Concept Paper

F*, an interpretable transformation of the F measure, equates to the critical success index

Gashirai K. Mbizvo and Andrew J. Larner *

Posted Date: 7 September 2023

doi: 10.20944/preprints202309.0556.v1

Keywords: Binary classification; Critical success index; F measure; F*; Jaccard index

Concept Paper

F*, an Interpretable Transformation of the F Measure, Equates to the Critical Success Index

Gashirai K. Mbizvo [1,2] and Andrew J. Larner [3]

[1] Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, United Kingdom
[2] Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom
[3] Cognitive Function Clinic, The Walton Centre NHS Foundation Trust, Liverpool, United Kingdom
* Correspondence: Cognitive Function Clinic, Walton Centre for Neurology and Neurosurgery, Liverpool, United Kingdom; andrew.larner2@nhs.net

Abstract: Recently a measure termed F* was described, as an interpretable transformation of the F measure, also known as the Dice coefficient. Using elementary mathematical methods, it is shown that F* is in fact identical to a previously described measure, monotonically related to the F measure, and variously termed in previous publications, dating from the late 19th to the late 20th century, as the ratio of verification, the Jaccard similarity measure or index, the threat score, the Tanimoto index, and the critical success index. The origins of these different terms in different disciplines (weather forecasting, ecology, machine learning) may explain the repeated independent redescription of this measure.

Keywords: binary classification; critical success index; F measure; F*; Jaccard index

1. Introduction

Many measures may be derived from the data in a 2x2 contingency table.[1] One of these is the F measure, defined as the harmonic mean of precision (or positive predictive value, PPV) and recall (or sensitivity, Sens).[2] This corresponds to the coefficient described by Dice[3] and independently by Sørensen,[4] sometimes known as the Dice coefficient or the Sørensen-Dice coefficient, and to the approach advocated by van Rijsbergen.[5]

In terms of the base data from a 2x2 contingency table containing N elements with four degrees of freedom (where TP = true positive, FP = false positive, FN = false negative, TN = true negative):

$$F = 2.TP/(2.TP + FP + FN)$$

or in terms of PPV and Sens:

$$F = 2.PPV.Sens/(PPV + Sens)$$
$$= 2/[1/Sens + 1/PPV]$$

that is, F is the harmonic mean of PPV and Sens.

More recently, Hand et al. (2021) have described "F*" as "an interpretable transformation of the F measure,"[6] where:

$$F^* = F/(2 - F)$$

As will be shown, these authors have in fact redescribed an already existing binary classification measure, first reported in the late nineteenth century as the ratio of verification in the context of forecasting tornadoes,[7] and subsequently as the Jaccard index or similarity coefficient (J),[8] the threat score,[9] the Tanimoto index,[10] and later still as the critical success index (CSI).[11,12] Here we use the latter terminology.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2023 doi:10.20944/preprints202309.0556.v1

2. Mathematical proofs of identity of F* and CSI

The identity of F* and CSI may be shown in several ways using elementary mathematical methods.

2a. From the base data of a 2x2 contingency table.

Hand et al.[6] showed that:

$$F^* = TP/(N - TN)$$

This also holds for CSI, since in terms of the base data:

$$CSI = TP/(TP + FP + FN)$$
$$= TP/(N - TN)$$

Hence F* = CSI, QED.

2b. From the monotonic relationship of F to CSI.

The monotonic relationship between F and CSI, as shown for example by Jolliffe[13] (modified), is given by:

$$F = 2.CSI/(1 + CSI)$$

The equivalence of F* and CSI may thus be shown. Since Hand et al.[6] showed that:

$$F^* = F/(2 - F)$$

Then rearranging:

$$F = F^*.(2 - F)$$
$$= 2.F^* - F^*.F$$

Dividing through by F and rearranging:

$$(2.F^*/F) - F^* = 1$$
$$F^* + 1 = 2.F^*/F$$

Hence:

$$F = 2.F^*/(F^* + 1)$$

Hence F* = CSI, QED.

2c. From the combination of PPV (or precision) and Sens (or recall).

Like F, CSI may be characterised in terms of PPV and Sens:

$$CSI = 1/[(1/PPV) + (1/Sens) - 1]$$

Again, the equivalence of F* and CSI may be shown. Hand et al.[6] found that:

$$F^* = (PPV \times Sens)/PPV + Sens - (PPV \times Sens)$$

Dividing through by (PPV x Sens) gives:

$$F^* = 1/[(1/Sens) + (1/PPV) - 1]$$

Hence F* = CSI, QED.

2d. From the combination of Sens, PPV, P, and Q.

In the 2x2 contingency table, prevalence or base rate P = (TP + FN)/N, and bias or threshold Q = (TP + FP)/N. Thus, from Powers[2]:

$$F = 2.Sens.P/(Q + P)$$
$$= 2.PPV.Q/(Q + P)$$

For CSI the equations are[1]:

$$CSI = 1/[(Q + P)/Sens.P] - 1$$
$$= 1/[(Q + P)/PPV.Q] - 1$$

 doi:10.20944/preprints202309.0556.v1

Since from Hand et al.[6]:

$$F^* = F/(2 - F)$$

Then substituting and rearranging:

$$F^* = [2.Sens.P/(Q + P)]/(2 - [2.Sens.P/(Q + P)])$$
$$= 1/[(Q + P)/Sens.P) - 1$$

$$F^* = [2.PPV.Q/(Q + P)]/(2 - [2.PPV.Q/(Q + P)])$$
$$= 1/[(Q + P)/PPV.Q) - 1$$

Hence F* = CSI, QED.

3. Conclusion

Hand et al. noted that "researchers may recognise this [i.e. F*] as the Jaccard coefficient widely used in areas where TN may not be relevant"[6] and they cite Jaccard's 1908 paper,[14] although others[2] cite his 1901 paper[15] as the forerunner of the 1912 English translation.[8]

We suggest that this is a parameter which, like F, has undergone periodic redescriptions (or convergent evolution). The first report of which we are aware is Gilbert's "ratio of verification" of 1884,[7] predating the Jaccard similarity coefficient.[8] This latter measure is equivalent in set theory to union over intersection, which was also proposed by Tanimoto in 1958 when working for IBM,[10] without reference to either Gilbert or Jaccard. The same measure has also been described by Palmer & Allen in 1949 as the threat score,[9] and as the critical success index by Donaldson et al.[11] in 1975 and by Schaefer[12] in 1990, and now as F* by Hand et al.[6] These multiple redescriptions may reflect use of this measure by researchers in different disciplines (weather forecasting, ecology, machine learning) unaware of prior authors and unbeknownst to later authors.

The critical success index has recently been exported to the domain of clinical medicine, for example to evaluate the accuracy of instruments used in day-to-day clinical practice for screening cognitive function in patients with possible dementia or mild cognitive impairment,[16] as well as in diagnostic accuracy studies of administrative epilepsy data.[17] In these studies the identity of F* and CSI has been confirmed using the respective datasets. We have also suggested possible application of CSI in assessing both NICE criteria for 2-week-wait suspected brain and CNS cancer referrals[18] and polygenic hazard scores.[19] These are all situations in which large numbers of TN may complicate the interpretation of more traditional measures such as PPV and Sens.

Funding: None.

Availability of data and materials: No datasets used and/or analysed during the current study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Larner AJ. *The 2x2 matrix. Contingency, confusion, and the metrics of binary classification* (2nd edition). London: Springer, 2024 (in press).
2. Powers DMW. What the F measure doesn't measure … Features, flaws, fallacies and fixes. *arXiv* 2015; doi:1503.06410.2015.
3. Dice LR. Measures of the amount of ecological association between species. *Ecology* 1945; 26(3): 297-302.
4. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 1948; 5(4): 1-34.
5. van Rijsbergen CJ. Foundation of evaluation. *Journal of Documentation* 1974; 30: 365-373.
6. Hand DJ, Christen P, Kirielle N. F*: an interpretable transformation of the F measure. *Machine Learning* 2021; 110(3): 451-456.
7. Gilbert GK. Finley's tornado predictions. *American Meteorological Journal* 1884; 1: 166-172.
8. Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist* 1912; 11: 37-50.

9. Palmer WC, Allen RA. *Note on the accuracy of forecasts concerning the rain problem*. U.S. Weather Bureau manuscript: Washington, DC., 1949.
10. Tanimoto TT. *An elementary mathematical theory of classification and prediction*. Internal IBM Technical Report 17th November 1958. http://dalkescientific.com/tanimoto.pdf
11. Donaldson RJ, Dyer RM, Kraus MJ. An objective evaluator of techniques for predicting severe weather events. *Preprints, 9th Conference on Severe Local Storms*. Norman, Oklahoma, 1975: 312-326.
12. Schaefer JT. The critical success index as an indicator of warning skill. *Weather Forecasting* 1990; 5, 570-575.
13. Jolliffe IT. The Dice co-efficient: a neglected verification performance measure for deterministic forecasts of binary events. *Meteorological Applications* 2016; 23(1): 89-90.
14. Jaccard, P. Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles* 1908; 44: 223-270.
15. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 1901; 37: 547-579.
16. Larner AJ. Assessing cognitive screening instruments with the critical success index. *Progress in Neurology and Psychiatry* 2021; 25(3): 33-37.
17. Mbizvo GK, Bennett KH, Simpson CR, Duncan SE, Chin RFM, Larner AJ. Using Critical Success Index or Gilbert Skill Score as composite measures of positive predictive value and sensitivity in diagnostic accuracy studies: weather forecasting informing epilepsy research. *Epilepsia* 2023; 64: 1466-1468.
18. Mbizvo GK, Larner AJ. Isolated headache is not a reliable indicator for brain cancer. *Clinical Medicine* 2022; 22(1): 92-93.
19. Mbizvo GK, Larner AJ. Re: Realistic expectations are key to realising the benefits of polygenic scores. *BMJ* https://www.bmj.com/content/380/bmj-2022-073149/rapid-responses (Published 11 March 2023).