Project 3.1

Association Algorithms

Apriori

Apriori is an association algorithm that generates n number of rules specified by the user by iteratively reducing the minimum support generating rules with the set minimum confidence. Multiple runs were done by varying the number of rules and minimum required support level. In the 3rd run 10 rules were specified as the first run but the minimum required support level was raised to 20% instead of 10%. This resulted in reducing the number of final generated rules to 8. The second run specified only 5 rules with a minimum support level of 10%, this resulted in no significant change generating the first 5 rules of the 1st run.

1st Run

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.15 (2 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 conf:(1)

2. temperature=cool 4 ==> humidity=normal 4 conf:(1)

3. humidity=normal windy=FALSE 4 ==> play=yes 4 conf:(1)

4. outlook=sunny play=no 3 ==> humidity=high 3 conf:(1)

5. outlook=sunny humidity=high 3 ==> play=no 3 conf:(1)

6. outlook=rainy play=yes 3 ==> windy=FALSE 3 conf:(1)

7. outlook=rainy windy=FALSE 3 ==> play=yes 3 conf:(1)

8. temperature=cool play=yes 3 ==> humidity=normal 3 conf:(1)

9. outlook=sunny temperature=hot 2 ==> humidity=high 2 conf:(1)

10. temperature=hot play=no 2 ==> outlook=sunny 2 conf:(1)

2nd Run

numRules = 5

Minimum support: 0.25 (3 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 26

Size of set of large itemsets L(3): 4

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 conf:(1)

2. temperature=cool 4 ==> humidity=normal 4 conf:(1)

3. humidity=normal windy=FALSE 4 ==> play=yes 4 conf:(1)

4. outlook=sunny play=no 3 ==> humidity=high 3 conf:(1)

5. outlook=sunny humidity=high 3 ==> play=no 3 conf:(1)

3rd Run

numRules 10 minConfidence 0.2

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 conf:(1)

2. temperature=cool 4 ==> humidity=normal 4 conf:(1)

3. humidity=normal windy=FALSE 4 ==> play=yes 4 conf:(1)

4. outlook=sunny play=no 3 ==> humidity=high 3 conf:(1)

5. outlook=sunny humidity=high 3 ==> play=no 3 conf:(1)

6. outlook=rainy play=yes 3 ==> windy=FALSE 3 conf:(1)

7. outlook=rainy windy=FALSE 3 ==> play=yes 3 conf:(1)

8. temperature=cool play=yes 3 ==> humidity=normal 3 conf:(1)

Predictive Apriori

An association algorithm similarly to Apriori that searches for the best association rules by applying an increasing support threshold and support-based corrected confidence value. Two runs were done with the Predictive Apriori class on the nominal data set. Both runs required 100 rules with the first run using general association rules and the second using mined rules. The fist run resulted in generating 99 rules while the second generated 30 rules. Using the mined therefore resulted in smaller more accurate rule set.

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

=== Associator model (full training set) ===

PredictiveApriori

===================

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 acc:(0.95323)

2. temperature=cool 4 ==> humidity=normal 4 acc:(0.95323)

3. humidity=normal windy=FALSE 4 ==> play=yes 4 acc:(0.95323)

4. outlook=sunny humidity=high 3 ==> play=no 3 acc:(0.92093)

5. outlook=sunny play=no 3 ==> humidity=high 3 acc:(0.92093)

6. outlook=rainy windy=FALSE 3 ==> play=yes 3 acc:(0.92093)

7. outlook=rainy play=yes 3 ==> windy=FALSE 3 acc:(0.92093)

8. outlook=sunny temperature=hot 2 ==> humidity=high play=no 2 acc:(0.86233)

9. outlook=sunny humidity=normal 2 ==> play=yes 2 acc:(0.86233)

10. outlook=sunny play=yes 2 ==> humidity=normal 2 acc:(0.86233)

11. outlook=overcast temperature=hot 2 ==> windy=FALSE play=yes 2 acc:(0.86233)

12. outlook=overcast windy=FALSE 2 ==> temperature=hot play=yes 2 acc:(0.86233)

13. outlook=rainy humidity=high 2 ==> temperature=mild 2 acc:(0.86233)

14. outlook=rainy windy=TRUE 2 ==> play=no 2 acc:(0.86233)

15. outlook=rainy play=no 2 ==> windy=TRUE 2 acc:(0.86233)

16. temperature=hot play=yes 2 ==> outlook=overcast windy=FALSE 2 acc:(0.86233)

17. temperature=hot play=no 2 ==> outlook=sunny humidity=high 2 acc:(0.86233)

18. temperature=mild humidity=normal 2 ==> play=yes 2 acc:(0.86233)

19. temperature=mild play=no 2 ==> humidity=high 2 acc:(0.86233)

20. temperature=cool windy=FALSE 2 ==> humidity=normal play=yes 2 acc:(0.86233)

21. windy=FALSE play=no 2 ==> outlook=sunny humidity=high 2 acc:(0.86233)

22. temperature=mild windy=FALSE play=yes 2 ==> outlook=rainy 2 acc:(0.86233)

23. humidity=normal 7 ==> play=yes 6 acc:(0.69497)

24. play=no 5 ==> humidity=high 4 acc:(0.59096)

25. windy=FALSE 8 ==> play=yes 6 acc:(0.56435)

26. temperature=hot 4 ==> humidity=high 3 acc:(0.54473)

27. temperature=hot 4 ==> windy=FALSE 3 acc:(0.54473)

28. temperature=cool 4 ==> humidity=normal play=yes 3 acc:(0.54473)

29. humidity=high play=no 4 ==> outlook=sunny 3 acc:(0.54473)

30. play=yes 9 ==> humidity=normal 6 acc:(0.53808)

31. play=yes 9 ==> windy=FALSE 6 acc:(0.53808)

32. temperature=mild 6 ==> humidity=high 4 acc:(0.51949)

33. temperature=mild 6 ==> play=yes 4 acc:(0.51949)

34. outlook=sunny humidity=high 3 ==> windy=FALSE play=no 2 acc:(0.49529)

35. outlook=sunny humidity=high 3 ==> temperature=hot play=no 2 acc:(0.49529)

36. outlook=sunny windy=FALSE 3 ==> humidity=high play=no 2 acc:(0.49529)

37. outlook=sunny play=no 3 ==> humidity=high windy=FALSE 2 acc:(0.49529)

38. outlook=sunny play=no 3 ==> temperature=hot humidity=high 2 acc:(0.49529)

39. outlook=rainy temperature=mild 3 ==> windy=FALSE play=yes 2 acc:(0.49529)

40. outlook=rainy humidity=normal 3 ==> temperature=cool 2 acc:(0.49529)

41. outlook=rainy humidity=normal 3 ==> windy=FALSE play=yes 2 acc:(0.49529)

42. outlook=rainy windy=FALSE 3 ==> humidity=normal play=yes 2 acc:(0.49529)

43. outlook=rainy windy=FALSE 3 ==> temperature=mild play=yes 2 acc:(0.49529)

44. outlook=rainy play=yes 3 ==> temperature=mild windy=FALSE 2 acc:(0.49529)

45. outlook=rainy play=yes 3 ==> humidity=normal windy=FALSE 2 acc:(0.49529)

46. temperature=hot humidity=high 3 ==> outlook=sunny play=no 2 acc:(0.49529)

47. temperature=hot windy=FALSE 3 ==> outlook=overcast play=yes 2 acc:(0.49529)

48. temperature=mild windy=FALSE 3 ==> outlook=rainy play=yes 2 acc:(0.49529)

49. temperature=cool play=yes 3 ==> humidity=normal windy=FALSE 2 acc:(0.49529)

50. humidity=high windy=TRUE 3 ==> temperature=mild 2 acc:(0.49529)

51. humidity=high windy=TRUE 3 ==> play=no 2 acc:(0.49529)

52. humidity=high play=yes 3 ==> outlook=overcast 2 acc:(0.49529)

53. humidity=high play=yes 3 ==> temperature=mild 2 acc:(0.49529)

54. humidity=normal windy=TRUE 3 ==> temperature=cool 2 acc:(0.49529)

55. windy=TRUE play=yes 3 ==> outlook=overcast 2 acc:(0.49529)

56. windy=TRUE play=yes 3 ==> temperature=mild 2 acc:(0.49529)

57. windy=TRUE play=no 3 ==> outlook=rainy 2 acc:(0.49529)

58. humidity=high 7 ==> temperature=mild 4 acc:(0.49376)

59. humidity=high 7 ==> windy=FALSE 4 acc:(0.49376)

60. humidity=normal 7 ==> temperature=cool 4 acc:(0.49376)

61. humidity=normal 7 ==> windy=FALSE play=yes 4 acc:(0.49376)

62. outlook=sunny 5 ==> windy=FALSE 3 acc:(0.48941)

63. outlook=sunny 5 ==> humidity=high play=no 3 acc:(0.48941)

64. outlook=rainy 5 ==> temperature=mild 3 acc:(0.48941)

65. outlook=rainy 5 ==> humidity=normal 3 acc:(0.48941)

66. outlook=rainy 5 ==> windy=FALSE play=yes 3 acc:(0.48941)

67. play=no 5 ==> windy=TRUE 3 acc:(0.48941)

68. play=no 5 ==> outlook=sunny humidity=high 3 acc:(0.48941)

69. windy=FALSE 8 ==> humidity=high 4 acc:(0.46991)

70. windy=FALSE 8 ==> humidity=normal play=yes 4 acc:(0.46991)

71. temperature=mild 6 ==> outlook=rainy 3 acc:(0.46058)

72. temperature=mild 6 ==> windy=TRUE 3 acc:(0.46058)

73. windy=TRUE 6 ==> temperature=mild 3 acc:(0.46058)

74. windy=TRUE 6 ==> humidity=high 3 acc:(0.46058)

75. windy=FALSE play=yes 6 ==> outlook=rainy 3 acc:(0.46058)

76. outlook=overcast 4 ==> humidity=high play=yes 2 acc:(0.44678)

77. outlook=overcast 4 ==> temperature=hot windy=FALSE 2 acc:(0.44678)

78. temperature=hot 4 ==> outlook=overcast windy=FALSE 2 acc:(0.44678)

79. temperature=hot 4 ==> outlook=sunny humidity=high 2 acc:(0.44678)

80. temperature=cool 4 ==> humidity=normal windy=TRUE 2 acc:(0.44678)

81. temperature=cool 4 ==> outlook=rainy humidity=normal 2 acc:(0.44678)

82. outlook=overcast play=yes 4 ==> windy=TRUE 2 acc:(0.44678)

83. temperature=mild humidity=high 4 ==> play=no 2 acc:(0.44678)

84. temperature=mild play=yes 4 ==> outlook=rainy windy=FALSE 2 acc:(0.44678)

85. humidity=high play=no 4 ==> temperature=hot 2 acc:(0.44678)

86. humidity=normal windy=FALSE 4 ==> outlook=rainy 2 acc:(0.44678)

87. play=yes 9 ==> outlook=overcast 4 acc:(0.44601)

88. play=yes 9 ==> temperature=mild 4 acc:(0.44601)

89. play=yes 9 ==> humidity=normal windy=FALSE 4 acc:(0.44601)

90. humidity=high 7 ==> temperature=hot 3 acc:(0.43225)

91. humidity=high 7 ==> outlook=sunny play=no 3 acc:(0.43225)

92. humidity=normal 7 ==> outlook=rainy 3 acc:(0.43225)

93. humidity=normal 7 ==> windy=TRUE 3 acc:(0.43225)

94. humidity=normal 7 ==> temperature=cool play=yes 3 acc:(0.43225)

95. outlook=sunny 5 ==> temperature=mild 2 acc:(0.41235)

96. outlook=sunny 5 ==> temperature=hot humidity=high 2 acc:(0.41235)

97. outlook=rainy 5 ==> temperature=mild humidity=high 2 acc:(0.41235)

98. outlook=rainy 5 ==> temperature=mild windy=FALSE 2 acc:(0.41235)

99. play=no 5 ==> outlook=sunny temperature=hot 2 acc:(0.41235)

2nd Run

Rules were Mined

Scheme: weka.associations.PredictiveApriori -N 100 -A -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

=== Associator model (full training set) ===

PredictiveApriori

===================

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 acc:(0.97594)

2. humidity=normal windy=FALSE 4 ==> play=yes 4 acc:(0.97594)

3. outlook=sunny humidity=high 3 ==> play=no 3 acc:(0.96426)

4. outlook=rainy windy=FALSE 3 ==> play=yes 3 acc:(0.96426)

5. outlook=sunny temperature=hot 2 ==> play=no 2 acc:(0.94406)

6. outlook=sunny humidity=normal 2 ==> play=yes 2 acc:(0.94406)

7. outlook=rainy windy=TRUE 2 ==> play=no 2 acc:(0.94406)

8. temperature=mild humidity=normal 2 ==> play=yes 2 acc:(0.94406)

9. temperature=cool windy=FALSE 2 ==> play=yes 2 acc:(0.94406)

10. humidity=normal 7 ==> play=yes 6 acc:(0.79098)

11. temperature=cool 4 ==> play=yes 3 acc:(0.64898)

12. windy=FALSE 8 ==> play=yes 6 acc:(0.6316)

13. outlook=sunny windy=FALSE 3 ==> play=no 2 acc:(0.60315)

14. outlook=rainy temperature=mild 3 ==> play=yes 2 acc:(0.60315)

15. temperature=hot humidity=high 3 ==> play=no 2 acc:(0.60315)

16. temperature=mild windy=TRUE 3 ==> play=yes 2 acc:(0.60315)

17. humidity=high windy=TRUE 3 ==> play=no 2 acc:(0.60315)

18. temperature=mild 6 ==> play=yes 4 acc:(0.58611)

19. outlook=sunny 5 ==> play=no 3 acc:(0.56112)

20. outlook=rainy 5 ==> play=yes 3 acc:(0.56112)

21. humidity=high 7 ==> play=no 4 acc:(0.5531)

22. temperature=hot 4 ==> play=yes 2 acc:(0.53349)

23. temperature=hot 4 ==> play=no 2 acc:(0.53349)

24. windy=TRUE 6 ==> play=yes 3 acc:(0.52917)

25. windy=TRUE 6 ==> play=no 3 acc:(0.52917)

26. humidity=high 7 ==> play=yes 3 acc:(0.50226)

27. outlook=sunny 5 ==> play=yes 2 acc:(0.5019)

28. outlook=rainy 5 ==> play=no 2 acc:(0.5019)

29. temperature=mild 6 ==> play=no 2 acc:(0.47442)

30. windy=FALSE 8 ==> play=no 2 acc:(0.425

Hotspot

The Hotspot association algorithm focuses on a target variable or value, maximizing or minimizing that target. Hotspot generates a set of rules in a tree-like structure. Hotspot will work on both a nominal or numeric target looking for outliers or values above/below average. Run 1 and 3 both used a minimum improvement of target of 1% but the 3rd run was chosen to minimize the target which resulted in a slightly different rule tree. A minimum improvement target of 10% was specified for run 2 which resulted in trimming the rule tree down to the first instance.

1st Run

Scheme: weka.associations.HotSpot -c 2 -V first -S 0.33 -M 2 -I 0.01

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

=== Associator model (full training set) ===

Hot Spot

========

Total population: 14 instances

Target attribute: temperature

Minimum segment size: 5 instances (33% of total population)

Maximum branching factor: 2

Minimum improvement in target: 1%

temperature (73.5714)

outlook = sunny (76.2 [5])

humidity > 80 (76.1429 [7])

| humidity <= 91 (78.2 [5])

2nd Run

Minimum improvement in target: 10%

temperature (73.5714)

3rd Run

Minimize Target

Minimum improvement in target: 1%

temperature (73.5714)

outlook = rainy (69.8 [5])

windy = TRUE (71.1667 [6])