Constructing concept lattices and compressed pseudo-lattices

by

F.J. (Dean) van der Merwe

Submitted in fulfilment of the requirements for the degree
M.Sc. Computer Science
in the Faculty Engineering, Built Environment
and Information Technology

University of Pretoria

Pretoria, South Africa

May 2003

Constructing concept lattices and compressed pseudo-lattices

By F.J. (Dean) van der Merwe
Supervisor: Prof. D.G. Kourie
Department of Computer Science
University of Pretoria
M.Sc. Computer Science

Summary

Concept lattices are well-defined mathematically structures used in a variety of areas of application. Concept lattices are exponential in size in the worst case and therefore constructing concept lattices may be time consuming. This dissertation approaches the efficiency problems associated with constructing concept lattices from two points of view and proposes complementary solutions for each. Firstly, a new lattice construction algorithm, called AddAtom, unrelated to well known and published algorithms is proposed. Secondly, a data structure called a compressed pseudo-lattice is proposed. Compressed pseudo-lattices involve the construction of sublattices instead of the complete lattice of all concepts.

The AddAtom algorithm efficiently constructs lattices and is proposed as a general purpose lattice construction algorithm. The algorithm is an incremental algorithm and has a theoretic complexity bound of $O(\|L\|.\|O\|^2.\max(\|O'\|))$ when constructing the concept lattice, L of a context with a set of objects O each of which can possess a maximum of $\max(\|O'\|)$ attributes from a set of attributes A. In experimental comparisons AddAtom outperformed other published algorithms in a range of contexts. In randomly generated contexts that had very high or very low densities its performance was the second best. It was however the best performing incremental lattice construction algorithm.

The notion of a compressed pseudo-lattice is introduced and suggested as a data structure in areas of application related to Formal Concept Analysis (FCA). It is closely related to the line diagram of a lattice and its use as a computational tool in applications such as machine learning, information retrieval and knowledge discovery in databases is discussed. The data structure, essentially a bipartite graph that incorporates an embedded sublattice, combines desirable features of concept lattices in a structure that allows for a flexible mechanism of scaling the size of the embedded sublattice, using defined operations that compress and expand it by removing or adding atoms and coatoms. A compressed pseudo-lattice essentially represents a complete sublattice from which a number of atoms and/or coatoms have been removed. Additionally the relation of the sublattice to the context from which it was derived is preserved. An application-dependent compression strategy or criterion is required to guide this process. The intent- and extent representative operations of a lattice are defined as substitutes for the infimum and supremum operations in compressed pseudo-lattices. It is argued that the removal of concepts from a concept lattice may hold advantages over traditional approaches.

Acknowledgements

My sincere thanks to:

- My family for their support and encouragement.
- Deon Oosthuizen for introducing me to the subject during many enthusiastic discussions and his support and assistance during his tenure as an associate professor at the University of Pretoria during which he also acted as supervisor.
- Prof. Derrick Kourie for his assistance, patience and invaluable advice as supervisor.
- Sergei Obiedkov for his collaborative assistance and for making available the data of his experimental comparisons of AddAtom and other lattice construction algorithms as well as helpful comments and suggestions.

Table of contents