Discovering Knowledge in Data Using Formal Concept Analysis

Simon Andrews, Conceptual Structures Research Group, Communication and Computing Research Centre, Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University, Sheffield, UK

Constantinos Orphanides, Conceptual Structures Research Group, Communication and Computing Research Centre, Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University, Sheffield, UK

ABSTRACT

Formal Concept Analysis (FCA) has been successfully applied to data in a number of problem domains. However, its use has tended to be on an ad hoc, bespoke basis, relying on FCA experts working closely with domain experts and requiring the production of specialised FCA software for the data analysis. The availability of generalised tools and techniques, that might allow FCA to be applied to data more widely, is limited. Two important issues provide barriers: raw data is not normally in a form suitable for FCA and requires undergoing a process of transformation to make it suitable, and even when converted into a suitable form for FCA, real data sets tend to produce a large number of results that can be difficult to manage and interpret. This article describes how some open-source tools and techniques have been developed and used to address these issues and make FCA more widely available and applicable. Three examples of real data sets, and real problems related to them, are used to illustrate the application of the tools and techniques and demonstrate how FCA can be used as a semantic technology to discover knowledge. Furthermore, it is shown how these tools and techniques enable FCA to deliver a visual and intuitive means of mining large data sets for association and implication rules that complements the semantic analysis. In fact, it transpires that FCA reveals hidden meaning in data that can then be examined in more detail using an FCA approach to traditional data mining methods.

Keywords: Business Intelligence, Concept Lattice, Formal Concept Analysis, Formal Concept Analysis (FCA), Formal Concepts, Formal Context, Guided Automation, Knowledge Discovery, Visualization

INTRODUCTION

Formal Concept Analysis (FCA) is a term introduced by Rudolf Wille in 1984, building on applied lattice and order theory developed by Birkhoff and others in the 1930’s. It was initially developed as a subsection of Applied Mathematics, based on the mathematisation of concepts and concepts hierarchy (Wille, 2005). These concepts and concepts hierarchy form an intuitive visualisation called a formal concept lattice that provides insight into the relationships between objects and attributes. FCA is carried out on a set of binary relations between objects and attributes, usually in the form of a cross-table called a formal context (Figure 1).
The cross-table above shows a formal context representing destinations for five airlines. The elements on the left side are formal objects; the elements at the top are formal attributes. If an object has a specific property (formal attribute), this is indicated by placing a cross in the corresponding cell of the table. An empty cell indicates that the corresponding object does not have the corresponding attribute. In the Airlines context above, Air New Zealand flies to Europe but does not fly to the Caribbean.

A formal context is defined as a triple $K := (G, M, I)$, with $G$ being a set of objects, $M$ a set of attributes and $I$ a binary (True/False) relation defined between $G$ and $M$. The relation $I$ is understood to be a subset of the cross product between the sets it relates, so $I \subseteq G \times M$. If an object $g$ has an attribute $m$, then $g \in G$ relates to $m$ by $I$, so we write $(g, m) \in I$, or $glm$. For a subset of objects $A \subseteq G$, a derivation operator $'$ is defined to obtain the set of attributes, common to the objects in $A$, as follows:

$$A' = \{ m \in M | \forall g \in A : glm \}$$

In the same way, for a subset of attributes $B \subseteq M$, the derivation operator $'$ is defined to obtain the set of objects, common to the attributes in $B$, as follows:

$$B' = \{ g \in G | \forall m \in B : glm \}$$

In FCA, a formal concept is constituted by its extension, which comprises of all objects belonging to the concept and its intension, comprising of all attributes which are shared by all objects of its extension. A pair $(A, B)$ is a formal concept in a given formal context $(G, M, I)$ only if $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The set $A$ is the extent of the concept and the set $B$ is the intent of the concept.

A formal concept is, therefore, a closed set of object/attribute relations, in that its extension contains all objects that have the attributes in its intension, and the intension contains all attributes shared by the objects in its extension. In the Airlines example, it can be seen from the cross-table that Air Canada, Nippon Airways and Austrian Airlines all fly to Europe and Asia Pacific. However, this does not constitute a formal concept because all three airlines also fly to USA. Adding this destination completes (closes) the formal concept:

$$(\{ \text{Air Canada, Nippon Airways, Austrian Airlines} \}, \{ \text{Europe, Asia Pacific, USA} \}).$$

Another central notion in FCA is a duality called a Galois connection, often observed be-

---

**Figure 1. A formal context of airliners and their destinations**

<table>
<thead>
<tr>
<th>Airlines</th>
<th>Latin America</th>
<th>Europe</th>
<th>Canada</th>
<th>Asia Pacific</th>
<th>Middle East</th>
<th>Africa</th>
<th>Mexico</th>
<th>Caribbean</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Canada</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air New Zealand</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nippon Airways</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ansett Australia</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austrian Airlines</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An Integrated Fuzzy-Based System for Cluster-Head Selection and Sensor Speed Control in Wireless Sensor Networks