Association Rules

Mining Massive Datasets

Prof. Carlos Castillo — https://chato.cl/teach

Universitat Pompeu Fabra
Barcelona

Sources

- Data Mining, The Textbook (2015) by Charu Aggarwal (Chapters 4, 5) – slides by Lijun Zhang

- Mining of Massive Datasets 2nd edition (2014) by Leskovec et al. (Chapter 6) - slides

- Data Mining Concepts and Techniques, 3rd edition (2011) by Han et al. (Chapter 6)

- Introduction to Data Mining 2nd edition (2019) by Tan et al. (Chapters 5, 6) – slides ch5, slides ch6

What is a rule

- A rule is of the form X⇒Y

    X and Y are itemsets

- X is the antecedent, Y is the consequent

- The **confidence** of the rule is:

$$\mathrm{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

Confidence of a rule

- The **confidence** of the rule $X \Rightarrow Y$ is:

$$\mathrm{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

- This is the conditional probability of X U Y occurring in a transaction, given that X occurs in the transaction

Confidence of a rule (cont.)

tid	Set of items
1	Bread, Jam, Juice
2	Tofu, Juice, Tomatoes
3	Bread, Strawberries, Tofu, Juice
4	Tofu, Juice, Tomatoes
5	Strawberries, Juice, Tomatoes

$$\mathrm{conf}(\{\mathrm{tofu}, \mathrm{juice}\} \Rightarrow \{\mathrm{tomatoes}\}) = ?$$

Lift of a rule

- The **lift** of the rule $X \Rightarrow Y$ is:
$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{sup}(X \cup Y)}{\mathrm{sup}(X)\,\mathrm{sup}(Y)}$$

- This is the ratio between the observed support and the expected support if X and Y were independent

Exercise

$$\text{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

$$\text{lift}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)\sup(Y)}$$

Rule	Support $\sup(X \cup Y)$	Confidence	Lift
$A \Rightarrow D$			
$C \Rightarrow A$			
$A \Rightarrow C$			
$B \& C \Rightarrow D$			

Answer in
Nearpod draw-it
Code to be given in class

Association rule (minsup, minconf)

- Let X, Y be two itemsets; the rule X⇒Y is an **association rule** of minimum support **minsup** and minimum confidence **minconf** if:

$$\text{sup}(X{\Rightarrow}Y) \geq \text{minsup}$$

and

$$\text{conf}(X{\Rightarrow}Y) \geq \text{minconf}$$

Summary

Things to remember

- Association rule of minsup and minconf

- The concepts of **confidence** and **lift**

Exercises for **TT11**-**TT12**

- Data Mining, The Textbook (2015) by Charu Aggarwal

  - Exercises 4.9 $\rightarrow$ 1-3, 5, 7-8

  - Exercises 5.7 $\rightarrow$ 1-5

- Mining of Massive Datasets 2nd edition (2014) by Leskovec et al.

  - Exercises 6.1.5 $\rightarrow$ 6.1.1-6.1.7

- Introduction to Data Mining 2nd edition (2019) by Tan et al.

  - Exercises 5.10 $\rightarrow$ 2-7