

Data Mining - Corso di Laurea Specialistica in  
Informatica per l'Economia e l'Azienda

Verifica 3 aprile 2008

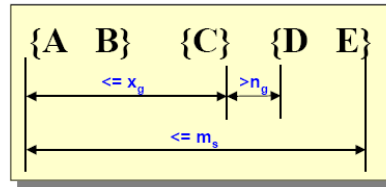
Esercizio 1 - Sequential Patterns (8 punti)

Si consideri la seguente sequenza  $W$  di input:

$$W = \langle \{a,b\} \{a,c\} \{c\} \{d,e\} \{a\} \{b,d\} \rangle$$

Si indichi quali delle seguenti sequenze sono sotto-sequenze semplici di  $W$  (senza vincoli temporali), quali rispettano il vincolo min-gap=1, quali il vincolo max-gap=2, quali il vincolo max-span=3 e quali li rispettano tutti insieme:

- $w_1 = \langle \{a\} \{e\} \{a\} \rangle$
- $w_2 = \langle \{a,b\} \{c\} \{d\} \rangle$
- $w_3 = \langle \{a\} \{e\} \{d\} \rangle$
- $w_4 = \langle \{a\} \{d\} \{c\} \rangle$
- $w_5 = \langle \{a\} \{b\} \rangle$



$x_g$ : max-gap  
 $n_g$ : min-gap  
 $m_s$ : maximum span

A tale scopo si utilizzi la seguente tabella:

	$w_i \leq W$	min-gap=1	max-gap=2	max-span=3	Tutti
$w_1 = \langle \{a\} \{e\} \{a\} \rangle$					
$w_2 = \langle \{a,b\} \{c\} \{d\} \rangle$					
$w_3 = \langle \{a\} \{e\} \{d\} \rangle$					
$w_4 = \langle \{a\} \{d\} \{c\} \rangle$					
$w_5 = \langle \{a\} \{b\} \rangle$					

Esercizio 2 – Regole associative (4 punti)

Si considerino le seguenti regole associative e relativo supporto e confidenza:

1	{beer} => {chips}	50%	100%
2	{beer} => {wine}	25%	50%
3	{chips} => {beer}	50%	66%
4	{pizza} => {chips}	25%	50%
5	{pizza} => {wine}	25%	50%
6	{wine} => {beer}	25%	50%
7	{wine} => {chips}	25%	50%
8	{wine} => {pizza}	25%	50%
9	{beer, chips} => {wine}	25%	50%

10	{beer, wine} => {chips} 25% 100%
11	{chips, wine} => {beer} 25% 100%
12	{beer} => {chips, wine} 25% 50%
13	{wine} => {beer, chips} 25% 50%

1. Determinare le regole valide rispetto al Min\_Support = 25%. Min\_conf= 60%
2. Determinare le regole valide rispetto al vincolo: "beer" è contenuto nell'antecedente. E' possibile utilizzare tale vincolo nella fase di generazione dei pattern frequenti?
3. Determinare le regole valide rispetto alle seguenti meta-regole:

a	W and X => Y and Z
b	X => Y and Z
c	X => Y
d	W and X => Y

### Esercizio 3 - Pattern Frequenti (12 punti)

---

Si consideri il seguente dataset:

	A	B	C	D	E
1	1	0	1	1	0
2	1	1	0	1	1
3	1	0	1	1	0
4	0	1	0	1	0
5	1	0	1	0	1
6	0	1	1	1	0

Assumendo un supporto del 50% e una confidenza dell'80%,

1. Trasformare il dataset in formato transazionale e applicare l'algoritmo Apriori per il calcolo degli itemset frequenti mostrando le varie iterazioni
2. Calcolare le regole associative a partire dagli itemset frequenti calcolati nel punto 1.
3. Calcolare gli itemset frequenti multidimensionali sul dataset originario.
4. Costruire la matrice di contingenza della regola con confidenza più alta e calcolare il lift.

### Esercizio 4 - Indici di similarità (8 punti)

---

Si consideri il seguente insieme di transazioni:

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

1. Si calcoli la matrice di similarità utilizzando l'indice di Jaccard.
2. Si calcoli la medesima matrice utilizzando l'indice Simple Matching (Nota: il dataset contiene solo 5 items distinti: Beer, Bread, Coke, Diaper, Milk).