# Reducing the Size of Concept Lattices: The JBOS Approach

Sérgio M. Dias and Newton J. Vieira

Department of Computer Science
Federal University of Minas Gerais (UFMG)
Av. Antônio Carlos 6627 - ICEx - 4010 - Pampulha ZIP: 31.270-010
Belo Horizonte, Minas Gerais, Brazil
{sergiomariano}@gmail.com
{nvieira}@dcc.ufmg.br

**Abstract.** Formal concept analysis (FCA) is used nowadays in a large number of applications in several different areas. Nevertheless, in some applications the size of a concept lattice may be prohibitively large, thus creating a need for new approaches and algorithms to reduce concept lattices to a manageable size. This paper presents a new method for reducing a concept lattice's complexity and introduces a new methodology for evaluating reduction methods. The main idea of the proposed approach, junction based on objects similarity (JBOS), is the substitution of groups of "similar" objects by prototypical ones. As a consequence, the resulting lattice is simplified in its size and structure. The user has control of the degree of simplification by means of parameters specification. Two measures are proposed for the evaluation of the final lattice with respect to the original one, the first related to the consistency of the reduced lattice and the other to the descriptive losses incurred by the reduction. The JBOS approach was applied to two data sets from the UCI machine learning repository. The results show that it is possible to reduce the size of a concept lattice in a controlled way. In particular, it makes possible the use of FCA in a data set (formal context) for which that use was previously impossible.

**Key words:** Formal Concept Analysis, Concept Lattices

## 1 Introduction

Formal concepts analysis (FCA) is a technique of applied mathematics based on the mathematization of the notion of concept, consisting of intention and extension, and on the organization of the concepts through a conceptual hierarchy[8]. Its formalization was born in 1982 with the work presented by Wille[18], who proposed to consider each lattice element as a formal concept and the lattice representing a conceptual hierarchy. Currently there is a growing interest in applications of FCA in information retrieval[12], data mining[16], neural networks[19], social networks[11], etc.

The great potential of FCA is provided by the organization of knowledge into a conceptual hierarchy. In fact, the main applications make use of this hierarchy representing it by means of a line diagram, a nested line diagram, a tree diagram etc. However, generating all formal concepts and classifying them hierarchically presents, in the worst case, an exponential behavior[14]. Although this behavior is rarely found in practical cases[10], the computational cost can be prohibitive for many applications.

In general, techniques for reducing the complexity of the concept lattices can be applied to the formal context or to the concept lattice itself. A simple way to reduce the size of the concept lattice would be, for example, the creation of a cutting threshold, from which only the formal concepts considered "relevant" would be selected and preserved (of course in this case it would be necessary to develop some criterion of relevance). A simple kind of reduction applied to the formal context would be the junction of lines and/or columns as suggested by Ganter and Wille[8]: if for two objects $g$ and $h$, $g' = h'$ then $g$ and $h$ can be replaced by one single object; dually, if for two attributes $m$ and $n$, $m' = n'$, then $m$ and $n$ can be replaced by one single attribute. Despite reducing the formal context, this method does not reduce the size of the concept lattice.

Another interesting possibility is to apply techniques of the KDD process (knowledge discovery in databases)[15] to reduce the size of the formal context. Gajdos et al.[7] and Karen et al.[4] propose the use of SVD (singular value decomposition), which is a technique capable of reducing the dimensionality of data masses. Another very interesting technique for reducing the concept lattice based on the formal context was presented by Belohlávek et al[2]. The authors propose to use the knowledge of the user to create restrictions on attributes of the formal context. These restrictions, called AD-formulas (attribute-dependency formulas), are used during the generation of formal concepts and concept lattice construction to retain only the formal concepts conforming to them.

Among the main techniques applied directly to the concept lattice, there is that of Arévalo et al.[1] which proposes the creation of a *Galois Subhierarchy*. In this case, the only formal concepts selected are $C^g \cup C^m$, where $C^g = \{(g'', g')|g \in G\}$ and $C^m = \{(m'', m')|m \in M\}$.

Another option to reduce the complexity of the concept lattice is to associate the concept of frequent items from the KDD process, to frequent formal concepts[15]. A formal concept $(X, Y)$ is considered frequent if, only if, $sup(Y, G) \geq minSup$, where $sup(Y, G)$ (the support) is the number of objects in $G$ which contains the attributes in $Y$ and $minSup$ is a minimal support provided as a parameter by the user. Another very interesting technique for reducing complexity is based on the concept of stability proposed by Kuznetsov[13]. Unlike other methods of reduction, the stability method seeks to create an index for concepts, indicating how much the intent of the concept depends on the set of objects. Using this index, one can assume a threshold at which all the formal concepts with lower values are removed. Recently, Jay et al.[11] applied the concept of stability and iceberg lattices in social network analysis.

Despite the range of techniques to reduce the complexity of concept lattices, no methodologies were found for the analysis of these techniques. Obviously, any proposal should be accompanied by measuring of how the resulting concept lattice relates to the original lattice. Thus, this paper presents a methodology based on two metrics. The first aims to evaluate the degree of consistency of the reduced lattice with respect to the original data. The second aims to assess the loss of the ability to discriminate between objects features.

We propose a new method for reducing the complexity of the concept lattice, called *junction based on objects similarity* (JBOS), which seeks to replace groups of similar objects by representative objects. The method was applied to two databases of the UCI machine learning repository, one which leads to a concept lattice of moderate size and the other which leads to a very large concept lattice. The results were evaluated through the proposed methodology, which showed for each of the databases, the most promising levels of reduction based on the values for the parameters of the JBOS method. In particular, the JBOS method was able to reduce the complexity of the large concept lattice, making plausible its analysis on the reduced form.

This paper is organized in five sections. In the next section a short review of the formal concept analysis is presented mainly in order to fix the terminology. In the third section, the method for reducing the complexity of concept lattices is presented and the methodology for the evaluation of reduction methods is proposed. In the fourth one, a case study is presented. Finally, the contributions and the conclusions of this work are presented.

## 2　Formal Concept Analysis - A short review

This very short review is mainly based on the notions and terminology exposed on [8]. Formal concept analysis is a field of mathematics which was born in the early eighties[3,8]. Its main characteristic is knowledge representation by means of specific diagrams called line diagrams (based on Hasse diagrams), which are what mathematicians call diagrams of (concept) lattices. In FCA, *formal contexts* are primordial elements. A formal context is a triplet $(G, M, I)$, where $G$ is a set whose elements are called *objects*, $M$ is a set whose members are called *attributes* and $I \subseteq G \times M$ is a relation termed *incidence relation*. If $(g, m) \in I$, one says that "*the object g has the attribute m*". A formal context is usually represented by a cross table where the objects are rows headers, the atributes are columns headers and there is a cross in row $g$ and column $m$ if and only if $(g, m) \in I$.

Given a set of objects $A \subseteq G$ from a formal context $(G, M, I)$, it could be asked which attributes from $M$ are common to all those objects. Similarly, it could be asked, for a set $B \subseteq M$, which objects have all the attributes from $B$. These questions are answered by the *derivation operators*, defined as: $A' = \{m \in M | \forall g \in A (g, m) \in I\}$ and $B' = \{g \in G | \forall m \in B (g, m) \in I\}$.

A *formal concept* is a pair $(A, B) \in \mathcal{P}(G) \times \mathcal{P}(M)$ such that $A' = B$ and $B' = A$, where $A$ is called the *extent* and $B$ the *intent* of the concept. The set of formal concepts is ordered by the partial order $\leq$ such that for any two formal

concepts $(A_1, B_1)$ and $(A_2, B_2)$, $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$ (and, in this case, $B_2 \subseteq B_1$). The set of concepts ordered by $\leq$ constitutes a complete lattice[5] termed *concept lattice*. The concept lattice obtained from a formal context $(G, M, I)$ is denoted by $\mathcal{B}(G, M, I)$.

The first part of the basic theorem on concept lattices says that a concept lattice $\mathcal{B}(G, M, I)$ is a complete lattice in which for any arbitrary set $C \subseteq \mathcal{B}(G, M, I)$ the *infimum* and *supremum* are given by $\bigwedge C = (\bigcap X, (\bigcup Y)'')$ and $\bigvee C = ((\bigcup X)'', \bigcap Y)$, where $X = \{A \,|\, (A, B) \in C\}$ and $Y = \{B \,|\, (A, B) \in C\}$[8].

## 3 Complexity reduction

In the development of the method of complexity reduction proposed in this work, it was assumed that the following requirements should be met:

1. avoid creating new objects, attributes or incidences;
2. seek to preserve original conceptual hierarchy at the most;
3. provide a high fidelity with respect to the original knowledge, in other words, the knowledge presented in the reduced concept lattice must be faithful (consistent) with the original; and
4. present low descriptive losses, in other words, seek to preserve the power of discrimination afforded by the range of attributes of the original formal context.

In order to measure the degree of satisfaction given to the last two requirements, a methodology for evaluation was created.

### 3.1 A methodology for evaluating methods of complexity reduction

In order to evaluate techniques for reducing the size of concept lattices, one must somehow compare the information in the formal context and/or in the obtained concept lattice with the information found in the original formal context and/or the respective concept lattice.

The methodology proposed here is based on two types of comparisons. In the first one, the obtained concept lattice, named *reduced concept lattice* is compared with the original formal context, but in an indirect way: from the reduced concept lattice it is produced a set of rules and the "capacity" of this set to take into account all the original objects is assessed. In the second one, the obtained formal context, named *reduced formal context* is evaluated with respect to the original formal context, seeking to assess the loss in the capacity to discriminate the characteristics of objects.

The first evaluation measure requires the production of implication rules from the reduced lattice. A valid implication rule for a certain formal context is an expression $P \rightarrow Q$, where $P$ and $Q$ are sets of attributes of the formal context and $P' \subseteq Q'$, expressing the fact that if an object has attributes in $P$, then it has

those of $Q$[9]. The set of all valid rules for a formal context usually has a high degree of redundancy and its extraction is impractical. On the other hand, the extraction of a minimum non-redundant set may also be impractical[3]. Thus, it is proposed to generate a set of rules possibly not minimal, but with little redundancy, for example, a set of rules *reduced to the left*[17]. A rule $P \to Q$ is said *reduced to the left* in a set of rules $Z$ if for any proper subset $S$ of $P$ the set $(Z - \{P \to Q\}) \cup \{S \to Q\}$ is not equivalent to $Z$[1].

To measure the degree of equivalence between the original formal context and the reduced formal context, two metrics are proposed: *fidelity* and *descriptive loss*. Let $G_o$ be the set of objects from the original formal context, $G_r$ be the set of objects from the reduced formal context and let us assume that $k$ rules are extracted from the reduced concept lattice, rules which will be denoted by $r_1$, $r_2, \ldots, r_k$.

The *fidelity*, $F$, aims to measure the percentage of successfull applications of rules to the original objects. If there is a rule $r_i = P \to Q$ for which $P \subset g'$ and $Q \not\subset g'$ for an object $g \in G_o$, then one says that the rule $r_i$ has *failed*. The fidelity is given by the following equation[2]:

$$F = \frac{\sum_{i=1}^{k} \left(1 - \frac{Nf_i}{|G_o|}\right)}{k} \times 100 \tag{1}$$

where $Nf_i$ is the total number of failures of the rule $r_i$ (i.e. the number of objects in $G_o$ for which the rule fails.). Thus, $F$ is the percentage of non failures considering the $|G_o| \times k$ applications of $k$ rules to $|G_o|$ objects. In order to avoid multiple contributions to $F$ from objects which have exactly the same attributes, such objects must be previously reduced to only one object.

The *descriptive loss*, $DL$, aims to measure the loss of ability to characterize the set of objects $G_o$ due to loss of attributes in the reduction. For its definition an onto function will be used $\nu : G_o \to G_r$ that maps to each object $g$ of $G_o$ an object $\nu(g)$ in $G_r$ as a result of a possible simplification of $g$ due to the reduction process. It's supposed that the object $\nu(g)$ will have as attributes a subset of the $g$ attributes, i.e. $\nu(g)' \subseteq g'$. In particular, if the object $g$ is "eliminated" in the reduction process, $\nu(g)' = \emptyset$. Note that it may happen that $\nu(g) = \nu(h)$ for two different objects $g$ and $h$, and this could be interpreted as saying that $g$ and $h$ stayed with the same attributes after the reduction process, becoming indistinguishable. In particular, all eliminated objects become indistinguishable. The descriptive loss is given by:

$$DL = \left(1 - \frac{\sum_{g \in G_o} \left(\frac{|\nu(g)'|}{|g'|}\right)}{|G_o|}\right) \times 100 \tag{2}$$

where $|\nu(g)'|$ is the number of attributes of the reduced object $\nu(g)$ and $|g'|$ is the number of attributes of the original object $g$. DL, as defined above, can

---

[1] Two sets of rules are said to be equivalent if they have the same set of consequences.
[2] $|C|$ denotes the number of elements of the set $C$.

report descriptive losses when attributes considered "redundant" in the actual application are eliminated. Therefore, such attributes must be removed before applying the reduction method.

### 3.2   Junction Based on Objects Similarity - JBOS

The JBOS method is applied from the formal context. It is based on the junction of "similar" objects, namely the replacement of such objects by a single representative of them. Naturally, the key entity here is the notion of similarity. For its definition, a weight is created for each attribute of the formal context, which seeks to represent the attribute's relevance, and the similarity is determined from the weights of the attributes. As the structure of the concept lattice is derived from the arrangements of attributes of objects, it is expected that the junction of similar objects will preserve at a certain degree (which will depend on the specific assigned weights) a portion of the original structure while providing a certain degree of simplification.

In the rest of this section, let $(G, M, I)$ be a formal context from which a reduced formal context must be produced.

It is assumed that to each attribute $m \in M$ is assigned a weight $w^m$ such that $0 \leq w^m \leq 1$. Such weight can be thought as measuring the significance of the attribute from 0 (no relevance) to 1 (absolute relevance). The weight should be defined by the user based on the semantic load of the attribute, i.e. based on its importance to the model represented by the formal context. The similarity between two objects $g, h \in G$ will also be expressed by a real number between 0 (completely dissimilar) and 1 (completely indistinguishable), so defined:

$$sim(g,h) = \frac{\sum_{m \in M} \mu(g,h,m)}{\sum_{m \in M} w^m}; \mu(g,h,m) = \begin{cases} w^m, \text{ if } (g,m) \in I \leftrightarrow (h,m) \in I \\ 0, \quad \text{otherwise.} \end{cases}$$

$$(3)$$

That is, the similarity between two objects $g$ and $h$ is given by the weighted sum of the weights of attributes in which both objects agree with each other (both have them or both do not have them).

From the similarity matrix it is possible to build an algorithm that makes clusters of similar objects, so that two objects can be in the same group only if they have a certain degree of "similarity". Accordingly, it is suggested the use of a similarity index $\epsilon$, and a maximum number of similar elements $\alpha$, both defined by the user. Two objects $g, h \in G$ will be considered similar by the algorithm, and therefore are likely to stay in the same group if and only if $sim(i, j) \geq \epsilon$. The number $\alpha$ is just one more cutting option for the user: it specifies that each group can contain at most $\alpha$ elements. Algorithm 1 shows how the groups of similar objects are determined. The auxiliary function $GreaterSimilarity(sim, g, G)$ returns an object of $G$ among the most similar to the object $g$; if there is more than one, it chooses any one. The function $ChooseObjectIn(G)$ selects any object from the set $G$. The set of groups of similar objects is denoted by $\gamma$ in the algo-

rithm. Notice that even though *similarity*[3] is a tolerance relation, the algorithm computes "maximal" sets $H$ such that $g$ is similar to $h$ for all $g, h \in H$.

---

**Algorithm 1** Function Group Objects

---

**Input:** a set of objects $G$, a similarity matrix *sim*, a similarity index $\epsilon$ and a maximum number of similar elements $\alpha$.

**Output:** the set of groups of similar objects $\gamma$.

```
 1: γ = ∅
 2: while  G ≠ ∅  do
 3:     g = ChooseObjectIn(G)
 4:     H = {g}
 5:     G = G - {g}
 6:     while  G ≠ ∅ and |H| < α do
 7:         h = GreaterSimilarity(sim,g,G)
 8:         if  sim(h,f) ≥ ε for all f ∈ H  then
 9:             H = H ∪ {h}
10:             G = G - {h}
11:         else
12:             break
13:         end if
14:     end while
15:     γ = γ ∪ {H}
16: end while
17: return  γ
```

---

Let $(G_o, M_o, I_o)$ be the original formal context. Let $\gamma$ be the set obtained by Algorithm 1 from $G_o$ and specific values for $\epsilon$ and $\alpha$. In the reduced formal context each set $H \in \gamma$ will be considered as an object and such object will have the attributes of $M_o$ which are common to all objects in $H$. That is, the attributes of $H$ will be those in $\bigcap\{g'|g \in H\}$, where the operator of derivation is applied from $I_o$. Therefore, by the JBOS method the reduced formal context $(G, M, I)$ is such that: $G = \gamma$; $M = \bigcup\{\bigcap\{g'|g \in H\}|H \in \gamma\}$; and $(H, m) \in I$ if, and only if, $m \in \bigcap\{g'|g \in H\}$.

The main merit of the JBOS method just described is to apply a consistent reduction based on the user's knowledge.If the relevance of the attributes is not known the user must discover it, for example, by correlation analysis, principal components analysis etc. By knowing the relative importance of attributes, the user can set weights that mirror that importance. This knowledge and the definition of the parameters $\epsilon$ and $\alpha$ are then used to reduce the number of objects, which in real applications is often higher than the number of attributes. Consequently, the number of formal concepts and the cost of performing the derivations are also reduced. Therefore, applications once inviable can eventually become viable.

---

[3] Two objects $g$ and $h$ are *similar* to each other iff $sim(g, h) \geq \epsilon$.

## 4   Case Study

The JBS method was analyzed using the proposed methodology on two classic databases from UCI machine learning repository: *Wine Data Set* and *Water Treatment Plant Data Set* (avaliable at: http://www.ics.uci.edu/~mlearn.). The data cleaning process and formal contexts modelling, including the choice of cut points for the discretization of the real parameters, are described in [6]. It is noteworthy that the number of formal concepts for the Water base, about 9 million, is intractable for the existing usual implementations.

**Table 1.** Characteristics of the databases used in the evaluation of the JBOS method.

|                             | Wine   | Water     |
| --------------------------- | ------ | --------- |
| **Number of objects**       | 132    | 316       |
| **Number of attributes**    | 29     | 67        |
| **Density**                 | 48.3%  | 52.2%     |
| **Number of formal concepts** | 29,032 | 8,940,648 |

The JBOS method and the calculations concerning the methodology for complexity reduction techniques evaluation have been implemented using the framework *EF-Concept Analysis* (*Educational Framework for Concept Analysis*) (available at: http://www.dcc.ufmg.br/~mariano)[6]. This framework consists of a set of interfaces, data structures and algorithms that ensures agile development of FCA algorithms.

### 4.1   Applying the JBOS method

As discussed in Section 3.2, the JBOS method requires a set of weights representing the significance of each attribute. In order to determine this set of weights, a deep study of the database or the assistance of a domain expert is necessary. That study is beyond the scope of this work, which seeks only to demonstrate that from a set of weights it is possible to *reduce the formal context* in order to obtain a concept lattice *consistent* with the original. Although the JBOS method is not validated here in terms of specific applications, which must be done in a future work, biased results or very specific results applying to particular databases are avoided. Thus, a set of weights for each of the databases, Wine and Water, is assigned based simply on the frequency of the attributes: *a higher frequency corresponds to a higher weight* (obviously, this criterion may not be appropriate depending on the application that will follow the reduction process.).

The method also needs values for the parameters $\epsilon$ and $\alpha$. The first was set to range from 1 to 0 for each of the bases. The second was set at 20% of the number of objects, which amounts to 26 and 63 for the Wine and Water bases, respectively.

The results for the Wine database can be seen in the graphs of Figures 1 and 2. It should be emphasized that all values are percentages of the respective elements of the original formal context. Initially, note on Figure 1 that for $\epsilon = 0.92$ the number of objects was reduced to 77% of the number of objects from the original formal context and the number of formal concepts to approximately 70% of the number of formal concepts of the original lattice. Note also that for the same value of $\epsilon$ the fidelity was approximately 99% and the descriptive loss about 3.8%, as shown in the graph in Figure 2. Thus, seting $\epsilon = 0.92$ can be an interesting option as it implies a significant reduction of the concept lattice combined with a high fidelity and a low descriptive loss. Indeed, for $\epsilon$ as low as 0.85 the numbers are relatively good. In particular, for $\epsilon = 0.85$ the number of formal concepts drops to 18.46%, the fidelity is 97% and the descriptive loss is approximately 18.8%. Note that despite the large reduction of the number of formal concepts, the measured quality indices may be deemed acceptable.
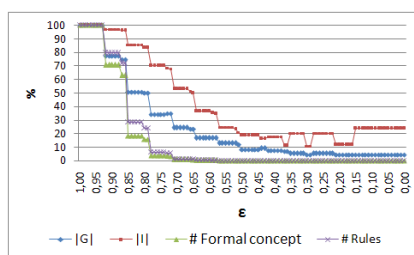


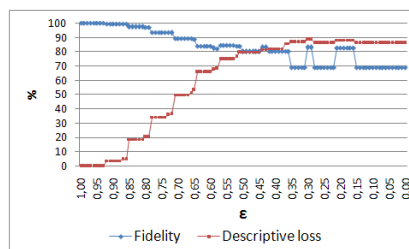**Fig. 1.** Objects, density, formal concepts and rules obtained for the Wine formal context.

**Fig. 2.** Fidelity and descriptive loss obtained for the Wine formal context.

Note too in the graph of Figure 1 that the reduction of the number of objects $|G|$ is higher, when compared with the reduction of the density $|I|$. One explanation for this is that the objects of each set $F \in \gamma$ have number of attributes relatively close to each other; and this induces a smaller change in density with respect to the number of objects. Note also that for $\epsilon \leq 0.15$ there were neither further reductions in the parameters under analysis nor changes in levels of fidelity and descriptive loss, since the number of formal concepts was reduced at the most for the given value of $\alpha$.

The results for the Water database can be seen in the graphs of Figures 3 and 4. First, note that using the framework *EF-Concept Analysis*, on the account of the huge number of formal concepts in this database, it was not possible to generate the formal concepts for values $\epsilon > 0.86$; consequently, it was not possible to get the number of rules for the original context. Thus, the graph in Figure 3 does not have the percentages of rules for such values of $\epsilon$. Note, however, that the calculation of fidelity after a reduction that leads to a treatable formal context is possible, since it does not depend on obtaining the formal concepts or the original concept lattice.

Figure 3 shows that for $\epsilon = 0.86$ the number of objects obtained was about 31% and the number of formal concepts about 2.14% (which results in about 191,800 formal concepts). For this same value of $\epsilon$, Figure 4 shows a fidelity of about 95% and a descriptive loss of about 21%. Note that even with a great reduction of formal concepts, the fidelity was still high, demonstrating a high degree of consistency of the JBOS method, and the descriptive loss was not high. Also note that for values $\epsilon < 0.42\%$ there is no reduction at all.
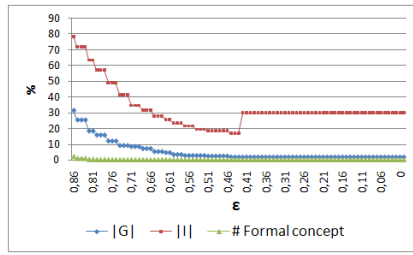


**Fig. 3.** Objects, density, formal concepts and rules obtained for the Water formal context.
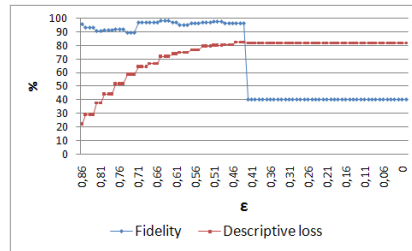


**Fig. 4.** Fidelity and descriptive loss obtained for the Water formal context.

The results show that the quality of the reduction achieved by the JBOS method, and even its viability, depend on a proper choice of values for $\epsilon$ after the weights of attributes are defined. Although the results do not show, as they don't consider particular uses of the reduced formal contexts, the definition of weights that really reflect the relevance of each attribute is very important because it is from these weights that the similarity calculations, which are the basis for the application of the method, are made. Regarding the parameter $\alpha$, it is not much important. It can, in general, be ignored or receive a high value as it was done above. In other words, decreasing values for $\alpha$ can be considered if experiments with values for weights of attributes and for $\epsilon$ are insufficient to produce acceptable reductions.

Regarding the measures of the reduction quality, a high fidelity and a low descriptive loss give evidence of good quality. The calculations of both are made from the original formal context, without the need of obtaining their respective formal concepts. After obtaining a reduced formal context through the JBOS method, it is always possible to calculate the descriptive loss incurred. The calculation of the fidelity is possible when the generation of rules from the resulting formal context is feasible, and this is what happens when the creation of formal concepts is achieved. This was the case, for example, for the Water database with $\epsilon \leq 0.86$. Another particularly important feature of these indices is that they are, as far as the authors know, the first application independent proposal for measuring quality of methods of complexity reduction.

## 5    Conclusions

The JBOS method obtains a reduced formal context based on a set of weights measuring the significance of the attributes. The requirements presented in Section 3, desirable for any method of complexity reduction, are met by JBOS. Regarding the first one, creating objects, attributes or incidences, besides going in the opposite direction of the reduction, it would alter the model described by the formal context and hence by the concept lattice. The JBOS method fully meets this requirement. Regarding the level of satisfaction for the last three requirements, it depends very much on a judicious choice of weights for the attributes and also on the allocation of an appropriate similarity threshold $\epsilon$. Anyway, such degree of judiciousness can be adjusted by means of experiments using different weight assignments and values for $\epsilon$ and $\alpha$.

It was also presented a methodology for evaluation of techniques for reducing the concept lattice's complexity, consisting of two metrics that seek to quantify precisely the degree of satisfaction to the latter two requirements refered to above. Because these two requirements are very dependent on the second, preservation of the original conceptual hierarchy, such metrics also indirectly quantify this second requirement. It is rarely found in the literature methodologies for evaluating methods of complexity reduction, and those which are found are designed based on specific reduction methods or applications. The methodology introduced in this paper is independent of reduction method and application. It is based on two metrics: an index of *fidelity* and a *descriptive loss*. The first, which measures the degree of consistency of the information present in the obtained formal context with respect to the original, is calculated from a set of rules extracted from the concept lattice. The second, which measures the loss of capacity to represent the details found in the model described in the original formal context, is calculated directly on the obtained formal context.

The results of the evaluation of the JBOS method applied to the two databases, Wine and Water, show that it is possible to reduce the complexity of the concept lattice preserving the consistency and with a good level of descriptive loss, both being measured by the methodology introduced. In particular, regarding the database Water, for which it was not possible to generate the formal concepts from the original context, it was possible to build the concept lattice with very significant reductions, yet maintaining a relatively high rate of fidelity and a not very big descriptive loss. The results showed that the method is able to reduce the complexity of the concept lattice (number of formal concepts) with a good level of control on the consistency and amount of knowledge detail produced. Besides, it is possible through adjustment of weights and parameters, to strive to meet the desired levels of quality.

As future work we intend to compare the JBOS method with other approaches and to analyze the impact of the discretization process (applied to build the formal context) on similarity and quality measures. The suitability of JBOS for reducing the structural complexity of concept lattices for some real applications will be tested.

# References

1. Gabriela Arévalo, Anne Berry, Marianne Huchard, Guillaume Perrot, and Alain Sigayret. Performances of galois sub-hierarchy-building algorithms. In *ICFCA'07: Proceedings of the 5th international conference on Formal concept analysis*, pages 166–180, Berlin, Heidelberg, 2007. Springer-Verlag.
2. Radim Belohlavek and Vilem Vychodil. Formal concept analysis with background knowledge: attribute priorities. *Trans. Sys. Man Cyber Part C*, 39(4):399–409, 2009.
3. C. Carpineto and G. Romano. *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, England, 2004.
4. Karen S. K. Cheung and Douglas Vogel. Complexity reduction in lattice-based information retrieval. *Information Retrieval*, 8(2):285–299, 2005.
5. B. Davey and H. Priestley. *Introduction to lattices and order.* Cambridge University Press, Cambridge, England, 1990.
6. Sérgio M. Dias. Algoritmos para geração de reticulados conceituais (in portuguese). Master's thesis, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, 2010.
7. Petr Gajdos, Pavel Moravec, and Václav Snásel. Concept lattice generation by singular value decomposition. In *CLA*, pages 102 –110, 2004.
8. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations.* Springer-Verlag, Germany, 1999.
9. Robert Godin and Rokia Missaoui. An incremental concept formation approach for learning from databases. *Theor. Comput. Sci.*, 133(2):387–419, 1994.
10. Robert Godin, Eugene Saunders, and Jan Gecsei. Lattice model of browsable data spaces. *Inf. Sci.*, 40(2):89–116, 1986.
11. Nicolas Jay, François Kohler, and Amedeo Napoli. Analysis of social communities with iceberg and stability-based concept lattices. In *ICFCA*, pages 258–272, 2008.
12. Bjoern Koester. *FooCA - Web Information Retrieval with Formal Concept Analysis.* Verlag Allgemeine Wissenschaft, 2006.
13. Sergei Kuznetsov. Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. *N.T.I.*, (12):21–29, 1990.
14. Sergei Kuznetsov. On computing the size of a lattice and related decision problems. *Order*, 18:313–321, 2001.
15. G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal. Computing iceberg concept lattices with titanic. *Data and Knowledge Eng.*, 42:189–222(34), 2002.
16. Petko Valtchev, Rokia Missaoui, and Robert Godin. Formal concept analysis for knowledge discovery and data mining: The new challenges. In *Concept Lattices*, volume 2961, pages 3901–3901. Springer Berlin / Heidelberg, 2004.
17. Renato Vimieiro. Um estudo de algoritmos para a extração de regras baseados em análise formal de conceitos (in portuguese). Master's thesis, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, 2007.
18. R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. *I. Rival (Ed.): Ordered Sets*, pages 445–470, 1982.
19. Luis E. Zárate and Sérgio M. Dias. Qualitative behavior rules for the cold rolling process extracted from trained ann via the fcann method. *Eng. Appl. Artif. Intell.*, 22(4-5):718–731, 2009.