Procesos de Explotación de Información – Casos de Estudio

En la siguiente tabla se plantean casos que pueden ser abordados con Explotación de Información. Marque con cruces aquellos procesos que estime sean necesarios aplicar para obtener la solución (puede haber más de una opción por caso).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N° DE CASO** | **DESCRIPCIÓN DEL CASO** | Descubrimiento de Reglas de Comportamiento | Descubrimiento  de Grupos | Ponderación de Interdependencia de Atributos | Descubrimiento de Reglas de Pertenencia a Grupos | Ponderación de Reglas de Comportamiento (o Pertenencia a Grupos) |
| **1** | Un banco ya tiene los usuarios categorizados en grupos y desea saber qué grupo asignarle cuando llega un nuevo cliente. |  |  | X |  |  |
| **2** | Una compañía de venta de celulares dispone de una base de datos de clientes, con datos del tipo: si posee recarga con tarjeta o abono, las recargas que realiza un cliente, fecha de renovación del plan, saldo actual en su cuenta, entre otras. Con base en esta información quiere saber cuándo viene un cliente nuevo, qué celular ofrecerle primero, y qué ventajas le puede ofrecer con ese celular. | X |  |  |  |  |
| **3** | Un banco cuenta con una base de datos de clientes que efectúan plazos fijos de manera periódica, otros que lo hacen aleatoriamente y otros que no realizan ningún tipo de plazo fijo. Desea caracterizar a los clientes para enfocar sus campañas de marketing. |  |  |  | X |  |
| **4** | Ante la aparición de un virus, una organización internacional, desea saber cuáles pueden ser los grupos de riesgo y los factores para la adquisición del virus. Se poseen datos de la población de infectados, lo que incluye edad, sexo, peso, estatura, estado de salud previo, zona geográfica de residencia, entre otros. |  |  |  |  | X |
| **5** | El área de control de calidad de una planta industrial desea distinguir los factores determinantes asociados a las fallas de las piezas que produce. La empresa posee datos sobre cada falla y el estado de las piezas. |  | X | X |  |  |
| **6** | Un club cuenta con un censo de socios en el que se registran datos personales y sus preferencias en cuanto a las actividades ofrecidas por el club. Se requiere poder ofrecer a los nuevos socios planes de actividades con una marcada como preferente. |  | X |  |  |  |
| **7** | Un Call Center de soporte técnico a cliente ha detectado un aumento significativo en incidentes de un producto determinado y la empresa necesita saber cuál es la razón para este incremento. Se cuenta con la BD de incidentes con los datos del cliente y datos del incidente, entre estos, tipificación de la falla y tipificación de la solución. | X |  |  |  |  |
| **8** | Una cadena de gimnasios desea determinar perfiles patrón de sus socios con el fin de ofrecerles servicios y productos exclusivos por perfil, como por ejemplo ropa o accesorios para determinadas disciplinas, viajes o excusiones para la tercera edad, campeonatos de futbol entre sedes, entre otros. El gimnasio cuenta con la base de datos de los clientes, donde se detallan sexo, año de nacimiento, actividades que realizan, frecuencia de asistencia, problemas físicos. |  |  |  |  | X |
| **9** | El dueño de un sitio web muy similar a Taringa, desea mejorar la calidad del sitio reduciendo la cantidad de posts que por distintos motivos terminan siendo denunciados y eliminados. Para ello, planea implementar cierta limitación para que los usuarios novatos que tengan una característica identificatoria no puedan crear posts potencialmente denunciados. | X |  |  |  | X |
| **10** | Una empresa dedicada al desarrollo de sistemas desea identificar qué características contribuyen al alto volumen de rotación de los empleados. | X |  |  |  | X |

UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

Guía de Preguntas

Procesos de Explotación de Información

1. ¿Qué relación existe entre Explotación de Información y Explotación de Información basada en Sistemas Inteligentes?

La Explotación de Información es la sub-disciplina Informática que aporta a la Inteligencia de Negocio, las herramientas para la transformación de información en conocimiento. Se ha definido como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información . Al hablar de explotación de información basada en sistemas inteligentes se refiere específicamente a la aplicación de métodos de sistemas inteligentes, para descubrir y enumerar patrones presentes en la información. Los métodos basados en sistemas inteligentes , permiten obtener resultados de análisis de la masa de información que los métodos convencionales no logran tales como: los algoritmos TDIDT (Top Down Induction Decision Trees), los mapas auto organizados (SOM) y las redes bayesianas. Los algoritmos TDIDT permiten el desarrollo de descripciones simbólicas de los datos para diferenciar entre distintas clases.

1. De una lista de Procesos de Explotación de Información aplicables a problemas de Inteligencia de Negocios.

La Explotación de Información toma de los sistemas inteligentes las siguientes tecnologías:

A continuaciòn se nombrar

**Algoritmos Genéricos**

Son métodos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las variables de estudio constituyen un segmento de información. Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la variable de respuesta, corresponderán a segmentos con mayor capacidad reproductiva. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Se puede además introducir elementos aleatorios para la modificación de las variables (mutaciones). Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización, pues las malas soluciones han ido descartándose, iteración tras iteración.

**Algoritmos TDIDT:**

Estos algoritmos (TDIDT - Top Down Induction Decisión Trees) pertenecen a los métodos inductivos del Aprendizaje Automático que aprenden a partir de ejemplos preclasificados. A esta familia pertenecen los algoritmos: ID3, C4.5 y C5. Estos algoritmos generan árboles y reglas de decisión a partir de ejemplos preclasificados. Para construir los árboles se utiliza el método de aprendizaje automático basado en la estrategia propuesta por Hunt en, que particiona el conjunto de ejemplos en subconjuntos a medida que avanza. Trabajar sobre cada subconjunto es más sencillo que trabajar sobre el total de los datos.

**Redes Neuronales BP:**

Son redes formadas por múltiples capas lo que les permite resolver problemas que no son linealmente separables. Pueden ser totalmente o localmente conectadas. En el primer caso cada salida de una neurona de la capa "i" es entrada de todas las neuronas de la capa "i+1", mientras que en el segundo caso, cada neurona de la capa "i" es entrada de una serie de neuronas (región) de la capa "i+1". Utilizan un algoritmo de aprendizaje llamado regla delta generalizada (ó regla de retropropagación del error), que consiste en minimizar el error (comúnmente cuadrático) por medio del método del gradiente descendente en los parámetros de entrenamiento de la red neuronal. Estas redes son conocidas como redes de retropropagación (Redes BP).

**Redes Neuronales SOM:**

Los mapas auto organizados o SOM (Self-Organizing Map), también llamados redes de Kohonen son un tipo de red neuronal no supervisada competitiva, con capacidad para formar mapas de características bidimensionales a partir del principio de formación de mapas topológicos. Se orientan a descubrir la estructura subyacente de los datos ingresados a partir de establecer características comunes entre los vectores de información de entrada a la red. A lo largo del entrenamiento de la red; los vectores de datos son introducidos en cada neurona y se comparan con el vector de peso característico de la misma. La neurona que presenta menor diferencia entre su vector de peso y el vector de datos es la neurona ganadora (o BMU) y ella y sus vecinas verán modificados sus vectores de pesos.

**Redes Bayesianas:**

Las redes bayesianas o probabilísticas se fundamentan en la teoría de la probabilidad y combinan la potencia del teorema de Bayes con la expresividad semántica de los grafos dirigidos; las mismas permiten representar un modelo causal por medio de una representación gráfica de las independencias / dependencias entre las variables que forman parte del dominio de aplicación. Se puede interpretar a una red bayesiana de dos formas: (a) distribución de probabilidad que representa la distribución de la probabilidad conjunta de las variables representadas en la red, ó (b) base de reglas en la que cada arco representa un conjunto de reglas que asocian a las variables involucradas y están cuantificadas por las probabilidades respectivas.

1. ¿Cuándo se aplican cada uno de los Procesos de Explotación de Información? Ejemplique.
2. Enuncie la técnica utilizada para cada Proceso de Explotación de Información, describa el proceso y de un esquema.

3 - **Descubrimiento de Reglas de Comportamiento :** Aplica cuando se requiere identificar cuales son las condiciones para obtener determinado resultado en el dominio del problema.

Ejemplos :

* Cuando se desea identificar las características del local más visitado por los clientes.
* Cuando se desea identificar los factores que inciden en el alza de las ventas de un producto dado.
* Establecer las características o rasgos de los clientes con alto grado de fidelidad a la marca

**Descubrimiento de Grupos:** Aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio del problema. No se dispone de ningún criterio de agrupamiento “a priori”.

Ejemplos :

* Cuando se desea identificar segmentos de clientes para bancos.
* Cuando se desea identificar tipos de llamadas de clientes para empresas de telecomunicaciones.
* Cuando se desea identificar grupos sociales con las mismas características.
* Cuando se desea identificar grupos de estudiantes con características homogéneas.

**Ponderación de Interdependencia de Atributos:** Aplica cuando se requiere identificar cuales son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema.

Ejemplos :

* Factores con incidencia sobre las ventas.
* Rasgos distintivos de clientes con alto grado de fidelidad a la marca.
* Atributos claves que convierten en vendible a un determinado producto.
* Características sobresalientes que tienen los visitantes de un Website.

**Descubrimiento de Reglas de Pertenencia a Grupos :** Aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio del problema.

Ejemplos :

* Para tipología de perfiles de clientes y caracterización de cada tipología.
* Para segmentación etaria de los estudiantes y comportamiento de cada segmento.
* Para clases de llamadas telefónicas en una región y caracterización de cada clase.

**Ponderación de Reglas de Comportamiento:**

Aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) en la determinación de una clase. - Existen clases/grupos identificados.

Ejemplos :

* Para rasgos con mayor presencia en los clientes con alto grado de fidelidad a la marca.
* Para frecuencia de ocurrencia de cada perfil de clientes.

**Ponderación de Pertenencia a Grupos:** Aplica cuando se requiere identificar cuales son las condiciones con mayor incidencia (o frecuencia de ocurrencia) en los atributos descriptores de un grupo. - No existen clases/grupos identificados.

Ejemplos :

* Identificar el factor dominante que incide en el alza las ventas de un producto dado.
* Identificar el tipo de llamada más frecuente en una región.

**Material de Referencia**

Britos, P., García-Martínez, R. (2009). *Propuesta de Procesos de Explotación de Información*. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.