

Introducción a Data Mining

Caso N.º 1

Integrantes: Pedro Allegro e Ignacio Dottore

Analisis del caso

Una cadena de hoteles que gestiona numerosos establecimientos en varios países registra información acerca de sus huéspedes. Esta información se genera mediante dos maneras:

- Durante la estadía: a la llegada se registran los datos demográficos del huésped y se le entrega una credencial que es utilizada para acceder a distintos puntos del hotel y también realizar pagos.
- Después de la estadía: se le solicita al huésped que complete un formulario de evaluación que cuenta también con la posibilidad de dejar mensajes o comentarios adicionales.

Problema a resolver

- Recomendación de eventos, actividades y promociones durante la estadía: por medio del canal de TV se desea realizar una selección a la medida de cada huésped.
 - Considerar que a priori solamente se cuenta con la información demográfica.
- Recomendación de hoteles a visitar luego de la estadía: por medio correos electrónicos
 - Considerar que generalmente un huésped no suele visitar muchas veces el mismo hotel.

Recorrer los diferentes pasos del proceso KDD, explicando cómo se aplican en este caso. Indicar para cada paso cuáles técnicas usaría y justificar su elección.

Proceso KDD

Knowledge Discovery from Data (KDD) es un proceso iterativo compuesto por los siguientes pasos:

1. Data cleaning: para eliminar ruido e inconsistencia de datos
2. Data integration: donde múltiples fuentes de datos son combinadas
3. Data selection: donde se seleccionan los datos relevante para el análisis
4. Data transformation: donde se realizan transformaciones y agregaciones de datos
5. Data mining: proceso donde un método es aplicado para extraer patrones de los datos
6. Pattern evaluation: identificación de los patrones realmente relevantes

7. Knowledge presentation: divulgación de los patrones relevantes a los interesados en el proyecto

Resolución del caso mediante KDD

Paso	Aplicacion
Data cleaning	<p><u>Durante la estadía</u>: la recolección de datos demográficos puede contener datos faltantes o erróneos. Se requiere que los mismos sean debidamente validados antes de ser incorporados al data warehouse.</p> <p><u>Después de la estadía</u>: el formulario completado por los clientes puede contener incongruencias (sobretudo si es completado de forma manual por los mismos). Los mensajes que dejen los clientes deben ser limpiados antes de ser incorporados al data warehouse.</p>
Data integration	Para ambos casos se requiere implementar un data warehouse que sea capaz de almacenar en forma tabular los registros existentes y añadir de forma periódica los nuevos registros que se generen
Data selection	Para ambos casos se requiere seleccionar del data warehouse las tablas y campos que se vayan a utilizar en el paso siguiente
Data transformation	Para ambos casos se requiere realizar las transformaciones necesarias antes de utilizar las tablas transformadas en el paso siguiente
Data mining	<p>Data Characterization Puede utilizarse para realizar distintos análisis exploratorios de los datos recolectados durante y luego de la estadía, como qué hoteles obtuvieron las mejores calificaciones, qué actividades fueron las más populares, o qué productos suelen entregar mayores ganancias.</p> <p>Patterns Se recomienda realizar este análisis para descubrir qué conjunto de actividades y productos suelen comprarse en simultáneo o secuencialmente. Asimismo como se relacionan con las N tipos de huéspedes del algoritmo de Clustering (ver siguiente punto).</p> <p>Clustering <u>Durante la estadía</u>: resulta interesante saber cómo se relacionan los huéspedes según sus datos demográficos junto con las actividades que realizan en los distintos hoteles de la empresa y las distintas alternativas que ofrecen los distintos hoteles. Dado que todos estos datos mencionados se encuentran disponibles en nuestro data warehouse, podemos utilizarlos para realizar una clusterización que nos permita identificar los N diferentes tipos de huéspedes que existen (ejemplo: familias con hijos pequeños, parejas, amigos, adultos mayores, etc)</p>

	<p>Clasificación</p> <p><u>Durante la estadía:</u> al ingresar un nuevo huésped solamente se pueden utilizar los datos demográficos del mismo para realizar una predicción sobre a cuál de las N clases de Clustering pertenece y así realizar las recomendaciones de actividades y productos que más se ajuste a las realizadas por los de su misma clase.</p> <p>Conforme pasen los días y se recolecten datos sobre la actividad del huésped, puede evaluarse que tan buena fue la predicción realizada, existiendo la posibilidad de reclasificar al huésped en una clase diferente (con la consiguiente modificación de las actividades y productos a recomendar). Asimismo en esta etapa sería aconsejable generar un modelo que defina a que huéspedes se asignen promociones según algún criterio (descuentos a aquellos que hayan adquirido actividades o productos en los días previos, o bien incentivos a aquellos que no hayan realizado ninguna o pocas actividades aun).</p> <p><u>Después de la estadía:</u> la clase resultante del huésped en combinación con los resultados de las encuestas de los demás huéspedes serán los inputs del modelo que desarrollemos para recomendar el siguiente hotel a visitar. El cual será el que más se ajuste a los gustos y preferencias de los sujetos de la clase a la que corresponda el huésped, eliminando del análisis el hotel que ya ha visitado.</p> <p>Regresion</p> <p>De acuerdo a la clase a la que pertenezca cada huésped pueden hacerse modelos que predigan el monto total que se espera que gasten durante cada estadía. Estos modelos pueden emplearse en combinación el algoritmo de recomendación de actividades y productos, como así también con el de recomendación de hoteles a fin de maximizar las utilidades esperadas por huésped.</p>
Pattern evaluation	Para ambos casos se deberán evaluar si los patrones encontrados resultan de utilidad para el negocio antes de pasar a la implementación.
Knowledge presentation	En esta fase final es donde se presentan los resultados hallados y se define la implementación de los mismos.