Association Rules

CSCI 347, Data Mining

Market Basket Analysis

- Used by retailers to uncover associations between items purchased

- Place items purchased together near each other

- This is not considering individual preferences. Recommender systems consider individual preferences.

Example Basket

Items

milk, broccoli, bread, cheese

bagel, cheese, potato chips

milk, bread, ham

bread, cheese

milk, potato chips

milk, bread, cheese

broccoli, bread, cheese

milk

broccoli, cheese, ham

milk, cheese

How to Use an Association Rule

Consider {bagels,...} → {potato chips}

- potato chips as consequent - determine what should be done to boost potato chip sales

- bagels in the antecedent - which products would be affected if the store discontinues selling bagels

- bagels in antecedent and Potato chips in the consequent - what products should be sold with bagels to promote sale of potato chips

1-Hot Encoding

Typically items are 1-hot encoded:

	Apple	Corn	Dill	Eggs	Ice cream	Kidney Beans	Milk	Nutmeg	Onion	Unicorn	Yogurt
0	0	0	0	1	0	1	1	1	1	0	1
1	0	0	1	1	0	1	0	1	1	0	1
2	1	0	0	1	0	1	1	0	0	0	0
3	0	1	0	0	0	1	1	0	0	1	1
4	0	1	0	1	1	1	0	0	1	0	0

Terms

Support – frequency of an item-set occurrence

Support(X → Y) =

transactions containing both X and Y /

total number of transactions

Confidence – likeliness of occurrence of the

consequent given that the cart already has the

antecedents

Confidence(X → Y) =

transactions containing both X and Y /

transactions containing X

Terms

Lift – tells the "lift" that the antecedent provides to our confidence of having the consequence.

Lift(X → Y) =

(confidence / fraction of transactions containing Y)

=

(transactions containing both X and Y /

transactions containing X)

/

(transactions containing y /

total number of transactions)

Example Basket

Items

milk, broccoli, bread, cheese

bagel, cheese, potato chips

milk, bread, ham

bread, cheese

milk, potato chips

milk, bread, cheese

broccoli, bread, cheese

milk

broccoli, cheese, ham

milk, cheese, ham

ham → milk

Support   2/10   or   .2
(probability rather than count)

Confidence –   2/3   or .67

Lift – (2/3) / (6/10)   or 1.11

Lift

Lift (X → Y) >1
 Y appears with X more often than expected
 X has a positive effect on Y

Lift (X → Y) near 1
X and Y appear together as often as expected X has little effect on Y

Lift (X → Y) <1
Y appears with X less often than expected
X has a negative effect on Y

Mining Association Rules

Can be highly computationally complex

One method:

- Determine item sets
- Build rules from those item sets

Input to Mining Association Rules

Two inputs

- Coverage (example - 2 instances)
- Accuracy (example - 100% accuracy)

Item Sets

Item: one attribute-value pair
   Example: outlook=rainy

Item set : set of items
   Example:
       outlook=rainy
       temperature = cool
       play = yes

Weather Data

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Item Sets for Weather Data

In total, 12 one-item sets, 47 two-item sets, 39 three-item sets, 6 four-item sets and 0 five-item sets (with minimum support of two)

One-item sets	Two-item sets	Three-item sets	Four-item sets
Outlook = Sunny (5)	Outlook = Sunny Temperature = Hot (2)	Outlook = Sunny Temperature = Hot Humidity = High (2)	Outlook = Sunny Temperature = Hot Humidity = High Play = No (2)
Temperature = Cool (4)	Outlook = Sunny Humidity = High (3)	Outlook = Sunny Humidity = High Windy = False (2)	Outlook = Rainy Temperature = Mild Windy = False Play = Yes (2)
...	...	...	...

Generating Rules from an Item Set

Once all item sets with minimum support have been generated, we can turn them into rules

`Humidity = Normal, Windy = False, Play = Yes (4)`

Example:

`If Humidity = Normal and Windy = False then Play = Yes`   `4/4`

`If Humidity = Normal and Play = Yes then Windy = False`   `4/6`

`If Windy = False and Play = Yes then Humidity = Normal`   `4/6`

`If Humidity = Normal then Windy = False and Play = Yes`   `4/7`

`If Windy = False then Humidity = Normal and Play = Yes`   `4/8`

`If Play = Yes then Humidity = Normal and Windy = False`   `4/9`

`If True then Humidity = Normal and Windy = False`
    `and Play = Yes`   `4/14`

Seven (2^N-1) potential rules:

15

Rules for Weather Data

Rules with support > 1 and confidence = 100%

	Association rule		Sup.	Conf.
1	Humidity=Normal Windy=False	$\Rightarrow$ Play=Yes	4	100%
2	Temperature=Cool	$\Rightarrow$ Humidity=Normal	4	100%
3	Outlook=Overcast	$\Rightarrow$ Play=Yes	4	100%
4	Temperature=Cold Play=Yes	$\Rightarrow$ Humidity=Normal	3	100%
	...	...	...	...
58	Outlook=Sunny Temperature=Hot	$\Rightarrow$ Humidity=High	2	100%

In total:
　　　　3 rules with support four
　　　　5 with support three
　　　　50 with support two

16

Example Rules from the Same Set

Item set:

Temperature = Cool, Humidity = Normal, Windy = False, Play = Yes (2)

Temperature = Cool, Windy = False $\Rightarrow$ Humidity = Normal, Play = Yes

Temperature = Cool, Windy = False, Humidity = Normal $\Rightarrow$ Play = Yes

Temperature = Cool, Windy = False, Play = Yes $\Rightarrow$ Humidity = Normal

Temperature = Cool, Windy = False   (2)

Temperature = Cool, Humidity = Normal, Windy = False   (2)

Temperature = Cool, Windy = False, Play = Yes   (2)