Concept Lattice Theory in Data Mining and its Applications*

Pascal SUNGU[a], KANINDA MUSUMBU[b] and Nathalie WANDJI[c]

[a] *Master student, Computational Mathematics, Pan African University of Science, Technology and Innovation. Assistant Université Nouveaux Horizons*

[b] *Maître de Conférences, Université Bordeaux, Université Nouveaux Horizons*

[c] *PhD student, Tutor at AIMS (African Institute of Mathematical Science) Cameroon*

ARTICLE INFO

Keywords:
Formal concept
Frequent pattern
Association rules

ABSTRACT

Concept lattice has been proven to be a very effective tool and architecture for data mining in general. It is widely used for data analysis and knowledge discovery and various concept lattice based approaches are used depending on the type of data. This paper aims at presenting one application of the lattice theory : the text mining. In this approach, we applied the notion of lattice theory by using one of its components mostly used in data mining, the formal concept analysis which has a powerful method, the association rule extraction which helps to find in a database patterns which appear frequently together.

1. Introduction

Nowadays, with the up growth of technology, the amount of data is constantly increasing and stored in databases. Extracting some element of knowledge becomes a real challenge. Therefore, the need of methods which can be useful to process and structure data in order to come out with interesting elements of knowledge becomes very important. *Data mining*, however, is defined as a technique of extracting information from data and constitutes one of the main step in the processing chain for discovering knowledge.

In this paper, we will present the specializations of data mining using the lattice theory to show the use of mathematics in other domains. However, before pursuing with the paper, some concepts have to be set:

1. *Formal Concept Analysis (FCA)* is a method of data analysis which can perform data mining tasks. It is a powerful process due to its capacity of generating diagrams which facilitate data representation and describe relationships between a particular set of objects and a particular set of attributes. FCA produces two kinds of results based on the input data: concept lattice and collection of attribute implication or association rules.
2. *Concept lattice* is a set of all the formal concepts which can be found in the data that are ordered by using the relation of subconcept and superconcept based on the inclusion relations on objects and attributes.
3. *Association rules* describes a particular dependency which is valid in data as in *every even number divisible by 2 and 3 are also divisible by 6*. It can be exact or approximative where their confident values is equal or different to 1 respectively and they formed a generic basis. This approach will be used with text mining to extract and present the result.

Concept lattice has been proven to be a very effective tool and architecture for data mining in general. It is widely used for data analysis and knowledge discovery and various concept lattice based approaches are used depending on the type of data.

*This document is the results of the research project done at AIMS/Cameroon.

pasklsungu@gmail.com (P. SUNGU); musumbu@gmail.com (K. MUSUMBU); nathalie.wandji@aims-cameroon.org (N. WANDJI)

http://www.labri.fr/Perso/~musumbu (K. MUSUMBU)

ORCID(s):

2. Mathematical Foundations : Lattice Theory

Fundamentally, a lattice has been defined as an algebraic structure consisting of a set of elements with two binary operations namely the sup boundary operation and inf boundary operation.

Definition 2.1. *(**Lattices as an algebraic structure**) Let $(S, \vee, \wedge)$ be an algebraic structure where S is a set, $\vee$ and $\wedge$ are two binary operations on S. Such a structure is called* lattice *if the following axioms hold for all elements x, y and z in S [?].*

a) ***Commutative laws:***

- $x \vee y = y \vee x$
- $x \wedge y = y \wedge x$

b) ***Associative laws:***

- $x \vee (y \vee z) = (x \vee y) \vee z$
- $x \vee (y \vee z) = (x \vee y) \vee z$

c) ***Absorption laws:***

- $x \vee (x \wedge y) = x$
- $x \wedge (x \vee y) = x$

d) ***Idempotent laws:***

- $x \vee x = x$
- $x \wedge x = x$

Lattice is also defined as a partially ordered set with 2 specific elements called: *sup boundary* and *inf boundary.*

Definition 2.2. *(**Lattices as a partially ordered sets**) Let $(S, \leq)$ be a pair of elements where $\leq$ is a partial order defined on the set S that means a binary relation such that [?]:*

a) ***Reflexivity**: $x \leq x$; for all $x \in S$*
b) ***Antisymmetry**: $x \leq y$ and $y \leq x$ implies that $x = y$ for all $x, y \in S$*
c) ***Transitivity**: $x \leq y$ and $y \leq z$ implies that $x \leq z$ for all $x, y, z \in S$*
d) ***Linearity**: $x \leq y$ or $y \leq x$ for all $x, y \in S$*

Definition 2.3. *(**Irreducible elements,Arrow relation** and **table of a lattice**)*

i) An element j of a lattice is said to be reducible *if it is the result of $x \wedge y$ or $x \vee y$. Otherwise, it will be called an* irreducible element.

ii) An element $j \in S$ is said to be join-reducible *(respectively* meet-reducible*) if:*

$$\exists\, x, y \in S \text{ such that } j = x \wedge y \text{ (respectively } j = x \vee y) \text{ where } j > x \text{ and } j > y \text{ (respectively } j < x \text{ and } j < y) \quad (1)$$

However, if $j \in S$ is such that $j \neq x \wedge y$ (respectively $j \neq x \vee y$), then j is said to be a join-irreducible *(respectively a* meet-irreducible*) element.*

The table of a lattice is defined on a particular relation called arrow relation *[?] which partitions the different relationships between join-irreducible and meet-irreducible elements into five binary relations defined on $J_L \times M_L$. The first one denoted by $P_{\leq}$ and $P_{\not\leq}$ is defined by:*

$$P_{\leq} = \{(j, m) \in J_L \times M_L : j \leq m\} \quad (2)$$
$$P_{\not\leq} = \{(j, m) \in J_L \times M_L : j \not\leq m\}. \quad (3)$$

iii) Let $L = (S, \leq)$ be a lattice, $j \in J_L$ and $m \in M_L$. An arrow relation is defined as:

- *$j \uparrow m$ if $j \not\leq m$ and $j < j^+$ with j^+ the unique successor of $m \in M_L$,*
- *$j \downarrow m$ if $j \not\leq m$ and $j^- < m$ with j^- the unique predecessor of $j \in J_L$.*

Table 1
Exemple of a table of lattice [?]

	a	b	c	d	e	f
b	$\downarrow$	$\times$	$\circ$	$\circ$	$\downarrow$	$\updownarrow$
c	$\times$	$\updownarrow$	$\times$	$\uparrow$	$\updownarrow$	$\updownarrow$
d	$\updownarrow$	$\downarrow$	$\circ$	$\times$	$\times$	$\times$
i	$\times$	$\updownarrow$	$\updownarrow$	$\times$	$\times$	$\times$
k	$\times$	$\downarrow$	$\downarrow$	$\updownarrow$	$\times$	$\times$
l	$\updownarrow$	$\times$	$\uparrow$	$\uparrow$	$\updownarrow$	$\times$
m	$\times$	$\downarrow$	$\downarrow$	$\circ$	$\updownarrow$	$\times$
n	$\times$	$\downarrow$	$\downarrow$	$\circ$	$\times$	$\updownarrow$

Therefore, we obtain four possible combination ($P_\uparrow, P_\downarrow, P_\updownarrow$ and $P_\circ$) from the pair $(j,m) \in P_{\not\leq}$ defined as:

$$P_\uparrow = \{(j,m) \in J_L \times M_L : j \uparrow m \text{ and } j \not\downarrow m\} \quad (4)$$
$$P_\downarrow = \{(j,m) \in J_L \times M_L : j \not\uparrow m \text{ and } j \downarrow m\} \quad (5)$$
$$P_\updownarrow = \{(j,m) \in J_L \times M_L : j \uparrow m \text{ and } j \downarrow m\} \quad (6)$$
$$P_\circ = \{(j,m) \in J_L \times M_L : j \not\uparrow m \text{ and } j \not\downarrow m\} \quad (7)$$

iv) Let $L = (S, \leq)$ be a lattice and T the table of L. The table T contains in columns the join-irreducible elements and in rows the meet-irreducible elements such that for all $j \in J_L$ and $m \in M_L$, $T[j,m]$ contains $\times, \uparrow, \downarrow, \updownarrow$ or $\circ$ where (j,m) belongs to $P_\uparrow, P_\downarrow, P_\updownarrow$ and $P_\circ$ respectively, as it is shown in Table 1.

A binary table *is a particular case of $T[j,m]$ which contains only $\times$ when $(j,m) \in P_\leq$. In Formal Concept Analysis, the binary table will be called* formal context *from which a* concept lattice *will be defined.*

3. Concept lattice and formal concept

Two notions are relevant to define the concept lattice: the *formal context* and the *Galois connexion.*

Definition 3.1. ***(Formal context)*** *A* formal context *$(\mathcal{O}, \mathcal{I}, R)$ consists of two sets $\mathcal{O}$ and $\mathcal{I}$ and a binary relation R between them as it is shown in Table 2. The elements of $\mathcal{O}$ are called* objects *and the elements of $\mathcal{I}$ are called* attributes.

Definition 3.2. ***(Galois connexion)*** *Let f and g be two functions on the set of object $\mathcal{O}$ and the set of attributes I respectively, such that:*

$$f(A) = \{x \in \mathcal{I} | yRx, \forall y \in A \subseteq \mathcal{O}\} \quad \text{and} \quad g(B) = \{y \in \mathcal{O} | yRx, \forall y \in B \subseteq \mathcal{I}\}$$

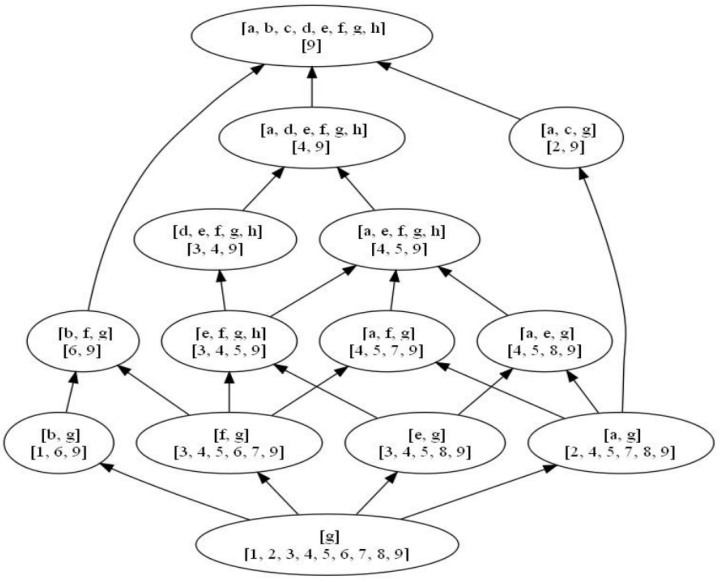
in order to obtain all the attributes which describe the same set of objects and share the same set of attributes respectively. The two functions form a Galois connexion *between objects and attributes.*

Definition 3.3. ***(Concept lattice)*** *The* concept lattice *or* Galois lattice *is defined on two sets ($\mathcal{O}$ and $\mathcal{I}$). Its elements are formed by all the possible connexions between the set of objects $\mathcal{O}$ and the set of attributes $\mathcal{I}$ and can be represented by a graph where the relations between elements (called concepts) are described by the formal context.*

Table 2
An example of a Formal context [?].

	a	b	c	d	e	f	g	h
1		×					×	
2	×		×				×	
3				×	×	×	×	×
4	×			×	×	×	×	×
5	×				×	×	×	×
6		×				×	×	
7	×					×	×	
8	×				×		×	
9	×	×	×	×	×	×	×	×

Figure 1: Concept lattice from the context shown in Table 2 [?]



Definition 3.4. ***(Formal concept)*** *A* formal concept *of a* formal context $(\mathcal{O}, \mathcal{I}, R)$ *is a pair* (A, B) *where* $A \subseteq \mathcal{O}$, $B \subseteq \mathcal{I}$ *and* $f(A) = B$, $g(B) = A$ *with* f, g *two maps from* $\mathcal{O}$ *to* $\mathcal{I}$ *and from* $\mathcal{I}$ *to* $\mathcal{O}$ *respectively. The sets* A *and* B *are called the* extent *and the* intent *of the formal concept* (A, B) *respectively [?].*

4. Implicational system

An *implicative system* or a *system of implication rules* is used to express the implication between data. The implication rule or the *exact rule* is a pair $(X, Y) \in R$ such that X implies Y where X is called *premise of the rule* and Y the *conclusion of the rule.*

Generally, an association rule on the set S is denoted by $X \to Y$ such that X and Y are subsets on S. We can define some interesting measures for association rules based on statistical validity measures such as the *support* and the *confidence* of the rule. Let $(\mathcal{O}, \mathcal{I}, R)$ be a formal context, let X and Y be the sets of attributes belonging to I.

Definition 4.1. ***(Support)*** *The support of* X *is defined as the ratio between the number objects which verify all the attributes of* X *and the total number of the objects:*

$$supp(X) = \frac{|g(X)|}{|\mathcal{O}|}$$

Table 3
Formal database

$\mathcal{R}$	a	b	c	d	e
o_1	1	0	1	1	0
o_2	0	1	1	0	1
o_3	1	1	1	0	1
o_4	0	1	0	0	1
o_5	1	1	1	0	1
o_6	0	1	1	0	1

Therefore the support of the association rule is defined as the percentage of records which contain $X \cup Y$ to the total number of records in the database :

$$support(X \to Y) = support(X \cup Y) = \frac{Number\ of\ possible\ case}{Total\ number\ of\ records}$$

Definition 4.2. ***(Confidence)*** *The* confidence *is defined as the percentage of the number of records which contain $X \cup Y$ to the total number of records that contain X*

$$Conf(X \to Y) = \frac{Supp(X \to Y)}{Supp(X)}$$

However, an association rule is said to be *valid* if its *support* and its *confidence* are greater than or equal to a certain *threshold* denoted by *minsup* and *minconf* which is the minimal support and the minimal confidence respectively [**?**].

5. Frequent pattern extraction

Frequent pattern extraction is a technique mostly used in data mining. Its purpose is to find patterns which frequently appear in a *formal database* the values of which are booleans indicating the presence or absence of a property.

Definition 5.1. ***(Formal database)*** *A formal database is defined by a set of three elements $(\mathcal{O}, \mathcal{P}, \mathcal{R})$ where:*

i) $\mathcal{O}$ is a set of finite objects;

ii) $\mathcal{P}$ is a set of finite properties;

iii) $\mathcal{R}$ is a relation on $\mathcal{O} \times \mathcal{P}$ that indicates if an object in $\mathcal{O}$ has a property in $\mathcal{P}$.

Example 5.2. *Let us consider a formal database shown in Table 3 where:*

i) $\mathcal{O}$ is a set of object denoted by $\{o_1, o_2, o_3, o_4, o_5, o_6\}$;

ii) $\mathcal{P}$ the set of attributes denoted by $\{a, b, c, d, e\}$.

iii) $o\ \mathcal{R}\ p$ if 1 is found on the intersection of the row of o and the column of p.

Therefore, we have from Table 3:

i) Pattern of length 0*: $\emptyset$*

ii) Pattern of length 1*: $\{a\}, \{b\}, \{c\}, \{d\}$ and $\{e\}$. For the sake of simplicity, we will denote by: $\underline{a}, \underline{b}, \underline{c}, \underline{d}, \underline{e}$*

iii) Pattern of length 2*: $\underline{ab}, \underline{ac}, \underline{ad}, \underline{ae}, \underline{bc}, \underline{bd}, \underline{be}, \underline{cd}, \underline{ce}, \underline{ed}$*

iii) Pattern of length 3*: $\underline{abc}, \underline{abd}, \underline{abe}, \underline{acd}, \underline{ace}, \underline{ade}, \underline{bcd}, \underline{bce}, \underline{bde}, \underline{cde}$*

iv) Pattern of length 4*: $\underline{abcd}, \underline{abce}, \underline{abde}, \underline{acde}, \underline{bcde}$*

v) Pattern of length 5*: $\underline{abcde}$*

If we consider o_1 we should notice that it has the following patterns: $\emptyset, \underline{a}, \underline{c}, \underline{d}, \underline{ac}, \underline{cd}, \underline{ad}, \underline{acd}$

6. Algorithm for extracting frequent patterns

One of the commonly used algorithms for extracting frequent patterns is the *A-priori*. It is used to cover all the power set $2^{|\mathcal{P}|}$ of the patterns, to compute their supports and keep only the most frequent among them. Let L_i be the set of all frequent patterns of length i. The approach used is described as follows:

i) Determine the set of all frequent patterns of length 1 denoted L_1;

ii) Build a set of C_2 frequent patterns of size 2 (obtained from the frequent patterns of size 1). Therefore, we obtain the list of frequent patterns of size 2 denoted L_2 by keeping only those which are greater or equal to the threshold (σ_s).

iii) Continue the process until $L_i = \emptyset$.

Therefore, we have the following proposition:

Proposition 6.1. *(Support decreasing principal [?])*

1. *Every sub frequent patterns set is also frequent.*
$$\textit{if } m' \subseteq m \textit{ and } support(m) \geq \sigma_s \textit{ then } support(m') \geq \sigma_s \tag{8}$$

2. *Every patterns above a non frequent pattern is also not frequent*
$$if\ m \subseteq m' \ and\ support(m) < \sigma_s\ then\ support(m') < \sigma_s \tag{9}$$

Therefore, we have the formal Algorithm for finding frequent patterns (see Algorithm 1)

Algorithm 1 A-priori Algorithm [?]

Require: $(\mathcal{O}, \mathcal{P}, \mathcal{R})$ *with* $\sigma_s \in [0, 1]$
Ensure: `frequent patterns`
  L_1 `list of motifs of` $support > \sigma_s$
  $i \leftarrow 1$
  repeat
    `i++`
    `from` L_{i-1} `determine the set` C_i `of frequent patterns`
    $L_i \leftarrow \emptyset$
    for all $m \in C_i$ **do**
      if $support(m) > \sigma_s$ **then**
        `add` m `to` L_i
      end if
    end for
  until $L_i = \emptyset$
  return $\bigcup_{i \geq 1} L_i$

7. Association rules extraction and analysis

The objective in text mining is to find, through a collection of data, relations between concepts and to locate them within the documents and eventually examine the set of documents obtained based on those relations. Therefore, those relations are expressed through *association rules* obtained from text. There are two main steps:

- The extraction of association rules;
- The classification of rules according some statistical indices.

The extraction of association rules is made through the formal concept analysis via frequent patterns generated by A-priori Algorithm 1. The obtained patterns will be used to generate association rules. However, statistical indices are weighting measures affected to the rule. The weight of the rule will help the analyst to classify each rule.

8. Association rule

The aims of association rules is to find correlation between data and can be presented as follows:

$$R : \underline{a} \wedge \underline{b} \Rightarrow \underline{c} \wedge \underline{d} \wedge \underline{e} \text{ where } \underline{a}, \underline{b}, \underline{c}, \underline{d}, \underline{e} \text{ are patterns.} \tag{10}$$

Rule (10) can be interpreted as : *if a document contains* $\{\underline{a}, \underline{b}\}$ *patterns, then it contains also* $\{\underline{c}, \underline{d}, \underline{e}\}$. In order to select interesting rules from the set of all possible rules, two concepts are introduced:

Definition 8.1. ***(Support)*** *The support of an association rule represent the number of documents which are described by terms which appeared at the left and right side of the rule [?]. Let* B *be the left side of the rule and* H *the right side. We have:*

$$supp\,[B \Rightarrow H] = number\ of\ documents\ containing \{\underline{a}, \underline{b}, \underline{c}, \underline{d}, \underline{e}\} \tag{11}$$

It can be also written in terms of probability as follows:

$$P(B, H) = \frac{supp\,[B \Rightarrow H]}{total\ number\ of\ documents\ considered}. \tag{12}$$

Definition 8.2. ***(Confidence)*** *The confidence of an association rule is defined by:*

$$conf\,[B \Rightarrow H] = \frac{number\ of\ documents\ containing\ B\ and\ H}{number\ of\ documents\ containing\ B}. \tag{13}$$

In terms of probability, it can be written as a conditional probability of H *knowing* B*:*

$$P(H|B) = \frac{supp\,[B \Rightarrow H]}{number\ of\ documents\ containing\ B.} \tag{14}$$

9. Statistical indices associated to the association rules

The support and the confidence are not the only indices which can give enough information to the analyst on the qualities of a rule. However, other indices such as *lift, dependence* of an association rule are also important.

Definition 9.1. ***(Lift)*** *The lift of a rule is defined as:*

$$lift(B \Rightarrow H) = \frac{supp\,[B \cup H]}{supp(B) \times supp(H)} \tag{15}$$

If a rule has a lift *of* 1*, it will imply that the probability of occurrence of the antecedent and of the consequence are independent. Therefore, no interesting association rule can be found. However, if the lift is* > 1*, the two occurrences are dependent and that makes the rule interesting for predicting the consequence in the future. Since it is important to know the degree of dependence between occurrence, let us also define the* dependence *of a rule.*

Definition 9.2. ***(Dependence)*** *The indice of dependence reinforces an association rule by giving the precision about the degree of dependence of the rule. It is defined as:*

$$dep\,[B \Rightarrow H] = |P(H|B) - P(H)| \tag{16}$$

10. Algorithm of generating association rules

Let $R = p_1 \longrightarrow p_2 \setminus p_1$ be the association rule with $p_1 \subseteq p_2$ (Note that $E \setminus F = \{x | x \in E \text{ and } x \notin F\}$). The principle of the algorithm is as follows:

- Let us consider rules of the form $p_1 \longrightarrow p_2 \setminus p_1$ where the conclusion is of length 1;
- Delete non valid rules;
- Combine the conclusions of the valid rules;

- Move on to conclusions of length 2, by deleting invalid rules and by combining the conclusions of the valid rules as well;
- Move on to inclusions of length 3 and so on

One is interested in the valid rules, that means, whose

$$\texttt{support}(R) = \texttt{support}(p_2) \geq \sigma_s$$

. Since $p_1 \subseteq p_2$, by the decreasing support principle,

$$\texttt{support}(p_1) \geq \sigma_s$$

11. Experimentations

Our database is formed with books of different domains in mathematical sciences (topology, algebra, artificial intelligence, data mining, machine learning, etc.) and a set of patterns that has been used to perform our research in the database. We constructed a formal context $K = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ where $\mathcal{O}$ represents a set of *objects* characterized by our sample of *book's titles*, $\mathcal{A}$ represents a set of *attributes* characterized by our sample of patterns and $\mathcal{R}$ the relation between *objects* and *attributes*. Therefore, to represent the presence of the patterns j in book i a cross $\times$ has to be found at the intersection of the i^{th} row and the j^{th} column.

12. Information extraction using association rules

Information extraction is a technique of finding useful information from a database. Association rule has been used to extract those pieces of information. However, association rules are interested to patterns that appear frequently in a database. Therefore, the process started by extracting all the patterns by keeping only the frequents with respect to a certain threshold.

13. Frequent pattern extraction

Apriori algorithm has been used to determine frequent patterns with a threshold $\sigma_s = \frac{6}{27}$. Therefore, we firstly determined the patterns of length 1 which are frequent denoted L_1. Then, constructed C_2 obtained by combining two by two the frequent patterns of length 1. That gave us the set of all the frequent patterns of length 2 denoted by L_3. We continued the process until we have obtained an empty set L_i.

14. Extracting association rules

Extracting association rules are based on frequent patterns. However, an association rule is said to be valid if its support is greater than or equal to a fixed threshold $\sigma_s = \frac{6}{27}$. Therefore the algorithm has been applied directly on those patterns of length 2 as an association rule is known to be of the form $B \Rightarrow H$.

15. More indices associated to the association rules:

The support and the confidence of an association are not sufficient to give enough informations about the quality of the rule. Furthermore, two other indices have been presented, namely as the *dependence* and the *lift* of an association rule in order to give supplementary information to our studies. However, only valid rules will be considered, that means where the threshold of the support σ_s and the threshold of the confidence σ_c are greater or equal to $\frac{6}{27}$ and $\frac{3}{4}$ respectively. The result is shown in tables below.

1. **Valid rules built on patterns of length 2**

	$\underline{a} \longrightarrow \underline{d}$	$\underline{a} \longrightarrow \underline{g}$	$\underline{g} \longrightarrow \underline{a}$	$\underline{a} \longrightarrow \underline{j}$	$\underline{b} \longrightarrow \underline{j}$	$\underline{c} \longrightarrow \underline{j}$	$\underline{c} \longrightarrow \underline{m}$	$\underline{g} \longrightarrow \underline{d}$
dependence	49%	66%	41%	3%	11%	14%	31%	41%
lift	**3.20**	**3.00**	**3.00**	0.96	0.89	0.84	**1.69**	**2.10**
	$\underline{d} \longrightarrow \underline{h}$	$\underline{d} \longrightarrow \underline{j}$	$\underline{g} \longrightarrow \underline{j}$	$\underline{h} \longrightarrow \underline{j}$	$\underline{i} \longrightarrow \underline{j}$	$\underline{l} \longrightarrow \underline{j}$	$\underline{m} \longrightarrow \underline{j}$	
dependence	25%	9%	48%	4%	7%	13%	2%	
lift	**1.44**	0.90	**1.13**	**1.05**	0.92	0.84	**1.03**	

2. **Valid rules built on patterns of length 3**

	$\underline{ad} \longrightarrow \underline{g}$	$\underline{ag} \longrightarrow \underline{d}$	$\underline{dg} \longrightarrow \underline{a}$	$\underline{a} \longrightarrow \underline{dg}$	$\underline{b} \longrightarrow \underline{jh}$	$\underline{bh} \longrightarrow \underline{j}$	$\underline{dg} \longrightarrow \underline{h}$
dependence	67%	49%	60%	61%	44%	6%	48%
lift	**3**	**2.3**	**3.3**	**3.3**	**1.93**	**1.35**	**1.54**
	$\underline{dh} \longrightarrow \underline{g}$	$\underline{gh} \longrightarrow \underline{d}$	$\underline{g} \longrightarrow \underline{dh}$	$\underline{dg} \longrightarrow \underline{j}$	$\underline{dj} \longrightarrow \underline{g}$	$\underline{gj} \longrightarrow \underline{d}$	$\underline{g} \longrightarrow \underline{dj}$
dependence	42%	63%	48%	11%	54%	41%	48%
lift	**2.25**	**2.7**	**2.25**	**1.13**	**2.63**	**2.1**	**2.63**
	$\underline{dj} \longrightarrow \underline{m}$	$\underline{dm} \longrightarrow \underline{j}$	$\underline{dj} \longrightarrow \underline{l}$	$\underline{dl} \longrightarrow \underline{j}$	$\underline{gh} \longrightarrow \underline{j}$	$\underline{hm} \longrightarrow \underline{j}$	$\underline{jl} \longrightarrow \underline{m}$
dependence	30%	31%	11%	31%	3%	11%	11%
lift	**1.69**	**1.13**	**1.69**	**2.25**	**1.13**	**1.13**	1
	$\underline{ad} \longrightarrow \underline{j}$	$\underline{aj} \longrightarrow \underline{d}$	$\underline{dj} \longrightarrow \underline{a}$	$\underline{a} \longrightarrow \underline{dj}$	$\underline{ag} \longrightarrow \underline{j}$	$\underline{aj} \longrightarrow \underline{g}$	$\underline{gj} \longrightarrow \underline{a}$
dependence	11%	63%	49%	56%	11%	83%	52%
lift	**1.13**	**2.7**	**2.89**	**2.89**	**1.13**	**3.5**	**3**
	$\underline{g} \longrightarrow \underline{aj}$	$\underline{ag} \longrightarrow \underline{h}$	$\underline{ah} \longrightarrow \underline{g}$	$\underline{gh} \longrightarrow \underline{a}$	$\underline{a} \longrightarrow \underline{gh}$	$\underline{bj} \longrightarrow \underline{d}$	$\underline{dj} \longrightarrow \underline{b}$
dependence	56%	30%	87%	74%	64%	63%	53%
lift	**3.5**	**1.54**	**3.6**	**3.9**	**3.9**	**2.7**	**3.4**
	$\underline{dg} \longrightarrow \underline{h}$	$\underline{dh} \longrightarrow \underline{i}$	$\underline{ml} \longrightarrow \underline{j}$	$\underline{a} \longrightarrow \underline{gj}$	$\underline{b} \longrightarrow \underline{jd}$		
dependence	30%	47%	33%	67%	70%		
lift	**1.54**	**2.15**	0.5	**3**	**3.38**		

3. **Valid rules built on patterns of length 4**

	$\underline{adg} \longrightarrow \underline{j}$	$\underline{dgj} \longrightarrow \underline{a}$	$\underline{adj} \longrightarrow \underline{g}$	$\underline{agj} \longrightarrow \underline{d}$	$\underline{ad} \longrightarrow \underline{gj}$	$\underline{ag} \longrightarrow \underline{dj}$
dependence	11%	74%	67%	63%	67%	56%
lift	**1.3**	**3.9**	**3.5**	**2.7**	**3.5**	**3.9**
	$\underline{a} \longrightarrow \underline{dgj}$	$\underline{dgh} \longrightarrow \underline{j}$	$\underline{dgj} \longrightarrow \underline{h}$	$\underline{ghj} \longrightarrow \underline{d}$	$\underline{dg} \longrightarrow \underline{hj}$	$\underline{hd} \longrightarrow \underline{gj}$
dependence	74%	11%	30%	63%	23%	42%
lift	**3.9**	**1.1**	**1.5**	**8.1**	**1.7**	**2.3**

4. **Valid rules built on patterns of length 5**

	$\underline{adgj} \longrightarrow \underline{h}$	$\underline{dghj} \longrightarrow \underline{a}$	$\underline{adg} \longrightarrow \underline{hj}$	$\underline{agh} \longrightarrow \underline{dj}$
dependence	44%	74%	48%	70%
lift	**1.5**	**3.4**	**1.9**	**3.4**
	$\underline{dgh} \longrightarrow \underline{aj}$	$\underline{dgj} \longrightarrow \underline{ah}$	$\underline{agj} \longrightarrow \underline{dh}$	$\underline{adj} \longrightarrow \underline{gh}$
dependence	78%	67%	56%	78%
lift	**4.5**	**10.5**	**2.9**	**4.5**
	$\underline{ad} \longrightarrow \underline{ghj}$	$\underline{ag} \longrightarrow \underline{dhj}$	$\underline{ah} \longrightarrow \underline{dgj}$	$\underline{aj} \longrightarrow \underline{dgh}$
dependence	78%	85%	74%	78%
lift	**4.5**	**5.8**	**4.6**	**4.5**
	$\underline{dj} \longrightarrow \underline{agh}$	$\underline{dh} \longrightarrow \underline{ghj}$	$\underline{gh} \longrightarrow \underline{adj}$	$\underline{a} \longrightarrow \underline{dghj}$
dependence	53%	53%	78%	78%
lift	**3.9**	**2.9**	**4.5**	**3.4**

16. Interpretation of the results

We have patterns that have been generated for our experiment on textual data base with electronic books picked from different domains and keywords. 48 patterns are frequents and 166 association rules with 83 valid according to the fixed threshold. However, as mentioned earlier, a good association is not only based on its *support* and *confidence* but we need to take in consideration the *dependence* and the *lift* of the rule to have a good interpretation of the result.

17. Effective association rules

An association rule is said to be interpretable if the value of the rule has a *lift* > 1 and has a certain degree of dependence. For our case we have fixed it at 75% approximatively. In this section, we present some of the interpreted rules which was useful for our study. However, each of the patterns $\underline{a}, \underline{b}, \underline{c}, \underline{d}, \underline{e}, \underline{f}, \underline{g}, \underline{h}, \underline{i}, \underline{j}, \underline{k}, \underline{l}, \underline{m}$

mean a particular keyword such as : *machine learning*, *topology*, *knowledge discovery*, etc. that has been picked in various domain for the sake of studies

Rule	Support	Lift	Dependence	Interpretation
$\underline{ag} \longrightarrow \underline{dhj}$	6	5.8	85%	This rule shows the support of 6 over 27 documents where the presence of *"information retrieval"* and *"text mining","machine learning"* depends at 85% of *"association rules", "knowledge discovery"* presence
$\underline{bh} \longrightarrow \underline{j}$	6	1.3	6%	This rule has the same support than the previous one. However, it has a dependence of 6% because the combination of words such *"topology"* and *"text mining"* has a weak rate of implication to *machine learning*
$\underline{ml} \longrightarrow \underline{j}$	7	0.5	33%	The rule forms by *"song writing"* and *"children care"* and *machine learning* does not have any dependence, their lift is less than 1

The interpretation could have been very interesting if we had taken into consideration the number of occurrence of each keyword in documents to obtain the formal context. However, based on the method used, the result presented in this section for some rules is what was expected.

18. Conclusion

In this paper, we presented concept lattice as very effective tool for data mining. In particular, we discuss some of its applications in text mining. For text mining, association rules discovery, mostly uses formal concept analysis to analyze the relations between patterns which appear at the same time. The method presented is performed on a database formed of electronic books from different domains and related keywords on which we applied some statistical analysis to find out interesting association rule. However, lattice theory can be used also for image sets characterization by using landmarks to enable a machine to automatically classify objects with respect to the image class they belong to. To be exploited by data mining methods, images must undergo a series of processing to obtain a digital table. That process aims at assigning to each set of images extracted from an object an identifier called *landmark* which aims at characterizing and distinguishing each set of images and, from there, each object. That may belong to our future work.

References

[1] K. Bertet ; Structure de treillis: contributions structurelles et algorithmiques: quelques usages pour des données images; 2010.
[2] K.I. Ignatov , Introduction to formal concept analysis and its applications in information retrieval and related fields; Russian Summer School in Information Retrieval; **42–141**; Springer; 2014.
[3] Zhao, Qiankun, Bhowmick and Sourav. ; Association rule mining: A survey; Nanyang Technological University, Singapore; 2003.
[4] Masseglia, Florent and Poncelet, Pascal and Teisseire, Maguelonne. ; Successes and new directions in data mining; IGI Global; 2008.
[5] Zhang, Chengqi and Zhang, Shichao. ; Association rule mining: models and algorithms, Springer-Verlag; 2002.
[6] Cherfi, Hacene and Toussaint, Yannick. ; Adéquation d'indices statistiques à l'interprétation de règles d'association; 6èmes Journées internationales d'Analyse statistique des Données Textuelles-JADT 2002; 233–244; 2002.
[7] G. Grätzer ; General lattice theory; Springer Science & Business Media; 2002.