures $(\sigma, \mathbb{S}), (\psi, \mathbb{T})$ that $(\psi, \mathbb{T})$ is coarser then $(\sigma, \mathbb{S})$, iff $\psi^{-1}(\mathrm{Ext}(\mathbb{T})) \subseteq \sigma^{-1}(\mathrm{Ext}(\mathbb{S}))$, denoted $(\psi, \mathbb{T}) \leq (\sigma, \mathbb{S})$. This also yields a natural equivalence relation $\sim$, which in turn, allowed for the definition [11] of a *scale-hierarchy* $\underline{\mathfrak{S}}(\mathbb{K}) = (\mathfrak{S}(\mathbb{K})/\sim, \leq)$. For any given context $\mathbb{K}$, the scale-hierarchy is lattice ordered [11] and isomorphic to the set of all sub-closure systems of $\mathrm{Ext}(\mathbb{K})$ ordered by $\subseteq$. To show this, we defined a *canonical representation* of scale-measures, using the so-called *canonical scale* $\mathbb{K}_{\mathcal{A}} := (G, \mathcal{A}, \in)$ for $\mathcal{A} \subseteq \mathrm{Ext}(\mathbb{K})$ with $\mathrm{Ext}(\mathbb{K}_{\mathcal{A}}) = \mathcal{A}$. In fact, for context $\mathbb{K}$ and $(\mathbb{S}, \sigma) \in \mathfrak{S}(\mathbb{K})$ a scale-measure, then $(\sigma, \mathbb{S}) \sim (\mathrm{id}, \mathbb{K}_{\sigma^{-1}(\mathrm{Ext}(\mathbb{S}))})$.

We argued in [11] that the canonical representation eludes human explanation to some degree. To remedy this issue by means of logical scaling [15] we used scales with logical attributes $M_\mathbb{S} \subseteq \mathcal{L}(M, \{\wedge, \vee, \neg\})$ ([11, Problem 1]). Let $m \in M$, then we define $I_m = I \cap G \times \{m\}$, $gI_{\phi_1 \wedge \phi_2}(\phi_1 \wedge \phi_2)$ iff $gI_{\phi_1}\phi_1 \wedge gI_{\phi_2}\phi_2$), $gI_{\phi_1 \vee \phi_2}(\phi_1 \vee \phi_2)$ iff $gI_{\phi_1}\phi_1 \vee gI_{\phi_2}\phi_2$, and $gI_{\neg\phi}(\neg\phi)$ iff not $gI_\phi\phi$.

Proposition 1 (Conjunctive Normalform (cf. Proposition 23, [11])). *Let $\mathbb{K}$ be a context, $(\sigma, \mathbb{S}) \in \mathfrak{S}(\mathbb{K})$. Then the scale-measure $(\psi, \mathbb{T}) \in \mathfrak{S}(\mathbb{K})$ given by $\psi = \mathrm{id}_G$ and $\mathbb{T} = |_{A \in \sigma^{-1}(Ext(\mathbb{S}))}(G, \{\phi = \wedge A^I\}, I_\phi)$ is equivalent to $(\sigma, \mathbb{S})$ and is called* conjunctive normalform *of $(\sigma, \mathbb{S})$.*

3   Ideals in the Lattice of Closure Systems

The goal for the main part is to identify *outstanding* and particularly interesting data scalings. This quest leads to the natural question for a structural understanding of the scale-hierarchy. In order to do this we rely on the isomorphism [11, Proposition 11] between a context's scale-hierarchy $\underline{\mathfrak{S}}(\mathbb{K})$ and the lattice of all sub-closure systems of $Ext(\mathbb{K})$. The latter forms an order ideal, denoted by $\downarrow_{\mathfrak{F}_G} Ext(\mathbb{K})$, in the lattice $\mathfrak{F}_G$ of all closure systems on G. This ideal is well-studied [1]. We may omit G in $\downarrow_{\mathfrak{F}_G} Ext(\mathbb{K})$ to improve the readability.

Equipped with this structure we have to recall a few notions and definitions for a complete lattice $(L, \leq)$. In the following, we denote by $\prec$ the *cover relation* of $\leq$. Furthermore, we say L is 1) *lower semi-modular* if and only if $\forall x, y \in L : x \prec x \vee y \implies x \wedge y \prec y$, 2) *join-semidistributive* iff $\forall x, y, z \in L : x \vee y = x \vee z \implies x \vee y = x \vee (y \wedge z)$, 3) *meet-distributive* (*lower locally distributive*, cf [1]) iff L is join-semidistributive and lower semi-modular, 4) *join-pseudocomplemented* iff for any $x \in L$ the set $\{y \in L \mid y \vee x = \top\}$ has a least, 6) *ranked* iff there is a function $\rho : L \mapsto \mathbb{N}$ with $x \prec y \implies \rho(x) + 1 = \rho(y)$, 7) *atomistic* iff every $x \in L$ can be written as the join of atoms in L. In addition to the just introduced lattice properties, there are properties for elements in L that we consider. An element $x \in L$ is 1) *neutral* iff every triple $\{x, y, z\} \subseteq L$ generates a distributive sublattice of L, 2) *distributive* iff the equalities $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ for every $y, z \in L$ hold, 3) *meet irreducible* iff $x \neq \top$ and $\bigwedge_{y \in Y} y$ for $Y \subseteq L$ implies $x \in Y$, 4) *join irreducible* iff $x \neq \bot$ and $\bigvee_{y \in Y} y$ for $Y \subseteq L$ implies $x \in Y$. For the rest of this work, we denote by $\mathcal{M}(L)$ the set of all meet-irreducible elements of L.

We can derive from literature [1, Proposition 19] the following statement.

Corollary 1. *For $\mathbb{K} = (G, M, I)$, $\downarrow Ext(\mathbb{K}) \subseteq \mathfrak{F}_G$ and $\mathcal{R}, \mathcal{R}' \in \downarrow Ext(\mathbb{K})$ that: $\mathcal{R}' \prec \mathcal{R} \iff \mathcal{R}' \cup \{A\} = \mathcal{R}$ with A is meet-irreducible in $\mathcal{R}$.*

Of special interest in lattices are the (meet-) join-irreducibles, since every element of a lattice can be represented as a (meet) join of these elements.

Proposition 2. *For $\mathbb{K}$, $\downarrow Ext(\mathbb{K}) \subseteq \mathfrak{F}_G$ and $\mathcal{R} \in \downarrow Ext(\mathbb{K})$: $\mathcal{R}$ is join-irreducible in $\downarrow Ext(\mathbb{K}) \iff \exists A \in Ext(\mathbb{K}) \setminus \{G\} : \mathcal{R} = \{G, A\}$*

Next, we investigate the meet-irreducibles of $\downarrow \text{Ext}(\mathbb{K})$ using a similar approach as done for $\mathfrak{F}_G$ [1] based on propositional logic. We recall, that an (object) implication for some context $\mathbb{K}$ is a pair $(A, B) \in \mathcal{P}(G) \times \mathcal{P}(G)$, shortly denoted by $A \to B$. We say $A \to B$ is valid in $\mathbb{K}$ iff $A' \subseteq B'$. The set $\mathcal{F}_{A,B} := \{D \subseteq G : A \nsubseteq D \vee B \subseteq D\}$ is the set of all *models* of $A \to B$. Additionally, $\mathcal{F}_{A,B}|_{\text{Ext}(\mathbb{K})} := \mathcal{F}_{A,B} \cap \text{Ext}(\mathbb{K})$ is the set of all extents $D \in \text{Ext}(\mathbb{K})$ that are models of $A \to B$. The set $\mathcal{F}_{A,B}$ is a closure system [1] and therefor $\mathcal{F}_{A,B}|_{\text{Ext}(\mathbb{K})}$, too. Furthermore, we can deduce that $\mathcal{F}_{A,B}|_{\text{Ext}(\mathbb{K})} \in \downarrow \text{Ext}(\mathbb{K})$.

Lemma 1. *For context $\mathbb{K}$, $\downarrow \text{Ext}(\mathbb{K}) \subseteq \mathfrak{F}_G$, $\mathcal{R} \in \downarrow \text{Ext}(\mathbb{K})$ with closure operator $\phi_\mathcal{R}$ we find $\mathcal{R} = \bigcap\{\mathcal{F}_{A,B}|_{\text{Ext}(\mathbb{K})} \mid A, B \subseteq G \wedge B \subseteq \phi_\mathcal{R}(A)\}$.*

For $\mathcal{R} \in \downarrow \text{Ext}(\mathbb{K})$ the set $\{\mathcal{F}_{A,B}|_{\text{Ext}(\mathbb{K})} \mid A, B \subseteq G \wedge B \subseteq \phi_\mathcal{R}(A)\}$ contains only closure systems in $\downarrow \text{Ext}(\mathbb{K})$ and thus possibly meet-irred. elements of $\downarrow \text{Ext}(\mathbb{K})$.

Proposition 3. *For context $\mathbb{K}$, $\downarrow \text{Ext}(\mathbb{K}) \subseteq \mathfrak{F}_G$ and $\mathcal{R} \in \downarrow \text{Ext}(\mathbb{K})$: 1. $\mathcal{R}$ is meet-irreducible in $\downarrow \text{Ext}(\mathbb{K})$ 2. $\exists A \in \text{Ext}(\mathbb{K}), i \in G$ with $A \prec_{\text{Ext}(\mathbb{K})} (A \cup \{i\})''$ such that $\mathcal{R} = \mathcal{F}_{A,\{i\}}|_{\text{Ext}(\mathbb{K})}$*

Propositions 2 and 3 provide a characterization of irreducible elements in $\downarrow\text{Ext}(\mathbb{K})$ and thereby in the scale-hierarchy of $\mathbb{K}$. Those may be of particular interest, since any element of $\downarrow\text{Ext}(\mathbb{K})$ is representable by irreducible elements. Equipped with this characterization we look into counting the irreducibles.

Proposition 4. *For context $\mathbb{K}$, the number of meet-irreducible elements in the lattice $\downarrow \text{Ext}(\mathbb{K}) \subseteq \mathfrak{F}_G$ is equal to $|\prec_{\downarrow \text{Ext}(\mathbb{K})}|$.*

Next, we turn ourselves to other lattice properties of $\downarrow \text{Ext}(\mathbb{K})$.

Lemma 2 (Join Complement). *For $\mathbb{K}$, $\downarrow \text{Ext}(\mathbb{K}) \subseteq \mathfrak{F}_G$ and $\mathcal{R} \in \downarrow \text{Ext}(\mathbb{K})$, the set $\hat{\mathcal{R}} = \bigvee_{A \in \mathcal{M}(\text{Ext}(\mathbb{K})) \backslash \mathcal{M}(\mathcal{R})} \{A, G\}$ is the inclusion minimum closure-system for which $\mathcal{R} \vee \hat{\mathcal{R}} = \text{Ext}(\mathbb{K})$.*

All the above results in the following statement about $\downarrow \text{Ext}(\mathbb{K})$:

Proposition 5. *For any context $\mathbb{K}$, the lattice $\downarrow \text{Ext}(\mathbb{K}) \subseteq \mathfrak{F}_G$ is: i) join-semidistributive ii) lower semi-modular iii) meet-distributive iv) join-pseudo-complemented v) ranked vi) atomistic*

4   Recommending Conceptual Scale-Measures

For the task of efficiently determining a scale-measure, based on human preferences, we propose the following approach. Motivated by the representation of meet-irreducible elements in the scale-hierarchy through object implications of the context ((Proposition 3), we employ the dual of the *attribute exploration* algorithm [6] by Ganter. We modified said algorithm toward exploring scale-measures

---

Algorithm 1: Scale-measure Exploration: A modified Exploration with Background Knowledge

---

Input   : Context $\mathbb{K} = (G, M, I)$
Output: $(id_G, \mathbb{S}) \in \mathfrak{S}(\mathbb{K})$ and optionally $\mathcal{L}_{\mathbb{S}}$
Init Scale $\mathbb{S} = (G, \emptyset, \in)$
Init $A = \emptyset$, $\mathcal{L}_{\mathbb{S}} = \mathrm{CanonicalBase}(\mathbb{K})$ (or $\mathcal{L}_{\mathbb{S}} = \{\}$ for larger contexts)
while $A \neq G$ **do**
    while $A \neq A^{I_{\mathbb{S}} I_{\mathbb{S}}}$ **do**
        if *Further differentiate objects having* $A^{I_{\mathbb{S}} I_{\mathbb{S}} I_{\mathbb{K}}}$
        by attributes in $A^{I_{\mathbb{K}}} \setminus A^{I_{\mathbb{S}} I_{\mathbb{S}} I_{\mathbb{K}}}$*?* **then**
            $\mathcal{L}_{\mathbb{S}} = \mathcal{L}_{\mathbb{S}} \cup \{A \to A^{I_{\mathbb{S}} I_{\mathbb{S}}}\}$
            Exit While
        else
            Enter $B \subseteq A^{I_{\mathbb{K}}} \setminus (A)^{I_{\mathbb{S}} I_{\mathbb{S}} I_{\mathbb{K}}}$ that should be considered
            Add attribute $B^{I_{\mathbb{K}}}$ to $\mathbb{S}$
    $A = \mathrm{Next\_Closure}(A, G, \mathcal{L}_{\mathbb{S}})$
return : $(id_G, \mathbb{S})$ and optionally $\mathcal{L}$

---

and present its pseudo-code in Algorithm 1. In this depiction we highlighted our modifications with respect to the original exploration algorithm (Algorithm 19, [7]) with darker print. This algorithm semi-automatically computes a scale context $\mathbb{S}$ and its canonical base. In each iteration of the inner loop of our algorithm the query that is stated to the *scaling expert* is if an object implication $A \to B$ is true in the closure system of user preferences. If the implication holds, it is added to the implicational base of $\mathbb{S}$ and the algorithm continues with the next query. Otherwise a counter example in the form of a closed set $C \in \mathrm{Ext}(\mathbb{K})$ with $A \subseteq C$ but $B \not\subseteq C$ has to be constructed. This closed set is then added as attribute to the scale context $\mathbb{S}$ with the incidence given by $\in$. If $C \notin \mathrm{Ext}(\mathbb{K})$ the scale $\mathbb{S}$ would contradict the scale-measure property (Proposition 20, [11]).

The object implicational theory $L_{\mathbb{S}}$ is initialized to the object canonical base of $\mathbb{K}$, which is an instance of according to attribute exploration with background knowledge [6]. This initialization can be neglected for larger contexts, however it may reduce the number of queries. The algorithm terminates when the implication premise of the query is equal to G. The returned scale-measure is in canonical form, i.e., the canonical representation $(id_G, (G, \mathrm{Ext}(\mathbb{S}), \in))$. The motivation behind exploration queries is to determine if an implication holds in the unknown representational context of the learning domain. In contrast, the exploration of scale-measures determines if a given $\mathrm{Ext}(\mathbb{K})$ can be coarsened by implications $A \implies B$, resulting in a smaller and thus more human-comprehensible concept lattice $\underline{\mathfrak{B}}(\mathbb{S})$, adjusted to the preferences of the scaling expert.

Querying object implications may be less intuitive compared to attribute implications, hence, we suggest to rather not test for $A \implies A^{I_{\mathbb{S}} I_{\mathbb{S}}}$ for $A \subseteq G$ but for the difference of the intents $A^{I_{\mathbb{K}}}$ and $(A^{I_{\mathbb{S}} I_{\mathbb{S}}})'$ in $\mathbb{K}$. Finally, as a post-processing, one may apply the *conjunctive normalform* [11, Proposition 23] of

scale-measures to improve human-comprehension. Yet, deriving other human-comprehensible representations of scale-measures is deemed to be future work.

(Semi-)Automatic Large Data Set Scaling. To demonstrate the applicability of the presented algorithm, we have implemented it in the `conexp-clj` ([9]) software suite. For this, we apply the scale-measure exploration Algorithm 1 on our running example $\mathbb{K}_W$, see Fig. 1. The evaluation steps of this algorithm are displayed in more detail in the long version of this work. One such intermediate step is for example row two where the implication $\{\} \implies \{D, FL, Br, F\}$ is true in the so far generated scale $\mathbb{S}$ and it is queried if it should hold. All objects of the implication do have at least the attributes *can move* and *needs water to live*. The answer of the scaling expert envisioned by us is the attribute *lives on land*. Thus, the object counter example is the attribute-derivation the union $\{M, W, LL\}^{I_W} = \{D, F\}$. In our example of the scale-measure exploration the algorithm terminates after the scaling expert provided in total nine counter examples and four accepts. The output is a scale context in canonical representation with twelve concepts as depicted in Fig. 2 (bottom right).

5   Related Work

Measurement is an important field of study in many (scientific) disciplines that involve the collection and analysis of data. A framework to describe and analyze the measurement for Boolean data sets has been introduced in [8] and [4], called *scale-measures*. It characterizes the measurement based on object clusters that are formed according to common feature (attribute) value combinations. An accompanied notion of dependency has been studied [19], which led to attribute selection based measurements of boolean data. The formalism includes a notion of consistency enabling the determination of different views and abstractions, called *scales*, to the data set. Despite the expressiveness of scale-measure framework, as demonstrated in this work, it is so far insufficiently studied in the literature. In particular algorithmical and practical calculation approaches are missing. Comparable and popular machine learning approaches, such as feature compression techniques, e.g., *Latent Semantic Analysis* [2,3], have the disadvantage that the newly compressed features are not interpretable by means of the original data and are not guaranteed to be consistent with said original data. The methods presented in this paper do not have these disadvantages, as they are based on meaningful and interpretable features with respect to the original features. In particular preserving consistency, as we did, is not a given, which was explicitly investigated in the realm scaling many-valued formal contexts [15] and implicitly studied for generalized attributes [12].

6   Conclusion

With this work we have shed light on the hierarchy of scale-measures. By applying multiple results from lattice theory, especially concerning ideals, to said hierarchy, we were able to give a more thorough structural description of $\downarrow_{\mathfrak{F}_G} \text{Ext}(\mathbb{K})$.

Our main theoretical result is Proposition 5, which in turn leads to our practical applications. In particular, based on this deeper understanding we were able to present an algorithm for exploring the scale-hierarchy of a binary data set $\mathbb{K}$. Equipped with this algorithm a data scaling expert may explore the lattice of scale-measures for a given data set with respect to her preferences and the requirements of the data analysis task. The practical evaluation and optimization of this algorithm is a promising goal for future investigations. Even more important, however, is the implementation and further development of the automatic scaling framework, as outlined in Sect. 4. This opens the door to empirical scale recommendation studies and a novel approach for data preprocessing.

References

1. Caspard, N., Monjardet, B.: The lattices of closure systems, closure operators, and implicational systems on a finite set: a survey. Discrete Appl. Math. **127**(2), 241–269 (2003). http://www.sciencedirect.com/science/article/B6TYW-46VBMT3-1/2/d6bbb12c7831e113c5e91bebc320bee1

2. Codocedo, V., Taramasco, C., Astudillo, A.: Cheating to achieve formal concept analysis over a large formal context. In: Napoli, A., Vychodil, V. (eds.) CLA, vol. 959, pp. 349–362. CEUR-WS.org (2011)

3. Dumais, S.T.: Latent semantic analysis. Annu. Rev. Inf. Sci. Technol. **38**(1), 188–230 (2004). https://doi.org/10.1002/aris.1440380105

4. Ganter, B., Wille, R.: Conceptual scaling. In: Robert, F. (ed.) Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. IMA, vol. 17, pp. 139–167. Springer-Verlag, Heidelberg (1989). https://doi.org/10.1007/978-1-4684-6381-1_6

5. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, Berlin (1999). https://doi.org/10.1007/978-3-642-59830-2

6. Ganter, Bernhard: Attribute exploration with background knowledge. Theor. Comput. Sci. **217**(2), 215–233 (1999). https://doi.org/10.1016/S0304-3975(98)00271-0

7. Ganter, B., Obiedkov, S.: More expressive variants of exploration. In: Conceptual Exploration, pp. 237–292. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49291-8_6

8. Ganter, B., Stahl, J., Wille, R.: Conceptual measurement and many-valued contexts. In: Gaul, W., Schader, M. (eds.) Classification as a Tool of Research, pp. 169–176. North-Holland (1986)

9. Hanika, T., Hirth, J.: Conexp-clj - A research tool for FCA. In: Cristea, D., Le Ber, F., Missaoui, R., Kwuida, L., Sertkaya, B. (eds.) Supplementary Proceedings of ICFCA 2019, volume 2378 of CEUR WS Proceedings, pp. 70–75. CEUR-WS.org (2019). http://ceur-ws.org/Vol-2378/shortAT8.pdf

10. Hanika, T., Hirth, J.: Exploring scale-measures of data sets (2021)

11. Hanika, T., Hirth, J.: On the lattice of conceptual measurements. arXiv preprint arXiv:2012.05287 (2020)

12. Kwuida, L., et al.: Generalized pattern extraction from concept lattices. Ann. Math. Artif. Intell. 72(1-2), 151–168 (2014)

13. Duncan Luce, R., Krantz, D.H., Suppes, P., Tversky, A.: Foundations of Measurement - Representation, Axiomatization, and Invariance, vol. 3. Academic Press (1990)

14. Pfanzagl, J.: Theory of Measurement. Physica, Heidelberg (1971)
15. Prediger, S., Stumme, G.: Theory-driven logical scaling: Conceptual information systems meet description logics. In Enrico Franconi and Michael Kifer, editors, Proc. KRDB'99, volume 21, pages 46–49. CEUR-WS.org, 1999. URL http://ceur-ws.org/Vol-21/prediger.ps
16. Fred, S.R.: Measurement Theory. Cambridge University Press, Cambridge (1984)
17. Stevens, S.S.: On the theory of scales of measurement. Science **103**(2684), 677–680 (1946). ISSN 0036–8075. https://doi.org/10.1126/science.103.2684.677, https://science.sciencemag.org/content/103/2684/677
18. Suppes, P., Krantz, D.H., Duncan Luce, R., Tversky, A.: Foundations of Measurement - Geometrical, Threshold, and Probabilistic Representations, vol. 2. Academic Press (1989)
19. Wille, R.: Dependencies of many valued attributes. In: Bock, H.-H. (ed.) Classification and related methods of data analysis, pp. 581–586. North-Holland (1988)