Extending FuzAtAnalyzer to approach the management of classical negation

Rodriguez-Jimenez, J.M., Cordero, P., Enciso, M., Mora, A.
jmrodriguez@ctima.uma.es, pcordero@uma.es, enciso@lcc.uma.es, amora@ctima.uma.es

Universidad de Málaga, Andalucia Tech, Spain
Published in CMMSE 2015

Abstract. FuzAtAnalyzer was conceived as a Java framework which goes beyond of classical tools in formal concept analysis. Specifically, it successfully incorporated the management of uncertainty by means of methods and tools from the area of fuzzy formal concept analysis. One limitation of formal concept analysis is that they only consider the presence of properties in the objects (positive attributes) as much in fuzzy as in crisp case. In this paper, a first step in the incorporation of negations is presented. Our aim is the treatment of the absence of properties (negative attributes). Specifically, we extend the framework by including specific tools for mining knowledge combining crisp positive and negative attributes.

1 Introduction

We create a tool called FuzAtAnalyzer since we have the need to explore fuzzy functional dependencies mining in datasets because there not exists other tools that covers the theory that we develope. Our first implementation was Simplification Logic with fuzzy functional dependencies [4] followed by the algorithms related with Fuzzy Attribute Tables [5, **?**].

Simplification Logic was adapted for working with crisp implications (without degrees)[8], that was the previous step before we study implications in Formal Concept Analysis.

In this section, the basic notions related with Formal Concept Analysis (FCA) [14] and attribute implications are briefly presented. See [10] for a more detailed explanation. A **formal context** is a triple $\mathbb{K} = \langle G, M, I \rangle$ where G and M are finite non-empty sets and $I \subseteq G \times M$ is a binary relation. The elements in G are named objects, the elements in M attributes and $\langle g, m \rangle \in I$ means that the object g has the attribute m. From this triple, two mappings $\uparrow: 2^G \to 2^M$ and $\downarrow: 2^M \to 2^G$, named concept-forming operators, are defined as follows: for any $X \subseteq G$ and $Y \subseteq M$,

$$X^{\uparrow} = \{m \in M \mid \text{ for each } g \in X : \langle g, m \rangle \in I\} \quad (1)$$

$$Y^{\downarrow} = \{g \in G \mid \text{ for each } m \in Y : \langle g, m \rangle \in I\} \quad (2)$$

$X^{\uparrow}$ is the subset of all attributes shared by all the objects in X and $Y^{\downarrow}$ is the subset of all objects that have the attributes in Y. The pair $(\uparrow, \downarrow)$ constitutes a Galois connection between 2^G and 2^M and, therefore, both compositions are closure operators.

A pair of subsets $\langle X, Y\rangle$ with $X \subseteq G$ and $Y \subseteq M$ such $X^{\uparrow} = Y$ and $Y^{\downarrow} = X$ is named a **formal concept**. X is named the **extent** and Y the **intent** of the concept. These extents and intents coincide with closed sets wrt the closure operators because $X^{\uparrow\downarrow} = X$ and $Y^{\downarrow\uparrow} = Y$. Thus, the set of all the formal concepts is a lattice, named **concept lattice**, with the relation

$$\langle X_1, Y_1\rangle \leq \langle X_2, Y_2\rangle \text{ if and only if } X_1 \subseteq X_2 \text{ (or equivalently, } Y_2 \subseteq Y_1) \quad (3)$$

Focusing the attention in relationships among sets of attributes is a second way in which the information can be summarized. Agrawal et al.[1] introduced **association rules** for discovering regularities between attributes. For this purpose, the concepts support and confidence were introduced. Support is defined for any subset $Y \subseteq M$, $supp(Y) = |Y^{\downarrow}|/|G|$ and confidence for association rule $Y_1 \to Y_2$ is defined as $conf(Y_1 \to Y_2) = supp(Y_1 \cup Y_2)/supp(Y_1)$. We have to remark that a fuzzy functional dependency and an association rule are different elements with different degrees. In first case, the degree indicates the relation between set of attributes, whereas association rules have a statistical degree.

These relationships among attribute sets where $confidence = 1$ are described in terms of *attribute implications*.

The concept lattice can be also characterized in terms of attribute implications. An **attribute implication** is an expression $A \to B$ where $A, B \subseteq M$ and it holds in a formal context if $A^{\downarrow} \subseteq B^{\downarrow}$. That is, any 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 satisfied by a context satisfies the Armstrong's Axioms:

`[Ref]` Reflexivity: If $B \subseteq A$ then $\vdash A \to B$.
`[Augm]` Augmentation: $A \to B \vdash A \cup C \to B \cup C$.
`[Trans]` Transitivity: $A \to B, B \to C \vdash A \to C$.

A set of implications $\mathfrak{B}$ is an **implicational basis** for $\mathbb{K}$ if: (1) any implication from $\mathfrak{B}$ holds in $\mathbb{K}$ and (2) any implication that $\mathbb{K}$ satisfies follows (can be inferred by using Armstrong's Axioms from $\mathfrak{B}$.

One of the most cited kind of basis is the so-called Duquenne-Guigues (or stem) base [11]. The premises of the implications in the Duquenne-Guigues basis are pseudo-intents: $P \subseteq M$ is a **pseudo-intent** if P is not an intent ($P^{\downarrow\uparrow} \neq P$) and $Q^{\downarrow\uparrow} \subseteq P$ holds for every pseudo-intent $Q \subsetneq P$. The Duquenne-Guigues base for $\mathbb{K}$ is

$$\{P \to (P^{\downarrow\uparrow} \smallsetminus P) \mid P \text{ is a pseudo-intent for } \mathbb{K}\} \quad (4)$$

and satisfies that its cardinality is minimum among all the bases. It is well-known the **NextClosure** Algorithm [10] that computes all the pseudo-intents and intents, and therefore the Duquenne-Guigues base for a context. This algorithm is based in the **lectic order** among sets of attributes that coincides with the usual order for binary numbers when set of attributes are represented by bit-maps.

Classical FCA only discover knowledge limited to positive attributes in the context, but it does not consider information relative to the absence of properties (attributes). Thus, the Duquenne-Guigues basis obtained from Table 1 is $\{e \to bc, d \to c, bc \to e, a \to b\}$. Moreover, the implications $b \to c$ either $b \to d$ do not hold in Table 1 and therefore they can not be derived from the basis by using the inference system. Nevertheless, both implications correspond with different situations. In the first case, some objects have attributes b and c (e.g. objects o_1 and o_3) whereas another objects (e.g. o_2) have the attribute b and do not have c. By the other side, in the second case, any object that has the attribute b does not have the attribute d.

I	a	b	c	d	e
o_1		×	×		×
o_2	×	×			
o_3		×	×		×
o_4			×	×	

Table 1. A formal context

A more general framework is necessary to deal with this kind of information. In [12], we have tackled this issue focusing on the problem of mining implication with positive and negative attributes from formal contexts. As a conclusion of that work we emphasized the necessity of a full development of an algebraic framework.

First, we begin with the introduction of an extended notation that allows us to consider the negation of attributes. From now on, the set of attributes is denoted by M, and its elements by the letter m, possibly with subindexes. That is, the lowercase character m is reserved for positive attributes. We use $\overline{m}$ to denote the negation of the attribute m and $\overline{M}$ to denote the set $\{\overline{m} \mid m \in M\}$ whose elements will be named negative attributes.

Arbitrary elements in $M \cup \overline{M}$ are going to be denoted by the first letters in the alphabet: a, b, c, etc. and $\overline{a}$ denotes the opposite of a. That is, the symbol a could represent a positive or a negative attribute and, if $a = m \in M$ then $\overline{a} = \overline{m}$ and if $a = \overline{m} \in \overline{M}$ then $\overline{a} = m$.

Capital letters A, B, C ... denote subsets of $M \cup \overline{M}$. If $A \subseteq M \cup \overline{M}$, then $\overline{A}$ denotes the set of the opposite of attributes $\{\overline{a} \mid a \in A\}$ and the following sets are defined:

- $\mathrm{Pos}(A) = \{m \in M \mid m \in A\}$
- $\mathrm{Neg}(A) = \{m \in M \mid \overline{m} \in A\}$
- $\mathrm{Tot}(A) = \mathrm{Pos}(A) \cup \mathrm{Neg}(A)$

Note that $\mathrm{Pos}(A), \mathrm{Neg}(A), \mathrm{Tot}(A) \subseteq M$.

Once we have introduced the notation, we are going to summarize some results concerning the mining of knowledge from contexts in terms of implications with negative and positive attributes. In [12], we have developed a method to mine mixed implications whose main goal has been to avoid the management of the large $(\mathbb{K}|\overline{\mathbb{K}})$ contexts, so that the performance of the corresponding method has a controlled cost.

First, we extend the definitions of derivation operators, formal concept, association rule and attribute implication.

Definition 1. *Let $\mathbb{K} = \langle G, M, I\rangle$ be a formal context. We define the operators $\Uparrow\colon 2^G \to 2^{M\cup\overline{M}}$ and $\Downarrow\colon 2^{M\cup\overline{M}} \to 2^G$ as follows: for $X \subseteq G$ and $Y \subseteq M \cup \overline{M}$,*

$$X^{\Uparrow} = \{m \in M \mid \langle g,m\rangle \in I \text{ for all } g \in X\} \\ \cup \{\overline{m} \in \overline{M} \mid \langle g,m\rangle \notin I \text{ for all } g \in X\} \quad (5)$$

$$Y^{\Downarrow} = \{g \in G \mid \langle g,m\rangle \in I \text{ for all } m \in Y\} \\ \cap \{g \in G \mid \langle g,m\rangle \notin I \text{ for all } \overline{m} \in Y\} \quad (6)$$

Definition 2. *Let $\mathbb{K} = \langle G, M, I\rangle$ be a formal context. A* mixed formal concept *in $\mathbb{K}$ is a pair of subsets $\langle X, Y\rangle$ with $X \subseteq G$ and $Y \subseteq M \cup \overline{M}$ such $X^{\Uparrow} = Y$ and $Y^{\Downarrow} = X$.*

To extending association rules to mixed attributes, we have to change the definition of support and, consequently, the confidence. For any subset $Y \subseteq M \cup \overline{M}$, $supp(Y) = |Y^{\Downarrow}|/|G|$.

Definition 3. *Let $\mathbb{K} = \langle G, M, I\rangle$ be a formal context and let $A, B \subseteq M \cup \overline{M}$, the context $\mathbb{K}$ satisfies a* mixed attribute implication *$A \to B$, denoted by $\mathbb{K} \models A \to B$, if $A^{\Downarrow} \subseteq B^{\Downarrow}$.*

For example, in Table 1, as we previously mentioned, two different situations were presented. Thus, in this new framework we have that $\mathbb{K} \not\models b \to d$ and $\mathbb{K} \models b \to \overline{d}$ whereas $\mathbb{K} \not\models b \to c$ either $\mathbb{K} \not\models b \to \overline{c}$.

In [12], we tackled this issue focusing on the problem of mining implication with positive and negative attributes from formal contexts. As a conclusion of that work we emphasized the necessity of a full development of the algebraic framework. A first step in this line was introduced in [13].

Although FuzAtAnalyzer deals with both, classical and fuzzy implications, in this paper we concentrate on classical implications as the target element to include negative attributes. Thus, in this first step our intention is to develop a tool which carries out the knowledge discovering of the full information in the system limited to the crisp case. Having said that, in this work we also allows the treatment of association rules. The main reason is to consider as much information as possible relaxing the implication semantics. Thus, association rules provides us more information without a significant change in the theoretical background model as fuzzy implications demands.

In this paper, we present how we incorporate these results to our tool, extending our previous work about fuzzy attributes implications and fuzzy data tables. This is a previous step in the extension of these results to fuzzy case.

2 FuzAtAnalyzer

Related to our research project, we develope a JAVA program called FuzAtAnalyzer that was initially used with fuzzy functional dependencies [**?**] and Fuzzy Attribute Tables [5]. Datasets are imported from Microsoft Excel archives (using Java Excel API[16]) where there is a degree between 0 and 1 that represents the relation of an object with an attribute. It is writed 1 if and object has an attribute and 0 if it has not in the case of Formal Concept Analysis with crisp data.

In the original version we focus our attention to obtaining implicational systems and the knowledge related to them implementing algorithms that appear in [5, **?**,**?**,**?**,**?**,**?**]

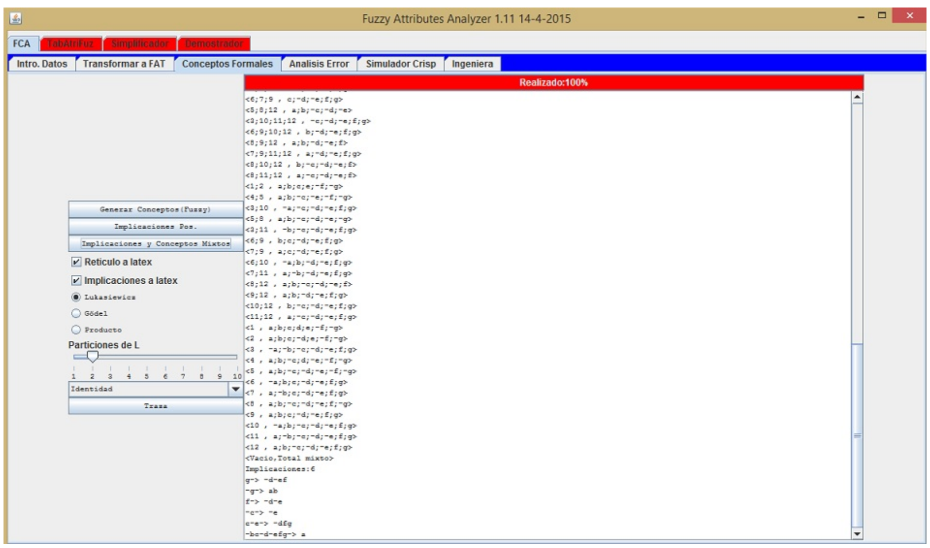

Fig. 1. Screen capture of FuzAtAnalyzer

We adapt this tool for working with datasets used in Formal Concept Analysis, implementing different algorithms like the former *NextClosure* [10]. In a first step, we use only positive attributes but, applying the theory that appears in [12], we extend the study to mixed attributes.

To the previous options that generates (fuzzy) formal concepts and implications with positive attributes, we add an option that generates mixed attributes implications and mixed concepts implementing the algorithm 3 in [12], adapting it to extract mixed concepts too.

This can be do with other programs like Concept Explorer [15] using the apposition $\{\mathbb{K}|\overline{\mathbb{K}}\}$ but this implementation allow us to generate the results in an easy and efficient way from an imported dataset.

Also, we add options that generate the mixed lattice and the implicational system in *latex* that allows to export the results to different documents. The lattice is generated using the package *XYPic*[17] of *latex*, distributing the concepts by levels based in the number of attributes, positives and negatives, in the intents of the concepts.

In a specific section for detecting errors, we adapt the methods for obtaining association rules that works with negative attributes too. This adaption allows us to compare these association rules with implications obtaining knowledge about the confidence and support for the subsets of mixed attributes. We can classify by relevance the mixed implications in order of the support of the subsets of mixed attributes that compose these implications.

3 Example

We are going to use an example with 9 objects and 5 attributes that corresponds to the dataset represented in the context in Table 2.

I	a	b	c	d	e
o_1	1	1	1	1	1
o_2	0	1	0	1	1
o_3	0	1	1	1	0
o_4	0	1	1	0	1
o_5	0	1	0	1	0
o_6	1	0	1	0	0
o_7	0	1	0	0	0
o_8	1	0	1	0	1
o_9	0	0	0	0	0

Table 2. A formal context

From this dataset, in a simple step, we can calculate the implicational system and the mixed concepts and we can export this data automatically to a *latex* archive with the representation of the lattice.

The mixed implicational system is $\{d \to b,\ \overline{d}e \to c,\ \overline{c} \to \overline{a},\ c\overline{d}\overline{e} \to a\overline{b},\ b\overline{e} \to \overline{a},\ b\overline{d} \to \overline{a},\ \overline{b}c\overline{d} \to a,\ bcde \to a,\ \overline{a}e \to b,\ \overline{a}c \to b\}$ that extends the original implicational system with positive attributes $\{d \to b,\ bcde \to a,\ a \to c,\ abc \to de\}$. The 2 first implications are included in the mixed implicational system and the others can be deduced from it.

The 48 concepts of the mixed concepts lattice associated with the context of Table 2 are represented in Figure 2. The file over the bottom concept is conformed with atoms, whereas the concepts represented with a square are the meet-irreducible elements. All the elements of the mixed concepts lattice can be represented as infimum of meeet-irreducible elements and supremum of join-irreducible elements (atoms).

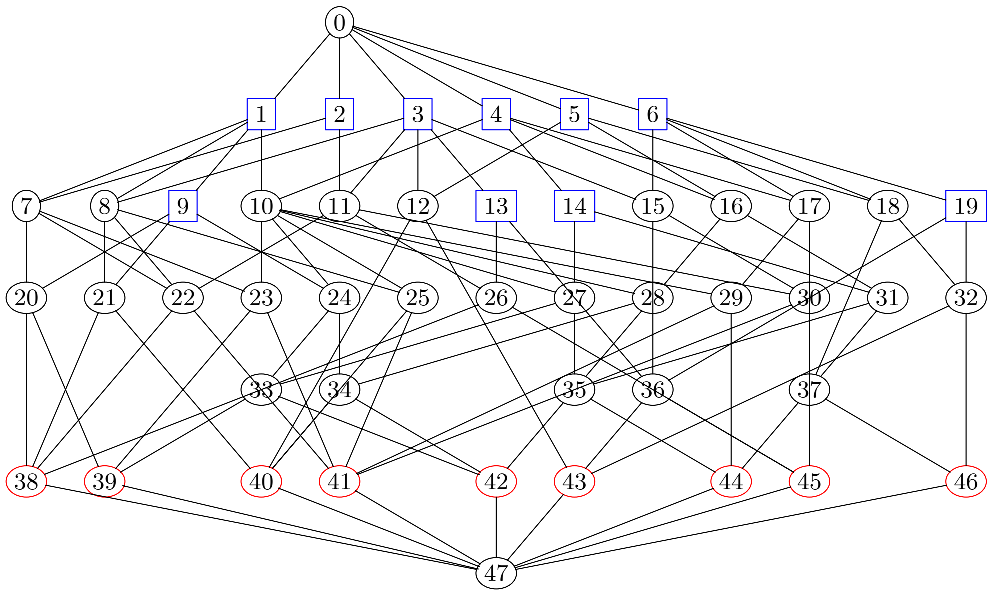

Fig. 2. Mixed concept lattice.

In a second phase, we examine association rules with a minimun value 0.4 of confidence. This means that, at least, one attribute or his negative appears in the subsets of single attributes.

In the case of considering only positives attributes, $\{X \subseteq M | supp(X) > 0.4\} = \{b, c, d, e, bd\}$. The association rules that we can obtain are $d \to b$ and $b \to d$ which confidences are 1 and 0.67 respectively. Adding the negative and mixed attributes, we have that $\{X \subseteq M \cup \overline{M} | supp(X) > 0.4\} = \{\overline{a}, b, c, \overline{c}, d, \overline{d}, e, \overline{e}, \overline{a}b, \overline{ae}, bd, \overline{ac}\}$ and the new mixed association rules are $b \to \overline{a}$, $\overline{a} \to b$ with confidences 0.83, $\overline{e} \to \overline{a}$ with confidence 0.8, $\overline{a} \to \overline{e}$ with confidence 0.67, $\overline{c} \to \overline{a}$ with confidence 1 and $\overline{a} \to \overline{c}$ with confidence 0.67.

With this study of minimun supports of antecedents and consequents of each association rule with confidence 1, we can define the relevance of the implication asociated to it and how a mixed rule could be more representative that a positive one.

Acknowledgements

References

1. R. Agrawal, T. Imieliński and A. Swami, *Mining Association Rules Between Sets of Items in Large Databases*, SIGMOD Rec.,**22**,207–216,1993
2. W. Armstrong,*Dependency structures of data base relationships*, Proc. IFIP Congress. North Holland, Amsterdam,580–583,1974

3. K. Bertet and B. Monjardet, *The multiple facets of the canonical direct unit implicational basis*, Theor. Comput. Sci.,**411**, 22-24,2010,2155-2166
4. P. Cordero, M. Enciso, A. Mora, I. Perez de Guzmán and J. M. Rodríguez-Jiménez, *An Efficient Algorithm for Reasoning about Fuzzy Functional Dependencies*, Advances in Computational Intelligence - International Work-Conference on Artificial Neural Networks (IWANN),**2**,412–420,2011
5. P. Cordero, M. Enciso, A. Mora, I. Perez de Guzman and J. M. Rodríguez-Jiménez, *Specification and inference of fuzzy attributes*, Proceedings of the IEEE Symposium on Foundations of Computational Intelligence, FOCI 2011, part of the IEEE Symposium Series on Computational Intelligence 2011, Paris, France,107–114
6. A. Mora, M. Enciso and P.Cordero, *Closure via functional dependence simplification*, International Journal of Computer Mathematics, **89** **4**, 510–526, 2012
7. P. Cordero, M. Enciso, A. Mora and M. Ojeda-Aciego, *Computing Minimal Generators from Implications: a Logic-guided Approach*, CLA,187–198,2012
8. P. Cordero, M. Enciso, A. Mora and J. M. Rodríguez-Jiménez, *Computing non-redundant sets of functional dependencies via simplification*, IEEE Symposium on Foundations of Computational Intelligence (FOCI),9–14,2013
9. P. Cordero, M. Enciso, A. Mora and J. M. Rodríguez-Jiménez, *Automated Inference with Fuzzy Functional Dependencies over Graded Data*, Advances in Computational Intelligence - International Work-Conference on Artificial Neural Networks (IWANN),**2** 254–265,2013
10. B. Ganter, *Two basic algorithms in concept analysis*, *Technische Hochschule, Darmstadt*, 1984
11. J.L. Guigues and V. Duquenne, *Familles minimales d implications informatives resultant d un tableau de donnees binaires*, *Mathematiques et Sciences Sociales*, **95**, 5–18, 1986.
12. J. M. Rodríguez-Jiménez, P. Cordero, M. Enciso, and A. Mora, *Negative attributes and implications in formal concept analysis*, *Procedia Computer Science*, 31(0):758 – 765, 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014
13. J. M. Rodríguez-Jiménez, P. Cordero, M. Enciso and A. Mora, *A generalized framework to consider positive and negative attributes in formal concept analysis*, CLA,267–278, 2014
14. R. Wille. *Restructuring lattice theory: an approach based on hierarchies of concepts*, Rival, I. (ed.): Ordered Sets, 445–470. Boston, 1982.
15. http://conexp.sourceforge.net/
16. http://jexcelapi.sourceforge.net/
17. https://tug.org/applications/Xy-pic/Xy-pic.html