Computing Left-Minimal Direct Basis of implications

Pablo Cordero, Manuel Enciso, Angel Mora, and Manuel Ojeda-Aciego

University of Málaga, Spain.
e-mail: {pcordero,enbciso}@uma.es, {amora,aciego}@ctima.uma.es

Abstract. The most popular basis in Formal Concept Analysis is the Duquenne-Guigues basis, which ensure minimality in the number of dependencies and it is built with pseudo-intents, and some method to calculate these basis from an arbitrary set of implications have been introduced. We propose in this paper, an automated method to calculate a left-minimal direct basis from the set of all implications built between a closed set and its corresponding minimal generators. The new basis also has the minimal property demanded in the Duquenne-Guigues basis. It is minimal in the cardinal of the set of implications, and minimal in the size of the left-hand side of the implications.

1 Introduction

Formal Concept Analysis [8] has shown to be a powerful framework to discover knowledge inside date sets. It provides a solid and formal theory which enhances other well known approaches. The main element of FCA community are binary relations between objects and attributes, which are described using matrixes (contexts) and represent the appearance of the attributes in the corresponding objects. One outstanding element in FCA is attribute implication, which is used to specify a constraint in the context. Thus we write an implication between attribute sets X and Y in the form $X \rightarrow Y$ whenever any object in the context which has all the attributes in X also has all the attributes in Y.

Attribute implication can be managed syntactically using Armstrong's Axioms [1], a sound and complete inference system. This axiomatic system allows us to derive new attribute implications that hold in a given context. This "inference" relation leads to the following problem: How to characterize the minimal set of implications for a given set of implications? Among the different basis notions, the Duquenne-Guigues basis [9] also called stem basis seems has to be cited because of their widely acceptation in the FCA area and because of its minimality notion (w.r.t. the number of implications). Nevertheless, minimality in the number of implications is a criteria that may be enhanced.

In [5] K. Bertet and B. Monjardet provided a set of orthogonal characteristics of the basis and established the equivalence of five definitions presented by different authors in several areas which correspond with the same notion of basis.

© Manuel Ojeda-Aciego, Jan Outrata (Eds.): CLA 2013, pp. 293–298,
ISBN 978–2–7466–6566–8, Laboratory L3i, University of La Rochelle, 2013.

In [6] we present a method to compute all the closed sets and its minimal generators from a context. This information allows us to built a set of implications whose left had side is a minimal generator and with its closed set in the right hand side. We propose in this paper a definition of basis with the good minimality property of Duquenne-Guigues basis and the characteristics of the above implications: minimal information in the left had side and a fast computation of attributes closures from them.

We also introduce an automated method to calculate from a set of implications a left-minimal direct basis. The new method is based on the Simplification Logic for Funcional Dependency $\mathbf{SL}_{\text{FD}}$ [?], a sound and complete inference system for implications. The main characteristics of $\mathbf{SL}_{\text{FD}}$ is that its inference system is not built around the transitivity rule, like other well known Armstrong-like axiomatic systems for implications.

The work is organized as follows. In Section 2 we summarize preliminary concepts and results on FCA concerning implications, basis, etc. In Section 3 we outline the automated method that we have proposed in [6] for the computation of minimal generators and we introduce a new method to calculate the left-minimal direct basis from the original set of implications. The paper ends with a Conclusion Section.

2 Background

We will use the well-kwon notation used on Formal Concept Analysis (FCA) [8]. For the analysis of the information contained in the context $\mathbf{K} = (G, M, I)$, a direction is the study of the pair (closed sets - minimal generators). The set of attributes A is said to be a minimal generator (**mingen**) if, for all set of attributes $X \subseteq A$ if $X'' = A''$ then $X = A$.

A relevant notion in the framework of Formal Concept Analysis is the concept of attribute implication [8]. This area is devoted to obtain knowledge from a context that is a table in which attributes and objects are related. An attribute implication is an expression $A \to B$ where A and B are sets of attributes. A context satisfies $A \to B$ if every object that has all the attributes in A has also all the attributes in B. It is well known that the sets of attribute implications that are valid in a context satisfies the Armstrong's Axioms. Although the two interpretations of formulas (functional dependency and attribute implication) are different, they have the same concept of semantic entailment. [2]

Alternatively, attribute implications allow us to capture all the information which can be deduced from a context. The set of all valid implications in a context may be syntactically managed by means of the following inference system known as Armstrong's axioms. An implication **basis** of $\mathbf{K}$ is defined as a set $\mathcal{L}$ of implications of $\mathbf{K}$ from which any valid implication for $\mathbf{K}$ can be deduced by using Armstrong rules.

The goal is to obtain an implication basis with minimal size. This condition is satisfied by the so-called Duquenne-Guigues (or stem) basis [9]. However, the definition of the Duquenne-Guigues basis refers to minimality only in the cardinality of the set of formulas, but as we have showed in [6] with an illustrative example, redundant attributes use to appear in this kind of minimal basis.

In the following, a summary of some interesting result of this survey are showed. More specifically we present the definitions of some characteristics studied in the survey that will be used to identify the kind of basis we introduce in this paper. In the practice, it is interesting considering some properties [5] related with minimality of them, in order to achieve efficiency.

Definition 1. *A set of implications Γ it is said*

- *minimal or non-redundant if $\Gamma \setminus \{X \to Y\}$ is not equivalent to Γ.*
- *minimum if if it is of least cardinality, that is, $\mid \Gamma \mid \leq \mid \Gamma' \mid$ for all set of implications Γ' equivalent to Γ.*
- *optimal if $|| \Gamma || \leq || \Gamma' ||$ for all set of implications Γ' equivalent to Γ, where the size of Γ is defined as*

$$|| \Gamma || = \sum_{\{X \to Y \in \Gamma\}} (\mid X \mid + \mid Y \mid)$$

And finally the two characteristics that constitutes the center of our basis are introduced:

Definition 2. *Let $\Gamma = \{X_0 \to Y_0, \dots X_n \to Y_n\}$ be a set of implications, it is said a left-minimal basis if there does not exist a $X_i \to Y_i$ and a subset $X'_i \subsetneq X_i$ such that $\Gamma \setminus \{X_i \to Y_i\} \cup \{X'_i \to Y_i\}$ is equivalent to Γ.*

A set of implications Γ and is said direct if for all implication $A \to B$ the set $A \cup B$ is a closed set w.r.t. Γ.

3 Obtaining basis from minimal generators

Some methods to obtain generators of closed sets have been studied in [7, 12, 13]. Moreover, minimal generators [10, 11] appear in the literature under different names in various fields, for instance they are the minimal keys of the tables in relational databases. In [13], the authors emphasize the importance of studying minimal generators although “they have been paid little attention so far in the FCA literature”.

We agree with these authors about the importance of the study of closed sets and minimal generators. They constitute a source of essential information to analyze a formal context. As we mention in the introduction, in [6] we illustrated the use of the Simplification paradigm to guide the search of all minimal generator sets. Thus, we introduce a method named `MinGen` [6] which computes a list of all pairs Φ =`<closed set, minimal generators>` from an arbitrary set of implications The goal of this paper can be considered as the reserve direction of the way we presented there. We will introduce a method to transform these set of pairs into a basis, preserving two good properties (fast computation of attribute closure and minimal left hand side in the implications) and providing minimality in the number of implications.Thus, our goal is to achieve from the minimal generators and closed sets a basis of implications fulfilling left-minimality and directness.

In the literature, the most cited algorithm to compute Duquenne-Guigues basis is the Ganter Algorithm [8]. Algorithm 1 is an adaptation of the algorithm showed in [3] to compute a Duquenne-Guigues basis from a set of LSI (labelled set of items) [6], and we obtain a Duquenne-Guigues basis.

Algorithm 1: Algorithm for computing a Duquenne-Guigues basis

```
input  : An LSI Φ
output: A Duquenne-Guigues basis
T := ∅;
foreach ⟨B, mg(B)⟩ ∈ Φ do
    foreach A ∈ mg(B) do  T := T ∪ {A → B}
repeat
    S := T;
    foreach A → B, C → D ∈ T such that A ⊊ C and B ⊈ C do
        T := T ∖ {C → D};
        if B ∪ C ≠ D then T := T ∪ {BC → D}
until T = S ;
S := ∅;
foreach A → B ∈ T do S := S ∪ {A → B ∖ A};
return S
```

In the following example, we show how to link the above algorithm with the work presented in [6].

Example 1. For the input $T = \{ab \to c, ac \to bd, b \to d, d \to c\}$, Algorithm MinGen_0 (see [6]) returns $\Phi = \{< abcd, \{ab, ac, ad\} >, < bcd, \{b\} >, < cd, \{d\} > \}$ and from here Algorithm 1 renders $\Gamma = \{d \to cd, b \to bcd, ac \to abcd\}$, which corresponds to a Duquenne-Guigues basis. □

The following theorem ensures the minimality (w.r.t. the cardinality) of the Duquenne-Guigues basis.

Theorem 1. *Any Duquenne-Guigues basis is a minimum basis.*

In the following, we describe the algorihtm 2 to calculate a Left-Minimal Direct Basis based on Algorithm 1. The algorithm is polynomial and it is described searching a good understanding. Of course, an implementation using lectic order would improve considerably its efficiency.

First, we consider the following equivalence rules.

Definition 3 (Aggregation rules). *Let A, B, C and D be sets of attributes.*

1. *If $A \subseteq C$ then $\{A \to B, C \to D\} \equiv \{A \to B, BC \to D \smallsetminus B\}$.*
2. *If $A \subseteq C \subseteq A \cup B$ then $\{A \to B, C \to D\} \equiv \{A \to BD\}$.*

Algorithm 2: Algorithm for computing a Left-Minimal Direct Basis

```
input  : An LSI Φ
output: A Minimal Direct Basis
T := ∅;
foreach ⟨B, mg(B)⟩ ∈ Φ do
    foreach A ∈ mg(B) do  T := T ∪ {(A, A → B)}
repeat
    S := T;
    foreach (M, A → B), (N, C → D) ∈ T such that A ⊊ C and B ⊈ C do
        T := T ∖ {(N, C → D)};
        if B ∪ C ≠ D then T := T ∪ {(N, BC → D)}
until T = S ;
S := ∅;
foreach (M, A → B) ∈ T do S := S ∪ {M → B ∖ M};
return S
```

We let for an extended version of this paper the proof that these equivalences rules are derived rules of equivalences presented in Theorem **??**.

We propose in the Algorithm 2 the use of the two aggregation rules for reducing the number of implications and also the consequent of implications.

Theorem 2. *Let $T = \{(M_1, A_1 \to B_1), \ldots\}$ be a set of pairs with minimal generators and implications obtained from minimal generators and closed sets. The exhaustive application of the two Aggregation rules produces a left-minimal direct basis.*

Example 2. Let $T = \{b \to agh, d \to a, bn \to h, ab \to defg, abc \to djk\}$ be a set of implications, Algorithm MinGen_0 ([6]) returns a list with closed sets and their minimal generators, $\Phi = \{< abdefgh, \{b\} >, < abcdefghkj, \{bc\} >, < abdefghn, \{bn\} >, < abcdefghkn, \{bcn\} >, < ad, \{d\} >\}$

In the first step of the Algorithm 2 with Φ we build $T = \{$(b, b→abdefgh), (bc,bc →abcdefghkj), $(bn, bn \to abdefghn), (bcn, bcn \to abcdefghkn), (d, d \to ad)\}$.

Then, we apply the Aggregation rules foreach couple of elements in T. At the end of these comparisons, we have $T = \{(b, b \to abdefgh), (bc, abcdefgh \to abcdefghkj), (d, d \to ad)\}$. And from this, in the last foreach of the Algorithm 2, it renders the following left-minimal direct basis $S = \{b \to adefgh, bc \to adefghkj, d \to a\}$.

By the other side, Algorithm 1 return the following Duquenne-Guigues basis $\{b \to adefgh, abcdefgh \to kj, d \to a\}$ which in a minimal basis, but not a left-minimal one.

4 Conclusion

In this paper we present an algorithm which allows the transformation of a set of all closed sets and their corresponding minimal generators into a left-minimal direct basis. The study about the soundness, completeness, and complexity of the algorithms proposed are left to a extended paper.

The new method uses some equivalences deduced in the $\mathbf{SL}_{\mathrm{FD}}$ Logic and follows the Lectic order to traverse the list of minimal generators and implications associated and return a set of implications but with good properties.

As future work we propose to extend the Duquenne-Guigues basis definition to consider a generalized fuzzy extension of implications. We propose the definition introduced in [2] that has been shown to be the most general one. In [4] a non trivial extension of the $\mathbf{SL}_{\mathrm{FD}}$ Logic for the generalized definition of fuzzy functional dependency was introduced. The generalized version of the $\mathbf{SL}_{\mathrm{FD}}$ Logic will be used to develop a method to get basis for the generalized fuzzy implications.

Acknowledgment

Supported by Grants TIN2011-28084 of the Science and Innovation Ministry of Spain, and No. P09-FQM-5233 of the Junta de Andalucía.

References

1. William W. Armstrong, *Dependency structures of data base relationships*, Proc. IFIP Congress. North Holland, Amsterdam: 580–583, 1974.
2. Radim Belohlávek, and Vilém Vychodil, *Functional Dependencies of Data Tables Over Domains with Similarity Relations*, Proc. of the 2nd Indian International Conference on Artificial Intelligence, 2005.
3. Radim Belohlávek, and Vilém Vychodil, *Formal Concept Analysis with Constraints by Closure Operators*, Lecture Notes in Computer Science, 4068:131–143, 2006.
4. Radim Belohlávek, Pablo Cordero, Manuel Enciso, Ángel Mora, and Vilém Vychodil, *An Efficient Reasoning Method for Dependencies over Similarity and Ordinal Data*, Lecture Notes in Computer Science, 7647: 408–419, 2012.
5. K. Bertet, B. Monjardet, *The multiple facets of the canonical direct unit implicational basis*, Theor. Comput. Sci., 411(22-24): 2155–2166, 2010.
6. Pablo Cordero, Manuel Enciso, Ángel Mora, Manuel Ojeda-Aciego, *Computing Minimal Generators from Implications: a Logic-guided Approach*, CLA 2012: 187–198, 2012.
7. A. Day, *The lattice theory of functional dependencies and normal decompositions*, International Journal of Algebra and Computation, 2: pp. 209–431, 1990.
8. B. Ganter, *Two basic algorithms in concept analysis*, Technische Hochschule, Darmstadt, 1984.
9. J.L. Guigues and V. Duquenne, *Familles minimales d'implications informatives résultant d'un tableau de données binaires*, Math. Sci. Humaines: 95, 5–18, 1986.
10. L. Szathmary and A. Napoli and S. O. Kuznetsov, ZART: A Multifunctional Itemset Mining Algorithm , Proc. of the 6th Intl. Conf. on Concept Lattices and Their Applications (CLA '08): 47–58, 2008.
11. L. Szathmary and P. Valtchev and A. Napoli and R. Godin, Efficient Vertical Mining of Frequent Closures and Generators, LNCS 5772: 393–404, 2009.
12. C. Frambourg, P. Valtchev, and R. Godin, Merge-based computation of minimal generators. Discrete Mathematics, LNCS 3596: 181–194, 2005.
13. K. Nehmé, P. Valtchev, M. H. Rouane, R. Godin, On Computing the Minimal Generator Family for Concept Lattices and Icebergs, LNCS 3403: 192–207, 2005.