CSC487: Data Mining - Homework #5

Cason Konzer

Refer to the table below for questions 1-4 and suppose the minimum support is 2.

TID	Basket
T1	A,B,E
T2	B,D
T3	B,C
T4	A,B,D
T5	A,C
T6	B,C
T7	A,C
T8	A,B,C,E
T9	A,B,C

1. Compute the confidences for the following rules. ***(20 points total)***

---

a.) $\{A, B\} \Rightarrow E$**.**

Given $\{A, B\}$ is observed, we have the following baskets bought:

$\langle\{A, B, E\}, \{A, B, D\}, \{A, B, C, E\}, \{A, B, C\}\rangle$

Of these four baskets two contain E. Our confidence is thus $P(E|\{A, B\}) = 2/4 = 0.5$

b.) $A \Rightarrow \{B, E\}$**.**

Given A is observed, we have the following baskets bought:

$\langle\{A, B, E\}, \{A, B, D\}, \{A, C\}, \{A, C\}, \{A, B, C, E\}, \{A, B, C\}\rangle$

Of these six baskets two contain $\{B, E\}$. Our confidence is thus $P(\{B, E\}|A) = 2/6 = 0.333$

---

2. Apply the Apriori procedure by using join operations as described on slide (see slide1) #15. You need to report all frequent k-itemsets. ***(25 points)***

k=1	k=2	k=3	k=4
{A} : 6	{A, B} : 4	{A, B, C} : 2	{A, B, C, E} : 1
{B} : 7	{A, C} : 4	{A, B, E} : 2	
{C} : 6	{A, D} : 1	{B, D, E} : 0	
{D} : 2	{A, E} : 2		
{E} : 2	{B, C} : 4		
	{B, D} : 2		
	{B, E} : 2		
	{C, D} : 0		
	{C, E} : 1		
	{D, E} : 0		

All Blue item sets are frequent.

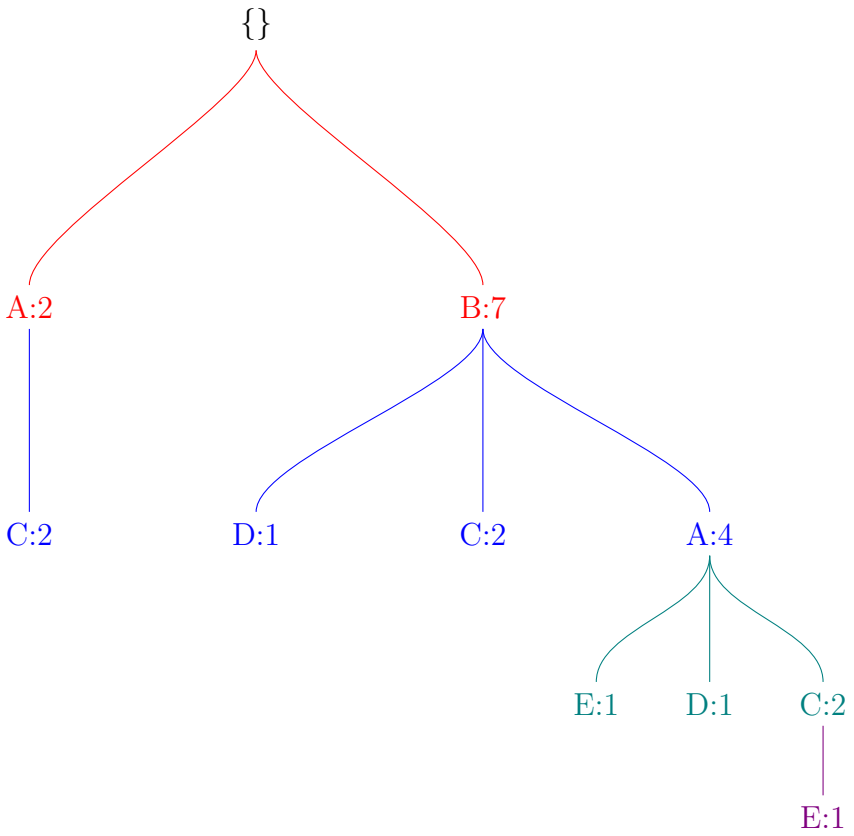
3. Draw the FP-Tree (see slide #28). *(25 points)*

We have the header table and ordered sets as follows:

Item	Frequency
B	7
A	6
C	6
D	2
E	2

TID	Ordered Set
T1	{B, A, E}
T2	{B, D}
T3	{B, C}
T4	{B, A, D}
T5	{A, C}
T6	{B, C}
T7	{A, C}
T8	{B, A, C, E}
T9	{B, A, C}

The growth tree can the be visualized as follows:



---

4. Import this table by preparing an appropriate input format for Weka and run Apriori algorithm. Please use either .arff or csv format by inspecting sample Weka files. Please report the association rules you find. ***(15 points)***

Best rules found:

1. B=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
2. E=y 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
3. D=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
4. E=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
5. B=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
6. B=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
7. B=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
8. D=y 2 ==> C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
9. E=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
10. D=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
11. A=y C=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
12. B=y E=y 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
13. A=y E=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
14. E=y 2 ==> A=y B=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
15. B=n C=y 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
16. A=y B=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
17. B=n 2 ==> A=y C=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
18. B=n D=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
19. A=y B=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
20. B=n 2 ==> A=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
21. B=n E=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
22. A=y B=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
23. B=n 2 ==> A=y E=n 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
24. D=n E=y 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
25. A=y E=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
26. E=y 2 ==> A=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
27. A=n C=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
28. A=n D=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)

29. A=n D=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
30. A=n C=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
31. A=n C=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
32. A=n D=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
33. C=n D=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
34. B=y D=y 2 ==> C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
35. D=y 2 ==> B=y C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
36. C=n E=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
37. D=n E=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
38. B=y E=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
39. E=y 2 ==> B=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
40. D=y E=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
41. B=y D=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
42. D=y 2 ==> B=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
43. B=n D=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
44. B=n C=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
45. B=n 2 ==> C=y D=n 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
46. B=n E=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
47. B=n C=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
48. B=n 2 ==> C=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
49. B=n E=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
50. B=n D=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
51. B=n 2 ==> D=n E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
52. D=y E=n 2 ==> C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
53. C=n E=n 2 ==> D=y 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
54. C=n D=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
55. D=y 2 ==> C=n E=n 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
56. A=y B=y C=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
57. B=y D=n E=y 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
58. A=y D=n E=y 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
59. A=y B=y E=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
60. D=n E=y 2 ==> A=y B=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
61. B=y E=y 2 ==> A=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)

62. A=y E=y 2 ==> B=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
63. E=y 2 ==> A=y B=y D=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
64. B=n C=y D=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
65. A=y B=n D=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
66. A=y B=n C=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
67. B=n D=n 2 ==> A=y C=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
68. B=n C=y 2 ==> A=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
69. A=y B=n 2 ==> C=y D=n 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
70. B=n 2 ==> A=y C=y D=n 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
71. B=n C=y E=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
72. A=y B=n E=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
73. A=y B=n C=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
74. B=n E=n 2 ==> A=y C=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
75. B=n C=y 2 ==> A=y E=n 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
76. A=y B=n 2 ==> C=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
77. B=n 2 ==> A=y C=y E=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
78. B=n D=n E=n 2 ==> A=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
79. A=y B=n E=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
80. A=y B=n D=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
81. B=n E=n 2 ==> A=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
82. B=n D=n 2 ==> A=y E=n 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
83. A=y B=n 2 ==> D=n E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
84. B=n 2 ==> A=y D=n E=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
85. A=n C=y D=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
86. A=n B=y D=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
87. A=n B=y C=y 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
88. A=n D=n 2 ==> B=y C=y 2 <conf:(1)> lift:(2.25) lev:(0.12) [1] conv:(1.11)
89. A=n C=y 2 ==> B=y D=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
90. A=n C=y E=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
91. A=n B=y C=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
92. A=n C=y 2 ==> B=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
93. A=n D=n E=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
94. A=n B=y D=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)

95. A=n D=n 2 ==> B=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
96. A=n D=n E=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
97. A=n C=y E=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
98. A=n C=y D=n 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
99. A=n D=n 2 ==> C=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
100. A=n C=y 2 ==> D=n E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
101. C=n D=y E=n 2 ==> B=y 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
102. B=y D=y E=n 2 ==> C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
103. B=y C=n E=n 2 ==> D=y 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
104. B=y C=n D=y 2 ==> E=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)
105. D=y E=n 2 ==> B=y C=n 2 <conf:(1)> lift:(3) lev:(0.15) [1] conv:(1.33)
106. C=n E=n 2 ==> B=y D=y 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
107. C=n D=y 2 ==> B=y E=n 2 <conf:(1)> lift:(1.8) lev:(0.1) [0] conv:(0.89)
108. B=y D=y 2 ==> C=n E=n 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
109. D=y 2 ==> B=y C=n E=n 2 <conf:(1)> lift:(4.5) lev:(0.17) [1] conv:(1.56)
110. B=n D=n E=n 2 ==> C=y 2 <conf:(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)
111. B=n C=y E=n 2 ==> D=n 2 <conf:(1)> lift:(1.29) lev:(0.05) [0] conv:(0.44)

The list goes on ...

---

5. Using Weka, implement Apriori and FP-Growth algorithms on Supermarket data, which is a sample data set coming with Weka installation. You can find it under Weka folder in your system. Please report your results with screen shots. You don't have to report all. Top of the results is enough for this question. ***(15 points)***