

Data Mining, CSCI 347, Fall 2019
Association Rules, Sept. 27

Consider the weather nominal data set.

Relation: weather.symbolic					
No.	outlook Nominal	temperature Nominal	humidity Nominal	windy Nominal	play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Find association rules with a minimum coverage of 2 instances and a minimum accuracy of 100%.

Step 1. Begin by determining the item-sets with the minimum amount of coverage..

There are 12 one-item sets:

Since outlook has 3 values, temperature has 3, humidity has 2, windy has 2 and play has 2. For each of these, the coverage is at least 2 (the dataset is pretty balanced with multiple values from each)

There are 47 two-item sets

Each of the 3 outlooks can combine with 3 temperature, etc. Once again, check that there are at least 2 instances to which the tests apply.

There are 39 three-item sets with a minimum coverage of 2 – Table 4.10, page 121 lists many of them

There are 4 four-item sets and no five-item sets with a minimum coverage of 2.

Step 2. For each item-set, determine all the rules which can be gotten and that have at least the minimum accuracy.

Nothing in the antecedent:

IF true THEN outlook=rainy AND humidity=high AND play=yes AND windy=false

Another:

IF outlook=rainy THEN humidity=high AND play=yes AND windy=false

IF play=yes AND outlook=rainy THEN humidity=high AND windy=false

IF TRUE THEN outlook=rainy AND humidity=high AND play=yes AND windy=false

When finding the accuracy of a rule:

IF TRUE THEN the denominator will be all of the instances.