THE MCC APPROACHES THE GEOMETRIC MEAN OF PRECISION AND RECALL AS TRUE NEGATIVES APPROACH INFINITY.

A PREPRINT

Jon Crall[*]
Kitware Inc.
jon.crall@kitware.com

November 28, 2023

ABSTRACT

The performance of a binary classifier is described by a confusion matrix with four entries: the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The Matthew's Correlation Coefficient (MCC), F1, and Fowlkes–Mallows (FM) scores are scalars that summarize a confusion matrix. Both the F1 and FM scores are based on only three of the four entries in a confusion matrix (they ignore TN). In contrast, the MCC takes into account all four entries of a confusion matrix and thus can be seen as providing a more representative picture.

However, in object detection problems, measuring the number of true negatives is so large it is often intractable. Thus we ask, what happens to the MCC as the number of true negatives approaches infinity? This paper provides insight into the relationship between the MCC and FM score by proving that the FM-measure is equal to the limit of the MCC as the number of true negatives approaches infinity.

Keywords Confusion Matrix · Binary Classification · Fowlkes–Mallows Index · Matthew's Correlation Coefficient · F1

1 Introduction

Evaluation of binary classifiers is central to the quantitative analysis of machine learning models [1]. Given a finite set of examples with known real labels, the quality of a set of corresponding predicted labels can quantified using a 2 × 2 confusion matrix. A confusion matrix counts the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) a model predicts with respect to the real labels. A confusion matrix is written as:

$$\begin{bmatrix} \texttt{TP} & \texttt{FP} \\ \texttt{FN} & \texttt{TN} \end{bmatrix} \tag{1}$$

This matrix provides a holistic view of classifier quality, however, it is often desirable to summarize performance using fewer numbers. Two popular metrics defined on a classification matrix are precision and recall.

Precision — also known as the positive-predictive-value (PPV) — is the fraction of positive predictions that are correct.

$$\texttt{PPV} = \frac{\texttt{TP}}{\texttt{TP} + \texttt{FP}} \tag{2}$$

Recall — also known as the true positive rate (TPR), sensitivity, or probability of detection (PD) — is the fraction of real positive cases that are correct.

$$\texttt{TPR} = \frac{\texttt{TP}}{\texttt{TP} + \texttt{FN}} \tag{3}$$

[*] https://github.com/Erotemic/ erotemic@gmail.com

One of the most popular confusion metrics is the F1 score. It can be defined as the harmonic mean of precision and recall.

$$\texttt{F1} = \frac{2\texttt{PPV} \cdot \texttt{TPR}}{\texttt{PPV} + \texttt{TPR}} = \frac{2\texttt{TP}}{2\texttt{TP} + \texttt{FP} + \texttt{FN}} \tag{4}$$

A similar, but less used metric is the Fowlkes–Mallows index [2], which was originally developed for measuring the similarity between two clusterings of a set of points. It can be defined as the geometric mean of precision and recall [3].

$$\texttt{FM} = \sqrt{\texttt{PPV} \cdot \texttt{TPR}} = \sqrt{\frac{\texttt{TP}}{\texttt{TP} + \texttt{FP}} \frac{\texttt{TP}}{\texttt{TP} + \texttt{FN}}} \tag{5}$$

In [1], Powers notes that the F1 score (and consequentially any metric that only includes precision and recall) only takes into account three of the four measures in a confusion matrix. Powers, introduces modifications of precision and recall he refers to as informedness and markedness. Additionally he advocates for the use of the MCC over the F1 measure.

The Matthews Correlation Coefficient (MCC) [4] accounts for all four terms in the confusion matrix and is defined as:

$$\texttt{MCC} = \frac{\texttt{TP} \cdot \texttt{TN} - \texttt{FP} \cdot \texttt{FN}}{\sqrt{(\texttt{TP} + \texttt{FP})(\texttt{TP} + \texttt{FN})(\texttt{TN} + \texttt{FP})(\texttt{TN} + \texttt{FN})}} \tag{6}$$

While the MCC is a desirable measure due to its balanced inclusion of all terms in a confusion matrix, it requires that the number of true negatives is measurable. In the case of object detection problems [5], this is often intractable as the number of the number of predicted boxes and missed true boxes is dwarfed by the total number of boxes that the system correctly did not predict. One can see this by considering the set of all $N \times M$ boxes centered at each pixel, most of which will be considered true negatives. If the width and height of the boxes are allowed to extend outside the image, then the number of predictable boxes actually is unbounded (and even if they must be contained in the image, there will still be a very large number of them in real world cases).

Because calculating the number of true negatives is difficult for open-world problems like object detection, it is conceptually simpler to ignore true negatives and simply focus on the much smaller set of true positives, false positives, and false negatives, which can be used to compute PPV, TPR, F1, and FM. While these measures have proven themselves to be effective, simply ignoring true negatives is somewhat unsatisfying. We seek to remedy this noting that in these open-world problems the number of true negatives is so large it is effectively infinite and thus we ask the question: what happens to the MCC as the number of true negatives approaches infinity?

The main contribution of this paper is to highlight a relationship between the MCC and the FM score. The MCC reduces to FM as the number of true negatives approaches infinity. While this is not a difficult result to show, to the best of the author's knowledge, this was first shown in a blog post [6], but has not yet been published. This paper is a more formal description of this result. Specifically, the contributions are:

- We informally (but rigorously) prove the statement $\lim_{\texttt{TN}\to\infty} \texttt{MCC} = \texttt{FM}$.

2 The Relationship Between MCC and FM

Taking the limit of the MCC Consider the limit of the MCC as the number of true negatives approaches infinity.

$$\lim_{\texttt{TN}\to\infty} \texttt{MCC} = \lim_{\texttt{TN}\to\infty} \frac{\texttt{TP} \cdot \texttt{TN} - \texttt{FP} \cdot \texttt{FN}}{\sqrt{(\texttt{TP} + \texttt{FP})(\texttt{TP} + \texttt{FN})(\texttt{TN} + \texttt{FP})(\texttt{TN} + \texttt{FN})}} \tag{7}$$

We can take this limit by applying some algebra to the body of the limit. We multiply the numerator and denominator by $\frac{1}{\texttt{TN}}$:

$$= \lim_{\texttt{TN}\to\infty} \frac{\frac{1}{\texttt{TN}}(\texttt{TP} \cdot \texttt{TN} - \texttt{FP} \cdot \texttt{FN})}{\frac{1}{\texttt{TN}}\sqrt{(\texttt{TP} + \texttt{FP})(\texttt{TP} + \texttt{FN})(\texttt{TN} + \texttt{FP})(\texttt{TN} + \texttt{FN})}} \tag{8}$$

We distribute the $\frac{1}{\texttt{TN}}$ term in the numerator and denominator:

$$= \lim_{\mathtt{TN}\to\infty} \frac{(\mathtt{TP} - \mathtt{FP} \cdot \frac{\mathtt{FN}}{\mathtt{TN}})}{\sqrt{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})(\frac{\mathtt{TN}+\mathtt{FP}}{\mathtt{TN}})(\frac{\mathtt{TN}+\mathtt{FN}}{\mathtt{TN}})}} \tag{9}$$

The $\frac{\mathtt{TN}}{\mathtt{TN}}$ terms in the denominator cancel:

$$= \lim_{\mathtt{TN}\to\infty} \frac{(\mathtt{TP} - \mathtt{FP} \cdot \frac{\mathtt{FN}}{\mathtt{TN}})}{\sqrt{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})(1+\frac{\mathtt{FP}}{\mathtt{TN}})(1+\frac{\mathtt{FN}}{\mathtt{TN}})}} \tag{10}$$

The terms involving $\mathtt{TN}$ are fractions of simple rational polynomials (w.r.t. $\mathtt{TN}$) and in each case the degree of the denominator is greater than that of the numerator, so in the limit each of these terms simplifies to 0. Thus, the entire equation simplifies to:

$$= \frac{(\mathtt{TP} - \mathtt{FP} \cdot 0)}{\sqrt{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})(1+0)(1+0)}} \tag{11}$$

Thus we find that the limit of the MCC as true negatives approach infinity is:

$$= \frac{\mathtt{TP}}{\sqrt{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})}} \tag{12}$$

Rearranging the FM Now rearranging the equation for FM, we find it is equivalent to the limit of the MCC as the number of true negatives approaches infinity.

$$\begin{aligned}
\mathtt{FM} &= \sqrt{\frac{\mathtt{TP}}{\mathtt{TP}+\mathtt{FP}} \frac{\mathtt{TP}}{\mathtt{TP}+\mathtt{FN}}} && (13)\\
&= \sqrt{\frac{\mathtt{TP}^2}{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})}} && (14)\\
&= \frac{\mathtt{TP}}{\sqrt{(\mathtt{TP}+\mathtt{FP})(\mathtt{TP}+\mathtt{FN})}} && (15)\\
&= \lim_{\mathtt{TN}\to\infty} \mathtt{MCC} && (16)
\end{aligned}$$

Verifying the proof The correctness of these claims can be verified using SymPy [7]. We define a symbolic expression for the definition of the MCC and FM score. We then use SymPy to determine the limit of the MCC as $\mathtt{TN} \to \infty$. Finally we subtract expressions that we claim are equal, which will result in zero only if they are equal.

```python

from sympy import sqrt, symbols, simplify
from sympy.series import limit

tp, tn, fp, fn = symbols("tp tn fp fn",
                         integer=True, negative=False)

# The definition of the MCC
numer = (tp * tn - fp * fn)
denom = sqrt((tp + fp) * (tp + fn) * (tn + fp) * (tn + fn))
mcc = numer / denom

# The definition of FM
FM = sqrt((tp / (tp + fn)) * (tp / (tp + fp)))

# Compute the limit of the MCC definition
mcc_lim = limit(mcc, tn, float('inf'))

# We claim the limit of the MCC and the FM are equivalant to:
```

```
mcc_lim_claim = tp / sqrt((tp + fn) * ((tp + fp)))

# Check the claim is equal to FM
assert simplify(FM - mcc_lim_claim) == 0
# Check the claim is equal to the MCC limit
assert simplify(mcc_lim - mcc_lim_claim) == 0
```

The above program does not raise an AssertionError, thus we have proven $\lim_{\texttt{TN}\to\infty} \texttt{MCC} = \texttt{FM}$, □.

3 Conclusion

This paper proves that the limit of the MCC as the number of true negatives goes to infinity is equivalent to the Fowlkes–Mallows index (i.e. the geometric mean of precision and recall).

This is a useful insight in open world problems where the number of true negative cases grows faster than the number of other confusion categories. It validates the use of precision and recall as a way of describing the quality of object detection results and hints that the FM score may be a preferable alternative to the more standard F1 score.

4 Acknowledgements

Thanks to Lee Newberg for the cleaner formulation of eqs. (8) to (10) to achieve eq. (11) without needing L'Hôpital's rule.

References

[1] David Martin Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. volume 2, pages 37–63, February 2011. URL http://dspace2.flinders.edu.au/xmlui/handle/2328/27165. 00000.

[2] E. B. Fowlkes and C. L. Mallows. A Method for Comparing Two Hierarchical Clusterings. volume 78, pages 553–569, 1983. doi:10.2307/2288117. URL https://www.jstor.org/stable/2288117. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[3] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 17(1):168–192, January 2020. ISSN 2210-8327. doi:10.1016/j.aci.2018.08.003. URL https://doi.org/10.1016/j.aci.2018.08.003. Publisher: Emerald Publishing Limited.

[4] B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. volume 405, pages 442–451, October 1975. doi:10.1016/0005-2795(75)90109-9. URL https://www.sciencedirect.com/science/article/pii/0005279575901099.

[5] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023.

[6] Jon Crall. Closed form of the mcc when tn->inf, 2019. URL https://erotemic.wordpress.com/2019/10/23/closed-form-of-the-mcc-when-tn-inf/. [Online; accessed 30-April-2023].

[7] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. Sympy: symbolic computing in python. volume 3, page e103, January 2017. doi:10.7717/peerj-cs.103. URL https://doi.org/10.7717/peerj-cs.103.