
Índice

0. PRÓLOGO: APRENDIZAJE [*MACHINE LEARNING*]

1. DATA MINING & KNOWLEDGE DISCOVERY IN DATABASES [*KDD*]

1. INTRODUCCIÓN	1-2
2. DATA MINING	1-4
3. MATERIAS RELACIONADAS	1-6
3.1 Estadística	1-6
3.2 Ingeniería del conocimiento	1-6
3.3 Bases de datos	1-6
4. REFERENCIAS	1-7

PRIMERA PARTE: REGLAS DE ASOCIACIÓN

2. REGLAS DE ASOCIACIÓN

1. INTRODUCCIÓN	2-2
2. DIVIDE Y VENCERÁS	2-5
3. ALGORITMOS PARA LA OBTENCIÓN DE CONJUNTOS RELEVANTES DE ITEMS	2-6
3.1 AIS	2-7
3.2 SETM	2-13
3.3 APRIORI, APRIORITID & APRIORIHYPBRID	2-19
3.3.1 Algoritmo Apriori	2-21
3.3.2 Algoritmo AprioriTID	2-27
3.3.3 Algoritmo AprioriHybrid	2-31
3.4 OCD	2-32
3.5 DHP	2-34
3.6 MAX-MINER	2-39

4. ALGORITMOS PARA LA OBTENCIÓN DE REGLAS DE ASOCIACIÓN	2-42
4.1 Algoritmo directo	2-43
4.2 Algoritmo mejorado	2-47
5. REFERENCIAS BIBLIOGRÁFICAS	2-52

3. REGLAS DE ASOCIACIÓN EN BASES DE DATOS RELACIONALES

1. INTRODUCCIÓN	3-2
2. ALGORITMO <i>t</i>	3-5
2.1 Itemsets relevantes	3-7
2.2 Reglas de asociación	3-18
3. ALGORITMO <i>T</i>	3-23
4. ALGORITMO <i>TBAR</i>	3-30
4.1 Visión general	3-31
4.2 Estructura de datos: Árbol de itemsets	3-33
4.3 Refinamientos del algoritmo	3-49
4.4 El algoritmo paso a paso	3-55
5. RESULTADOS EXPERIMENTALES	3-60

4. LAS REGLAS DE ASOCIACIÓN Y SU USO

1. EXTENSIONES DEL MODELO BÁSICO	4-2
1.1 Jerarquías de conceptos	4-2
1.2 Manejo de atributos numéricos	4-3
1.2.1 MAQA	4-5
1.3 Medidas alternativas de relevancia	4-6
1.4 Obtención de reglas de asociación con restricciones	4-7
1.5 Mantenimiento de bases de datos de reglas de asociación	4-7
2. POST-PROCESAMIENTO DE LAS REGLAS DE ASOCIACIÓN	4-8
2.1 Técnicas de visualización	4-9
3. OTRAS APLICACIONES	4-10
3.1 Análisis de secuencias temporales	4-10
3.2 Análisis de imágenes	4-10
3.3 Clasificación	4-11
3.3.1 ARTs (© Juan Carlos Cubero)	4-13
4. REFERENCIAS BIBLIOGRÁFICAS	4-16

SEGUNDA PARTE: APRENDIZAJE SUPERVISADO

5. CLASIFICACIÓN

1. INTRODUCCIÓN	5-2
2. MODELOS DE CLASIFICACIÓN	5-3
2.1 Árboles de decisión	5-3
2.2 Reglas de producción	5-4
2.3 Reglas de asociación	5-5
2.4 Clasificadores basados en medidas de similaridad	5-6
2.5 Técnicas estadísticas	5-7
2.6 Redes neuronales	5-8
2.6.1 LVQ	5-8
2.7 Algoritmos genéticos	5-9
3. BIBLIOGRAFÍA	5-10

6. ÁRBOLES DE DECISIÓN

1. INTRODUCCIÓN	6-2
2. CONSTRUCCIÓN DE ÁRBOLES DE CLASIFICACIÓN	6-4
2.1 REGLAS DE DIVISIÓN	6-5
2.1.1 La ganancia de información	6-6
2.1.2 El criterio de proporción de ganancia	6-7
2.1.3 El índice de diversidad de Gini	6-8
2.1.4 MAX (© Juan Carlos Cubero)	6-9
2.1.5 MAXDIF	6-12
2.2 REGLAS DE PARADA	6-13
2.3 REGLAS DE PODA	6-14
2.3.1 Poda por estimación del error	6-15
2.3.2 Poda por coste-complejidad	6-15
2.3.3 Poda pesimista	6-16
2.4 TESTS	6-17
2.5 INFORMACIÓN INCOMPLETA	6-18
2.6 ALGORITMOS ESTÁNDAR	6-19
2.6.1 CLS	6-19
2.6.2 ID3	6-19
2.6.3 ACLS	6-20
2.6.4 ASSISTANT	6-20
2.6.5 CART	6-20
2.6.6 C4	6-20
2.6.7 ID4 & ID5	6-21
2.6.8 C4.5	6-21
2.6.9 SLIQ & SPRINT	6-21

3. <i>GENERACIÓN DE REGLAS</i>	6-22
3.1 <i>GENERALIZACIÓN DE REGLAS</i>	6-23
4. <i>RESULTADOS</i>	6-24
5. <i>BIBLIOGRAFÍA</i>	6-29

7. STAR: GENERACIÓN PROGRESIVA DE REGLAS

1. <i>INTRODUCCIÓN</i>	7-2
2. <i>ALGORITMOS STAR</i>	7-4
2.1 <i>INDUCE</i>	7-7
2.2 <i>AQ</i>	7-8
2.3 <i>CN2</i>	7-11
3. <i>RESULTADOS Y LIMITACIONES</i>	7-15
4. <i>BIBLIOGRAFÍA</i>	7-17

TERCERA PARTE: APRENDIZAJE NO SUPERVISADO

8. MÉTODOS DE AGRUPAMIENTO [*CLUSTERING*]

1. <i>INTRODUCCIÓN</i>	8-2
1.1 <i>Formulación matemática del problema</i>	8-3
2. <i>ALGORITMOS DE AGRUPAMIENTO</i>	8-4
2.1 <i>Agrupamiento mediante enfriamiento simulado</i>	8-6
2.2 <i>Método adaptativo</i>	8-8
2.3 <i>Algoritmo de Batchelor y Wilkins</i>	8-10
2.4 <i>Algoritmo de las K medias</i>	8-11
2.5 <i>Algoritmo GRASP</i>	8-12
2.6 <i>Algoritmo de agrupamiento secuencial</i>	8-17
2.7 <i>Algoritmo ISODATA</i>	8-18
2.8 <i>Métodos basados en grafos</i>	8-19
3. <i>BIBLIOGRAFÍA</i>	8-21

9. TÉCNICAS DE UMBRALIZACIÓN AUTOMÁTICA

1. INTRODUCCIÓN	9-2
2. MÉTODOS BÁSICOS DE UMBRALIZACIÓN AUTOMÁTICA	9-3
2.1 Media aritmética	9-3
2.2 Máximos [<i>two-peaks thresholding</i>]	9-4
2.3 Selección iterativa	9-5
2.4 El método de Otsu	9-6
2.5 Métodos basados en la entropía	9-7
2.6 Conjuntos difusos	9-9
2.7 Error mínimo	9-11
3. BIBLIOGRAFÍA	9-12

CUARTA PARTE: APÉNDICES

A. CLIS [*CALL-LEVEL INTERFACES*]

1. INTRODUCCIÓN	A-2
2. SQL [<i>STRUCTURED QUERY LANGUAGE</i>]	A-3
3. ODBC [<i>OPEN DATABASE CONNECTIVITY</i>] DE MICROSOFT	A-8
4. BDE [<i>BORLAND DATABASE ENGINE</i>] DE BORLAND	A-9
5. JDBC [<i>JAVA DATABASE CONNECTIVITY</i>]	A-10
6. ORACLE	A-17
6.1 PL/SQL	A-18
6.2 Pro*C (Embedded SQL)	A-22
7. ESTUDIO COMPARATIVO	A-23
7.1 Pruebas	A-25
7.2 Conclusiones	A-32
8. REFERENCIAS BIBLIOGRÁFICAS	A-33

B. JAVA

1. <i>ORIGEN DE JAVA</i>	B-2
2. <i>PROGRAMACIÓN EN JAVA</i>	B-3
2.1 <i>Internet & Java</i>	B-3
2.2 <i>Características de Java</i>	B-4
2.3 <i>Rendimiento</i>	B-8
2.4 <i>¿Originalidad?</i>	B-8
3. <i>EL LENGUAJE DE PROGRAMACIÓN JAVA</i>	B-9
4. <i>BIBLIOGRAFÍA</i>	B-22

C. CONTROLADORES JDBC

1. <i>INTRODUCCIÓN</i>	C-2
2. <i>ARQUITECTURA DE JDBC</i>	C-3
2.1 <i>La interfaz de JDBC para el programador de aplicaciones</i>	C-3
2.2 <i>La interfaz JDBC para los controladores JDBC</i>	C-5
3. <i>IMPLEMENTACIÓN DE CONTROLADORES JDBC</i>	C-6
4. <i>GUÍA RÁPIDA DE ACCESO A BASES DE DATOS A TRAVÉS DE JDBC</i>	C-9
5. <i>REFERENCIAS</i>	C-11

D. ACRÓNIMOS

E. BIBLIOGRAFÍA