A Taxonomy of Concept Lattice Construction Algorithms

Chris Schulz

University of Pretoria

Espresso Workshop 2007

Outline of Topics

TABASCO

- TAxonomy BAsed Software COnstruction
- Classification of algorithms or data structures according to their similarities and differences
- Domain Engineering Method
  - Focus on a family of closely related algorithms or data structures within a domain.
  - Creation of Reusable Toolkits - Domain Specific Toolkits (DSTs).
  - Toolkits are then used within application engineering.

Advantages

- Brings order to the domain
  - Taxonomy shows relationships between different algorithms.
  - Algorithms are specified using a single notation.
  - Correctness arguments for each algorithm.
- The discovery of new algorithms
- Improvements on existing algorithms
- Reusable DSTs

General Process

1. Selection of domain
2. Literature Study
3. Taxonomy Construction
4. Toolkit Design
5. Domain Specific Language (DSL) design
6. Toolkit Implementation
7. Benchmarking
8. DSL Implementation

Definitions

Lattice

A lattice is a partially ordered set (poset) denoted by $< L, \leq >$ in which each pair of elements has a unique least upper bound (supremum) and a unique greatest lower bound (infimum).

Complete Lattice

A lattice is complete if there exists a supremum and infimum for every one of its subsets.

Note: All non-empty finite lattices are complete

Definitions

Context

A context is a triple (G, M, I) consisting of two sets G and M and a relation I between G and M ($I \subseteq GxM$). The elements of G are called objects and the elements of M are called attributes. I is called the incidence relation and describes whether an object in G has a specific attribute in M.

Cross-table

$$G = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$
$$M = \{\text{Even}, \text{Odd}, \text{Prime}, \text{Square}\}$$

Number	Even	Odd	Prime	Square
1		X		X
2	X		X	
3		X	X	
4	X			X
5		X	X	
6	X			
7		X	X	
8	X			
9		X		X

Table: Cross-table representing an example Context

Definitions

Concept

For $A \subseteq G$ and $B \subseteq M$, define

$$\begin{aligned} A' &= \{m \in M | (\forall g \in A)(g, m) \in I\} \\ B' &= \{g \in G | (\forall m \in B)(g, m) \in I\} \end{aligned}$$

A concept of a context (G, M, I) is a pair (A, B) where $A' = B$ and $B' = A$.

A is called the **extent** of the concept and B is called the **intent**.

Definitions

Concept Lattice

For concepts (A_1, B_1) and (A_2, B_2), we write $(A_1, B_1) \leqslant (A_2, B_2)$ iff $A_1 \subseteq A_2$ (and dually $B_1 \supseteq B_2$). The set of concepts ordered by the relation $\leqslant$ form a complete lattice called a **concept lattice**.

Context

$$\{1\}' = \{\text{Odd}, \text{Square}\}$$
$$\{1\} = \{\text{Odd}, \text{Square}\}'$$

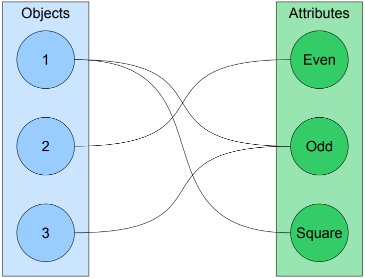

Figure: Example Context

Concepts

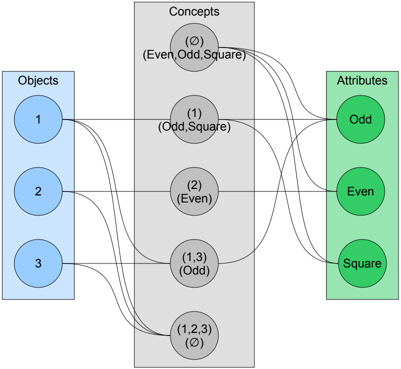

Figure: Example Concepts

Concept Lattice

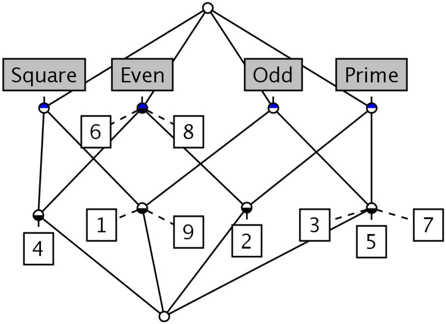

Figure: Example Concept Lattice (Line Diagram)

Concept Lattice Construction

- Incremental Algorithms
- Batch Algorithms
- Algorithms to determine:
  - Concepts
  - Lattice Structure
- Parallel Algorithms

Definitions

Itemset

A subset $S \subseteq M$ is called an **itemset**.

Support

$$sup(S) = |\{g \in G | (\forall s \in S)(g, s) \in I\}|$$

Definitions

Frequent Itemsets

$$F = \{S \subseteq M | sup(S) \geqslant minsup\}$$

minsup is the minimum support required for an itemset to be considered frequent.

Closed Itemset

An itemset S is closed if these exists no proper superset of S that has the same support as S.

Frequent Closed Itemsets (FCIs)

- Condensed Representation of Frequent Itemsets
- Each subset of a FCI is again a frequent itemset
- Support of a frequent itemset is equal to the support of the smallest FCI containing the itemset
- Closed Itemsets map to the concepts of the context (G,M,I)
- Algorithms to determine:
    - Closed itemsets
    - Support for each frequent itemset

Research Contribution

- A Taxonomy of concept lattice construction algorithms
- Comparison of Algorithms in FCA with those in FCI
- Possible improvement on current algorithms as well as the discovery of new ones

Questions?