

5. Conclusiones y Líneas Futuras

En este capítulo se exponen las conclusiones y líneas futuras derivadas de este proyecto.

La minería de datos permite extraer información no evidente a partir de grandes cantidades de datos que no serían procesables por un humano. En este trabajo se describen las fases del diseño y el desarrollo de un proyecto de minería de datos para la toma de decisiones por parte de un operador, a partir de datos provenientes de un CRM.

Si se analiza todo el proceso de principio a fin, puede observarse que el diseño de la solución al problema se va moldeando levemente para adaptarse lo mejor posible al procesamiento requerido de los datos. El diseño debe ser flexible para amoldarse al problema a medida que se acerca a la solución. La solución depende fuertemente del problema a resolver. Este hecho es el que ha propiciado la existencia de gran cantidad de algoritmos de Data Mining distintos, cada uno de ellos diseñado inicialmente para un cometido específico.

Sin embargo, el resultado final de este trabajo es un sistema que, apoyado sobre el algoritmo Apriori, a pesar de ser ideado para un propósito concreto, es configurable y reutilizable para otros cometidos. Esto se debe a que en el paso previo que se realiza previamente a la ejecución del algoritmo, el operador puede transformar la información a su gusto con una variedad razonablemente completa de opciones.

El componente humano a la hora de analizar los datos no se puede eliminar. Sin minería de datos, el operador debería revisar a mano miles de registros; en la reutilización de este trabajo, el operador debería preparar la transformación de los datos para cada distinta aplicación; y en todo caso, la calidad de los resultados finales, el conocimiento, solo puede ser realmente evaluada como útil por un humano.

El sistema diseñado, a pesar de estar orientado a una aplicación concreta, resulta ser moldeable. Una ampliación de este trabajo es su reutilización para otros problemas que requieran el análisis de datos de naturaleza distinta. Esto implica, en lugar de rediseñar la aplicación de Data Mining por completo, cambiar el tratamiento previo de los datos antes de pasar por el algoritmo. Es decir, se ahorra el diseño de la aplicación de minería de datos, para centrarse en la elección de las propiedades importantes a tratar y el cómo tratarlas.

Además, es importante en una aplicación de este tipo el ensayo y error. Cada nueva ejecución sirve para refinar el resultado obtenido hasta que el operador lo considere útil, o descarte que tras el tratamiento realizado pueda descubrir nueva información. Grandes cantidades de datos manipuladas de forma incorrecta pueden conducir a que no se descubra ningún nuevo conocimiento nuevo. Cantidades no tan grandes de datos, que aparentemente no aportan información, pueden en realidad contener conocimiento no evidente. Para llegar a ese nuevo conocimiento es necesario identificar las propiedades de interés, ya sea de forma razonada o intuitiva, y ejecutar el algoritmo. Aunque jamás se puede saber a priori si se descubrirá nuevo conocimiento o no.