Using Concept Lattices as a Visual Assistance for Attribute Selection

Jean Villerd, Sylvie Ranwez, and Michel Crampes

LGI2P - École des Mines d'Alès,
Parc scientifique Georges Besse, F-30035 Nîmes Cedex 1, France
{jean.villerd,sylvie.ranwez,michel.crampes}@ema.fr
http://www.lgi2p.ema.fr/

Abstract. The increasing size of structured data that are digitally available emphasizes the crucial need for more suitable representation tools than the traditional textual list of results. A suitable visual representation should both reflect the database's structure for navigation purpose and allow performing visual analytical tasks for knowledge extraction purpose. In [13] we presented a visual navigation method that uses a Galois lattice to represent the database's structure. Moreover, beyond the navigation task, we aim to propose a visual assistance for more analytical tasks. We show how this representation, combined with data analysis techniques, can be used both for navigation and attribute selection while keeping users' mental map.

Key words: Formal Concept Analysis, Information Visualization

1 Introduction

The size of digitally available indexed document sets increases every day. However, associated exploring tools are often based on the same traditional model: users send their query and are then answered back with huge lists of results. There is a crucial need for more suitable representation tools where the semantics of the documents are better exploited and may be used as a guideline during navigation through the database. Formal concept analysis (FCA) helps to form conceptual structures from data. Such structures may be used to visualize inherent properties in data sets and to dynamically explore a collection of documents. We introduced a visual navigation method in [13]. Furthermore the associated mathematical formalization is useful not only to organize the database but also to perform analytical tasks during the information retrieval process and in this paper we aim to propose a visual assistance for one of these analytical tasks. The rest of this paper is organized as follows. Related works are presented in section 2. Section 3 deals with our proposal for visual attribute selection method. The last section concludes with some of the limits to and perspectives of our approach.

2 State of the Art

Searching for a solution to assist the navigation through a large database, we focus particularly on two aspects in the following state of the art: applications that use FCA techniques for information retrieval and visualization techniques that may be used to graphically parse large sets of data.

The powerful classification skills of Formal Concept Analysis have found many applications for information retrieval. Some of them have been listed in [9]. Since the early works of [5] on an information retrieval system based on document/term lattices, a lot of research leading to significant results has been done. In [1], Carpineto and Romano argue that, in addition to their classification behaviors for information retrieval tasks, concept lattices can also support an integration of querying and browsing by allowing users to navigate into search results. Nowadays several industrial FCA-based applications like Credo [1] or Mail-Sleuth [4] are available. Mail-Sleuth is an e-mail management system providing classification and query tools based on FCA. This tool allows users to navigate into data and intervene in the term classification by displaying concept lattices. Upstream research has studied the understandability of a lattice representation by novice users [4][12]. Image-Sleuth [3] proposes an interactive FCA-based image retrieval system in which subjacent lattices are hidden. Although users do not interact with an explicit representation of a lattice, they navigate from one concept to another by adding or removing terms suggested by the system. This ensures a progressive navigation into the lattice.

3 Using Concept Lattice for Attribute Selection

This section aims to provide a visual answer to the following question: "*I have identified a set of instances of particular interest in the database. I would like to find its location in the database structure and which attributes have the ability to put theses instances together*". Databases have increased not only in size but also in complexity. [6] reports that while as of 1997 only few papers in the attribute selection community were dealing with domains described by more than 40 attributes, most papers were exploring domains with hundreds to tens of thousands of features five years later. Consequently a preprocess called attribute selection is often needed in order to reduce dimensionality before starting data analysis techniques. It consists in selecting attributes that are relevant according to the future data mining task. Beyond the technical purpose of reducing dimensionality for data mining processes, attribute selection constitutes an interesting process as itself. When used to select attributes that are relevant according to a classification task, its results give information about which attributes can be used to separate or describe classes. Detailed reviews of attribute selection techniques can be found in [6] and [7]. In particular, IGLUE [8] is an instance-based learning system that uses Galois lattices to perform attribute selection. All techniques share the following skeleton which sums up the process in four key steps, namely *subset generation*, *subset evaluation*, *stopping criterion*, and *result validation.*

First an attribute subset is generated according to a certain search strategy, the second step evaluates the subset's relevance and consequently selects or discards the subset, then if the stopping criterion is not satisfied a new subset is generated and the process is repeated. Finally the selected best subset needs to be validated by prior knowledge. Attribute selection is used for many data mining tasks, we will focus on its application for classification.
Our attribute selection process is supervised. It means that the objects' membership to the considered class is known *a priori*. This prior knowledge may come from an additional class attribute or from an additional numerical attribute with a threshold value. These additional attributes do not belong to the formal context, and thus are not involved in the lattice computation, because we assume that attributes used to build the lattice reflect the persistant database structure, while class membership attributes are related to a particular exploitation of the database. Objects are partitioned into two classes with respect to an additional class attribute, positive (objects that belong to the class) and negative objects and we propose to use the database's lattice to perform attribute selection. The search strategy consists in browsing the Galois lattice using breadth-first traversal from top to bottom, the generated subset being the current node's intent. The intent is evaluated considering the value for Shannon's entropy on the current node's extent. The entropy will be minimal if all objects in the extent belong to the same class. If entropy is below a given threshold and most of the objects in extent positive, the intent is selected. The stopping criterion is satisfied when all nodes have been evaluated.
In the following, A (resp. O) denotes the attribute (resp. object) set, $I \subseteq O \times A$ a binary relation and L the associated Galois lattice. (O_1, A_1) denotes a formal concept and $\leq_L$ the partial order between L's concepts such that $(O_2, A_2) \leq_L (O_1, A_1) \Leftrightarrow O_2 \subseteq O_1 \Leftrightarrow A_2 \supseteq A_1$.

3.1 Subset Generation

Considering a context with n attributes, there exist 2^n candidate subsets. An exhaustive search is therefore computationally prohibitive. In order to reduce the search space, two main strategies have been designed: complete and sequential search. Complete search strategies, such as *branch and bound*, ensures that all optimal subsets will be explored. The space search is still in $O(2^n)$ but in practice fewer subsets are explored. Concerning sequential search strategies, mostly based on the greedy hill climbing approach, they explores a search space in $O(n^2)$ or less but completeness is not garanteed. Randomness may be introduced in sequential approaches in order to avoid local optima. Since our main goal is to maintain users' mental map and thus to use the same visual structure, the lattice, for both navigation and display of attribute selection results, we use the lattice as the search space. The explored subsets are the nodes' intents. The number of explored subsets is then the size of the lattice, i.e. $O(2^{\min(|A|,|O|)})$. In practice the size of the lattice is smaller since only attribute subsets that are meaningful according to the two closure operators lead to the creation of a node.

3.2 Subset Evaluation

Shannon entropy [10] is used to evaluate the relevance of a node's intent. It measures the ability of the intent to discriminate the positive with the negative objects that appear in the node's extent. Note that this evaluation does not take into account the number of objects in the extent. Therefore nodes containing very few objects may be selected. Formally, considering a node (O_1, A_1) its associated entropy is computed as follows:

$$H(O_1, A_1) = -\left(\frac{|O_1^+|}{|O_1|} \cdot \log_2\left(\frac{|O_1^+|}{|O_1|}\right) + \frac{|O_1^-|}{|O_1|} \cdot \log_2\left(\frac{|O_1^-|}{|O_1|}\right)\right)$$

where $|O_1^+|$ (resp. $|O_1^-|$) is the number of positive (resp. negative) objects in the extent. A null entropy occurs when objets in the extent are either all positive or all negative. Since the goal is to select attributes according to the class, i.e. according to positive objects, a node is said optimal if its entropy is below a given threshold α and if positive objects represent more than half of the extent, formally: (O_1, A_1) is optimal if $H(O_1, A_1) \leq \alpha$ and $\frac{|O_1^-|}{|O_1|} < \frac{1}{2}$. In the following example, we set $\alpha = 0$. Note that if $H(O_1, A_1) = 0$ then $\frac{|O_1^-|}{|O_1|} < \frac{1}{2} \Leftrightarrow O_1^- = \emptyset$.

3.3 Example

This section illustrates the attribute selection process on a small example taken from UCI Repository [11]. The lenses data set [2] contains 24 instances and four nominal attributes. An instance is a patient profile described by the four attributes *age*, *spectacle prescription*, *astigmatism*, and *tear-drop rate*, i.e. factors that have to be taken into account in the choice of a type of contact lenses for a particular patient.

A fifth attribute, *decision*, gives for each profile the recommended type of lenses: *hard*, *soft*, or *none*, dividing patient profiles into three classes. The scenario applied on this example consists in identifying which of the four medical factors are associated with the decision to contraindicate contact lenses. A nominal scale is applied in order to discretize the four attributes. The resulting formal context has seven binary attributes, namely *age:young*, *age:pre-presbyopic*, *age:presbyopic*, *prescription:myope*, *prescription:hypermetrope*, *astigmatism*, and *tear-drop:reduced*. We assume that these binary attributes reflect the persistant structure of the database and the associated Galois lattice, computed using Galicia, has 50 nodes. Positive objects are those which have the value *none* for the additional attribute *decision*, negative ones are the others. Figure 1 shows the resulting lattice where square nodes denote optimal nodes with null entropy. During the breadth-first traversal, the first optimal node found is the one labelled 3. Its intent is $A_1 = \{tear - drop : reduced\}$ and its extent O_1 contains 12 positive objects and no negative one. Its associated entropy is then:

$$H(O_1, A_1) = -\left(\frac{12}{12} \cdot \log_2\left(\frac{12}{12}\right) + \frac{0}{12} \cdot \log_2\left(\frac{0}{12}\right)\right) = 0$$

assuming that $\log_2 0 = 0$ by applying L'Hôpital's rule. The fact that the node (O_1, A_1) is optimal can be interpreted as: "only positive objects own A_1". In the present case it means that "only positive objects own $\{tear{-}drop : reduced\}$", or in other words $\forall o \in O, \{tear{-}drop : reduced\} \in f(o) \Rightarrow o \in O^+$. If $O^+ - O_1 = \emptyset$, i.e. if all positive objects belong to the optimal node's extent, the converse is also satisfied.

An interesting point is that, thanks to the Galois lattice structure, all the child nodes of an optimal node are also optimal. Hence, considering two concepts $(O_2, A_2) \leq_L (O_1, A_1)$, if (O_1, A_1) is optimal then $O_1 \subseteq O^+$. Since $O_2 \subseteq O_1$ thanks to $\leq_L$ we have $O_2 \subseteq O^+$. When an optimal node is found, this result allows to discard all its child nodes from the search space. Note that this property is only true for an entropy threshold $\alpha = 0$.

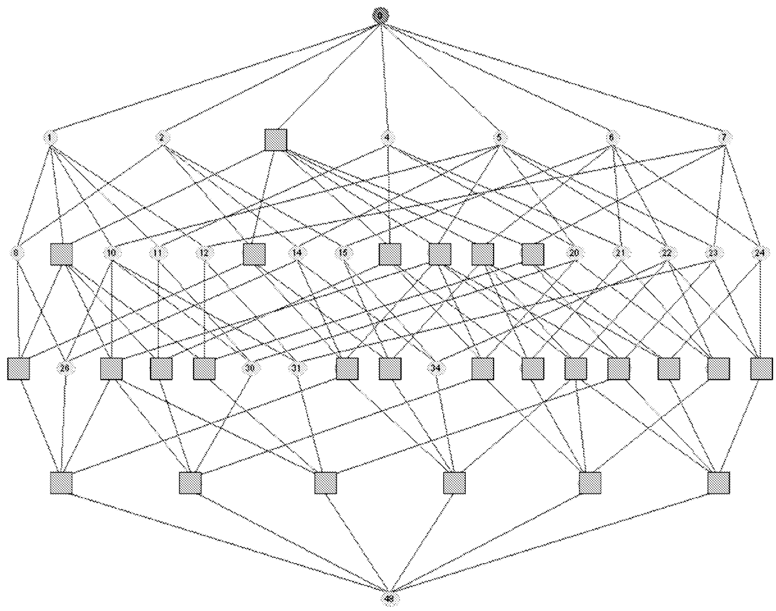

Fig. 1. Galois Lattice computed from the contact-lenses database binary context. Square nodes are optimal and their intents form the resulting selected attribute subsets with respect to the *no lenses* class.

3.4 Results Interpretation and Association Rules

The resulting lattice answers the original question: "*where are my instances of interest in the database structure and what are the related relevant attributes?*". Users can see at first sight how considered instances are dispatched with respect

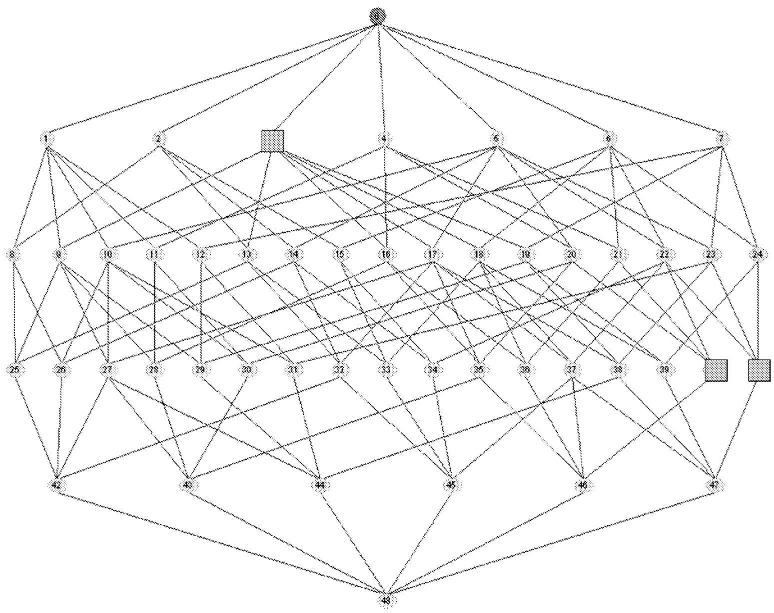

Fig. 2. Redundant squares have been removed from the lattice on Figure 1

to the database structure representation used for navigation. Related attribute subsets are the emphasized nodes' intents. Optimal nodes can also be interpreted as association rules between their intent and the class membership. These rules have a maximal confidence since all objects in optimal nodes are positive. Their support is the number of objects in the extent. Note that thanks to the lattice based representation, users can identify optimal nodes with best supports with respect to their relative position. Hence, considering two optimal nodes $(O_2, A_2) \leq_L (O_1, A_1)$ and their related rules $c_2 : A_2 \rightarrow class$ and $c_1 : A_1 \rightarrow class$, then $support(c2) \leq support(c1)$ since $|O_2| \leq |O1|$. Also note that c_2 is redundant compared to c_1 since $A_2 \subseteq A_1$.

Since child nodes of an optimal node are also optimal (with an entropy threshold $\alpha = 0$), when an optimal node appears among the top node's direct child nodes like in the present example, the resulting lattice may be overcrowded by redundant square nodes. It is not visually easy to separate these redundant nodes from those that are not child nodes of a higher optimal node. For this reason we propose to emphasize optimal nodes that are not child nodes of an optimal node (see Fig. 2). Only three optimal nodes remain: one node labelled 3 which intent is {*tear-drop:reduced*} and two nodes labelled 40 and 41 which respective intents are {*age:pre-presbyopic, astigmatism, prescription:hypermetrope*} and {*age:presbyopic, astigmatism, prescription:hypermetrope*}. These two last nodes were hidden among child nodes of the first one in Fig. 1. Finally, the attribute selection process visually provides to users the following results: the patient profiles for which contact lenses are contraindicated are either those who have a reduced tear-drop rate or those whose have one of the two particular attributes combination listed above.

4 Conclusion

Research presented in this paper only deals with assisting users in interpreting results of an attribute selection process, and it does not actually infer information that present techniques would not be able to infer. This is a proper problem of information visualization. Indeed, noticing Robert Spence's definitions "*to visualize is to form a mental model or mental image of something. Visualization is a human cognitive activity, not something that a computer does*", our goal is not to produce formal results from raw data because data analysis techniques such as FCA succeed without any visualization need. We try to explore new techniques to provide bootstraps for the cognitive activity of users.

References

1. Carpineto, C., Romano, G.: Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. In: Journal of Universal Computing, vol. 10, n8 (2004) 985-1013
2. Cendrowska, J.: PRISM: An Algorithm for Inducing Modular Rules. International Journal of Man-Machine Studies 27, 349–370 (1987)
3. Ducrou, J., Vormbrock, B., Eklund, P.: FCA-based Browsing and Searching of a Collection of Images. In: Proceedings of the 14th International Conference on Conceptual Structures. LNAI 4068, Springer Verlag (2006) 203-214
4. Eklund, P., Ducrou, J., Brawn, P.: Concept Lattices for Information Visualization: Can Novices Read Line Diagrams? In: P. Eklund (Ed.), Concept Lattices: Second International Conference on Formal Concept Analysis, LNCS 2961. Berlin: Springer, (2004) 14-27
5. Godin, R., Gecsei, J., Pichet, C.: Design of Browsing Interface for Information Retrieval. In N. J. Belkin, C. J. van Rijsbergen (Eds.), Proc.SIGIR '89, (1989) 32-39
6. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 1157–1182 (2003)
7. Liu, H., Yu, L.: Towards Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Transactions on Knowledge and Data Engineering 17, 491–502 (2005)
8. Nguifo, E.M., Njiwoua, P., IGLUE: A lattice-based constructive induction system. Intelligent Data Analysis 5, 73–91 (2001)
9. Priss, U.: Formal Concept Analysis in Information Science. In: Annual Review of Information Science and Technology, vol. 40. American Society for Information Science, Washington, DC (2006) 521-543
10. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423 623–656 (1948)
11. UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/
12. Tane, J., Cimiano, P., Hitzler, P.: Query-Based Multicontexts for Knowledge Base Browsing: An Evaluation. In: Proceedings of the 14th International Conference on Conceptual Structures. LNAI 4068, Springer Verlag, 413–426 (2006)

13. Villerd, J., Ranwez S., Crampes, M., Carteret, D.: Using Concept Lattices for Visual Navigation Assistance in Large Databases: Application to a Patent Database. In: Fifth International Conference on Concept Lattices and Their Applications, pp. 88–99. (2007)