Session 6: External Measures III: Pairwise Measures

In-video Quiz

❑ The following table summarizes the clustering result of two clustering algorithms, algorithm A and B.

❑ Which of the following statements are correct?

   1) Based on Jaccard coefficient, Algorithm A is better;

   2) Based on Rand Statistic, Algorithm A is better;

   3) Fowlkes-Mallow Measure, Algorithm A is better;

❑ Answer: 1st and 3rd

❑ Explanation:

   ❑ For algorithm A, we have that

Algorithm A			
$C\backslash T$	T_1	T_2	Sum
C_1	8	3	11
C_2	1	6	7
m_j	9	9	18

Algorithm B			
$C\backslash T$	T_1	T_2	Sum
C_1	7	2	9
C_2	2	7	9
m_j	9	9	18

$$TP = \binom{8}{2} + \binom{3}{2} + \binom{6}{2} = 46$$

$$FN = \binom{9}{2} + \binom{9}{2} - TP = 26$$

$$FP = \binom{11}{2} + \binom{7}{2} - TP = 30$$

$$TN = \binom{18}{2} - TP - FN - FP = 51$$

   ❑ For algorithm B, we have that

$$TP = \binom{7}{2} + \binom{2}{2} + \binom{2}{2} + \binom{7}{2} = 44$$

$$FN = \binom{9}{2} + \binom{9}{2} - TP = 28$$

$$FP = \binom{9}{2} + \binom{9}{2} - TP = 28$$

$$TN = \binom{18}{2} - TP - FN - FP = 53$$

❑ For Jaccard coefficient
   ❑ Algorithm A: *Jaccard* = 0.451
   ❑ Algorithm B: *Jaccard* = 0.44
   ❑ Thus, Algorithm A is better

❑ For **Rand Statistic**:
   ❑ Algorithm A: 0.634
   ❑ Algorithm B: 0.634
   ❑ Thus, A and B are comparable;

❑ For **Fowlkes-Mallow Measure**
   ❑ Algorithm A: 0.622
   ❑ Algorithm B: 0.611
   ❑ Thus, Algorithm A is better