La gestión de la información es uno de los pilares esenciales de la Ingeniería Informática. Esta tesis doctoral toma como principal base teórica el Análisis Formal de Conceptos (FCA, por sus siglas en inglés: Formal Concept Analysis) [1], y más concretamente, una de sus herramientas fundamentales: los conjuntos de implicaciones. La gestión inteligente de estos elementos mediante técnicas lógicas y computacionales confiere una alternativa para superar obstáculos en campos de la Ingeniería Informática como las bases de datos y los sistemas de recomendación (SRs).

FCA parte de una representación de conjuntos de objetos y atributos por medio de tablas de datos. A partir de ahí, se generan dos herramientas básicas para representar el conocimiento: los retículos de conceptos y los conjuntos de implicaciones.

Trabajar con implicaciones permite utilizar técnicas de razonamiento automático basadas en la lógica por medio de sistemas axiomáticos correctos y completos, como los axiomas de Armstrong [2] y la Lógica de Simplificación [3]. Estos métodos se utilizan en esta tesis doctoral sobre tres áreas de investigación: claves minimales, generadores minimales y sistemas de recomendación conversacionales, donde además, la computación paralela llevada a cabo en entornos de supercomputación ha desempeñado un papel crucial.

Claves Minimales

Una clave de un esquema relacional está compuesta por un subconjunto de atributos que identifican a cada uno de los elementos de una relación [4]. La identificación de las claves es una tarea crucial para muchas áreas de tratamiento de la información.

El problema de la búsqueda de claves consiste en encontrar todos los subconjuntos de atributos que componen una clave a partir de un conjunto de dependencias funcionales. El estudio se ha centrado en el diseño e implementación de los algoritmos de búsqueda de claves basados en la lógica, y más específicamente, en aquellos que utilizan el paradigma de tableaux [5] como sistema de inferencia.

En concreto, se ha diseñado un nuevo método, denominado Closure Keys, que incorpora un mecanismo eficiente de poda de atributos e implicaciones mediante el método del Cierre de la Lógica de Simplificación [6]. Se han desarrollado las implementaciones de los algoritmos en sus versiones secuenciales y también paralelas, mediante computación de alto rendimiento en entornos de supercomputación. Utilizar este tipo de recursos computacionales ha permitido alcanzar resultados que no eran alcanzables mediante computación secuencial.

Generadores Minimales

Los conjuntos cerrados son la base para la generación del retículo de conceptos. El concepto de generador minimal nace como la representación canónica de cada conjunto cerrado [1]. El trabajo de esta parte de la tesis ha consistido en el estudio y diseño de métodos para la enumeración de todos los conjuntos cerrados y sus generadores minimales a partir del conjunto de implicaciones.

Se han estudiado, diseñado e implementado los métodos de generadores minimales referentes en la literatura y se ha hecho una clasificación de las ventajas e inconvenientes obtenidos por cada uno de ellos.

Asimismo, la capacidad computacional de una máquina convencional actual no es suficiente para solucionar el problema de la enumeración de los generadores minimales. Por tanto, se vuelve a utilizar el paralelismo como estrategia para abordar el problema.

Para verificar el rendimiento y la idoneidad de los métodos, se ha realizado una amplia batería de pruebas tanto sobre información sintética como información real.

Sistemas de Recomendación Conversacionales

De forma muy simplificada, se podría considerar que un SR es un sistema inteligente que proporciona a los usuarios una serie de sugerencias personalizadas a partir de un conjunto de elementos [7]. Abordar la generación de recomendaciones haciendo uso de FCA es una aproximación existente en la literatura desde hace años [8].

En esta tesis se ha orientado el trabajo a abordar el problema de la dimensionalidad [9] en los SRs, (el cual aparece cuando es necesario trabajar sobre datasets con un alto número de características) mediante un SR híbrido (basado en contenido, conocimiento y conversacional) haciendo uso de los conjuntos de implicaciones. Gracias a éstas y a la aplicación del Cierre de la Lógica de Simplificación, el sistema reduce la sobrecarga de información consiguiendo una reducción del número de pasos necesarios para obtener una recomendación. Se han realizado numerosas pruebas de aplicación, tanto con información real como sintética, que contrastan su validez.

Bibliografía

[1] Ganter, B., and Wille, R. Formal Concept Analysis: Mathematical Foundations, 1st ed. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1997.

[2] Armstrong, W. W. Dependency structures of data base relationships. In Proceedings of the International Federation for Information Processing Congress (1974), pp. 580-583.

[3] Cordero, P., Enciso, M., Mora, A., and de Guzmán, I. P. SLFD logic: Elimination of data redundancy in knowledge representation. In IBERAMIA 2002: Proceedings of the 8th Ibero-American Conference on AI (London, UK, 2002), Springer-Verlag, pp. 141-150.

[4] Codd, E. F. The Relational Model for Database Management: Version 2. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.

[5] Risch, V., and Schwind, C. Tableaux-based theorem proving and non-standard reasoning. In Workshop Theorem Proving with Analytic Tableaux and Related Methods, Lautenbach. Universitat Karlsruhe, Fakultat fur Informatik, Institut fur Logik, Komplexitat und Deduktionssysteme, Interner Bericht 8/92, March 18-20, 1992 (1992), pp. 76-78.

[6] Mora, A., Cordero, P., Enciso, M., Fortes, I., and Aguilera, G. Closure via functional dependence simplification. International Journal of Computer Mathematics 89, 4 (2012), 510-526.

[7] Ricci, F., Rokach, L., and Shapira, B. Recommender Systems Handbook, 2nd ed. Springer Publishing Company, Incorporated, 2015.

[8] du Boucher-Ryan, P., and Bridge, D. Collaborative recommending using formal concept analysis. Knowledge-Based Systems 19, 5 (2006), 309-315.

[9] Nagler, T., and Czado, C. Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. Journal of Multivariate Analysis 151 (2016), 69-89.