Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων (ISE709)

ΑΚΑΔ. ΕΤΟΣ 2024-25

Διδάσκουσα: Γ. Κολωνιάρη

HW2

Εξόρυξη Γνώσης σε Δεδομένα - Εργασία 2 (Association Rules)

Φοιτητής: Ερρίκος Ματεβοσιάν
ΑΜ: iis23018

(A) Apriori με το χέρι

Transaction ID	Items
T1	apples, bananas
T2	apples, oranges, grapes
T3	pears, apples, berries
T4	pears, berries, bananas
T5	pears, apples, bananas, berries, kiwis

Item	Transactions	Support	(support >= 40%)
apples	T1,T2,T3,T5	4/5 = 80%	Ναι
bananas	T1, T4, T5	3/5 = 60%	Ναι
oranges	T2	1/5 = 20%	Όχι
grapes	T2	1/5 = 20%	Όχι
pears	T3, T4, T5	3/5 = 60%	Ναι
berries	T3, T4, T5	3/5 = 60%	Ναι
kiwis	T5	1/5 = 20%	Όχι

1-στοιχειοσύνολα: {apples}, {bananas}, {pears}, {berries}

	Transactions	Support	(support ≥ 40%)
{apples, bananas}	T1, T5	2/5 = 40%	Ναι
{apples, pears}	T3, T5	2/5 = 40%	Ναι
{apples, berries}	T3, T5	2/5 = 40%	Ναι
{bananas, pears}	T4, T5	2/5 = 40%	Ναι
{bananas, berries}	T4, T5	2/5 = 40%	Ναι
{berries, pears}	T3, T4, T5	3/5 = 60%	Ναι

2- στοιχειοσύνολα = {apples, bananas}, {apples, pears}, {apples, berries}, {bananas, pears}, {bananas, berries}, {berries, pears}

	Transaction	Support	(support ≥ 40%)
{apples, berries, pears}	T3, T5	2/5 = 40%	Ναι
{bananas, berries, pears}	T4, T5	2/5 = 40%	Ναι

3-στοιχειοσύνολα = {apples, berries, pears}, {bananas, berries, pears}

4- στοιχειοσύνολα = no

Ισχυροί κανόνες:

{berries} → {pears}
Support = Support({berries, pears}) / Support({berries} = 60%
Confidence = Support({berries, pears}) / Support({berries} = 0.6/0.6 = 1
Lift = Confidence/Support = 1.67

{pears} → {berries}
Support = 0.6
Confidence = 1
Lift = 1.67

{apples, berries} → {pears}
Support({apples, berries, pears}) = 40%
Confidence = Support({apples, berries, pears}) / Support({apples, berries}) = 0.4/0.4 =1
Lift = 1/0.6 = 1.67

{apples, pears} → {berries}
Support = 0.6
Confidence = 1
Lift = 1.67

{bananas, berries} → {pears}
Support = 0.6
Confidence = 1
Lift = 1.67

{bananas, pears} → {berries}
Support = 0.6
Confidence = 1
Lift = 1.67

b) Weka

Τα αποτελέσματα που εμφανίζει το weka βρίσκονται παρακάτω

```
Apriori
=======

Minimum support: 0.7 (3 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 6

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 7

Size of set of large itemsets L(4): 1

Best rules found:

 1. grapes=0 4 ==> oranges=0 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
 2. oranges=0 4 ==> grapes=0 4    <conf:(1)> lift:(1.25) lev:(0.16) [0] conv:(0.8)
 3. bananas=1 3 ==> oranges=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 4. bananas=1 3 ==> grapes=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 5. pears=1 3 ==> oranges=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 6. berries=1 3 ==> oranges=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 7. pears=1 3 ==> grapes=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 8. berries=1 3 ==> grapes=0 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
 9. berries=1 3 ==> pears=1 3    <conf:(1)> lift:(1.67) lev:(0.24) [1] conv:(1.2)
10. pears=1 3 ==> berries=1 3    <conf:(1)> lift:(1.67) lev:(0.24) [1] conv:(1.2)

Apriori
=======

Minimum support: 0.7 (3 instances)
Minimum metric <lift>: 1.5
Number of cycles performed: 6

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 7

Size of set of large itemsets L(4): 1

Best rules found:

 1. pears=1 3 ==> berries=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 2. berries=1 3 ==> pears=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 3. pears=1 3 ==> oranges=0 berries=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 4. oranges=0 pears=1 3 ==> berries=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 5. berries=1 3 ==> oranges=0 pears=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 6. oranges=0 berries=1 3 ==> pears=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 7. pears=1 3 ==> grapes=0 berries=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 8. grapes=0 pears=1 3 ==> berries=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
 9. berries=1 3 ==> grapes=0 pears=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
10. grapes=0 berries=1 3 ==> pears=1 3    conf:(1) < lift:(1.67)> lev:(0.24) [1] conv:(1.2)
```

Για Lift = 1.67
{berries=1} → {pears=1}
{pears=1} → {berries=1}

Αυτοί οι κανόνες ταιριάζουν ακριβώς με αυτούς που υπολογίσαμε προηγουμένως, επιβεβαιώνοντας ότι οι ισχυρότεροι συσχετισμοί αφορούν berries και τα pears.

(B) Μελέτη περίπτωσης με το WEKA

Αφού δημιουργήσουμε το arff αρχείο και το ανοίξουμε στο weka, μεταβαίνουμε στο preprocess tab, επιλέγουμε στα φίλτρα unsupervised → attribute → Remove, πληκτρολογούμε στις ρυθμίσεις του φίλτρου στο πεδίο attributeIndices το 1 και πατάμε apply. Για τα χαρακτηριστικά legs και type επιλέγουμε το φίλτρο Discretize και στο αντίστοιχο πεδίο πληκτρολογούμε πρώτα το 13 (αφού εκεί αντιστοιχεί το χαρακτηρστικό legs) και έπειτα την ίδια διαδικασία, απλώς πληκτρολογούμε το 17. Έπειτα αφού επιλέξουμε τον Apriori στο Associate Tab, επεξεργαζόμαστε τους παραμέτρους του όπως θέλουμε και τον τρέχουμε.

```
Best rules found:

    1. venomous=0 tail=1 71 ==> backbone=1 71    <conf:(1)> lift:(1.22) lev:(0.13) [12] conv:(12.65)
    2. aquatic=0 65 ==> fins=0 65    <conf:(1)> lift:(1.2) lev:(0.11) [10] conv:(10.94)
    3. aquatic=0 breathes=1 64 ==> fins=0 64    <conf:(1)> lift:(1.2) lev:(0.11) [10] conv:(10.77)
    4. backbone=1 venomous=0 fins=0 63 ==> breathes=1 63    <conf:(1)> lift:(1.26) lev:(0.13) [13] conv:(13.1)
    5. toothed=1 61 ==> feathers=0 61    <conf:(1)> lift:(1.25) lev:(0.12) [12] conv:(12.08)
    6. toothed=1 61 ==> backbone=1 61    <conf:(1)> lift:(1.22) lev:(0.11) [10] conv:(10.87)
    7. toothed=1 backbone=1 61 ==> feathers=0 61    <conf:(1)> lift:(1.25) lev:(0.12) [12] conv:(12.08)
    8. feathers=0 toothed=1 61 ==> backbone=1 61    <conf:(1)> lift:(1.22) lev:(0.11) [10] conv:(10.87)
    9. toothed=1 61 ==> feathers=0 backbone=1 61    <conf:(1)> lift:(1.6) lev:(0.23) [22] conv:(22.95)
   10. aquatic=0 venomous=0 61 ==> fins=0 61    <conf:(1)> lift:(1.2) lev:(0.1) [10] conv:(10.27)
   11. venomous=0 tail=1 domestic=0 61 ==> backbone=1 61    <conf:(1)> lift:(1.22) lev:(0.11) [10] conv:(10.87)
   12. aquatic=0 breathes=1 venomous=0 60 ==> fins=0 60    <conf:(1)> lift:(1.2) lev:(0.1) [10] conv:(10.1)
   13. airborne=0 toothed=1 59 ==> feathers=0 59    <conf:(1)> lift:(1.25) lev:(0.12) [11] conv:(11.68)
   14. airborne=0 toothed=1 59 ==> backbone=1 59    <conf:(1)> lift:(1.22) lev:(0.1) [10] conv:(10.51)
   15. airborne=0 toothed=1 backbone=1 59 ==> feathers=0 59    <conf:(1)> lift:(1.25) lev:(0.12) [11] conv:(11.68)
   16. feathers=0 airborne=0 toothed=1 59 ==> backbone=1 59    <conf:(1)> lift:(1.22) lev:(0.1) [10] conv:(10.51)
   17. airborne=0 toothed=1 59 ==> feathers=0 backbone=1 59    <conf:(1)> lift:(1.6) lev:(0.22) [22] conv:(22.2)
   18. breathes=1 venomous=0 tail=1 59 ==> backbone=1 59    <conf:(1)> lift:(1.22) lev:(0.1) [10] conv:(10.51)
   19. toothed=1 venomous=0 57 ==> feathers=0 57    <conf:(1)> lift:(1.25) lev:(0.11) [11] conv:(11.29)
   20. toothed=1 venomous=0 57 ==> backbone=1 57    <conf:(1)> lift:(1.22) lev:(0.1) [10] conv:(10.16)
   21. toothed=1 backbone=1 venomous=0 57 ==> feathers=0 57    <conf:(1)> lift:(1.25) lev:(0.11) [11] conv:(11.29)
   22. feathers=0 toothed=1 venomous=0 57 ==> backbone=1 57    <conf:(1)> lift:(1.22) lev:(0.1) [10] conv:(10.16)
   23. toothed=1 venomous=0 57 ==> feathers=0 backbone=1 57    <conf:(1)> lift:(1.6) lev:(0.21) [21] conv:(21.45)
   24. venomous=0 fins=0 tail=1 56 ==> backbone=1 56    <conf:(1)> lift:(1.22) lev:(0.1) [9] conv:(9.98)
   25. venomous=0 fins=0 tail=1 56 ==> breathes=1 56    <conf:(1)> lift:(1.26) lev:(0.12) [11] conv:(11.64)
   26. breathes=1 venomous=0 fins=0 tail=1 56 ==> backbone=1 56    <conf:(1)> lift:(1.22) lev:(0.1) [9] conv:(9.98)
   27. backbone=1 venomous=0 fins=0 tail=1 56 ==> breathes=1 56    <conf:(1)> lift:(1.26) lev:(0.12) [11] conv:(11.64)
   28. venomous=0 fins=0 tail=1 56 ==> backbone=1 breathes=1 56    <conf:(1)> lift:(1.46) lev:(0.18) [17] conv:(17.74)
   29. airborne=0 toothed=1 venomous=0 55 ==> feathers=0 55    <conf:(1)> lift:(1.25) lev:(0.11) [10] conv:(10.89)
   30. airborne=0 toothed=1 venomous=0 55 ==> backbone=1 55    <conf:(1)> lift:(1.22) lev:(0.1) [9] conv:(9.8)
   31. airborne=0 toothed=1 backbone=1 venomous=0 55 ==> feathers=0 55    <conf:(1)> lift:(1.25) lev:(0.11) [10] conv:(10.89)
   32. feathers=0 airborne=0 toothed=1 venomous=0 55 ==> backbone=1 55    <conf:(1)> lift:(1.22) lev:(0.1) [9] conv:(9.8)
   33. airborne=0 toothed=1 venomous=0 55 ==> feathers=0 backbone=1 55    <conf:(1)> lift:(1.6) lev:(0.2) [20] conv:(20.69)
   34. hair=0 eggs=1 54 ==> milk=0 54    <conf:(1)> lift:(1.68) lev:(0.22) [21] conv:(21.92)
```

Επιλέγουμε στη παράμετρο N τον δυνάτον περισσότερους κανόνες, ενώ παράλληλα θέτουμε lowerBoundMinSupport ίσο με 0.2 και minMetric ίσο με 0.9.

Ένας ισχυρός κανόνας που βλέπουμε είναι milk=1 ==> type=1. Όσα ζώα παράγουν γάλα και θηλάζουν είναι θηλαστικά