INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Mining Association Rules Using Formal Concept Analysis

Bhavana Jamalpur
Department of CSE, S.R. Engineering College, Warangal, A.P., India
R. Vijaya Prakash
Department of CSE, S.R. Engineering College, Warangal, A.P., India
S. S. V. N. Sarma
Dean, Department of Computer Science & Engineering
Vaagdevi College of Engineering, Warangal, A.P., India

Abstract:
In this paper, we present a methodology based on formal concept analysis (FCA) for the Knowledge Discovery process. We show that FCA can be useful for understanding conceptual model. We focus on Association Rule Mining and formal concept analysis for Booksales transaction database. Formal concept analysis deals with formal mathematical tools and techniques to develop and analyze the relationship between concepts and to develop concept structures.
Given a set of transactions, the problem of mining association rules is to discover all the rules that have user specified minimum support Minimum Confidence by implementing lattice concepts and implications rules.

Keywords*: Association Rulemining, minimum support, minimum confidence, formal concept analysis, lattice.*

1. Introduction

An association rule discovers dependencies among values of an attribute grouped by some other attributes in a given relation. A specific case of discovering association's concerns with a concrete problem that focuses on the analysis of the market-basket-data (or, simply, basket relation) and in the end the solution of the market-basket problem helps a retail store to learn about its customers' purchasing trends.

2. Basic Concepts of Association Rules

Data considered is transactional or relational. Each transaction or row consists of an identifier and a set of items.
In Boolean association rules:
Each item can be seen as a Boolean variable presenting the presence or absence of that item in the transaction/row..
Typical representation formats for association rules:
diapers ⇒ *beer* [0.5%, 60%]
buys: diapers ⇒ *buys :beer* [0.5%, 60%]

"IF buys diapers, THEN buys beer in 60% of the cases. Diapers and beer are bought together in 0.5% of the rows in the database."
A support of an item set *I* is the number of transactions/rows containing *I*.
A minimum support s is a threshold for support.
A general form of an association rule is:
Body⇒⇒⇒Head [Support, Confidence]

Parts of association rules:
diapers⇒ *beer* [0.5%, 60%]

- Antecedent, left-hand side (LHS), body
- Consequent, right-hand side (RHS), head
- Support, frequency
- Confidence, strength

Support of the rule $A \Rightarrow B$: denotes the frequency of the rule within all transactions in the database,
i.e., the probability that a transaction contains both *A* and *B*.
support $(A \Rightarrow B\ [\ s, c\ ]) = p(A \cup B) =$ *support* $(\{A,B\})$

Confidence of the rule $A \Rightarrow B$:
denotes the percentage of transactions containing *A* which also contain *B,*
i.e., the probability that a transaction containing *A* also contains *B.*
Confidence $(A \Rightarrow\Rightarrow B\ [\ s, c\ ]) = p(B|A) = p(A\Rightarrow B) / p(A)$
$=$ *support* $(\{A,B\}) /$ *support*$(\{A\})$.

Apriori algorithm is a basic algorithm for finding frequent item sets of boolean association rules based on level wise search iteratively find frequent itemsets with size from 1 to *k* (*k*-item set)
Basic idea is to reduce the search space by using the Apriori principle: any subset of a frequent itemset must be frequent that is, if {*AB*} is a frequent itemset, both {*A*} and {*B*} should be frequent itemsets

2.1. Association rule generation
Association rule mining is a two-step process:

- Find the frequent itemsets, i.e., the sets of items that have at least the minimum supports.
- Use the frequent itemsets to generate (strong) association rules that satisfy the minimum support **s** and minimum confidence $\Rightarrow$**.**

2.2. Pseudo-code
For every frequent itemset *l* generate all nonempty subsets *s* of *l;* **for** every nonempty subset *s* of *l* output the rule "$s \Rightarrow \Rightarrow \Rightarrow (l\text{-}s)$",
if *support(l)/support(s)* $\Rightarrow\Rightarrow\Rightarrow$
For example, a frequent set $l = \{abc\}$ and its subsets $s = \{a, b, c, ab, ac, bc\}$ can give rules
$a \Rightarrow\Rightarrow b$, $a \Rightarrow\Rightarrow c$, $b \Rightarrow\Rightarrow c$
$a \Rightarrow\Rightarrow bc$, $b \Rightarrow\Rightarrow ac$, $c \Rightarrow\Rightarrow ab$
$ab \Rightarrow\Rightarrow c$, $ac \Rightarrow\Rightarrow b$, $bc \Rightarrow\Rightarrow a$
if their confidences $\geq \gamma$

3. Formal Concept Analysis
Formal Concept Analysis (FCA) is a theoretical method for the mathematical analysis of scientific data and was found by Wille in the middle of 80s during the development of a framework to carry out the lattice theory applications. FCA models the real world as objects and attributes. FCA will define concepts in their given content and study the inter-concept relationship regarding the structure of the lattice that corresponds to the content. The mathematical notion of concept has its origin in formal logic.
This common definition can be made by two routes, extent and intent. The intent provides the attributes of context while extent covers the objects that are included in the concept. Many applications of FCA to real-life problems in intelligent data analysis, data mining, knowledge representation and acquisition, software engineering, database systems and information retrieval and may other disciplines.

3.1. Order-Theoretic Preliminaries
In this subsection, the main concepts that constitute FCA will be defined briefly.
Definition 1 (Partial Order): A binary relation *R* (often use the symbol $\leq$) on a set *M* is called an partial order relation, if it satisfies the following conditions for all
elements $x, y, z \in M$;
1. $x \leq x$
2. $x \leq y$ and $y \leq x$® $x = y$
3. $x \leq y$ and $y \leq z$ ® $x \leq z$
These conditions are referred to, respectively as reflexivity, anti-symmetry and transitivity. A partially ordered set (poset) is a pair $(M,\leq)$, with *M* being a set and $\leq$ an order relation on *M*. A relation $\leq$ on a set *M* which is reflexive and transitive but not necessarily anti-symmetric is called quasi-order.

Definition 2 (Infimum, Supremum): Let $(M,\leq)$ be a partially ordered set (poset) and *A* a subset of *M* . A lower bound of *A* is an element *s* of *M* with $s \leq a$ for all $a \in A$. An upper bound of *A* is defined dually. If there

is a largest element in the set of all lower bounds of A , it is called the infimum of A and is denoted by inf A or $\wedge A$;dually, a least upper bound is called supremum and denoted by sup A or $\vee A$. If $A = \{x, y\}$, we also write $x \wedge y$ for inf A and $x \vee y$ for sup A. Infimum and supremum are frequently also called meet and join.

Definition 3 (Lattice, Complete Lattice): A partially ordered set (poset) $\nu := (V,\leq)$ is called a lattice, if for any two elements x and y in V the supremum $x \vee y$ and the
infimum $x \wedge y$ always exist. ν is called a complete lattice, if the supremum $\vee X$ and the infimum $\wedge X$ exist for any subset X of V. Every complete lattice ν has a largest
element, $\vee V$, called the unit element of the lattice, denoted by $V1$. Dually, the smallest element $V0$ is called the zero
element. The definition of a complete lattice assumes that supremum and infimum exist for every subset X , in
particular for $X = \varnothing$. We have $V \wedge \varnothing = 1$
and $V \vee \varnothing = 0$, from which it follows that $V \neq \varnothing$ for every complete lattice. Every non-empty complete lattice is a complete lattice.

Definition 4: Let S be a set and φ a mapping from the power set of S into the power set of S. Then φ is called a closure operator on S if it is,
1. Extensive: $A \subseteq \varphi(A)$ for all $A \subseteq S$;
2. Monotone: $A \subseteq B \rightarrow \varphi(A) \subseteq \varphi(B)$ for all $A, B \subseteq S$; and
3. Idempotent: $\varphi(\varphi(A)) = \varphi(A)$.

3.2. Line Diagram

If V is finite there is a unique smallest relation p known as the cover or neighbor relation, whose transitive, reflexive, closure is $\leq$. A Hasse diagram of ν is a diagram of the acyclic graph $(V, \nu)$where the edges are straight line segments and, if $a < b$ inν , then the vertical coordinate for
a is less than the one for b . Because of this second condition arrows are omitted from the edges in the diagram. A lattice is an partially ordered set (poset) in which every
pair of elements a and b has a least upper bound, $a \vee b$ and a greatest lower bound, $a \wedge b$ and so also has a Hasse diagram. These Hasse diagrams are an important tool for researchers in lattice theory and ordered set theory and are now used to visualize data. The concept lattices can be
graphically represented by line diagrams which have been proven to be useful representations for the understanding of
conceptual relationship in data. A line diagram is a specialized Hasse diagram with several notational extensions. Line diagram contains vertices and edges.
In the line diagram, the name of an object g is always attached to the circle representing the smallest concept with g in its extent; dually, the name of an attribute m is always attached to the circle representing the largest concept with m in its intent. This allows us to read the context relation from the diagram because an object g has an attribute m if and only if there is an ascending path from the circle labeled by g to the circle labeled by m . The extent of a concept consists of all objects whose labels are

3.3. Implications between Attributes

Given a formal context K, one other common method to analyze it is to find (a canonical base of) the implications between the attributes of this context. They are statements of the form
"Every object that satisfies the attributes M_i *also satisfies* M_j "
Formally, an implication between attributes is defined as follows:
An association rule having confidence is equal to 1 is called implication (exact association rule), otherwise, this rule is called approximate association rule.
The support and confidence values of
$X \Rightarrow Y$ rule are defined as below

$$Supp = \frac{|(X \cup Y)`|}{|G|}$$

$$Conf = \frac{|(X \cup Y)`|}{|X|}$$

Consider the following set of transactions in a bookshop. A set of 6 transactions of purchases of books. Purchases were made of books on Compiler Design CD, Distributed Databases DD, Theory of Computation TOC, Computer Graphics CG, and Neutral Networks NN

Tid	Books Purchased				
	CD	DD	TOC	CG	NN
100	X		X	X	X
200	X	X		X	
300	X		X	X	
400	X	X		X	X
500	X	X	X	X	X
600	*X*	*X*	*X*		

Table 1: Formal Context $T_{Candidates}$

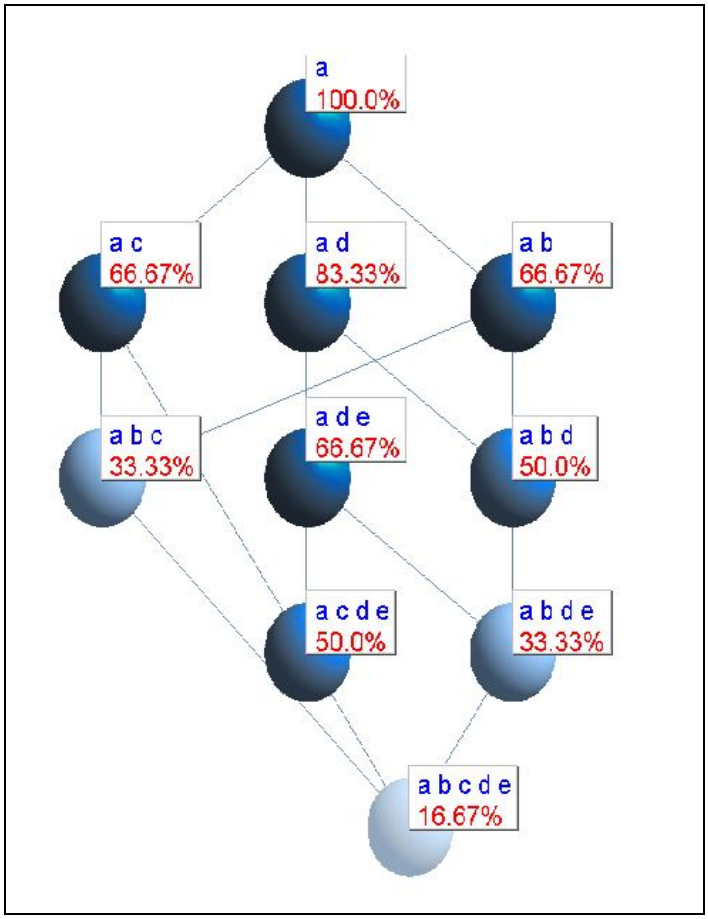

Figure 1: The formal concept lattice corresponding to the candidate formal context in Table 1

Each rule has a left-hand side i.e antecedent and right-hand side as consequent. Finding all the items whose support is greater than the user-specified minimum support such items are frequent itemsets.

Rules generated with user-specified Minimum support and Minimum confidence

Minsupp 20%, Minconf 100%
{d}->{a}
{b}->{a}
{c}->{a}
{b,d}->{a}
{e}->{a,d}
{c,d}->{a,e}
{c,e}->{a,d}

Table 2: Association Rules Generated

4. Conclusion

In this study, the mathematical background and definitions of FCA which is one of the symbolic data mining methods are explained. By that way, the implications and association rules are obtained. These rules can help the decision makers to make the best decision as which books are frequent so as to analyses the customer transactions.

5. References

1. R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts", in graph and order.
2. J. Lieber, A. Napoli, L. Szathamary and Y. Toussaint, "First elements on knowledge discovery guided by domain knowledge (KDDK)".
3. D. Li, J. Han, X. Shi and M. C. Chan, "Knowledge Representation and Discovery based on Linguistic Atoms", Knowledge-Based Systems
4. B. Sertkaya, "Formal Concept Analysis Methods for Description Logics"
5. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules",
6. W. J. Frawley, G. P.-Shapiro and C. J. Matheus, "Knowledge Discovery in Databases: An Overview", AI Magazine
7. K. H. Yang, D. Olson, and J. Kim, "Comparison of First Order Predicate Logic, Fuzzy Logic and Non-Monotonic Logic as Knowledge Representation Methodology"
8. G. Stumme, "Conceptual Knowledge Discovery with Frequent Concept Lattices"
9. J. S. Deogun and J. Saquer, "Monotone Concepts for Formal concept Analysis", Discrete Applied Mathematics
10. B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundations"
11. G. Stumme, "Acquiring expert knowledge for the design of conceptual information systems"
12. G. Stumme, R. Taouil, Y. Bastide, N. Pasquier and L. Lakhal, "Computing Iceberg Concept Lattices wit TITANIC"
13. B. Ganter and R. Wille. (1997). Applied Lattice Theory: Formal Concept Analysis
14. J. H. Correia, G. Stumme, R. Wille and U. Wille, "Conceptual knowledge discovery and data analysis"