Browsing Large Concept Lattices through Tree Extraction and Reduction Methods

Cassio Melo, École Centrale Paris, Paris, France
Bénédicte Le-Grand, Centre de Recherche en Informatique, Université Paris 1 Panthéon, Sorbonne, France
Marie-Aude Aufaure, Mas Laboratory, École Centrale Paris, Paris, France

ABSTRACT

Browsing concept lattices from Formal Concept Analysis (FCA) becomes a problem as the number of concepts can grow significantly with the number of objects and attributes. Interpreting the lattice through direct graph-based visualisation of the Hasse diagram rapidly becomes difficult and more synthetic representations are needed. In this work the authors propose an approach to simplify concept lattices by extracting and visualising trees derived from them. The authors further simplify the browse-able trees with two reduction methods: fault-tolerance and concept clustering.

Keywords: Concept Clustering, Fault-Tolerance, Formal Concept Analysis (FCA), Hasse Diagram, Reduction Methods, Tree, Visual Analytics

1. INTRODUCTION

The vast amount of data generated over the last decades has brought new challenges to the analytics science. Visual data analysis and knowledge representation employ methods such as Formal Concept Analysis (FCA) in order to identify groupings of patterns from the analysis process (Ganter & Wille, 1999). FCA provides an intuitive understanding of generalization and specialization relationships among data elements (objects) and their features (attributes or properties) in a structure known as a concept lattice. For this reason, FCA is extensively used in Information Retrieval Systems (Carpineto & Romano, 1996, Carpineto & Romano, 2005; Ducrou & Eklund, 2006) to refine queries, restrict the search space, recommend related documents, and in particular, to browse document collections (Carpineto & Romano, 2005). For example, in a collection of documents containing the word “Paris”, some documents may also refer to “Restaurants”, “Flight tickets” or “Concerts”. In the concept represented by {“Paris”, “Restaurants”}, one may find documents about “Best view”, “Famous”, “Cheap”,

DOI: 10.4018/ijiit.2013100102
and so forth. Each navigation step is a formal concept and the relationships expressed in the concept lattice allow navigating from one concept to the other.

Browsing concept lattices becomes a problem as the number of clusters grows significantly with the number of objects and attributes. Interpreting the lattice through a direct visualisation of the Hasse diagram rapidly becomes difficult and more synthetic representations are needed. A common approach is to show or hide parts of the lattice via interactive exploration of subsets of terms or neighbours of a focus concept (Ducrou & Eklund, 2006). Carpineto et al. (Carpineto & Romano, 2005) defined constraints to be applied to the concept lattice in order to simplify lattice querying and navigation.

Trees are good alternatives to represent concept lattices in this context, because they do not suffer from edges crossings and they are natural metaphors for navigation history since there is only one parent per child. Additionally, users are familiar with that structure as trees are used to navigate through folders and files in most operating systems. In Carpineto and Romano (2005) it was noted that trees are particularly interesting structures to represent concept lattices for browsing, however authors pointed that their main disadvantage is the amount of replicated information when concepts have multiple parents. In the current work the authors avoid duplication by selecting only one parent for each concept in the lattice. One inherent challenge is to formally define the notion of “best” parent among the potentially numerous parents of a concept. Previous work (Le Grand et al., 2009) used an approach to extract trees from lattices based on the assignment of weights to attributes. Concepts with higher average weights were selected as parents while the other concepts were removed from the resulting tree.

In this work the authors propose a tree-based approach for large lattice browsing, firstly by extracting and visualising trees derived from the lattice structure; Secondly, by enhancing the readability of the derived trees through colouring and distortion techniques. Finally, further simplification of the tree is done by using lattice reduction techniques. In order to extract trees from lattices, a single parent node needs to be chosen for each concept of the initial lattice: the authors therefore define a set of (unique) parent concept selection criteria, including the stability and support indexes (Kuznetsoy, 1990, Stumme et al., 2002) provided by FCA literature, as well as confidence and similarity measures. Before proceeding, the authors would like to briefly introduce FCA terminology and basic definitions used throughout the paper.

1.1. A Brief Introduction to FCA

Rudolf Wille and Bernhard Ganter introduced formal Concept Analysis in the, 1990s based on the mathematisation of the philosophical notion of concepts and concepts hierarchy (Ganter & Wille, 1999). A formal concept has an extent, composed of all objects that belong to the concept, and an intent, comprising all attributes (or properties) which apply to all objects of the extent. The set of objects and attributes, together with their relation to each other, form a formal context which is represented by a cross table, like the one illustrated in Table 1. Using the FCA terminology (Ganter & Wille, 1999), given a (formal) context \( K = (G, M, I) \), where \( G \) is called a set of objects or extent, \( M \) is called a set of attributes or intent, and the binary relation \( I \subseteq G \times M \) specifies each object’s attributes. Derivation operators \((\cdot)^\prime\) are defined for \( A \subseteq G \) and \( B \subseteq M \) as:

\[
A^\prime = \{ m \in M \mid \forall g \in A: g I m \}
\]

\[
B^\prime = \{ g \in G \mid \forall m \in B: g I m \}
\]

A pair \( \{A, B\} \) where \( A \subseteq G \) (extent), \( B \subseteq M \) (intent) is a formal concept in a given formal context \( K \) only if \( A^\prime = B \) and \( B^\prime = A \). A formal concept is, therefore, a closed set of object/attribute relations, in that its extent contains all objects that have the attributes in its intent, and its intent contains all attributes shared by the objects of its extent. In the example illustrated in Table 1, it can be seen that \( \{ \text{Flying,} \)
Related Content

Research Commentaries on Cangelosi’s “Solutions and Open Challenges for the Symbol Grounding Problem”
[www.igi-global.com/article/research-commentaries-cangelosi-solutions-open/52604?camid=4v1a](www.igi-global.com/article/research-commentaries-cangelosi-solutions-open/52604?camid=4v1a)

DBN Models for Visual Tracking and Prediction
[www.igi-global.com/chapter/dbn-models-visual-tracking-prediction/5501?camid=4v1a](www.igi-global.com/chapter/dbn-models-visual-tracking-prediction/5501?camid=4v1a)
Optimal Strategy for the Smartphone Industry in Taiwan: HTC Case Study
www.igi-global.com/article/optimal-strategy-smartphone-industry-taiwan/74830?camid=4v1a

EnOntoModel: A Semantically-Enriched Model for Ontologies
www.igi-global.com/article/enontomodel-semantically-enriched-model-ontologies/2428?camid=4v1a