

Tarea 2 - Market Basket Analysis

Reconocimiento de Patrones en Minería de Datos

Profesor: Marcelo Mendoza - mmendoza@inf.utfsm.cl

Ayudante: Ignacio Espinoza - ignacio.espinoza@alumnos.inf.utfsm.cl

2 de mayo de 2017

1. Introducción

En esta tarea trabajará buscando reglas de asociación en el dataset *groceries.csv* y con el software Weka versión 3.8.1, disponible en <http://www.cs.waikato.ac.nz/ml/weka/index.html>.

Los algoritmos que debe probar son los siguientes:

1. Apriori
2. FP-growth

Lo que debe hacer es lo siguiente:

1. Transformar el dataset de formato csv a arff¹, extensión utilizada por Weka.
2. Para el algoritmo Apriori debe seleccionar las top-500 reglas más interesantes (*numRules*). Utilice un soporte mínimo (*lowerBoundMinSupport*) de 0.001. Debe encontrar las reglas para valores de confianza (*minMetric*) 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94 y 0.95. Para cada valor de confianza, contabilice el número de reglas encontradas. Luego, plotee el número de reglas (eje y) en función de la confianza (eje x). Comente sus resultados. Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas, identificando su valor de confianza y cantidad de apariciones.
3. Usando la misma configuración anterior sobre el algoritmo Apriori, cambie el valor del soporte mínimo a 0.0013. Repita el experimento anterior (encontrar las reglas para valores de confianza) y plotee el número de reglas (eje y) en función de la confianza (eje x). Comente sus resultados. Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas. Justifique su elección.
4. Para el algoritmo FP-growth, mantenga el soporte mínimo en 0.001 y busque reglas con confianza mínima en el rango 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94 y 0.95. Para cada valor de confianza, contabilice el número de reglas encontradas. Repita el experimento con FP-growth subiendo el soporte mínimo en 1 % cada vez, hasta llegar a 0.0018. Para cada valor de confianza, contabilice el número de reglas encontradas. Luego, plotee el número de reglas (eje y) en función de la confianza (eje x). Comente sobre el comportamiento de las curvas. Finalmente, seleccione 5 reglas que le parezcan inesperadas y coméntelas. Justifique su elección.

¹<http://www.cs.waikato.ac.nz/ml/weka/arff.html>

2. Entregable y consideraciones:

- Los equipos de trabajo pueden ser de 2 a 3 personas.
- Informe: subir un informe con los resultados obtenidos en su tarea a Moodle. Muestre todos los plots encontrados. Indique la configuración de parámetros usada. Comente las reglas seleccionadas como inesperadas. Indique los integrantes del grupo.
- **FECHA DE ENTREGA: Martes 23 de Mayo hasta las 23:55. Pasada la hora tendrá 20 puntos de descuento por día de atraso.**