

## Introducción

El análisis formal de conceptos, FCA por sus siglas en inglés, es una teoría matemática basada en conceptos y jerarquías de conceptos que tiene como objetivo el descubrir estructuras conceptuales dentro de un conjunto de datos. Esta teoría queda definida por Rudolf Wille en 1982 y actualmente su uso está bastante extendido especialmente en problemas lingüísticos, como el procesamiento del lenguaje natural y la creación de bases de datos lingüísticas.

## Conceptos y contextos formales

Como se ha dicho anteriormente, el FCA se basa en conceptos y jerarquías, desde un punto de vista lingüístico, un concepto se compone de dos partes, una de ellas es conjunto de objetos que pertenecen al concepto (Extensión) y la otra es una serie de atributos o propiedades que tienen todos los objetos anteriores (Intensión). En FCA se crea un modelo matemático del concepto que permite hablar de objetos, atributos y las relaciones entre ellos que nos permiten expresar que un objeto tiene un atributo. Este modelo es el contexto formal.

El contexto formal es la forma utilizada en FCA para la representación de los datos y se define como un conjunto de la forma  $K := (G, M, I)$  donde  $G$ , del alemán Gegenstände, es el conjunto de objetos.  $M$ , del alemán Merkmale, son los atributos e  $I$  son las relaciones binarias entre  $G$  y  $M$  ( $g$  tiene el atributo  $m$ ). La representación más sencilla para un contexto formal es una matriz donde las filas se corresponden con el conjunto de objetos y las columnas se corresponden con el conjunto de atributos.

Un concepto formal del contexto formal  $K := (G, M, I)$  queda definido como una tupla  $(A, B)$ , donde  $A$  es un conjunto de objetos (Extensión) tal que  $A \subseteq G$  y  $B$  es un conjunto de atributos (Intensión) tal que  $B \subseteq M$ , además se debe cumplir que  $A = B^I$  y  $B = A^I$  donde  $A^I$  y  $B^I$  son operadores de derivación tal que:

$$A^I := \{m \in M \mid gIm \text{ para todo } g \in A\}$$

$$B^I := \{g \in M \mid gIm \text{ para todo } m \in B\}$$

Como estos operadores forman una conexión de Galois, se cumplen las siguientes propiedades tanto para el conjunto  $A$  como para el conjunto  $B$ :

$$Z_1 \subseteq Z_2 \implies Z_1^I \supseteq Z_2^I$$

$$Z \subseteq Z^{II}$$

$$Z^{III} = Z^I$$

(1)

Adems los conceptos se pueden ordenar, se considera que un objeto es mayor que otro cuando:

$$(A_1, B_1) \leq (A_2, B_2) :\Leftrightarrow A_1 \subseteq A_2$$

De esta forma decimos que  $(A_2, B_2)$  es un superconcepto de  $(A_1, B_1)$  y que  $(A_1, B_1)$  es un subconcepto de  $(A_2, B_2)$ .

## Retcula

Al conjunto ordenado de todos los objetos de un contexto  $(B(G, M, I))$  se lo conoce como retcula del contexto. La retcula se define con el siguiente teorema:

Para un conjunto  $\{(A_i, B_i) | i \in I\} \subseteq B(G, M, I)$  de conceptos formales, el supremo ( $\top$ ) viene dado por:

$$\bigvee(A_i, B_i) = ((\bigcup A_i)^n, \bigcap B_i)$$

y el nfimo ( $\perp$ ) por:

$$\bigwedge(A_i, B_i) = (\bigcap A_i, (\bigcup B_i)^n)$$

## Reglas de asociacin

### Concept Explorer

Concept Explorer es la herramienta que hemos utilizado cargar el contexto que vamos a generar y obtener las reglas de asociacin que posteriormente utilizaremos en la fase de clasificacin. Adems podremos visualizar la jerarquía de nuestro contexto.

### Creacin del contexto formal

Para la creacin de un contexto  $K := (G, M, I)$  necesitamos definir un conjunto de atributos (M) y un conjunto de objetos (G). Como atributos hemos tomado un conjunto de palabras extraídas de las noticias formado por las palabras ms comunes de cada categoría de noticias. Como conjunto de objetos tenemos el texto de un número determinado de noticias de cada categoría.

El proceso que hemos seguido para especificar las relaciones entre objetos y atributos (I) es el siguiente, lo primero es generar el conjunto de atributos como est descrito en el capítulo 131541648674864654156486. Para obtener el conjunto de objetos hacemos consultas a la base de datos para obtener, de cada categoría, un subconjunto de noticias seleccionadas de manera aleatoria. Una vez hemos determinado ambos conjuntos, tomamos el texto y la lista de atributos y recorremos la lista comprobando si el elemento est en el texto si est lo sealamos en el fichero con una X y sino con un punto.

El contexto creado se guarda con un fichero de extensión ".cxt" generado de manera automática.

### Creacin del retculo

Otra de las posibilidades que ofrece FCA es la representacin del contexto y de las relaciones entre objetos y atributos, estas relaciones son conocidas como relaciones de Galois. Como durante la realizacin del proyecto hemos generado contexto con un nmero alto de atributos y objetos la representacin del retculo no se puede visualizar de forma clara. An as es posible obtener una representacin simplificada reduciendo el nmero de objetos, el retculo obtenido es el siguiente:

fca/lattice.png

### Resultados obtenidos

Para determinar como afecta el tamao de los conjuntos de objetos y atributos a la precisin del clasificador hemos realizado pruebas con diferentes tamaos de conjuntos con los valores indicados en la siguiente tabla:

Experimento	Nmero de objetos	Nmero de atributos
Experimento 1	133	145
Experimento 2	258	208
Experimento 3	372	257

Como resultado obtenemos un conjunto de reglas de asociacin que transformaremos en un fichero clp de CLIPS y el grafo de la jerarquía.

### Extraccin de reglas

Para generar el fichero clp vamos procesando cada regla filtrando aquellas con un soporte mayor a 0, ya que aquellas reglas que no cumplen esta condición no tienen ningún objeto que las cumpla. También eliminamos aquellas reglas que en su parte izquierda no tengan una etiqueta de alguna categoría ya que no nos aportan información en nuestro estudio.