**BUSINESS DATA MINING**

**(IDS 572)**

**Solutions to Homework 10**

**Association Rules**

**Group Members**

* Amey Pophali (apopha2@uic.edu)
* Karthik Varanasi (vvaran3@uic.edu)
* Mrinal Dhawan (mdhawa3@uic.edu)

1. Solution –
   1. Fraction of all customers who have a sunroof = 30/74 = 0.40
   2. Confidence, Lift and Support of the rule sunroof ⇒ improved audio

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Support | 22/74 = 0.29 |
| Confidence | 22/30 = 0.73 |
| Lift | (22/74) / (30/74) = 1.88 |

* 1. Confidence, Lift and Support of the rule (M and A) ⇒ S

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Support | 18/74 = 0.24 |
| Confidence | 18/32 = 0.56 |
| Lift | (18/32) / (30/74) = 1.38 |

* 1. Confidence, Lift and Support of the rule M ⇒ S

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Support | 26/74 = 0.35 |
| Confidence | 26/45 = 0.57 |
| Lift | (26/45) / (30/74) = 1.42 |

* 1. Confidence, Lift and Support of the rule (not M) ⇒ S

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Support | 4/74 |
| Confidence | 4/29 |
| Lift | (4/29) / (74/30) = 0.34 |

1. Given -

P - Pears, A- Apples, B - Banana

Confidence = P(P|A) = 0.5 (Equation 1)

Support = P(P^A) = 0.3 (Equation 2)

Lift = Confidence / P(P) = 1.11 (Equation 3)

From equation (3)

* P(P|A) / P(P) = 1.11
* P(P) = P(P|A) / 1.11 = 0.5/1.11

Hence fraction who did not buy Pears = 1-P(P) = 1 - 0.5/1.11 = 0.61/1.11

Confidence = P(B|A) = 0.6666 (Equation 5)

Support = P(B^A) = 0.4 (Equation 6)

Lift = Confidence / P(B) = 1.33 (Equation 7)

From equation (6)

* P(B|A) / P(P) = 1.33
* P(B) = P(B|A) / 1.33 = 0.6666/1.33
* P(B) = 0.6666/1.33

P(B|A) = P(B^A) / P(A)

* 0.6666 = 0.4/P(A)
* P(A) = 0.4/0.6666
  1. Fraction of all your customers who bought both Apples and Bananas

= P(B^A)

= 0.4

* 1. Of those customers who bought Apples, the fraction who did not buy Pears.

Customers who bought Apples but not pears

P(|A) = P(A) - P(A^P)

= (0.4/0.6666) - 0.3

= 0.6 - 0.3

= 0.3

* 1. Fraction of your customers who did not buy Bananas.

= 1-P(B) = 1 - (0.6666/1.33) = 0.6634/1.33

* 1. Fraction who bought apples = P(A) = 0.4/0.6666 = 0.6
  2. Calculations for Apriori algorithm are attached in the excel sheet provided here–



* 1. The 4 and 3 item set for with maximum support from L3 and L4 respectively are
     1. Undergraduate, Young, US
     2. Undergraduate, Engineering, Young, US

The support of the first option is 64% and the second option is 30%

Now generating rules for the above conditions.

The rule selection needs to be done with a minimum support of 20% and a confidence greater than 90%.

Hence for the three item datasets, the rules go as follows -

For L3

Undergraduate, Young, US

1. Young 🡪 Undergraduate, U.S. Confidence = 84%
2. Undergraduate 🡪 Young, U.S Confidence = 80%
3. U.S. 🡪 Young, Undergraduate Confidence = 88%
4. Young, Undergraduate 🡪 U.S. Confidence = 84%
5. **Young, U.S. 🡪 Undergraduate Confidence = 100%**
6. **Undergraduate, U.S. 🡪 Young Confidence = 95%**

For L4

Undergraduate, Engineering, Young, US

1. Engineering->Undergraduate, Young, US Confidence = 58%
2. Undergraduate->Engineering, Young, US Confidence = 38%
3. Young->Engineering, Young, US Confidence = 40%
4. US->Engineering, Undergraduate, Young Confidence = 42%
5. Engineering, Undergraduate->Young, US Confidence = 71%
6. Engineering, Young->Undergraduate, US Confidence = 71%
7. **Engineering, US->Undergraduate, Young Confidence = 100%**
8. Undergraduate, Young->Engineering, US Confidence = 40%
9. Undergraduate, US->Engineering, Young Confidence = 45%
10. Young, US->Engineering, Undergraduate Confidence = 47%
11. Engineering, Undergraduate, Young->US Confidence = 71%
12. **Engineering, Undergraduate, US->Young Confidence = 100%**
13. **Engineering, Young, US->Undergraduate Confidence = 100%**
14. Undergraduate, Young, US->Engineering Confidence = 47%