The F-Measure Paradox

Presentation at ADVCOMP 2020

Tim vor der Brück
Fernfachhochschule Schweiz (FFHS)
Brig, Switzerland
tim.vorderbrueck@ffhs.ch

October 2020

Lucerne University of Applied Sciences and Arts HOCHSCHULE LUZERN

FFHS Fernfachhochschule Schweiz Zürich | Basel | Bern | Brig
Mitglied der SUPSI

CV

- 2001: Diploma Degree at Saarland University
- 2001-2003: Scientific Assistant at the German Research Center for Artificial Intelligence (DFKI)
- 2003-2006: Scientist at the German Meteorological Service
- 2006-2010: PhD Student at the Distance University of Hagen
- 2011-2015: Postdoc at Goethe University Frankfurt am Main
- 2015-now: Research Associate at Lucerne University of Applied Science and Arts
- 2019-now: Lecturer at FFHS (Fernfachhochschule Schweiz)

Paradoxes in Computer Science and Mathematics

- Paradoxes have always fascinated people
- Typical characteristics: They exhibit a surprising behavior that is contrary to people's believes.
- There are quite a few identified paradoxes in mathematic and computer science.

Example: Proposition of Russel

There is not set that contains exactly the sets that does not contain itself
Proof by contradiction: Assume such a set exist. Does it contain itself?

Example: Proposition of Russel

- Case 1: It contains itself. This would contradict the assumption, that it can only contains sets that does not contain itself.
- Case 2: It does not contain itself. Then this must must contain it, since it contains all sets that does not contains itself.
  Both case 1 and case 2 lead to a contradiction. Therefore such a set cannot exists.

Banach-Tarski-Paradox

- published in 1924 by Stefan Banach and Alfred Tarski
- First, a sphere is decomposed into parts
- By putting these parts together, one obtains two spheres of the same volume as the original
- It is named a paradox since it contradicts geometric intuition

Accuracy Paradox

		Prediction outcome p	Prediction outcome n	total
Actual value	p′	8	2	P′
	n′	12	9978	N′
	total	P	N	

- Obtained Accuracy of model above: 0.9986
- Predicting always the majority class: 0.999
- ⇒ A machine learning model with lower accuracy can have higher predictory performance

Properties of the harmonic mean

- harmonic mean (HM) of two input values a,b always assumes a value inside the interval [a,b]
- HM is drawn to the smaller one of the two input values
- HM is zero, if one the input values is zero
- If the HM coincides with one of the input values and is non-zero, then the second argument must also assume this value
- the sign of both input values must coincide
- formula: $H(a,b) = \frac{2ab}{a+b} = \frac{2}{\frac{1}{a}+\frac{1}{b}}$

Properties of the Harmonic Mean

What is $H(0,0)$? Actually $H(0,0)=\frac{2 \cdot 0 \cdot 0}{0+0}=\frac{0}{0}$
However, $H(0,0)=0$ is a sensible definition considering limits, since:
$\lim_{a\to 0, b\to 0, sign(a)=sign(b)} \frac{2}{\frac{1}{a}+\frac{1}{b}}=\frac{2}{\infty}=0$
Therefore, in the remainder we assume $H(0,0)=0$

Harmonic Mean

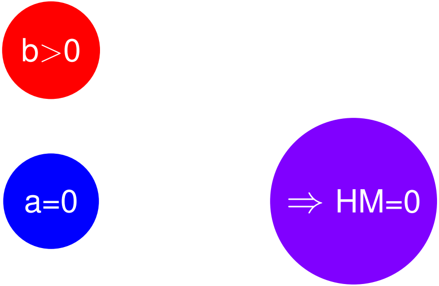
Definitions:

a: Argument 1 — a

b: Argument 2 — b

Harmonic mean of a and b — HM

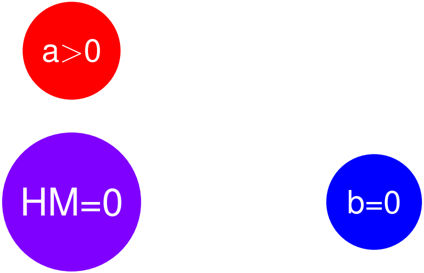
Case: One of the inputs (a) is zero



$$a = 0 \Rightarrow H(a, b) = 0$$

Properties of the harmonic mean

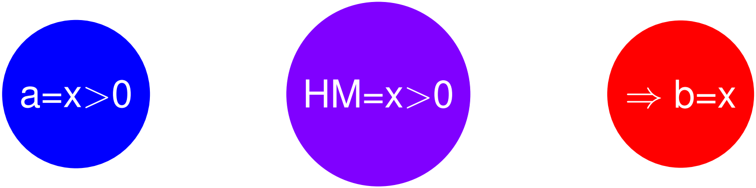
Case: Harmonic mean is zero



$$H(a,b) = 0 \not\Rightarrow a = 0$$

Properties of the harmonic mean

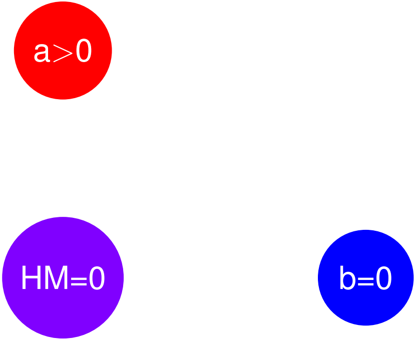
Case: HM equals to one of its inputs and greater zero



$$a \neq 0, a = H(a,b) \Rightarrow b = H(a,b)$$

Properties of the harmonic mean

Case: HM equals to one of its inputs and the other one is greater zero



$$a \neq 0 \wedge b = H(a, b) = 0 \not\Rightarrow a = 0$$

Summary

Proposition	HM	F1
$a = 0 \Rightarrow H(a,b) = 0$	✓	
$H(a,b) = 0 \Rightarrow a = 0$	✗	
$a \neq 0 \wedge H(a,b) = a \Rightarrow b = a$	✓	
$a \neq 0 \wedge H(a,b) = b \Rightarrow b = a$	✗	

Proposition	HM	F1
$a = 0 \Rightarrow H(a,b) = 0$	✓	✗
$H(a,b) = 0 \Rightarrow a = 0$	✗	✓
$a \neq 0 \wedge H(a,b) = a \Rightarrow b = a$	✓	✗
$a \neq 0 \wedge H(a,b) = b \Rightarrow b = a$	✗	✓

We give here proof for line 1 and 2 (3 and 4 see paper). For column F1: a:=prec(ision),b:=rec(all)
Note that *a* and *b* are interchangeable

Bochvar extension: NaN

- Precision (recall) can assume 0/0=NaN (Not a Number), if $TP = 0 \wedge FP = 0 (FN = 0)$
- NaN: absorbing element
- $\mathbb{R} \cup \{NaN\}, +)$ is a semi-group

Computation rules: $a \in \mathbb{R}$

$$\begin{aligned} aNaN =& NaN \\ a + NaN =& NaN \\ a/NaN =& NaN \\ a \cdot NaN =& NaN \end{aligned} \tag{1}$$

Proofs

Proof of properties for F1-Score:
to be shown:

$$prec = 0 \not\Rightarrow F1(prec, rec) = 0$$

Counter-Example:

$$\begin{aligned} TP = 0, FN = 0, FP \neq 0 \\ \Rightarrow rec = NaN, prec = 0 \\ \Rightarrow F1(prec, rec) = NaN \neq 0 \end{aligned} \tag{2}$$

Proof.

$$F1(prec, rec) = 0 \Rightarrow prec \neq NaN, rec \neq NaN$$
$$\Rightarrow TP + FP \neq 0 \qquad (*)$$
$$\Rightarrow \frac{2prec \cdot rec}{prec + rec} = 0$$
$$\Rightarrow prec = 0 \vee rec = 0$$
$$\text{Case 1 :} prec = 0 \Rightarrow \text{Claim}$$
$$\text{Case2 :} rec = 0 \Rightarrow TP = 0$$
$$\Rightarrow \frac{TP}{TP + FP} \overset{*}{=} 0 \Rightarrow prec = 0$$

□

Conclusion

- Properties of the F1-Score contradicts properties of Harmonic mean
- Caused by special relationship between recall and precision and necessary inclusion of NaN

Implications

- Practical implications: basic assumptions about F1 score, precision, recall can be incorrect in certain cases (usually if NaN shows up)
- Shortcomings of proofs in general: NaN-case usually not covered. However, not so rare in practice due to
  - missing observations
  - potential uncomputability of values (NaN)