

MINERIA DE DATOS

por BRITOS
Isbn 9871104308

Indice del Contenido

Capítulo 1. INTRODUCCIÓN A LA EXPLOTACIÓN DE DATOS INTELIGENTES

1. Introducción
2. Concepto de Data Warehousing
 - 2.1. Análisis de la definición de Bill Inmon
3. Usos y aplicaciones de un Data Warehousing
4. Las bases para el desarrollo de un data warehouse
 - 4.1. Arquitectura del data warehouse
 - 4.2. Diferencias entre OLTP y OLAP
 - 4.3. Construcción de un data warehouse
 - 4.4. Metodología para construir un data warehouse
5. Costos del data warehouse
6. Impactos de implementación del data warehouse
7. El data mining y su relación con el data warehouse
8. El Data Warehouse dota a las organizaciones de memoria, y el data mining de inteligencia
9. Minería de datos
 - 9.1. Descubrimiento de conocimientos
 - 9.2. Tareas realizadas por un sistema de Minería de Datos
10. Aprendizaje Automático
 - 10.1. Ejemplos de sistemas de aprendizaje automático
 - 10.2. Aprendizaje supervisado y no supervisado

Capítulo 2. MINERÍA DE DATOS CLÁSICA

- 1.1. Introducción
- 1.2. Evolución conjunta de los requerimientos de los usuarios y de las herramientas utilizadas en Minería de Datos
- 1.3. Técnicas Clásicas para el campo de la Minería de Datos

Capítulo 3. PREPROCESAMIENTO DE DATOS

1. Introducción
2. ¿Por qué preprocesar los Datos?
3. Limpieza de Datos
 - 3.1. Valores Faltantes
 - 3.2. Datos con Ruido
 - 3.3. Datos inconsistentes
4. Integración y Transformación de Datos
 - 4.1. Integración de Datos
 - 4.2. Transformación de Datos
5. Reducción de datos
 - 5.1. Agregado a cubo de datos
 - 5.2. Reducción de dimensión
 - 5.3. Compresión de datos
6. Discretización y generación de jerarquías conceptuales
 - 6.1. Discretización y generación de jerarquías conceptuales para datos numéricos
 - 6.2. Generación de jerarquías conceptuales y datos categóricos
7. Metodología de preprocesamiento de datos
 - 7.1. Detalle para el diseño del controlador de tareas
 - 7.2. Características que no se contemplan en el diseño del controlador de tareas
 - 7.3. Detalle de las herramientas que se utilizarán para las transformaciones

- 7.4. Componentes para el controlador de tareas
- 7.5. Método de transformación

Capítulo 4. DESCRIPCIÓN CONCEPTUAL

- 1. Introducción
- 2. ¿Qué se entiende por Descripción Conceptual?
- 3. Generalización y sumarización de datos basada en la caracterización
 - 3.1. Inducción Orientada a los Atributos
 - 3.2. Implementación eficiente de la inducción Orientada a los Atributos
 - 3.3. Presentación derivada de la generalización
- 4. Caracterización Analítica: Análisis de Atributos Relevantes
 - 4.1. ¿Por qué realizar el Análisis de Atributos Relevantes?
 - 4.2. Métodos de Análisis de Relevancia de Atributos
 - 4.3. Caracterización Analítica: Un Ejemplo
- 5. Resumen

Capítulo 5. REGLAS DE ASOCIACIÓN EN MINERÍA DE GRANDES BASES DE DATOS

- 1. Introducción
 - 1.1. Reglas de Asociación de Minería
- 2. Reglas de Asociación de Minería Monodimensionales Booleanas Para Bases De Datos Transaccionales
 - 2.1. El Algoritmo Apriori: Encontrar Itemsets Frecuentes Usando la Generación de Candidatos
 - 2.2. Generación de Reglas de Asociación para Itemsets Frecuentes
 - 2.3. Mejorando la Eficiencia del Apriori
 - 2.4. Minería de Itemsets Frecuentes Sin Generación de Candidatos
 - 2.5. Consultas Témpano
- 3. Reglas de asociación de minería multinivel para bases de datos transaccionales
 - 3.1. Reglas de Asociación Multinivel
 - 3.2. Aproximaciones a las reglas de asociación multinivel
 - 3.3. Chequeo de Redundancia de Reglas de Asociación Multinivel
- 4. Reglas de Asociación de Minería Multidimensional Para Bases de Datos Relacionales y Data Warehouses
 - 4.1. Reglas de Asociación Multidimensional
 - 4.2. Reglas de la Asociación Multidimensional de Minería Usando Discretización Estática de Atributos Cuantitativos
 - 4.3. Generación de Reglas de Asociación Cuantitativa
 - 4.4. Reglas de Asociación Basadas en la Distancia

Capítulo 6. APRENDIZAJE AUTOMÁTICO

- 1. Introducción
- 2. Aprendizaje Automático
 - 2.1. Esquema general de un sistema de Aprendizaje Automático
 - 2.2. Aprendizaje de conceptos
 - 2.3. Aprendizaje supervisado y no supervisado
 - 2.4. Tipos de aprendizaje automático
- 3. Métodos clásicos de aprendizaje
 - 3.1. Aprendizaje AQ
 - 3.2. Aprendizaje divide y reinarás
- 4. Minería de Datos
 - 4.1. Descubrimiento de conocimientos
 - 4.2. Problemas inherentes al proceso de aprendizaje
 - 4.3. Tareas realizadas por un sistema de Minería de Datos
 - 4.4. Métodos de Minería de Datos
 - 4.5. Componentes de la Minería de Datos
- 5. Aprendizaje Automático & Minería de Datos
 - 5.1. Aplicaciones

- 5.2. Inferencia de Reglas Simples
- 5.3. Construcción de reglas de decisión
- 6. La Familia TDIDT
 - 6.1. Construcción de los árboles de decisión
 - 6.2. Atributos Desconocidas
 - 6.3. Transformación a Reglas de Decisión
 - 6.4. Resolución de un ejemplo utilizando el ID3
- 7. Evaluación de los Métodos de Aprendizaje
 - 7.1. Evaluación en la familia TDIDT
 - 7.2. Métodos de evaluación
 - 7.3. Estimación del costo

Capítulo 7. SEDES NEURONALES

- 1. Fundamentos de las redes neuronales
 - 1.1. El modelo biológico
 - 1.2. Elementos de una red neuronal artificial (RNA)
- 2. Regla de aprendizaje
 - 2.1. Representación vectorial
 - 2.2. Estructura de una RNA
 - 2.3. Mecanismo de aprendizaje
 - 2.4. Redes neuronales con conexiones hacia adelante

Capítulo 8. ALGORITMOS GENÉTICOS

- 1. Introducción
 - 1.1. Algoritmo simple
 - 1.2. Sistemas naturales y sistemas artificiales
 - 1.3. Aplicaciones
- 2. Operadores básicos: variantes
 - 2.1. Métodos de selección
 - 2.2. Selección por ruleta
 - 2.3. Selección con control sobre el número esperado
 - 2.4. Selección elitista
 - 2.5. Selección por ranking
 - 2.6. Métodos de cruza
 - 2.7. Métodos de mutación
- 3. Algoritmos genéticos en profundidad
 - 3.1. Esquemas
 - 3.2. Teorema Fundamental
 - 3.3. Efecto De la Selección
 - 3.4. Efecto De La Cruza
 - 3.5. Efecto De la Mutación
 - 3.6. Mecanismos De Selección
 - 3.7. Función De Aptitud
 - 3.8. Necesidad De Escalado
 - 3.9. Mecanismos De Cruza
 - 3.10. Mecanismos De Mutación
- 4. Resolución de problemas con AG
 - 4.1. El Problema De la Representación
 - 4.2. Codificación De Parámetros
 - 4.3. Parámetros Binarios
 - 4.4. Parámetros No Binarios
 - 4.5. Números Enteros
 - 4.6. Codificación Binaria de Parámetros Enteros
 - 4.7. Codificación Gray
 - 4.8. Números Reales
 - 4.9. Distribución De Los Genes Dentro Del Cromosoma

- 4.10. Elección De La Función De Aptitud
- 4.11. Metodología De Diseño De AG
- 4.12. Conclusión
- 5. Algoritmos evolucionarios
- 5.1. Programación Evolucionaria
- 5.2. Estrategia Evolucionaria
- 5.3. Algoritmos Genéticos Secuenciales
- 5.4. Algoritmos Genéticos Paralelos
- 5.5. Global
- 5.6. Grano Grueso
- 5.7. Grano Fino
- 5.8. Conclusión

Capítulo 9. REDES BAYESIANAS

- 1. Introducción
- 2. Definición formal de las redes bayesianas
- 3. Representación del conocimiento
- 3.1. Independencia condicional
- 3.2. Inferencia
- 3.3. Algoritmos de propagación
- 4. El aprendizaje en las redes bayesianas
- 4.1. Aprendizaje paramétrico
- 4.2. Aprendizaje estructural
- 4.3. Ventajas de las redes bayesianas

Capítulo 10. CASO DE ALGORITMO DE INDUCCIÓN

- 1. Contexto del Problema
- 2. Aplicación de los Algoritmos de la Familia TDIDT ala Minería de Datos
- 3. Solución propuesta
- 3.1. Interpretación de los resultados
- 3.2. Descripción de los dominios
- 3.3. Resultados Obtenidos con el ID3
- 3.4. Resultados Obtenidos con el C4.5
- 3.5. Comparación de los resultados obtenidos con el ID3 y con el C4.5
- 3.6. Análisis general de los resultados obtenidos
- 4. Conclusiones Generales
- 4.1. Conceptos destacables
- 4.2. Análisis de los Resultados Obtenidos
- 4.3. Análisis de la solución propuesta
- 4.4. Una mirada al futuro

Capítulo 11. CASO DE APLICACIÓN DE REDES NEURONALES

- 1. Introducción
- 2. Definición de fraude en telefonía celular
- 2.1. Clasificación de tipos de fraude
- 2.2. Detección de fraude en telefonía celular
- 3. Redes neuronales
- 4. Descripción del problema
- 4.1. El problema de la construcción y mantenimiento de "perfiles de usuario"
- 4.2. El problema de la detección de cambios de comportamiento
- 4.3. El problema de la performance
- 4.4. Representación gráfica del problema a resolver
- 5. Solución propuesta
- 5.1. Procesamiento previo de la información-Mediación
- 5.2. Solución a la construcción y mantenimiento de "perfiles de usuario"
- 5.3. Solución a la detección de cambios de comportamiento

- 5.4. Solución a las cuestiones de performance
- 5.5. Restricciones de la solución propuesta
- 6. Prueba experimental
 - 6.1. Conjunto de datos utilizados
 - 6.2. Variables a observar
 - 6.3. Realización de los experimentos
 - 6.4. Resultados
 - 6.5. Análisis de los resultados
- 7. Conclusiones

Capítulo 12. CASO DE APLICACIÓN DE ALGORITMOS GENÉTICOS

- 1. Descripción del problema
 - 1.1. La búsqueda en el espacio de soluciones
 - 1.2. Utilización de algoritmos genéticos
- 2. Solución propuesta
 - 2.1. Adaptaciones existentes
 - 2.2. Algoritmo propuesto
- 3. Prueba experimental
 - 3.1. Conjunto de datos utilizado
 - 3.2. Variables a observar
 - 3.3. Realización de los experimentos
- 4. Resultados
 - 4.1. Experimentos variando la cantidad de grupos
 - 4.2. Experimentos variando la cantidad de documentos
 - 4.3. Análisis de los resultados
- 5. Conclusiones
 - 5.1. Respuesta a las cuestiones planteadas

Capítulo 13. CASO DE USO DE REDES BAYESIANAS

- 1. Descripción del problema
- 2. Solución propuesta
 - 2.1. Datos de entrada
- 3. Sistema integrador
 - 3.1. Descripción general
- 3. Prueba experimental
 - 3.1. Descripción de los dominios
 - 3.2. Metodología utilizada
 - 3.3. Análisis estadístico de los resultados
- 4. El test de Wilcoxon para la comparación de muestras apareadas
 - 4.1. Introducción
- 5. Resultados
- 6. Conclusiones

Capítulo 14. HERRAMIENTA DE ALGORITMO DE INDUCCIÓN

- 1. Introducción
 - 1.1. Hoja ReadMe
 - 1.2. Hoja UserInput
 - 1.3. Hoja Data
 - 1.4. Hoja Tree
 - 1.5. Hoja NodeView
 - 1.6. Hoja Result
 - 1.7. Hoja Rules

Capítulo 15. HERRAMIENTA DE AGRUPAMIENTO A TRAVÉS DE REDES NEURONALES

- 1. Introducción
 - 1.1. Hoja ReadMe

- 1.2. Hoja Input
- 1.3. Hoja Data
- 1.4. Hoja Weights
- 1.5. Hoja Output
- 1.6. Hoja Junk y Hoja Plot
- 1.7. Base de Datos de ejemplo

Capítulo 16. HERRAMIENTA DE PREDICCIÓN A TRAVÉS DE REDES NEURONALES

1. Introducción
- 1.1. Hoja ReadMe
- 1.2. Hoja UserInput
- 1.3. Hoja Data
- 1.4. Hoja Calc
- 1.5. Hoja Output
- 1.6. Hoja Profile
- 1.7. Hoja LifChart

Capítulo 17. HERRAMIENTA DE ALGORITMOS GENÉTICOS

1. Introducción
- 1.1. Requisitos Del Sistema
2. Operadores Provistos
3. Menús
4. Ventanas
5. Escribiendo Funciones de Evaluación
- 5.1. Pasaje de Información
- 5.2. Información debe proveer la DII
- 5.3. Definiciones del FDL
6. Comienzo Rápido
7. Problema del viajante
- 7.1. Definición del problema
- 7.2. Espacio de Búsqueda
- 7.3. Aproximación de la solución
8. Usando el programa Wingo

Capítulo 18. HERRAMIENTA DE REDES BAYESIANAS

1. Introducción
- 1.1. Descripción de los algoritmos de aprendizaje
2. Obtención de redes bayesianas a través de Elvira
- 2.1. Uso de Elvira
3. Comparación de los resultados obtenidos con cada Clasificador
- 3.1. Comparación de las redes obtenidas
- 3.2. Comparación de los tiempos de proceso
- 3.3. Respuestas de las distintas redes al ingreso de evidencia
4. Conclusión

BIBLIOGRAFÍA