

INTELIGENCIA DE NEGOCIO

2019 - 2020



- **Tema 1. Introducción a la Inteligencia de Negocio**
- **Tema 2. Minería de Datos. Ciencia de Datos**
- **Tema 3. Modelos de Predicción: Clasificación, regresión y series temporales**
- **Tema 4. Preparación de Datos**
- **Tema 5. Modelos de Agrupamiento o Segmentación**
- **Tema 6. Modelos de Asociación**
- **Tema 7. Modelos Avanzados de Minería de Datos.**
- **Tema 8. Big Data**

Modelos de asociación

Objetivos:

- Analizar el descubrimiento de asociaciones y las reglas de asociación
- Describir el algoritmo Apriori y explicar sus fases
- Definir medidas alternativas para el proceso de descubrimiento de asociaciones
- Introducir la extensión al descubrimiento de subgrupos.

Inteligencia de Negocio

TEMA 7. Modelos de Asociación

- 1. Descubrimiento de asociaciones**
2. Reglas de asociación
3. Algoritmo Apriori
4. Medidas de interés
5. Descubrimiento de Subgrupos: Caracterización de clases mediante asociaciones

1. Descubrimiento de asociaciones

- Descubrimiento de reglas de asociación:
 - Búsqueda de patrones frecuentes, asociaciones, correlaciones, o estructuras causales entre conjuntos de artículos u objetos (datos) a partir de bases de datos transaccionales, relacionales y otros conjuntos de datos
 - Búsqueda de secuencias o patrones temporales
 - Aplicaciones:
 - análisis de cestas de la compra (*Market Basket analysis*)
 - diseño de catálogos,...
 - ¿Qué hay en la cesta? Libros de Jazz
 - ¿Qué podría haber en la cesta? El último CD de Jazz
 - ¿Cómo motivar al cliente a comprar los artículos que es probable que le gusten?

1. Descubrimiento de asociaciones:

Market Basket Analysis

- ¿Qué es el análisis de cestas de la compra?
 - **Análisis de clientes:**
 - Se utiliza información sobre lo que ha comprado un cliente para ofrecernos una aproximación sobre quién es y por qué hace ciertas compras.
 - **Análisis de productos:**
 - Aporta información sobre qué productos tienden a ser comprados juntos.

1. Descubrimiento de asociaciones:

Market Basket Analysis

Compra: zumo de naranja, plátanos, detergente para vajillas, limpia cristales, gaseosa, ...

¿Cómo afecta la demografía de la vecindad a la compra de los clientes?

¿Es típico comprar gaseosa y plátanos? ¿Es importante la marca de la gaseosa?

¿Aumenta la compra del limpia cristales cuando se compran a la vez detergente para vajillas y zumo de naranja?

¿Dónde deberían colocarse los detergentes para maximizar sus ventas?



1. Descubrimiento de asociaciones:

Market Basket Analysis

Ejemplo: Asociación y Pañales y Cervezas

- Los clientes que compran cerveza también compran patatas
¡Para eso no es necesario el uso de técnicas de Minería de Datos!
- Los viernes por la tarde, con frecuencia, quienes compran pañales, compran también cerveza.

- ✓ ¿Qué significa?
- ✓ ¿A qué se debe?
- ✓ Acciones a realizar



1. Descubrimiento de asociaciones: *Market Basket Analysis*

Ejemplo: Compra conjunta de pañales y cerveza.
Ficción y leyenda para ilustrar el análisis de transacciones



Si compro **cerveza**, entonces compro **pañales**

60%

Si compro **pañales**, entonces compro **cerveza**

100% ✓

1. Descubrimiento de asociaciones:

Market Basket Analysis

Análisis de cestas de la compra (asociaciones)

Explicación más probable

- Se acerca el fin de semana
- Hay un bebé en casa
- No quedan pañales
- El padre/madre compra pañales al salir del trabajo
- ¡No pueden salir!
- Comprar cervezas para el fin de semana (y un partido/película PPV)

- Se acerca el fin de semana
- Hay un bebé en casa luego nada de ir fuera
- Hay que comprar pañales
- Quedarse en casa → ver partido/película
- Comprar cervezas para el partido/película

Pañales → Cerveza



Inteligencia de Negocio

TEMA 7. Modelos de Asociación

1. Descubrimiento de asociaciones
- 2. Reglas de asociación**
3. Algoritmo Apriori
4. Medidas de interés
5. Descubrimiento de Subgrupos: Caracterización de clases mediante asociaciones

2. Reglas de Asociación

- Se ha desarrollado una gran cantidad de investigación en torno al área de análisis de cestas de la compra, debido a la claridad y utilidad de sus resultados, que se expresan en forma de ***reglas de asociación***
- Objetivo de los algoritmos de extracción de reglas de asociación:
 - Dada una **base de datos de transacciones**, donde cada transacción es una lista de artículos (comprados por un cliente en la misma visita)
 - Encontrar **todas las reglas** que co-relacionen la presencia de un conjunto de artículos con otro conjunto de artículos.
 - Ejemplo: *98% de la gente que compra neumáticos y accesorios para el automóvil, también adquiere servicios (cambio de neumáticos, ...)*

2. Reglas de Asociación: Conceptos básicos

- La idea es obtener reglas del tipo:
 - “Antecedente => Consecuente [soporte, confianza]”
 - compra(x, “pañales”) => compra(x, “cervezas”) [0.5%, 60%]

2. Reglas de Asociación: Conceptos básicos

Transacción:

Formato relacional

<Tid, item>

<1, item1>

<1, item2>

<2, item3>

Formato compacto

<Tid, itemset>

<1, {item1,item2}>

<2, {item3}>

- Item (o artículo) : elemento individual
- Itemset (o conjunto): conjunto de items/artículos
- Soporte de un conjunto I: n° de transacciones conteniendo I
- Soporte mínimo m_s : umbral de soporte
- Conjunto frecuente: con soporte $\geq m_s$

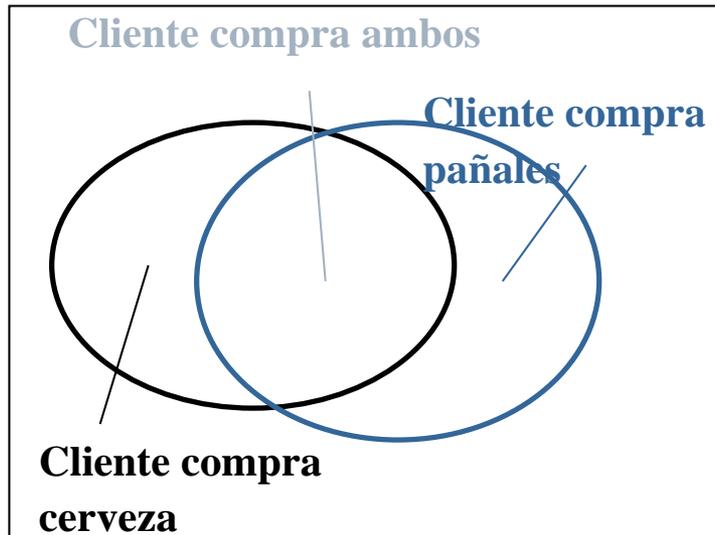
Los conjuntos frecuentes representan conjuntos de artículos que están correlacionados positivamente

2. Reglas de Asociación:

Distintos tipos de reglas de asociación

- Asociaciones **Booleanas vs Cuantitativas** dependiendo del tipo de los valores que se manejan)
 - compra (x, "SQLServer") ^ compra (x, "Libro de MD") → compra (x, "DBMiner") [0.2%, 60%]
 - Edad (x, '30..39') ^ ingresos (x, '42..48K') → compra (x, 'PC') [1%,75%]
- Asociaciones **unidimensionales vs. Multidimensionales**
A -> B A & B & ... & N -> D
- Análisis con distintos niveles de abstracción:
 - Edad (x, '30..39') → compra (x, cerveza)
 - Edad (x, '30..39') → compra (x, cerveza alemana)
- Posibles Extensiones:
 - Correlaciones, análisis de causalidad
 - Asociación no implica necesariamente correlación o causalidad

2. Reglas de Asociación: Medidas de soporte y confianza



- Encontrar todas las reglas $X \& Y \Rightarrow Z$ con un mínimo de confianza y soporte
 - **Soporte** (s): probabilidad de que una transacción contenga $\{X \& Y \& Z\}$
 - **Confianza** (c): probabilidad condicional $P(Z|X\&Y)$

# transacción	artículos
2000	A,B,C
1000	A,C
4000	A,D
5000	B,E,F

Sea el valor mínimo para confianza y soporte 50%,

$A \Rightarrow C$ (50%, 66.6%)

$C \Rightarrow A$ (50%, 100%)

2. Reglas de Asociación: Proceso de extracción

- Al tratar con bases de datos grandes, el proceso se descompone en **dos pasos**:
 - ***Encontrar conjuntos de artículos frecuentes***
 - Mayor ocurrencia que el soporte mínimo fijado.
 - ***Generar reglas de asociación "fuerte" a partir de los conjuntos de artículos frecuentes***
 - Deben satisfacer el mínimo fijado tanto para soporte como para confianza.

2. Reglas de Asociación:

Proceso de extracción. Ejemplo

Transaction ID	Items Bought
2000	A,B,C
1000	A,C
4000	A,D
5000	B,E,F

Min. soporte 50%
Min. confianza 50%

Frequent Itemset	Support
{A}	75%
{B}	50%
{C}	50%
{A,C}	50%

Para la regla $A \rightarrow C$:

soporte = soporte($\{ A \&C \}$) = 50%

confianza = soporte($\{ A \&C \}$)/soporte($\{ A \}$) = 66.6%

2. Reglas de Asociación:

Proceso de extracción. Ejemplo

Transaction	Items
t_1	Bread,Jelly,PeanutButter
t_2	Bread,PeanutButter
t_3	Bread,Milk,PeanutButter
t_4	Beer,Bread
t_5	Beer,Milk

$I = \{\text{Beer, Bread, Jelly, Milk, PeanutButter}\}$

Soporte de $\{\text{Bread,PeanutButter}\}$ es 60%

2. Reglas de Asociación: Proceso de extracción. Ejemplo

$X \Rightarrow Y$	s	α
Bread \Rightarrow PeanutButter	60%	75%
PeanutButter \Rightarrow Bread	60%	100%
Beer \Rightarrow Bread	20%	50%
PeanutButter \Rightarrow Jelly	20%	33.3%
Jelly \Rightarrow PeanutButter	20%	100%
Jelly \Rightarrow Milk	0%	0%

Inteligencia de Negocio

TEMA 7. Modelos de Asociación

1. Descubrimiento de asociaciones
2. Reglas de asociación
3. **Algoritmo Apriori**
4. Medidas de interés
5. Descubrimiento de Subgrupos: Caracterización de clases mediante asociaciones

3. El algoritmo APRIORI

- Encuentra las asociaciones más frecuentes
- Itera sobre la base de datos hasta que las asociaciones obtenidas no tienen el soporte mínimo
- Método simple pero robusto
- Salida intuitiva

- Requisitos:
 - No necesita decir los atributos de los lados derecho (consecuente) e izquierdo (antecedente) de las reglas pues se generan de manera automática.
 - Existen variedades para tratar todo tipo de datos.
 - Especificar mínimo soporte.
 - Especificar máximo número de reglas.

3. El algoritmo APRIORI

- El algoritmo busca **iterativamente** conjuntos frecuentes con cardinalidad 1 hasta k (k-conjunto), y después
- Usa los conjuntos frecuentes para generar las reglas de asociación
- En el paso clave del descubrimiento de **conjuntos frecuentes**, se basa en el principio "Apriori" :
 - **Cualquier subconjunto de un conjunto de artículos frecuente debe ser frecuente**

Ejemplo: si $\{AB\}$ es un conjunto frecuente, entonces tanto $\{A\}$ y $\{B\}$ deberían ser frecuentes
- Esto permite definir el **principio de poda** en Apriori: Si existe **algún** conjunto "infrecuente", entonces no hay necesidad de generar sus superconjuntos

3. El algoritmo APRIORI

- **Unión:** C_k es generado uniendo conjuntos de L_{k-1} (se asume orden lexicográfico en las transacciones y que los prefijos son comunes)
- **Poda:** cualquier $(k-1)$ -conjunto que no es frecuente, no puede ser un subconjunto de un k -conjunto frecuente

- Pseudo-código:

C_k : conjunto candidato de cardinalidad k

L_k : conjunto frecuente de cardinalidad k

$L_1 = \{\text{artículos frecuentes}\};$

for ($k = 1; L_k \neq \emptyset; k++$) **do begin**

C_{k+1} = candidatos generados desde L_k ;

for each transacción t en la base de datos **do**

 incrementar el contador de todos los candidatos en C_{k+1}
 que están contenidos en t

L_{k+1} = candidatos en C_{k+1} con `min_support`

end

return $\cup_k L_k$;

3. El algoritmo APRIORI

Generación de candidatos

- Suponemos que los items de L_{k-1} están ordenados
- Paso 1: Unión sobre L_{k-1}
 - Insertar en C_k
 - Seleccionar $p.item1, p.item2, \dots, p.item_{k-1}, q.item_{k-1}$
 - Partiendo de $L_{k-1} p$ y $L_{k-1} q$
 - Donde $p.item1=q.item1, \dots, p.item_{k-2}=q.item_{k-2}$ y $p.item_{k-1} < q.item_{k-1}$
- Paso 2: Poda
 - Para todos los *itemsets* c de C_k hacer
 - Para todos los $(k-1)$ -subconjuntos de c hacer
 - Si (s no pertenece a L_{k-1}) entonces eliminar c de C_k

3. El algoritmo APRIORI

Ejemplo de generación de candidatos

- $L_3 = \{abc, abd, acd, ace, bcd\}$
- Unión en L_3 : $L_3 * L_3$
 - $abcd$ de $ab-c$ y $ab-d$
 - $acde$ de $ac-d$ y $ac-e$
- Poda:
 - $acde$ es eliminado porque ade no está en L_3
- $C_4 = \{abcd\}$

3. El algoritmo APRIORI. Ejemplo

Database D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Scan D

itemset	sup.
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3

L_1

itemset	sup.
{1}	2
{2}	3
{3}	3
{5}	3

C_2

itemset	sup.
{1 2}	1
{1 3}	2
{1 5}	1
{2 3}	2
{2 5}	3
{3 5}	2

C_2

itemset
{1 2}
{1 3}
{1 5}
{2 3}
{2 5}
{3 5}

Scan D

L_2

itemset	sup.
{1 3}	2
{2 3}	2
{2 5}	3
{3 5}	2

C_3

itemset
{2 3 5}

Scan D

L_3

itemset	sup.
{2 3 5}	2

3. El algoritmo APRIORI

Generación de reglas a partir de conjuntos frecuentes

- Una vez disponemos de los conjuntos frecuentes basta calcular la confianza y añadir las reglas que cumplan con los umbrales mínimos.

Base de datos

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Conjunto frecuente

itemset	sup
{2 3 5}	2

Si los valores mínimos son 50%, entonces:

2,3 -> 5 [50,100]

2,5 -> 3 [50,66]

3,5 -> 2 [50,100]

2 -> 3,5 [50,66]

3 -> 2,5 [50,66]

5 -> 2,3 [50,66]

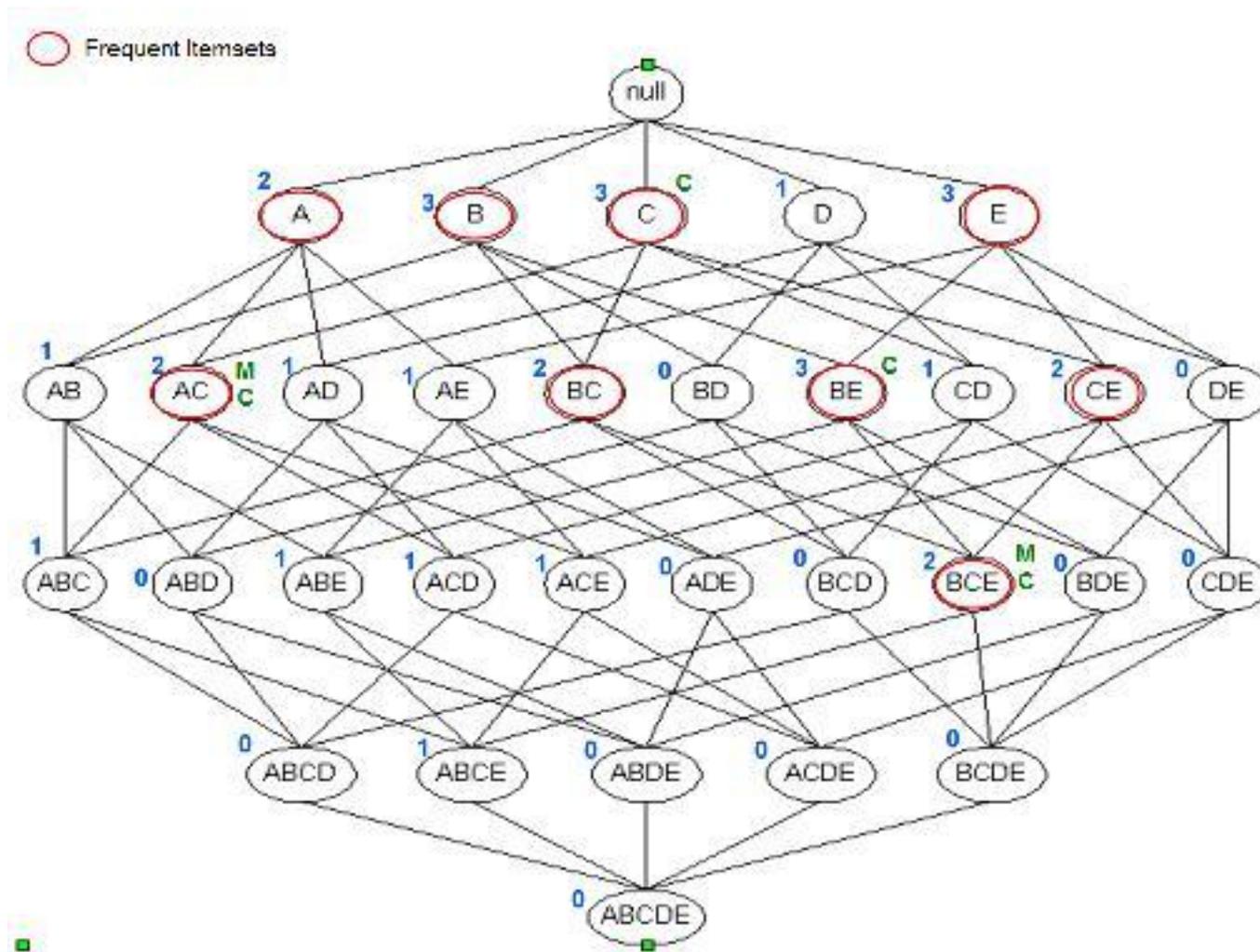
3. El algoritmo APRIORI

Parámetros soporte y confianza

- **Soporte mínimo (s):**
 - Alto \Rightarrow **pocos** conjuntos frecuentes
 - \Rightarrow **pocas** reglas válidas que ocurren **con frecuencia**
 - bajo \Rightarrow **muchas** reglas válidas que ocurren **raramente**
- **Confianza mínima (c):**
 - Alta \Rightarrow **pocas** reglas, pero todas "**casi ciertas lógicamente**"
 - Baja \Rightarrow **muchas** reglas, pero muchas de ellas muy **inciertas**
- **Valores típicos:** **soporte** = 2-10 % **confianza** = 70-90 %

3. El algoritmo APRIORI

Espacio de búsqueda de reglas



Inteligencia de Negocio

TEMA 7. Modelos de Asociación

1. Descubrimiento de asociaciones
2. Reglas de asociación
3. Algoritmo Apriori
- 4. Medidas de interés**
5. Descubrimiento de Subgrupos: Caracterización de clases mediante asociaciones

4. Medidas de interés

- Medidas objetivas (dos de las más populares):
 - *soporte* y
 - *confianza*
- Medidas subjetivas:
 - Una regla (patrón) es interesante si
 - es *inesperada* (sorprendente para el usuario) y/o
 - *útil* (el usuario puede hacer algo con ella)

4. Medidas de interés

Críticas a Soporte y Confianza

	basketball	not basketball	sum(row)
cereal	2000	1750	3750
not cereal	1000	250	1250
sum(col.)	3000	2000	5000

- Ejemplo 1:
 - Entre 5000 alumnos
 - 3000 juegan baloncesto
 - 3750 desayunan cereales
 - 2000 juegan al baloncesto y comen cereales
 - *juega baloncesto* \Rightarrow *come cereales* [40%, 66.7%] es errónea porque el % global de alumnos que comen cereales es 75%, que es mayor que 66.7%
 - *juega baloncesto* \Rightarrow *no come cereales* [20%, 33.3%] es mucho más exacta, aunque con menor soporte y confianza

4. Medidas de interés

Críticas a Soporte y Confianza

X	1	1	1	1	0	0	0	0
Y	1	1	0	0	0	0	0	0
Z	0	1	1	1	1	1	1	1

Rule	Support	Confidence
X=>Y	25%	50%
X=>Z	37.50%	75%

- Ejemplo 2:
 - X & Y: + correlacionadas
 - X & Z: - relacionadas
 - Soporte y confianza de X=>Z domina
- Se necesita una medida de dependencia o sucesos correlacionados
- $P(B|A)/P(B)$ se conoce como el **empuje** de la regla $A \Rightarrow B$

4. Medidas de interés

- **Interés** (correlación, empuje)

- $P(B|A)/P(B) = P(B \wedge A)/P(B)*P(A)$
- Toma $P(A)$ y $P(B)$ en consideración
- $P(A \text{ y } B)=P(B)*P(A)$, si A y B son independientes
- A y B neg. correlacionadas, si el valor es menor que 1; en otro caso A y B positivamente correlacionadas

X	1	1	1	1	0	0	0	0
Y	1	1	0	0	0	0	0	0
Z	0	1	1	1	1	1	1	1

Itemset	Support	Interest
X,Y	25%	2
X,Z	37.50%	0.9
Y,Z	12.50%	0.57

Otros algoritmos

- FP-Growth (FP tree based model)

FP-Growth: allows frequent itemset discovery without candidate itemset generation. Two step approach:

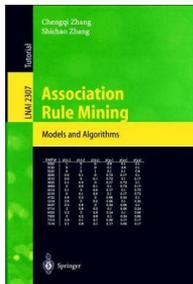
Step 1: Build a compact data structure called the FP-tree
Built using 2 passes over the data-set.

Step 2: Extracts frequent itemsets directly from the FP-tree

- OPUS search based algorithms

- Negative association rules

- Quantitative association rules (QAR)



Chengqi Zhang, Shichao Zhang. **Association rule mining : models and algorithms.** Springer, 2002.

Ejemplo: Análisis de asociaciones en transacciones de tarjetas de crédito



Ejemplo: Asociaciones en Salud



Ejemplo: Análisis de cesta de compra

 TARGET


Disculpas Empresa.
Disculpas del padre,
confirmación del
embarazo de la hija

Modelo de predicción de
clientes embarazadas por
medio de sus patrones de
compra.


Enfado de un padre: Su Hija
recibe publicidad de
productos para embarazadas


Acción: Envío de
cupones para cada
fase del embarazo

Descubrimiento: Cremas
sin perfume al tercer mes.
Dos docenas de productos
Predicción de fecha parto



Inteligencia de Negocio

TEMA 7. Modelos de Asociación

1. Descubrimiento de asociaciones
2. Reglas de asociación
3. Algoritmo A priori
4. Medidas de interés
5. **Descubrimiento de Subgrupos:
Caracterización de clases mediante asociaciones**

5. Descubrimiento de subgrupos

Supervised vs. unsupervised learning: A rule learning perspective

- **Supervised learning:** Rules are induced from labeled instances (training examples with class assignment) - usually used in **predictive induction**
- **Unsupervised learning:** Rules are induced from unlabeled instances (training examples with no class assignment) - usually used in **descriptive induction**

Exception: Subgroup discovery

Discovers **individual rules** describing interesting regularities in the data induced from **labeled** examples

5. Descubrimiento de subgrupos

Subgroup discovery is a process to identify relations between a dependent variable (target variable) and usually many explaining, independent variables.

For example, consider the subgroup described by "smoker=true AND family history=positive" for the target variable coronary heart disease=true.

Subgroup discovery does not necessarily focus on finding complete relations; instead partial relations, i.e., (small) subgroups with "interesting" characteristics can be sufficient.

5. Descubrimiento de subgrupos

W. Klösgen , 1996:

“Given a population of individuals and a property of those individuals we are interested in, find population subgroups that are statistically ‘most interesting’, e.g., are as large as possible and have the most unusual statistical characteristics with respect to the property of interest”.

W. Klösgen, Explora: A multipattern and multistrategy discovery assistant, Advance in Knowledge Discovery and Data Mining, MIT Press, 249-271, 1996.

5. Descubrimiento de subgrupos

Find subgroups of members of a population that exhibit interesting deviations from overall population behavior

Interestingness criteria:

- As large as possible
- Class distribution as different as possible from the distribution in the entire data set
- Significant
- Surprising to the user
- Non-redundant
- Simple
- Useful - actionable

5. Descubrimiento de subgrupos

- Assume 100 % of all calls made to Australia from a mobile phone at night are fraudulent (total of 0.01% of all calls)
- but fraudulence does not otherwise depend on country, phone or time

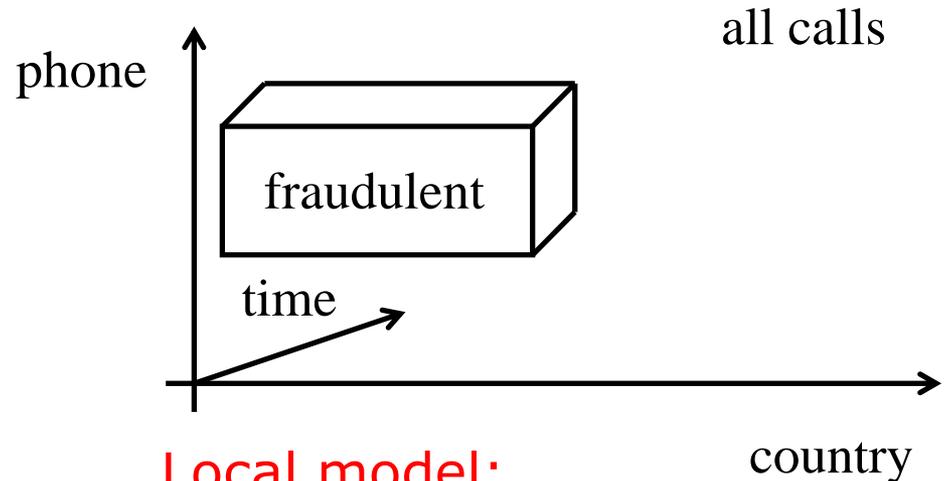


Global model:

decision tree

Constructed with C4.5

(Simply always answers majority class!)



Local model:

Subgroup discovery

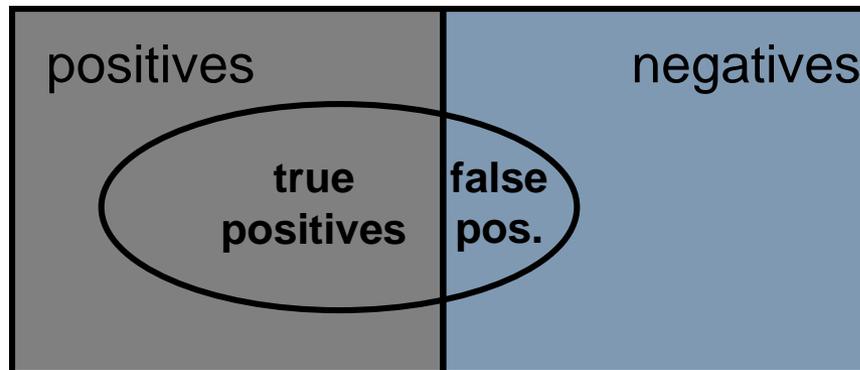


How to find *comprehensible* subgroups in large amounts of data for fraudulent class?

5. Descubrimiento de subgrupos

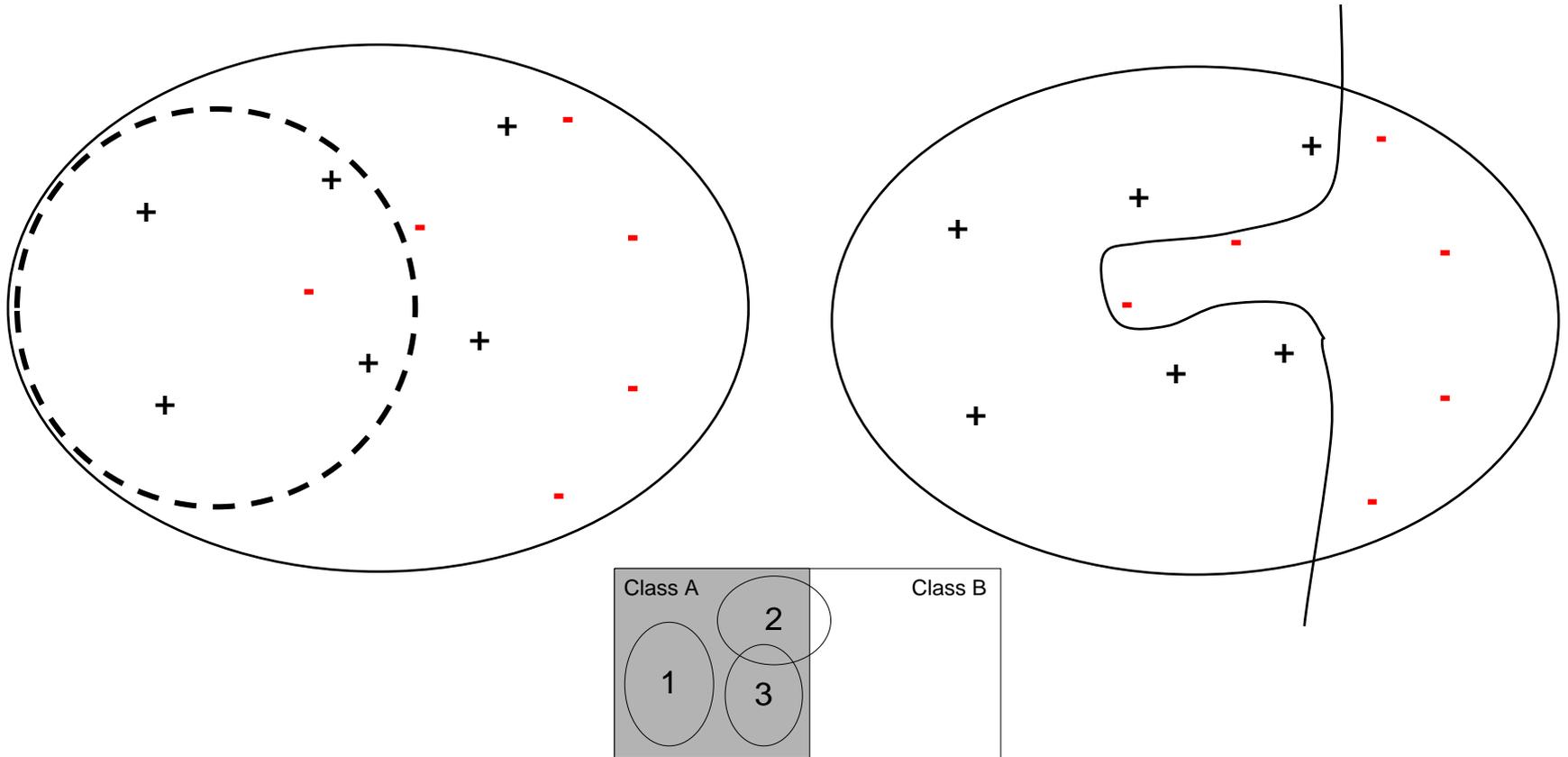
Subgroups vs. classifiers

- Classification rules aim at pure subgroups
 - A set of rules forms a domain model – high prediction algorithm
- Each rule is an **independent chunk of knowledge**
 - Subgroups aim at significantly higher (or different) proportion of positives



5. Descubrimiento de subgrupos

Subgroups vs. classifiers



5. Descubrimiento de subgrupos

Subgroups vs. classifiers

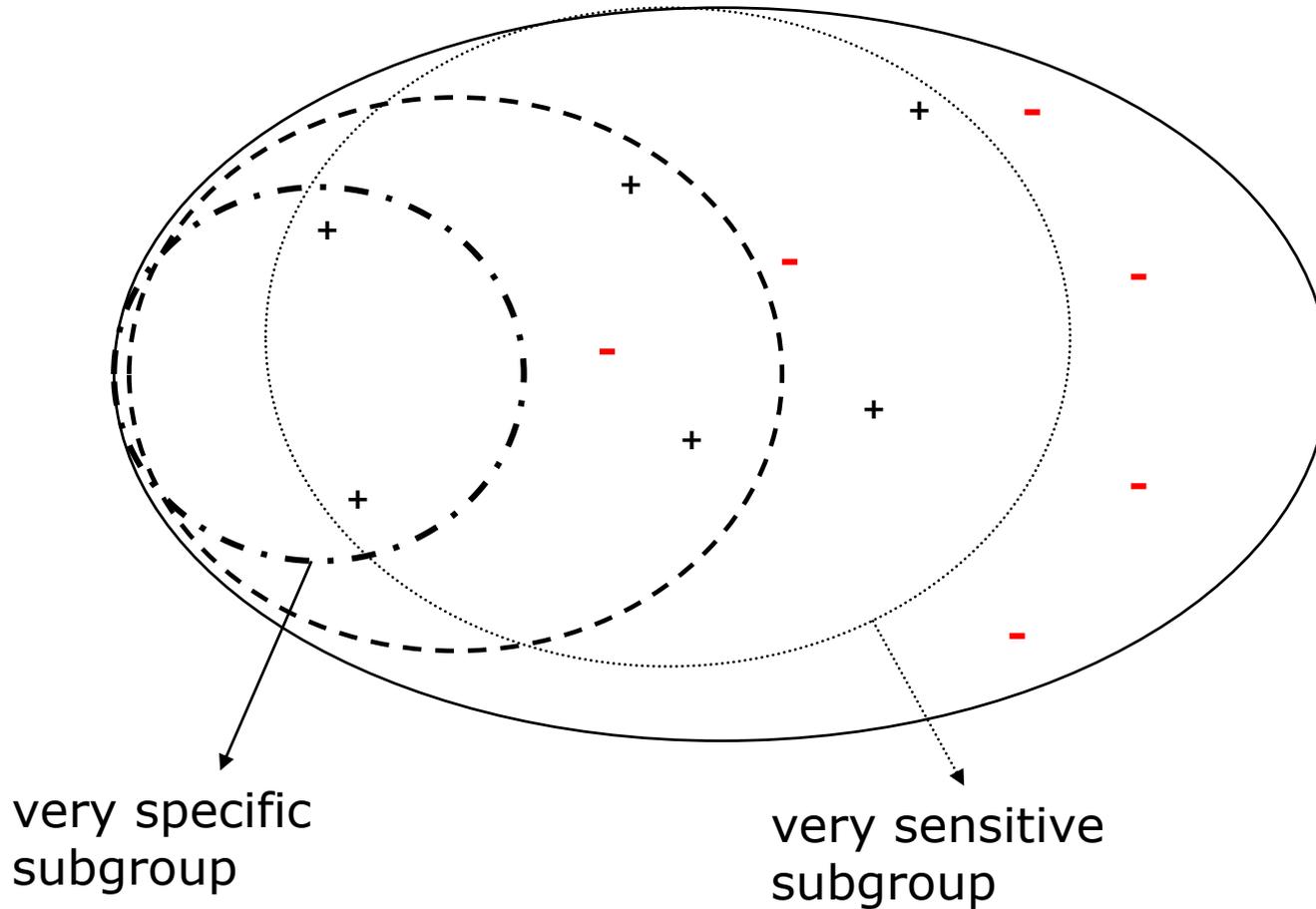
■ Classification

- predictive induction
- constructing sets of classification rules
- aimed at learning a model for classification or prediction
- rules are dependent

■ Subgroup Discovery

- descriptive induction
- constructing individual subgroup-describing rules
- aimed at finding interesting patterns in target class examples

5. Descubrimiento de subgrupos



generality – the main parameter of the subgroup induction process

5. Descubrimiento de subgrupos

Probabilistic distribution and model extraction

Each *rule* returns a **probability distribution**, instead of class distribution in terms of the number of examples covered.

Weighted Relative Accuracy: Weighted WRAcc Search Heuristic, with added example weights

Balance between coverage of the rule and accuracy gain

5. Descubrimiento de subgrupos

Probabilistic distribution and model extraction

Weighted Relative Accuracy

$$\begin{aligned} &WRAcc(\text{Class} \leftarrow \text{Condition}) \\ &= p(\text{Condition})[p(\text{Class} \mid \text{Condition}) - p(\text{Class})] \end{aligned}$$

$$WRAcc(\text{Cond} \rightarrow \text{Class}) = \frac{n'(\text{Cond})}{N'} \cdot \left(\frac{n'(\text{Cond}, \text{Class})}{n'(\text{Cond})} - \frac{n'(\text{Class})}{N'} \right)$$

increased coverage, decreased # of rules

Balance between coverage of the rule and accuracy gain

N' : sum of weights of examples

$n'(\text{Cond})$: sum of weights of all covered examples

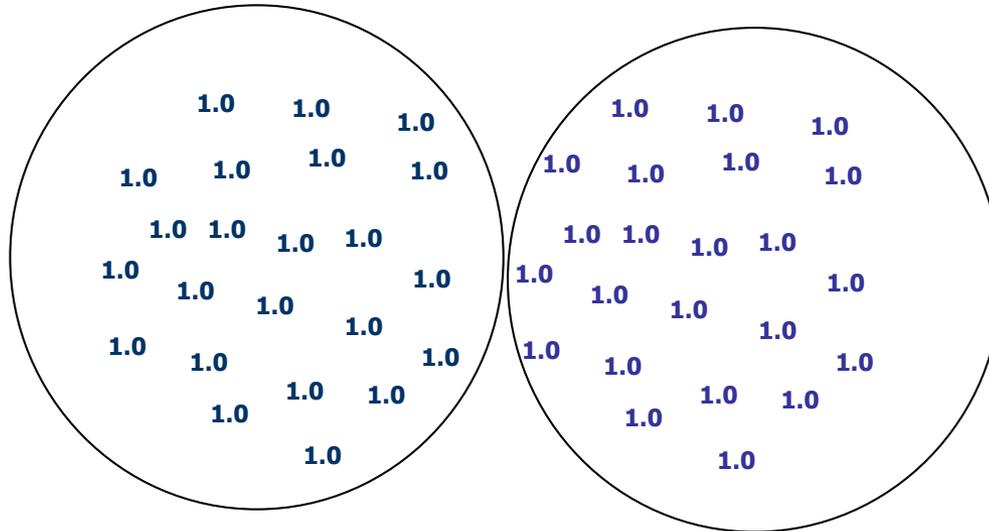
$n'(\text{Cl. Cond})$: sum of weights of all correctly covered examples

5. Descubrimiento de subgrupos

Weighted covering algorithm for rule set construction

Disease patients

other patients



For learning a set of subgroup describing rules, SD implements an iterative weighted covering algorithm.

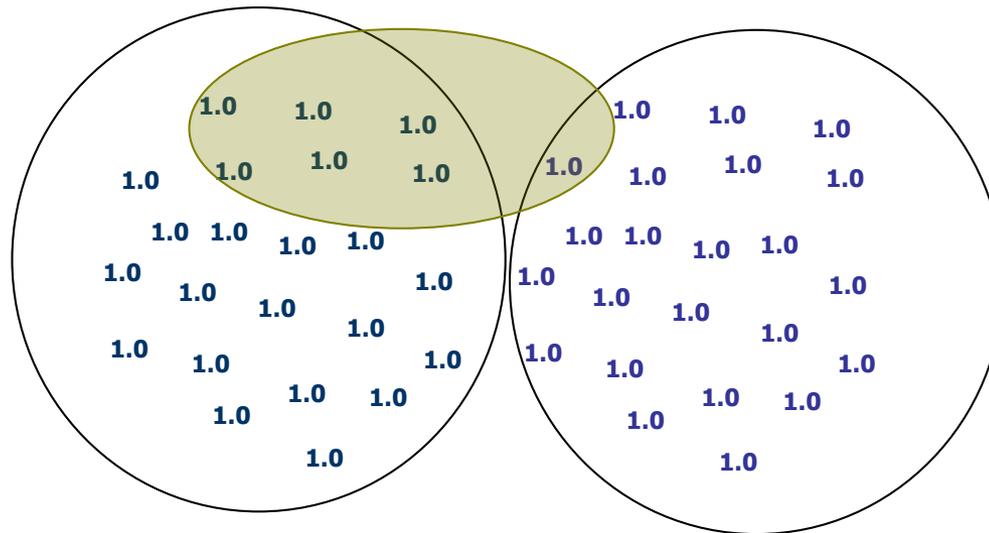
Quality of a rule is measured by trading-off coverage and precision.

5. Descubrimiento de subgrupos

Weighted covering algorithm for rule set construction

Disease patients

other patients

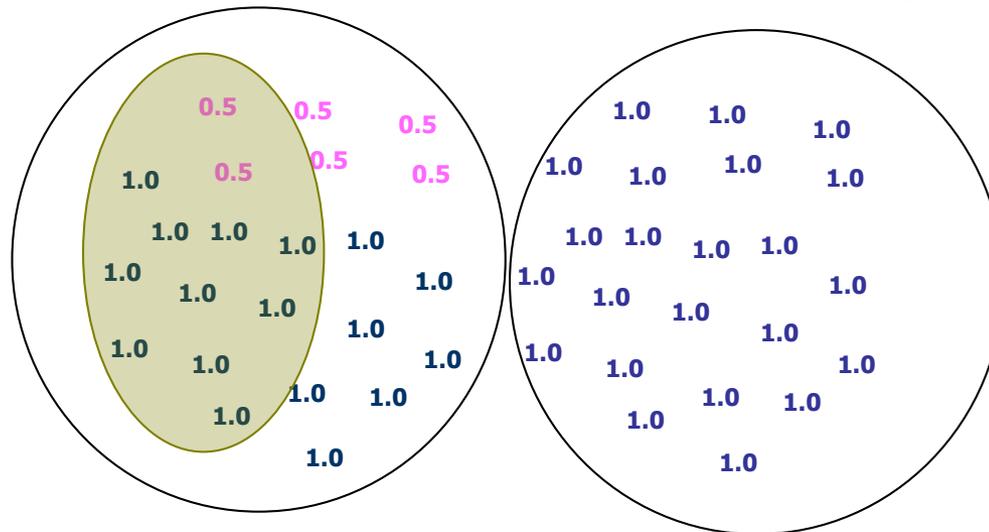


5. Descubrimiento de subgrupos

Weighted covering algorithm for rule set construction

Disease patients

other patients



In contrast with classification rule learning algorithms (e.g. CN2), the covered positive examples are not deleted from the training set in the next rule learning iteration; they are re-weighted, and a next 'best' rule is learned.

5. Descubrimiento de subgrupos

Rule extraction and Post-process:

Rule subset selection by a weighted covering approach

Create a set of rules

Take the best rule w.r.t. WRAcc

Decrease the weights of covered examples

Reorder the remaining rules and repeat until stopping
criterion is satisfied

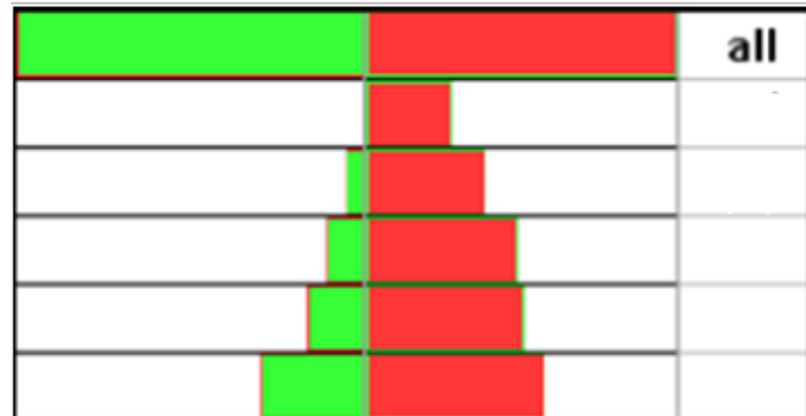
significance threshold

WRAcc threshold

5. Descubrimiento de subgrupos

Interestingness criteria:

- As large as possible
- Class distribution as different as possible from the distribution in the entire data set
- Significant
- Surprising to the user
- Non-redundant
- Simple
- Useful



5. Descubrimiento de subgrupos

Fig. 1. Pseudocode of Apriori-SD algorithm

1. algorithm *APRIORI – SD*(*Examples, Classes, minSup, minConf, k*)
 2. *Ruleset* = *APRIORI – C*(*Examples, Classes, minSup, minConf*)
set all example weights of *Examples* to 1)
 3. *Majority* = the majority class in *Examples*
 4. *Resultset* = {}
 5. Repeat
 6. *BestRule* = rule with the highest weighted relative accuracy in *Ruleset*.
 7. *Resultset* = *Resultset* \cup *BestRule*
 8. *Ruleset* = *Ruleset* \ decrease the weights of examples covered by *BestRule* remove from *Examples* the examples covered more than *k*-times
 9. until *Examples* = {} or *Ruleset* = {}
 10. return *Resultset* = *Resultset* \cup true \rightarrow *Majority*
-

parameter *k* determines the threshold for covered example elimination in rule post-processing ensuring the convergence of the algorithm

Kavsek, B., Lavrac, N., APRIORI-SD: Adapting association rule learning to subgroup discovery. *Applied Artificial Intelligence*, 20(7) (2006) 543-583.

5. Descubrimiento de subgrupos

Association Rule Learning for Subgroup Discovery: Advantages

- May be used to create rules of the form
Class ← Conditions
- Each rule is an independent “chunk” of knowledge, with
 - high support and coverage
($p(\text{Class}.\text{Cond}) > \text{MinSup}$, $p(\text{Cond}) > \text{MinSup}$)
 - high confidence $p(\text{Class}|\text{Cond}) > \text{MinConf}$
 - all interesting rules found (complete search)
- Building small and easy-to-understand classifiers
- Appropriate for unbalanced class distributions

5. Descubrimiento de subgrupos

Ejemplo de Aplicación:

Bank business. Applications for a loan from the bank

A mortgage loan is a very common type of debt instrument, used by many individuals to purchase housing.

Quality data control can provide direct services to customer.

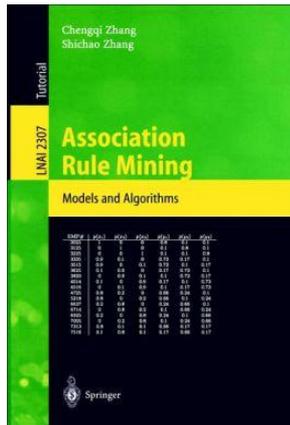


IF "bank account for the payroll" = YES
and Base SALARY \geq 2500 and Profession = "teacher"
and Variable-salary \leq 300
THEN low risk (run automatic credit process)

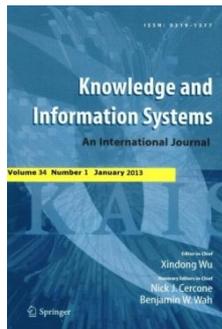
IF Variable-salary \geq 1000
and Profession = "working-class without qualification"
THEN high risk (no run automatic credit process)

To control the defaulter risk

Bibliografía



Chengqi Zhang, Shichao Zhang. **Association rule mining : models and algorithms.** Springer, 2002.



F. Herrera, C.J. Carmona, P. González and M.J. del Jesus, **An overview on Subgroup Discovery: Foundations and Applications** . *Knowledge and Information Systems 29:3 (2011) 495-525*, [doi: 10.1007/s10115-010-0356-2](https://doi.org/10.1007/s10115-010-0356-2).



M. Atzmueller. **Subgroup discovery.** *WIREs Data Mining Knowl Discov 2015, 5:35–49*

INTELIGENCIA DE NEGOCIO

2019 - 2020



- **Tema 1. Introducción a la Inteligencia de Negocio**
- **Tema 2. Minería de Datos. Ciencia de Datos**
- **Tema 3. Modelos de Predicción: Clasificación, regresión y series temporales**
- **Tema 4. Preparación de Datos**
- **Tema 5. Modelos de Agrupamiento o Segmentación**
- **Tema 6. Modelos de Asociación**
- **Tema 7. Modelos Avanzados de Minería de Datos**
- **Tema 8. Big Data**