A Note on Approximate Randomization Test of F-measure

Hiroya Takamura
takamura@pi.titech.ac.jp

February 18, 2015

Abstract

approximate randomization test

1 Preparation

case number	baseline	proposed	answer
0	yes	no	yes
1	yes	no	no
2	no	yes	yes
3	no	yes	no
4	yes	yes	yes
5	yes	yes	no
6	no	no	yes
7	no	no	no

Let t_i denote the number of instances in case i.

2 Approximate Randomization Test for F-measure

Approximate Randomization Test consists of multiple shuffles. Let ns (say, 10000) denote the number of shuffles. We compute the number nge of shuffles in which the difference between two pseudo F-measures is greater than or equal to the difference between the actual F-measures.

At each shuffle, we need to randomly determine "exchange or not" for each instance. But, the exchange of instances in the same case have the same effect on the randomized F-measure. So, we randomly draw a number k_i which indicates "how many instances in each case i are to be exchanged".

$$\Delta_{actual} = |F^{baseline} - F^{proposed}| \tag{1}$$

For each iteration i, we compute the following:

$$k_0 \sim {}_{t_0}C_{k_0}(\frac{1}{2})^{t_0} \tag{2}$$

$$k_1 \sim {}_{t_1}C_{k_1}(\frac{1}{2})^{t_1} \tag{3}$$

$$k_2 \sim {}_{t_2}C_{k_2}(\frac{1}{2})^{t_2} \tag{4}$$

$$k_3 \sim {}_{t_3}C_{k_3}(\frac{1}{2})^{t_3}. \tag{5}$$

$$F^{baseline}_{pseudo}(k_0, k_1, k_2, k_3) = \frac{2 \times \frac{t_0+t_4-k_0+k_2}{t_0+t_1+t_4+t_5-(k_0+k_1)+(k_2+k_3)} \times \frac{t_0+t_4-k_0+k_2}{t_0+t_2+t_4+t_6}}{\frac{t_0+t_4-k_0+k_2}{t_0+t_1+t_4+t_5-(k_0+k_1)+(k_2+k_3)} + \frac{t_0+t_4-k_0+k_2}{t_0+t_2+t_4+t_6}} \tag{6}$$

$$F_{pseudo}^{proposed}(k_0, k_1, k_2, k_3) = \frac{2 \times \frac{t_2+t_4-k_2+k_0}{t_2+t_3+t_4+t_5-(k_2+k_3)+(k_0+k_1)} \times \frac{t_2+t_4-k_2+k_0}{t_0+t_2+t_4+t_6}}{\frac{t_2+t_4-k_2+k_0}{t_2+t_3+t_4+t_5-(k_2+k_3)+(k_0+k_1)} + \frac{t_2+t_4-k_2+k_0}{t_0+t_2+t_4+t_6}} \quad (7)$$

$$\Delta_{pseudo} = |F_{pseudo}^{baseline} - F_{pseudo}^{proposed}| \quad (8)$$

$$\text{If} \quad \Delta_{pseudo} \leq \Delta_{actual} \quad (9)$$

$$\text{then} \quad nge++ \quad (10)$$

This way we obtain nge. Then, the fraction

$$\frac{nge+1}{ns+1} \quad (11)$$

gives the significance level.

3 Conclusion

randomization test.

This note was written with reference to the papers by Chinchor et al. [1] and Efron and Tibshirani [2].

References

[1] Nancy Chinchor, Lynette Hirschman, and David D. Lewis. Evaluating message understanding systems: An analysis of the third Message Understanding Conference (MUC-3). *Computational Linguistics*, 19(3):409–449, 1993.

[2] Bradley Efron and Robert Tibshirani. Statistical data analysis in the computer age. *Science*, 253:390–395, 1991.