



The Minimal Description of Formal Concept Analysis

Li Fu

Department of Mathematics, Qinghai University for Nationalities

Xining Ba Yi middle road, Qinghai 810000, China

Tel: 86-971-829-2658 E-mail: 0971@163.com

The research is financed by The Education department chunhui program. No. Z2006-1-81001 (Sponsoring information)

Abstract

In rough sets, there were some of results of minimal description formed by covering, this paper gives some remarks about them. Using the idea of minimal description in formal concept analysis, discusses the corresponding relation between the covering of objects set and the attributes set.

Keywords: Rough set, Formal concept analysis, Minimal description, Covering, Partition

In 1982, R.Wille (1999) proposed a new model to represent the formal concepts associated to a context (G, M, I) , named formal concept analysis based on the formal context, which is a pair consisting of a set of objects (the extension) and a set of attributes (the intension) such that the intension consists of exactly those attributes that the objects in the extension have in common, and the extension contains exactly those objects that share all attributes in the intension. The extension and the intension of a concept uniquely determine each other. The main goal is to reveal the hierarchical structure of formal concepts and to investigate the dependencies among attributes. It provides a theoretical framework for the discovery and design of concept hierarchies from relational information systems. It serves as a basis for conceptual data analysis and knowledge processing. The family of all formal concepts is a complete lattices, which is an effective method for several real-world applications in data analysis, such as object-oriented databases, inheritance lattices, mining for association rules, generating frequent sets (F.M.Baltasar, 1998; J.S. Deogun, 2004; So Kuznetsov, 2002; R.Godin, 1995) etc, one of the important challenges in data handling is generating or navigating the concept lattice of binary relation.

The theory of rough sets as a tool for processing uncertain and incomplete information, proposed by Z.Pawlak (1991) in 1982, in which the lower approximation and upper approximation of an arbitrary subset of universe U are the basic operators.

The concepts of the lower and upper approximations in rough sets theory are fundamental to the examination of granularity in knowledge, and are an effective way of studying imprecision, vagueness, and uncertainty, and rough set has been succeeded using all kinds of artificial intelligence fields, such as: data mining, machine learning. There are many authors studying rough sets and formal concept analysis, the readers can see (Zhang wenxiu, 2006; A.Burusco, 1994; A.Burusco, 2000; R. Belohavek, 1995). In (A.Skowron, 1992; Z.Bonikowski, 1998; William Zhu, 2003; J.A.Pomykala, 1987), the concept of covering of a universe was presented to construct the upper and the lower approximations of an arbitrary set. In (Z.Bonikowski, 1998), the authors studied the reduct of covering of generalized rough sets; In (Z.Bonikowski, 1998; Z.Bonikowski, 2003), the authors mainly studied the structure of cover, authors proposed a representative approximation spaces, that is, discussed the point x of universe U (*i.e.* $x \in U$) by covering, they defined the description of x and the minimal description of x ; In (Xu WeiHua, 2007), the authors considered some new concepts and main results in generalized rough sets induced by a covering; In (Qiu Weigen, 2006), the author discussed fined covering fuzzy generalized rough sets; In (Eric C.C.Tsang, 2008), the authors defined the induced cover and the intersection of a family of covers by using the induced cover in covering generalized rough sets.

The rest of the paper is organized as following, in section 1, we give some basic knowledge of concept lattices and rough sets. In section 2, give some remarks of the minimal description which depict by covering in covering approximation space. In section 3, we use minimal description to objects set and attributes set of formal context, and discuss the corresponding relation of the covering of objects set and the covering of attributes set, and conclusions are given in section 4.

1. Preliminaries

Definition 1.1 (B.Ganter, 1999) Let (G, M, I) be a formal context, where $G = \{a_1, \dots, a_m\}$ is an objects set, $M =$

$\{x_1, \dots, x_n\}$ is an attributes set, a binary relation $I \subseteq G \times M$. If object $g \in G$ has attribute $m \in M$ under relation I , denote gIm or $(g, m) \in I$. Typically, 1 means $(g, m) \in I$, and 0 means $(g, m) \notin I$. (In the following, we will use this denotation).

Definition 1.2 (B.Ganter, 1999) In formal context $K = (G, M, I)$, $\forall A \subseteq G, \forall B \subseteq M$, define:

$A' = \{m \in M \mid gIm, \forall g \in A\}$, i.e., the set of all attributes shared by all objects from A .

$B' = \{g \in G \mid gIm, \forall m \in B\}$, i.e., the set of all objects sharing all attributes from B .

Definition 1.3 (B.Ganter, 1999) The formal concept of formal context $K = (G, M, I)$ is the set pair (A, B) , where $A \subseteq G, B \subseteq M$, and satisfying $A' = B, B' = A$. A, B is called the extent and intent of concept (A, B) , respectively, and using $\mathcal{B}(G, M, I)$ denote the set of all concepts of $K = (G, M, I)$.

Definition 1.4 (Z.Bonikowski, 1998) Let U be a finite nonempty set of objects, called the **universe**, let C be a covering of U , i.e. C is a family of nonempty subsets of U whose union is U , i.e. $U = \bigcup_{X \in C} X$. The ordered pair $\mathcal{A} = (U, C)$ is called an **approximation space**.

Remark 1.1 It should be noticed that the definition of covering, the join of members of C need not be an empty set, that is, $\exists X_i, X_j \in C, s.t. X_i \cap X_j \neq \emptyset$. In other word, the covering of U can not be a partition of U , conversely, a partition of U is also a covering of U .

Definition 1.5 (Z.Bonikowski, 1998) Let $x \in U$, the family $\{X \in C \mid x \in X\}$ is called the **description of x** .

Definition 1.6 (Z.Bonikowski, 1998) Let $x \in U$, the family $Md(x) = \{X \in C \mid x \in X \wedge \forall Y \in C(x \in Y \wedge Y \subseteq X \Rightarrow X = Y)\}$ is called the **minimal description of x** .

Definition 1.7 (Z.Bonikowski, 1998) An object $x \in U$ such that $Md(x) = \{X\}$ is called **representative element of X** .

Definition 1.8 (Xu WeiHu, 2007) Let U be the universe, C be a subsets family of U , if $\bigcup C = U$, then C is the covering of U , the ordered pair (U, C) is called a **covering approximation space**. $\forall X, Y \in C, \exists Z_1, \dots, Z_k \in C$, such that, $X \cap Y = Z_1 \cup \dots \cup Z_k$, then called C is a **finer covering of U** , the ordered pair (U, C) is called a **finer covering approximation space**.

Remark 1.2 It is important to note that in the definition of finer covering, C includes the empty set, that is, $\emptyset \in C$.

Definition 1.9 (Qiu Weigen, 2006) Suppose U is a finite universe, and $C = \{C_1, \dots, C_n\}$ is a cover of U . For every $x \in U$, let $C_x = \{C_j \mid C_j \in C, x \in C_j\}$, then $Cov(C) = \{C_x \mid x \in U\}$ is also a cover of U , we call it the **induced cover of C** .

Remark 1.3 For all $C_x, C_y \in C, C_x \neq C_y, C_x \cap C_y$ can not be empty, for example:

$U = \{1, \dots, 8\}, C = \{\{1, 3, 4\}, \{2, 5, 7\}, \{1, 2, 8\}, \{3, 4, 6, 7\}, \{2, 6, 7\}, \{1, 7, 8\}, \{2, 5, 7, 8\}, \{4, 5, 7\}\}$, in here, $C_3 = \{3, 4\}, C_4 = \{4\}, C_3 \cap C_4 \neq \emptyset$, so C_x is a covering of U , not a partition of U , however, for all $C_x \in Cov(C)$, if $x \neq y, C_x \cap C_y = \emptyset$, then $Cov(C)$ is a partition of U ; if $C_x \cap C_y \neq \emptyset$, and $C_x \subseteq C_y$, or $C_y \subseteq C_x$, then $Md(x)$ is a singleton set, and is a special covering of U .

Definition 1.10 (Qiu Weigen, 2006) Suppose U is a finite universe, and $\Delta = \{C_1, \dots, C_m\}$ is a family of cover of U . For every $x \in U$, let $\Delta_x = \{C_{ix} \mid C_{ix} \in Cov(C_i), x \in C_{ix}\}$, then $Cov(\Delta) = \{\Delta_x \mid x \in U\}$ is also a cover of U , we call it the **induced cover of Δ** .

Remark 14 For all $x \in U$, if $x \in Cov(\Delta)$, then must exist C_i , such that $x \in Cov(C_i)$, that is, $Cov(\Delta) \subseteq Cov(C_i)$. In other words, $Cov(\Delta)$ is the smaller minimum than $Cov(C_i)$.

2. Some remarks about the existed results

Remark 2.1 In the definition of covering, $X_i \neq X_j \in C$, it can exist $X_i \cap X_j \neq \emptyset, \forall x \in U$, at least exist one $X \in C, s.t. x \in X$, so, the minimal description of x must exist, if x belongs to the two or more than two subsets which do not include each other, then the elements of $Md(x)$ more than 1, that is $|Md(x)| \geq 2$, for example:

Example 1 Let universe $U = \{1, 2, \dots, 5\}, C = \{\{1, 3\}, \{1, 2, 3\}, \{2, 3, 4\}, \{3, 5\}, \{1, 3, 5\}, \{4, 5\}\}$.

Obviously, C is the covering of U . Using $Dp(x)$ denote the description of x , then $Dp(1) = \{\{1, 3\}, \{1, 2, 3\}, \{1, 3, 5\}\}, Dp(2) = \{\{1, 2, 3\}, \{2, 3, 4\}\}, Dp(3) = \{\{1, 2, 3\}, \{2, 3, 4\}, \{3, 5\}, \{1, 3, 5\}\}, Dp(4) = \{\{4, 5\}, \{2, 3, 4\}\}, Dp(5) = \{\{1, 3, 5\}, \{3, 5\}, \{4, 5\}\}$. And $Md(1) = \{\{1, 3\}, \{1, 3, 5\}\}, Md(2) = \{\{1, 2, 3\}, \{2, 3, 4\}\}, Md(3) = \{\{1, 2, 3\}, \{2, 3, 4\}, \{3, 5\}\}, Md(4) = \{\{4, 5\}, \{2, 3, 4\}\}, Md(5) = \{\{3, 5\}, \{4, 5\}\}$.

Proportion 2.1 Let C be nonempty family of $U, C = \{X_i \mid \emptyset \neq X_i \subseteq U, i = 1, \dots, k\}$, if $U = \bigcup_{i=1}^k X_i$, and $\forall x \in U$, exist unique minimal description of x , that is, $\exists! X \in \{X \in C \mid x \in X\}$, such that $Md(x) = \{X\}$, then $\forall i \neq j, X_i, X_j \in C, X_i \cap X_j = \emptyset$, or $X_i \subseteq X_j$. Furthermore, if $x, y \in X$, then $Md(x) = Md(y)$.

Proof If covering C is the partition of the universe U , then $\forall x \in U$, at least one $X_i \in C, s.t. x \in X_i$, if x only belongs to the unique X_i , then $\forall i \neq j, X_i, X_j \in C, X_i \cap X_j = \emptyset$, if x belongs to more than one X_i , then $\forall i \neq j, X_i, X_j \in C, X_i \subseteq X_j$, otherwise, $Md(x) = \{X_i, X_j \mid X_i, X_j \in C\}$, it is contradict with the minimal description of x .

Remark 2.2 In (Z.Bonikowski, 1998), the authors defined *fin* covering, where $\forall X, Y \in C, \exists Z_1, \dots, Z_k \in C$, such that $X \cap Y = Z_1 \cup \dots \cup Z_k$, however, $X \cap Y$ can be empty, if $X \cap Y \neq \emptyset$, although there exists in theory, it is difficult to achieve in practice, even in (Z.Bonikowski, 1998), the example that the authors gave did not satisfy $\forall X, Y \in C, X \cap Y \neq \emptyset$, for example: $X_2 = \{a, c\}, X_5 = \{e, b\}, X_2 \cap X_5 = \emptyset$, so, it must permit $X \cap Y = \emptyset$, that is, in the definition of covering includes empty set, in addition, the *fin* covering formed by non-empty subsets class almost is singleton set, I don't think it has too much meaning.

Remark 2.3 From the definition of the representative element, we can declare, if the every element x of the universe U is a representative element, that is $\forall x \in U, \exists X \in \{X \in C \mid x \in X\}$, such that, $Md(x) = \{X\}, \forall X, Y \in C$, then it must have $X \cap Y = \emptyset$ or $X \subseteq Y$, if $\forall X, Y, X \cap Y = \emptyset$, then the covering C is formed the partition of U , it is identical with the result of proposition 2.1.

3. The minimal description of formal context

The formal context (G, M, I) is studying the concept pair (A, B) which extent and intent are objects subset $A \subseteq G$ and attributes subset $B \subseteq M$, respectively, so, we can consider the covering formed by their subset family, and the minimal description of elements.

Example 2 The following table 1 is formal context (G, M, I) , where $G = \{1, 2, \dots, 8\}, M = \{a, b, \dots, i\}$.

Given $C_G = \{X_1, X_2, X_3, X_4, X_5\}$, where $X_1 = \{1, 3\}, X_2 = \{1, 2, 4\}, X_3 = \{3, 5, 7\}, X_4 = \{6, 8\}, X_5 = \{6, 7, 8\}$. Obviously, $\bigcup_{i=1}^5 X_i = G$, and: $Dp(1) = Md(1) = \{X_1, X_2\}, Dp(2) = Md(2) = \{X_2\}, Dp(3) = Md(3) = \{X_1, X_3\}, Dp(4) = Md(4) = \{X_2\}, Dp(5) = Md(5) = \{X_3\}, Dp(6) = Dp(8) = \{X_4, X_5\}, Md(6) = Md(8) = \{X_4\}, Dp(7) = Md(7) = \{X_3, X_5\}$.

$C_M = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$, where $Y_1 = \{a, f, g, h\}, Y_2 = \{a, c, d\}, Y_3 = \{b, e, i\}, Y_4 = \{f, g\}$. Obviously, $\bigcup_{i=1}^4 Y_i = M$, and: $Dp(a) = Md(a) = \{Y_1, Y_2\}, Dp(b) = Md(b) = Dp(e) = Md(e) = Dp(i) = Md(i) = \{Y_3\}, Dp(c) = Md(c) = Dp(d) = Md(d) = \{Y_2\}, Dp(f) = Dp(g) = \{Y_4, Y_1\}, Md(f) = Md(g) = \{Y_4\}, Dp(h) = Md(h) = \{Y_1\}$.

We all know, the all objects in G share all attributes of M , conversely, all attributes in M common to the all objects of G , so, we have:

Theorem 3.1 The covering C_G is formed by the subset of objects set G , every objects subset which possesses maximal attributes set is corresponding a covering of the attributes set.

A nature question is whether a pair of the objects subset and its corresponding attributes subset is formed a concept? The answer is negative, for example:

$X_1 = \{1, 3\} \subseteq G$, its corresponding the maximal attributes subset is $\{a, b, c, g, h\} \subseteq M$, but $(\{1, 3\}, \{a, b, c, g, h\})$ is not a concept, because $\{a, b, c, g, h\}' = \{3\}$. In here, we consider objects subset corresponding the common attributes, that is, the intent of the objects subset, it also does not form the concept. Such as, $X_1 = \{1, 3\}, \{1, 3\}' = \{a, b, g\}$, but $\{a, b, g\}' = \{1, 2, 3\}$, hence, $(\{1, 3\}, \{a, b, g\})$ is not a concept.

Remark 3.1 If replace "objects subset and its corresponding the maximal attributes subset" for "objects subset and its corresponding the common attributes subset" in theorem 3.1, it can occur the attributes subsets can not formed the covering of the attributes set. For example, consider example 2, given the covering of object set, $X'_1 = \{a, b, g\}, X'_2 = \{a, g\}, X'_3 = \{a\}, X'_4 = \{a, c, d, f\}, X'_5 = \{a, c, d\}, \bigcup_{i=1}^5 X'_i = \{a, b, c, d, f, g\} \neq M$, the intents of object subsets are not formed the covering of the attribute set.

It is similar to the objects set, the covering of the attributes set has similar results.

Theorem 3.2 The covering C_M is formed by the subset of attributes set M , every attributes subset which shares maximal objects set is corresponding a covering of the objects set.

A similar question is whether a pair of the attributes subset and its corresponding objects subset is formed a concept? Furthermore, attributes subset corresponding the common objects, that is, the extent of the attributes subset whether is formed a concept? The answer is also negative, for example: in example 2, let $Y = \{a, c, g\} \subseteq M$, and its corresponding objects subset is $\{1, 3, 4\} \subseteq G$, but $(\{1, 3, 4\}, \{a, c, g\})$ is not a concept, because $\{1, 3, 4\}' = \{a, b, g, h\}$.

Remark 3.2 If replace "attributes subset and its corresponding the maximal objects subset" for "attributes subset and its corresponding the sharing common objects subset" in theorem 3.2, it can occur the objects subsets are not formed the covering of the objects set. For example:

Example 3 The following table 2 is a formal context, where $G = \{1, 2, \dots, 5\}, M = \{a, b, \dots, i\}$, the figure 1 is Hasse figure of its concept lattice.

Given the covering of attributes subset $C_M = \{Y_1, Y_2, Y_3, Y_4\}$, where $Y_1 = \{a, c, f, h\}, Y_2 = \{d, g, i\}, Y_3 = \{b, d, e, f, g\}, Y_4 = \{b, f, h\}, Y'_1 = \{1\}, Y'_2 = \{3\}, Y'_3 = \emptyset, Y'_4 = \{4\}, \bigcup_{i=1}^4 Y'_i = \{1, 3, 4\} \neq G$, that is the extent of attributes subset is not formed the covering of the objects set.

A natural problem, what conditions can make the binary ordered pair of the objects subset and its corresponding attributes

subset formed a concept?

In fact, this is not difficult, we only need choose the covering subset from the point of Hasse figure of formal context, which can assure objects subset and its corresponding common attributes subset or attributes subset and its corresponding sharing objects subset formed a concept, that is, the extent of concept is the objects subset, and the intent is the attributes subset. For example, consider the Hasse figure (figure 1), let $C_M = \{\{a, c\}, \{g, i\}, \{c, f, h\}, \{b, e, g\}\}$, and $\{a, c\}' = \{1, 2\}$, $\{g, i\}' = \{2, 3\}$, $\{c, f, h\}' = \{1, 4\}$, $\{b, e, g\}' = \{5\}$, from Hasse figure, we can see, $(\{1, 2\}, \{a, c\})$, $(\{2, 3\}, \{g, i\})$, $(\{1, 4\}, \{c, f, h\})$, $(\{5\}, \{b, e, g\})$ are all concepts. It is similar to the objects set. However, there is a flaw, like as the remark 3.1, 3.2, it will cause that the objects subset or the attributes subset can not formed the covering of objects set and attributes set. Such as, in example 3, $C_M = \{\{1, 2, 3\}, \{1, 2, 4\}, \{2, 3, 5\}, \{4, 5\}\}$, and $\{1, 2, 3\}' = \{a\}$, $\{1, 2, 4\}' = \{c\}$, $\{2, 3, 5\}' = \{g\}$, $\{4, 5\}' = \{b\}$, from Hasse figure, we can see, $(\{1, 2, 3\}, \{a\})$, $(\{1, 2, 4\}, \{c\})$, $(\{2, 3, 5\}, \{g\})$, $(\{4, 5\}, \{b\})$ are all concepts, but $\{a\} \cup \{c\} \cup \{g\} \cup \{b\} = \{a, b, c, g\} \neq M$.

In order to overcome this drawback, we can take covering subsets from the bottom upward of Hasse figure .

Remark 3.3 In (Qiu Weigen, 2006), the object x of C_x shares the common attribute that can't be formed a covering of attributes set, and if consider the maximum attributes set, then it must be a covering of attributes set, for the covering of attributes set has similar results.

4. Conclusion

In this paper, we further discuss the minimal description of rough sets, in order to understand some the former results, we give some remarks, at the same time, we use the minimal description under covering to the formal concepts of formal context, we study the covering that formed by the concept pair of objects subset and attributes subset, discuss their corresponding relation, and illustrate them by examples.

References

- A.Burusco Juandeburre, R.Fuentes-Gonzalez. (1994). The study of L-Fuzzy concept lattice. *Math and Soft Computing*, 3, 208-218.
- A.Burusco, R.Fuentes-Gonzalez. (2000). Concept lattices defined from implication operators. *Fuzzy Sets and Systems*, 114(3), 431-436.
- A.Skowron, C.Rauszer. (1992). The discernibility matrices and functions in information systems, Intelligent Decision Support, in: R. Slowinski(Ed.), *Handbook of applications and advances of rough sets theory*, Kluwer Academic Publishers.
- B.Ganter, R.Wille. (1999). Formal concept Analysis. *Mathematic Foundations*, Berlin:Springer-Verlag.
- Eric C.C.Tsang, Chen Degang, Daniel S Yeung. (2008). Approximations and reducts with covering generalized rough Sets. *Computers and Mathematics with Applications*, 56(1),279-289.
- F.M.Baltasar, F.V. Alfredo. (1998). Building educational tools based on formal concept analysis. *Education and Information Technologies*, 3-4: 187-201.
- J.A.Pomykala. (1987). Approximation operations in approximation space. *Bulletin of the Polish Academy of Science*, 9-10, 653-662.
- J.S. Deogun, J. Saqer. (2004). Monotone concepts for formal concept analysis. *Discrete Appl. Math.* 144, 70C78.
- Qiu, Weigen. (2006). Luo Zhongliang Some Results on Fined Covering Fuzzy Generalized rough Sets. *Computer Engineering and Applications*, 42(14), 86-88.
- R.Belohavek. (1999). Fuzzy Galois connections. *Math. Logic Quarterly*, 45(4), 497-504.
- R.Godin, R.Missaou, H.Alaoui. (1995). Incremental concept formation algorithms based on Galois (concept) lattice. *Computational Intelligence*, 11(2): 246-267.
- So Kuznetsov. (2002). Machine learning on the basis of formal concept analysis. *Automation and Remote Control*, 62:1543-1564.
- William Zhu, Fei Yuewang. (2003). Reduct and axiomization of covering generalized rough sets. *Information Science*, 152, 217-230.
- Xu, WeiHua, Zhang, WenXiu. (2007). Measuring roughness of generalized rough sets induced by a covering. *Fuzzy Sets and Systems*, 158, 2443-2455.
- Z.Bonikowski, E.Bryniarski, U.Wybraniec. (1998). Extensions and intentions in the rough set theory. *Information Science*, 107, 146-167.
- Z.Bonikowski, E. Bryniarski, U.Wybraniec. (1998). Extensions and intensions in the rough set theory. *Information Science*, 107, 149-167.

Z.Bonikowski. (2003). *Algebraic structures of rough set in representative approximation space*. J Publishers by Elsevier Science B.V, 1-12.

Z.Pawlak. (1991). *Rough sets, Theoretical aspects of reasoning about data Dordrecht*. Kluwer Academic Published.

Zhang, Wenxiu, Yao, Yiyu, Liang, Yi. (2006). *Rough set and concept lattice*. XI'an JiaoTong University Press.

Table 1.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
1	1	1	0	0	0	0	1	0	0
2	1	1	0	0	0	0	1	1	0
3	1	1	1	0	0	0	1	1	0
4	1	0	1	0	0	0	1	1	1
5	1	1	0	1	0	1	0	0	0
6	1	1	1	1	0	1	0	0	0
7	1	0	1	1	1	0	0	0	0
8	1	0	1	1	0	1	0	0	0

Table 2.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
1	1	0	1	0	0	1	0	1	0
2	1	0	1	0	0	0	1	0	1
3	1	0	0	1	0	0	1	0	1
4	0	1	1	0	0	1	0	1	0
5	0	1	0	0	1	0	1	0	0

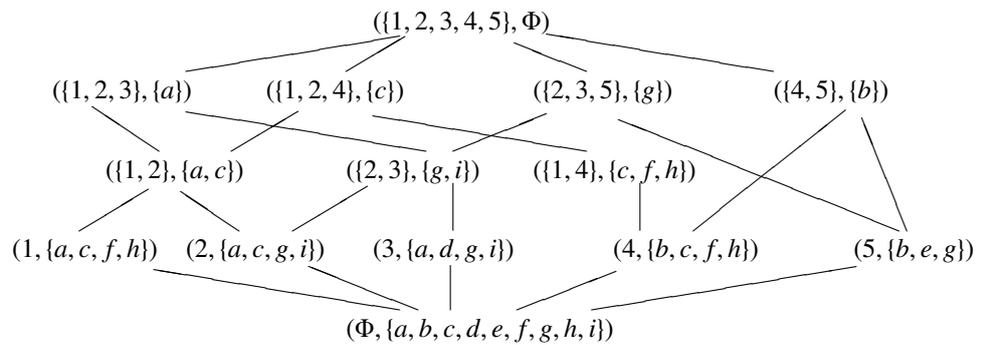


Figure 1