Title: Formal concept analysis

URL: https://en.wikipedia.org/wiki/Formal_concept_analysis

PageID: 313845

Categories: Category:Data mining, Category:Formal semantics (natural language), Category:Lattice theory, Category:Machine learning, Category:Ontology (information science), Category:Semantic relations

---

In information science , formal concept analysis (FCA) is a principled way of deriving a concept hierarchy or formal ontology from a collection of objects and their properties . Each concept in the hierarchy represents the objects sharing some set of properties; and each sub-concept in the hierarchy represents a subset of the objects (as well as a superset of the properties) in the concepts above it. The term was introduced by Rudolf Wille in 1981, and builds on the mathematical theory of lattices and ordered sets that was developed by Garrett Birkhoff and others in the 1930s.

Formal concept analysis finds practical application in fields including data mining , text mining , machine learning , knowledge management , semantic web , software development , chemistry and biology .

Overview and history

The original motivation of formal concept analysis was the search for real-world meaning of mathematical order theory . One such possibility of very general nature is that data tables can be transformed into algebraic structures called complete lattices , and that these can be utilized for data visualization and interpretation. A data table that represents a heterogeneous relation between objects and attributes, tabulating pairs of the form "object g has attribute m ", is considered as a basic data type. It is referred to as a formal context . In this theory, a formal concept is defined to be a pair (A , B), where A is a set of objects (called the extent) and B is a set of attributes (the intent) such that

the extent A consists of all objects that share the attributes in B , and dually

the intent B consists of all attributes shared by the objects in A .

In this way, formal concept analysis formalizes the semantic notions of extension and intension .

The formal concepts of any formal context can—as explained below—be ordered in a hierarchy called more formally the context's "concept lattice". The concept lattice can be graphically visualized as a "line diagram", which then may be helpful for understanding the data. Often however these lattices get too large for visualization. Then the mathematical theory of formal concept analysis may be helpful, e.g., for decomposing the lattice into smaller pieces without information loss, or for embedding it into another structure that is easier to interpret.

The theory in its present form goes back to the early 1980s and a research group led by Rudolf Wille , Bernhard Ganter and Peter Burmeister at the Technische Universität Darmstadt . Its basic mathematical definitions, however, were already introduced in the 1930s by Garrett Birkhoff as part of general lattice theory. Other previous approaches to the same idea arose from various French research groups, but the Darmstadt group normalised the field and systematically worked out both its mathematical theory and its philosophical foundations. The latter refer in particular to Charles S. Peirce , but also to the Port-Royal Logic .

Motivation and philosophical background

In his article "Restructuring Lattice Theory" (1982), initiating formal concept analysis as a mathematical discipline, Wille starts from a discontent with the current lattice theory and pure mathematics in general: The production of theoretical results—often achieved by "elaborate mental gymnastics"—were impressive, but the connections between neighboring domains, even parts of a

theory were getting weaker.

Restructuring lattice theory is an attempt to reinvigorate connections with our general culture by interpreting the theory as concretely as possible, and in this way to promote better communication between lattice theorists and potential users of lattice theory

— Rudolf Wille,

This aim traces back to the educationalist Hartmut von Hentig, who in 1972 pleaded for restructuring sciences in view of better teaching and in order to make sciences mutually available and more generally (i.e. also without specialized knowledge) critiqueable. Hence, by its origins formal concept analysis aims at interdisciplinarity and democratic control of research.

It corrects the starting point of lattice theory during the development of formal logic in the 19th century. Then—and later in model theory —a concept as unary predicate had been reduced to its extent. Now again, the philosophy of concepts should become less abstract by considering the intent. Hence, formal concept analysis is oriented towards the categories extension and intension of linguistics and classical conceptual logic.

Formal concept analysis aims at the clarity of concepts according to Charles S. Peirce's pragmatic maxim by unfolding observable, elementary properties of the subsumed objects. In his late philosophy, Peirce assumed that logical thinking aims at perceiving reality , by the triade concept, judgement and conclusion . Mathematics is an abstraction of logic, develops patterns of possible realities and therefore may support rational communication . On this background, Wille defines:

The aim and meaning of Formal Concept Analysis as mathematical theory of concepts and concept hierarchies is to support the rational communication of humans by mathematically developing appropriate conceptual structures which can be logically activated.

— Rudolf Wille,

Example

The data in the example is taken from a semantic field study, where different kinds of bodies of water were systematically categorized by their attributes. For the purpose here it has been simplified.

The data table represents a formal context , the line diagram next to it shows its concept lattice . Formal definitions follow below.

The above line diagram consists of circles, connecting line segments, and labels. Circles represent formal concepts . The lines allow to read off the subconcept-superconcept hierarchy. Each object and attribute name is used as a label exactly once in the diagram, with objects below and attributes above concept circles. This is done in a way that an attribute can be reached from an object via an ascending path if and only if the object has the attribute.

In the diagram shown, e.g. the object reservoir has the attributes stagnant and constant , but not the attributes temporary, running, natural, maritime . Accordingly, puddle has exactly the characteristics temporary, stagnant and natural .

The original formal context can be reconstructed from the labelled diagram, as well as the formal concepts. The extent of a concept consists of those objects from which an ascending path leads to the circle representing the concept. The intent consists of those attributes to which there is an ascending path from that concept circle (in the diagram). In this diagram the concept immediately to the left of the label reservoir has the intent stagnant and natural and the extent puddle, maar, lake, pond, tarn, pool, lagoon, and sea .

Formal contexts and concepts

A formal context is a triple $K = (G , M , I)$, where G is a set of objects , M is a set of attributes , and $I \subseteq G \times M$ is a binary relation called incidence that expresses which objects have which attributes. For subsets $A \subseteq G$ of objects and subsets $B \subseteq M$ of attributes, one defines two derivation operators as follows:

Applying either derivation operator and then the other constitutes two closure operators :

The derivation operators define a Galois connection between sets of objects and of attributes. This is why in French a concept lattice is sometimes called a treillis de Galois (Galois lattice).

With these derivation operators, Wille gave an elegant definition of a formal concept:

a pair (A , B) is a formal concept of a context (G , M , I) provided that:

Equivalently and more intuitively, (A , B) is a formal concept precisely when:

every object in A has every attribute in B ,

for every object in G that is not in A , there is some attribute in B that the object does not have,

for every attribute in M that is not in B , there is some object in A that does not have that attribute.

For computing purposes, a formal context may be naturally represented as a (0,1)-matrix K in which the rows correspond to the objects, the columns correspond to the attributes, and each entry k i , j equals to 1 if "object i has attribute j ." In this matrix representation, each formal concept corresponds to a maximal submatrix (not necessarily contiguous) all of whose elements equal 1. It is however misleading to consider a formal context as boolean , because the negated incidence ("object g does not have attribute m ") is not concept forming in the same way as defined above. For this reason, the values 1 and 0 or TRUE and FALSE are usually avoided when representing formal contexts, and a symbol like × is used to express incidence.

Concept lattice of a formal context

The concepts (A i , B i) of a context K can be (partially) ordered by the inclusion of extents, or, equivalently, by the dual inclusion of intents. An order $\leq$ on the concepts is defined as follows: for any two concepts (A 1 , B 1) and (A 2 , B 2) of K , we say that (A 1 , B 1) $\leq$ (A 2 , B 2) precisely when A 1 $\subseteq$ A 2 . Equivalently, (A 1 , B 1) $\leq$ (A 2 , B 2) whenever B 1 $\supseteq$ B 2 .

In this order, every set of formal concepts has a greatest common subconcept , or meet. Its extent consists of those objects that are common to all extents of the set. Dually , every set of formal concepts has a least common superconcept , the intent of which comprises all attributes which all objects of that set of concepts have.

These meet and join operations satisfy the axioms defining a lattice , in fact a complete lattice . Conversely, it can be shown that every complete lattice is the concept lattice of some formal context (up to isomorphism).

Attribute values and negation

Real-world data is often given in the form of an object-attribute table, where the attributes have "values". Formal concept analysis handles such data by transforming them into the basic type of a ("one-valued") formal context. The method is called conceptual scaling .

The negation of an attribute m is an attribute ¬ m , the extent of which is just the complement of the extent of m , i.e., with (¬ m) ′ = G \ m ′ . It is in general not assumed that negated attributes are available for concept formation. But pairs of attributes which are negations of each other often naturally occur, for example in contexts derived from conceptual scaling.

For possible negations of formal concepts see the section concept algebras below.

Implications

An implication A $\rightarrow$ B relates two sets A and B of attributes and expresses that every object possessing each attribute from A also has each attribute from B . When (G , M , I) is a formal context and A , B are subsets of the set M of attributes (i.e., A,B $\subseteq$ M), then the implication A $\rightarrow$ B is valid if A ′ $\subseteq$ B ′ . For each finite formal context, the set of all valid implications has a canonical basis , an irredundant set of implications from which all valid implications can be derived by the natural inference (Armstrong rules). This is used in attribute exploration , a knowledge acquisition method based on implications.

Arrow relations

Formal concept analysis has elaborate mathematical foundations, making the field versatile. As a basic example we mention the arrow relations , which are simple and easy to compute, but very useful. They are defined as follows: For $g \in G$ and $m \in M$ let

and dually

Since only non-incident object-attribute pairs can be related, these relations can conveniently be recorded in the table representing a formal context. Many lattice properties can be read off from the arrow relations, including distributivity and several of its generalizations. They also reveal structural information and can be used for determining, e.g., the congruence relations of the lattice.

Extensions of the theory

Triadic concept analysis replaces the binary incidence relation between objects and attributes by a ternary relation between objects, attributes, and conditions. An incidence ■ (g , m , c) {\displaystyle (g,m,c)} ■ then expresses that the object g has the attribute m under the condition c . Although triadic concepts can be defined in analogy to the formal concepts above, the theory of the trilattices formed by them is much less developed than that of concept lattices, and seems to be difficult. Voutsadakis has studied the n -ary case.

Fuzzy concept analysis : Extensive work has been done on a fuzzy version of formal concept analysis.

Concept algebras : Modelling negation of formal concepts is somewhat problematic because the complement (G \ A , M \ B) of a formal concept (A , B) is in general not a concept. However, since the concept lattice is complete one can consider the join (A , B) Δ of all concepts (C , D) that satisfy $C \subseteq G \setminus A$; or dually the meet (A , B) ■ of all concepts satisfying $D \subseteq M \setminus B$. These two operations are known as weak negation and weak opposition , respectively. This can be expressed in terms of the derivation operators . Weak negation can be written as (A , B) Δ = ((G \ A)″, (G \ A)') , and weak opposition can be written as (A , B) ■ = ((M \ B)', (M \ B)″) . The concept lattice equipped with the two additional operations Δ and ■ is known as the concept algebra of a context. Concept algebras generalize power sets . Weak negation on a concept lattice L is a weak complementation , i.e. an order-reversing map Δ: $L \to L$ which satisfies the axioms $x \Delta\Delta \leq x$ and (x ■ y) ■ (x ■ y Δ) = x . Weak opposition is a dual weak complementation. A (bounded) lattice such as a concept algebra, which is equipped with a weak complementation and a dual weak complementation, is called a weakly dicomplemented lattice . Weakly dicomplemented lattices generalize distributive orthocomplemented lattices , i.e. Boolean algebras .

Temporal concept analysis

Temporal concept analysis (TCA) is an extension of Formal Concept Analysis (FCA) aiming at a conceptual description of temporal phenomena. It provides animations in concept lattices obtained from data about changing objects. It offers a general way of understanding change of concrete or abstract objects in continuous, discrete or hybrid space and time. TCA applies conceptual scaling to temporal data bases.

In the simplest case TCA considers objects that change in time like a particle in physics, which, at each time, is at exactly one place. That happens in those temporal data where the attributes 'temporal object' and 'time' together form a key of the data base. Then the state (of a temporal object at a time in a view) is formalized as a certain object concept of the formal context describing the chosen view. In this simple case, a typical visualization of a temporal system is a line diagram of the concept lattice of the view into which trajectories of temporal objects are embedded.

TCA generalizes the above mentioned case by considering temporal data bases with an arbitrary key. That leads to the notion of distributed objects which are at any given time at possibly many places, as for example, a high pressure zone on a weather map. The notions of 'temporal objects', 'time' and 'place' are represented as formal concepts in scales. A state is formalized as a set of object concepts.

That leads to a conceptual interpretation of the ideas of particles and waves in physics.

Algorithms and tools

There are a number of simple and fast algorithms for generating formal concepts and for constructing and navigating concept lattices. For a survey, see Kuznetsov and Obiedkov or the book by Ganter and Obiedkov, where also some pseudo-code can be found. Since the number of formal concepts may be exponential in the size of the formal context, the complexity of the algorithms usually is given with respect to the output size. Concept lattices with a few million elements can be handled without problems.

Many FCA software applications are available today. The main purpose of these tools varies from formal context creation to formal concept mining and generating the concepts lattice of a given formal context and the corresponding implications and association rules . Most of these tools are academic open-source applications, such as:

ConExp

ToscanaJ

Lattice Miner

Coron

FcaBedrock

GALACTIC

Related analytical techniques

Bicliques

A formal context can naturally be interpreted as a bipartite graph . The formal concepts then correspond to the maximal bicliques in that graph. The mathematical and algorithmic results of formal concept analysis may thus be used for the theory of maximal bicliques. The notion of bipartite dimension (of the complemented bipartite graph) translates to that of Ferrers dimension (of the formal context) and of order dimension (of the concept lattice) and has applications e.g. for Boolean matrix factorization.

Biclustering and multidimensional clustering

Given an object-attribute numerical data-table, the goal of biclustering is to group together some objects having similar values of some attributes. For example, in gene expression data, it is known that genes (objects) may share a common behavior for a subset of biological situations (attributes) only: one should accordingly produce local patterns to characterize biological processes, the latter should possibly overlap, since a gene may be involved in several processes. The same remark applies for recommender systems where one is interested in local patterns characterizing groups of users that strongly share almost the same tastes for a subset of items.

A bicluster in a binary object-attribute data-table is a pair (A,B) consisting of an inclusion-maximal set of objects A and an inclusion-maximal set of attributes B such that almost all objects from A have almost all attributes from B and vice versa.

Of course, formal concepts can be considered as "rigid" biclusters where all objects have all attributes and vice versa. Hence, it is not surprising that some bicluster definitions coming from practice are just definitions of a formal concept. Relaxed FCA-based versions of biclustering and triclustering include OA-biclustering and OAC-triclustering (here O stands for object, A for attribute, C for condition); to generate patterns these methods use prime operators only once being applied to a single entity (e.g. object) or a pair of entities (e.g. attribute-condition), respectively.

A bicluster of similar values in a numerical object-attribute data-table is usually defined as a pair consisting of an inclusion-maximal set of objects and an inclusion-maximal set of attributes having similar values for the objects. Such a pair can be represented as an inclusion-maximal rectangle in the numerical table, modulo rows and columns permutations. In it was shown that biclusters of similar values correspond to triconcepts of a triadic context where the third dimension is given by a scale that represents numerical attribute values by binary attributes.

This fact can be generalized to n -dimensional case, where n -dimensional clusters of similar values in n -dimensional data are represented by n+1 -dimensional concepts. This reduction allows one to

use standard definitions and algorithms from multidimensional concept analysis for computing multidimensional clusters.

Knowledge spaces

In the theory of knowledge spaces it is assumed that in any knowledge space the family of knowledge states is union-closed. The complements of knowledge states therefore form a closure system and may be represented as the extents of some formal context.

Hands-on experience with formal concept analysis

The formal concept analysis can be used as a qualitative method for data analysis. Since the early beginnings of FCA in the early 1980s, the FCA research group at TU Darmstadt has gained experience from more than 200 projects using the FCA (as of 2005). Including the fields of: medicine and cell biology , genetics , ecology , software engineering , ontology , information and library sciences , office administration , law , linguistics , political science .

Many more examples are e.g. described in: Formal Concept Analysis. Foundations and Applications , conference papers at regular conferences such as: International Conference on Formal Concept Analysis (ICFCA), Concept Lattices and their Applications (CLA), or International Conference on Conceptual Structures (ICCS).

See also

Association rule learning

Cluster analysis

Commonsense reasoning

Conceptual analysis

Conceptual clustering

Conceptual space

Concept learning

Correspondence analysis

Description logic

Factor analysis

Formal semantics (natural language)

General Concept Lattice

Graphical model

Grounded theory

Inductive logic programming

Pattern theory

Statistical relational learning

Schema (genetic algorithms)

Notes

References

Ganter, Bernhard; Stumme, Gerd; Wille, Rudolf, eds. (2005), Formal Concept Analysis: Foundations and Applications , Lecture Notes in Artificial Intelligence, vol. 3626, Springer, doi : 10.1007/978-3-540-31881-1 , ISBN 3-540-27891-5

Ganter, Bernhard; Wille, Rudolf (1998), Formal Concept Analysis: Mathematical Foundations , translated by C. Franzke, Springer-Verlag, Berlin, ISBN 3-540-62771-5

Carpineto, Claudio; Romano, Giovanni (2004), Concept Data Analysis: Theory and Applications , Wiley, ISBN 978-0-470-85055-8

Wolff, Karl Erich (1994), "A first course in Formal Concept Analysis" (PDF) , in F. Faulbaum (ed.), SoftStat'93: Advances in Statistical Software 4. , Gustav Fischer Verlag, pp. 429– 438

Davey, B.A.; Priestley, H. A. (2002), "Chapter 3. Formal Concept Analysis", Introduction to Lattices and Order , Cambridge University Press , ISBN 978-0-521-78451-1

External links

A Formal Concept Analysis Homepage

Demo

Formal Concept Analysis. ICFCA International Conference Proceedings doi : 10.1007/978-3-540-70901-5 2007 5th doi : 10.1007/978-3-540-78137-0 2008 6th doi : 10.1007/978-3-642-01815-2 2009 7th doi : 10.1007/978-3-642-11928-6 2010 8th doi : 10.1007/978-3-642-20514-9 2011 9th doi : 10.1007/978-3-642-29892-9 2012 10th doi : 10.1007/978-3-642-38317-5 2013 11th doi : 10.1007/978-3-319-07248-7 2014 12th doi : 10.1007/978-3-319-19545-2 2015 13th doi : 10.1007/978-3-319-59271-8 2017 14th doi : 10.1007/978-3-030-21462-3 2019 15th doi : 10.1007/978-3-030-77867-5 2021 16th

doi : 10.1007/978-3-540-70901-5 2007 5th

doi : 10.1007/978-3-540-78137-0 2008 6th

doi : 10.1007/978-3-642-01815-2 2009 7th

doi : 10.1007/978-3-642-11928-6 2010 8th

doi : 10.1007/978-3-642-20514-9 2011 9th

doi : 10.1007/978-3-642-29892-9 2012 10th

doi : 10.1007/978-3-642-38317-5 2013 11th

doi : 10.1007/978-3-319-07248-7 2014 12th

doi : 10.1007/978-3-319-19545-2 2015 13th

doi : 10.1007/978-3-319-59271-8 2017 14th

doi : 10.1007/978-3-030-21462-3 2019 15th

doi : 10.1007/978-3-030-77867-5 2021 16th