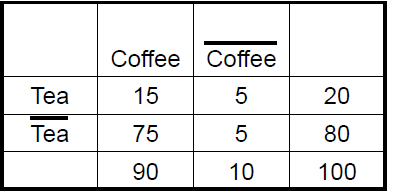
Q1.

沿用課本的例子



P(Coffee|Tea)=0.75>0.5具有高confidence,使得Tea->Coffee看起來很合理,但是P(Coffee|Tea)<P(Coffee)=0.9,反而是在一般情況之下,喝咖啡的機會更高,如此一來此association rule將會產生矛盾。

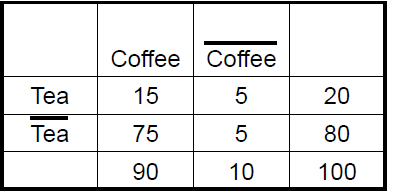
Lift不會有這個問題是因為Lift有把P(Y)考慮進去。

Q2.

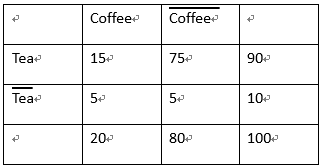
Confidence:asymmetric

Lift:symmetric

反例:



Confidence=0.75



Confidence:0.17

由此可知confidence為asymmetric

Q3.

Setting:

Lowerboundminsup:0.02

Upperboundminsup:1

Delta:0.05

Metric: confidence

minMetric: 0.5

Top 10 rules:

1. ROSES REGENCY TEACUP AND SAUCER =True PINK REGENCY TEACUP AND SAUCER=True ==> GREEN REGENCY TEACUP AND SAUCER=True

2. GREEN REGENCY TEACUP AND SAUCER=True PINK REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True

3. PINK REGENCY TEACUP AND SAUCER=True ==> GREEN REGENCY TEACUP AND SAUCER=True

4. PINK REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True

5. GREEN REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True

6. GARDENERS KNEELING PAD CUP OF TEA =True ==> GARDENERS KNEELING PAD KEEP CALM =True

7. ROSES REGENCY TEACUP AND SAUCER =True GREEN REGENCY TEACUP AND SAUCER=True ==> PINK REGENCY TEACUP AND SAUCER=True

8. DOTCOM POSTAGE=True ==> JUMBO BAG RED RETROSPOT=True

9. CHARLOTTE BAG PINK POLKADOT=True ==> RED RETROSPOT CHARLOTTE BAG=True

10. ROSES REGENCY TEACUP AND SAUCER =True ==> GREEN REGENCY TEACUP AND SAUCER=True

“TEACUP AND SAUCER”這類型得商品很常會出現在前10條rule當中,可能原因是當消費者經常一次購買大量不同顏色的此商品,導致其彼此之間得association rule 的confidence很高。

Q4.

Setting:

Lowerboundminsup:0.01~0.04

Upperboundminsup:1

Delta:0.05

Metric: confidence

minMetric: 0

0.04,

代表所有能形成association rule的item之support都<0.04。

Q5.

Setting:

Lowerboundminsup:0.01

Upperboundminsup:1

Delta:0.05

Metric: confidence

minMetric: 0.9

item quadruple (REGENCY CAKESTAND 3 TIER, ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER)

符合此定義之組合中confidence>0.9的只有以下一條rule:

REGENCY CAKESTAND 3 TIER=TRUE ROSES REGENCY TEACUP AND SAUCER =TRUE PINK REGENCY TEACUP AND SAUCER=TRUE 140 ==> GREEN REGENCY TEACUP AND SAUCER=TRUE 129 <conf:(0.92)> lift:(19.86) lev:(0.01) [122] conv:(11.13)

此即為confidence最高的rule

Q6.

Setting:

Lowerboundminsup:0.02

Upperboundminsup:1

Delta:0.05

Metric: Lift

minMetric: 1

Top 10 rules:

1. ROSES REGENCY TEACUP AND SAUCER =True GREEN REGENCY TEACUP AND SAUCER=True ==> PINK REGENCY TEACUP AND SAUCER=True

2. PINK REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True GREEN REGENCY TEACUP AND SAUCER=True

3. GREEN REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True PINK REGENCY TEACUP AND SAUCER=True

4. ROSES REGENCY TEACUP AND SAUCER =True PINK REGENCY TEACUP AND SAUCER=True ==> GREEN REGENCY TEACUP AND SAUCER=True

5. GARDENERS KNEELING PAD KEEP CALM =True ==> GARDENERS KNEELING PAD CUP OF TEA =True

6. GARDENERS KNEELING PAD CUP OF TEA =True ==> GARDENERS KNEELING PAD KEEP CALM =True

7. GREEN REGENCY TEACUP AND SAUCER=True ==> PINK REGENCY TEACUP AND SAUCER=True

8. PINK REGENCY TEACUP AND SAUCER=True ==> GREEN REGENCY TEACUP AND SAUCER=True

9. ROSES REGENCY TEACUP AND SAUCER =True ==> GREEN REGENCY TEACUP AND SAUCER=True PINK REGENCY TEACUP AND SAUCER=True

10. GREEN REGENCY TEACUP AND SAUCER=True PINK REGENCY TEACUP AND SAUCER=True ==> ROSES REGENCY TEACUP AND SAUCER =True

為何挑選這個metric:

Confidence只考慮到P(Y|X),而未考慮P(Y),Lift則P(X)和P(Y)皆考慮到了,為更好的metric因此使用此metrc。

與Q3差異:

Item的部分有許多重複,rule的部分只有4個 是一樣的。Q6前10名的rule普遍confidence都比Q3低上許多,可見得,Lift高不代表Confidence一定就會高,反之亦然。Q6中有許多Confidence較低但是Lift較高者,代表其P(Y)較低。

Q7.

工具:Weka

參數設定:見每題setting部分

前處理:將資料csv檔案中’Fasle’取代成’?’

實驗步驟:

1. 匯入檔案
2. 調整參數
3. Lowerboundminsup
4. Upperboundminsup
5. Delta
6. Metric
7. minMetric
8. Start