CI603 Data Mining

Tutorial 3

(solution)

1. Consider the data set shown below:

    a.

    $$s(\{e\}) = \frac{7}{10} = 0.7$$

    $$s(\{b, d\}) = \frac{2}{10} = 0.2$$

    $$s(\{b, d, e\}) = \frac{2}{10} = 0.2$$

    b.

    $$c(\{b, d\} \rightarrow \{e\}) = \frac{2}{2} = 1 = 100\,\%$$

    $$c(\{e\} \rightarrow \{b, d\}) = \frac{2}{7} = 0.29 = 29\,\%$$

    It quite obvious that **confidence** is **not a symmetric measure**.

    c.

    $$s(\{e\}) = \frac{4}{5} = 0.8$$

    $$s(\{b, d\}) = \frac{5}{5} = 1$$

    $$s(\{b, d, e\}) = \frac{4}{5} = 0.8$$

Customer ID	Transaction ID	Items Bought
1	0001	{a, b, c, d, e}
2	0002	{a, b, c, d, e}
3	0003	{b, c, d, e}
4	0004	{a, b, c, d}
5	0005	{a, b, c, d, e}

    d.

    $$c(\{b, d\} \rightarrow \{e\}) = \frac{4}{5} = 0.8 = 80\,\%$$

    $$c(\{e\} \rightarrow \{b, d\}) = \frac{4}{4} = 1 = 100\,\%$$

2. Consider the data set shown below

a.

itemset	Support Count	Support
{Bread}	7	7/9
{Butter}	2	2/9
{Cheese}	2	2/9
{Jam}	6	6/9
{Milk}	6	6/9

b.

minsup ≥ 2

Frequent 1-itemset

itemset	Support Count
{Bread}	7
{Butter}	2
{Cheese}	2
{Jam}	6
{Milk}	6

Merge K-1 to generate **2-itemset**

itemset
{Bread}
{Butter}
{Cheese}
{Jam}
{Milk}

itemset
{Bread}
{Butter}
{Cheese}
{Jam}
{Milk}

2-itemset candidates

itemset	Support Count
{Bread, Butter}	2
{Bread, Cheese}	2
{Bread, Jam}	4
{Bread, Milk}	4
{Butter, Cheese}	0
{Butter, Jam}	1
{Butter, Milk}	2
{Cheese, Jam}	0
{Cheese, Milk}	1
{Jam, Milk}	4

Frequent 2-itemset

itemset	Support Count
{Bread, Butter}	2
{Bread, Cheese}	2
{Bread, Jam}	4
{Bread, Milk}	4
{Butter, Milk}	2
{Jam, Milk}	4

3-itemset candidates

itemset	Support Count
{Bread, Butter, Cheese}	0
{Bread, Butter, Jam}	1
{Bread, Butter, Milk}	2
{Bread, Cheese, Jam}	0
{Bread, Cheese, Milk}	1
{Bread, Jam, Milk}	2

Frequent 3-itemset

itemset	Support Count
{Bread, Butter, Milk}	2
{Bread, Jam, Milk}	2

c.

Rules from {Bread, Butter, Milk} with consequents that contain only 1-itemset:

1. {Bread, Butter} → {Milk}

2. {Bread, Milk} → {Butter}

3. {Butter, Milk} → {Bread}

Rules from {Bread, Jam, Milk} with consequents that contain only 1-itemset:

4. {Bread, Jam} → {Milk}

5. {Bread, Milk} → {Jam}

6. {Jam, Milk} → {Bread}

d.

(1 ∩ 2) {Bread} →{Butter, Milk}

(1 ∩ 3) {Butter} →{Bread, Milk}

(2 ∩ 3) {Milk} →{Bread, Butter}

(4 ∩ 5) {Bread} →{Jam, Milk}

(4 ∩ 6) {Jam} → {Bread, Milk}

(5 ∩ 6) {Milk} →{Bread, Jam}

e. and f.

itemset	Confidence	Lift
{Bread, Butter} → {Milk}	$\frac{2}{2} = 1$	$\frac{1}{6/9} = 1.5$
{Bread, Milk} → {Butter}	$\frac{2}{4} = 0.5$	$\frac{1/2}{2/9} = 2.25$
{Butter, Milk} → {Bread}	$\frac{2}{2} = 1$	$\frac{1}{7/9} = 1.29$
{Bread, Jam} → {Milk}	$\frac{2}{4} = 0.5$	$\frac{1/2}{6/9} = 0.75$
{Bread, Milk} → {Jam}	$\frac{1}{2} = 0.5$	$\frac{1/2}{2/3} = 0.75$
{Jam, Milk} → {Bread}	$\frac{1}{2} = 0.5$	$\frac{2/7}{4/9} = 0.64$

itemset	Confidence	Lift
{Bread} →{Butter, Milk}	$\frac{2}{7} = 0.29$	$\frac{2/7}{2/9} = 1.29$
{Butter} →{Bread, Milk}	$\frac{2}{2} = 1$	$\frac{1}{4/9} = 2.25$
{Milk} →{Bread, Butter}	$\frac{2}{6} = 0.33$	$\frac{1/3}{2/9} = 1.5$
{Bread} →{Jam, Milk}	$\frac{2}{7} = 0.29$	$\frac{2/7}{4/9} = 0.64$
{Jam} → {Bread, Milk}	$\frac{2}{6} = 0.33$	$\frac{1/3}{4/9} = 0.75$
{Milk} →{Bread, Jam}	$\frac{2}{6} = 0.33$	$\frac{1/3}{4/9} = 0.75$