The truth of the F-measure

1 author:

Yutaka Sasaki
Toyota Technological Institute

The truth of the F-measure

Yutaka Sasaki
Research Fellow
School of Computer Science, University of Manchester
MIB, 131 Princess Street, Manchester, M1 7DN
Yutaka.Sasaki@manchester.ac.uk

October 26, 2007

Abstract

It has been past more than 15 years since the F-measure was first introduced to evaluation tasks of information extraction technology at the Fourth Message Understanding Conference (MUC-4) in 1992. Recently, sometimes I see some confusion with the definition of the F-measure, which seems to be triggered by lack of background knowledge about how the F-measure was derived. Since I was not involved in the process of the introduction or device of the F-measure, I might not be the best person to explain this but I hope this note would be a little help for those who are wondering what the F-measure really is. This introduction is devoted to provide brief but sufficient information on the F-measure.

1 Overview

Definition of the F-measure

The F-measure is defined as a harmonic mean of precision (P) and recall (R):[1]

$$F = \frac{2PR}{P+R}.$$

[1] In biomedicine, precision is called *positive predictive value (PPV)* and recall is called *sensitivity* but to my knowledge, there is nothing corresponding to the F-measure in the domain.

If you are satisfied with this definition and need no further information, that's it. However, if you are deeply interested in the definition of the F-measure, you should recap the definitions of the arithmetic and harmonic means.

Arithmetic and harmonic means

The arithmetic mean A (an average in a usual sense) and the harmonic mean H are defined as follows.

$$\begin{aligned} A &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + ... + x_n). \\ H &= \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + ... + \frac{1}{x_n}}. \end{aligned}$$

When $x_1 = P$ and $x_2 = R$, A and H will be:

$$\begin{aligned} A &= \frac{1}{2}(P + R). \\ H &= \frac{2}{\frac{1}{P} + \frac{1}{R}} \\ &= \frac{2}{\frac{P+R}{PR}} \\ &= \frac{2PR}{P + R}. \end{aligned}$$

The harmonic mean is more intuitive than the arithmetic mean when computing a mean of ratios.

Suppose that you have a finger print recognition system and its precision and recall be 1.0 and 0.2, respectively. Intuitively, the total performance of the system should be very low because the system covers only 20% of the registered finger prints, which means it is almost useless.

The arithmetic mean of 1 and 0.2 is 0.6 whereas the harmonic mean of them is

$$\frac{2 \cdot 1 \cdot \frac{2}{10}}{1 + \frac{2}{10}} = \frac{4}{12} = \frac{1}{3}.$$

As you see in this example, the harmonic mean (0.333...) is a more reasonable score than the arithmetic mean (0.6).

2 Derivation of the F-measure

Some researchers call the definition of the F-measure in the previous section F_1-measure. What is 1 of F_1?

The full definition of the F-measure is given as follows.[Chinchor, 1992]

$$F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P + R} \qquad (0 \leq \beta \leq +\infty).$$

β is a parameter that controls a balance between P and R. When $\beta = 1$, F_1 comes to be equivalent to the harmonic mean of P and R. If $\beta > 1$, F becomes more recall-oriented and if $\beta < 1$, it becomes more precision-oriented, e.g., $F_0 = P$.

While it seems that van Rijsbergen did not define the formula of the F-measure per se, the origin of the definition of the F-measure is van Rijsbergen's E (effectiveness) function [van Rijsbergen, 1979]:

$$E = 1 - \frac{1}{\alpha \dfrac{1}{P} + (1-\alpha)\dfrac{1}{R}},$$

where $\alpha = \frac{1}{\beta^2+1}$.

Let's remove α using β.

$$\begin{aligned} E &= 1 - \frac{1}{\dfrac{1}{\beta^2+1}\dfrac{1}{P} + \left(1 - \dfrac{1}{\beta^2+1}\right)\dfrac{1}{R}}, \\ &= 1 - \frac{PR}{\frac{1}{\beta^2+1}R + \frac{\beta^2+1-1}{\beta^2+1}P}. \\ &= 1 - \frac{(\beta^2+1)PR}{R + \beta^2 P}. \end{aligned}$$

Now you see that

$$E = 1 - F_\beta.$$

Note that F rises if R or P gets better whereas E becomes small if R or P improves. This seems the reason why F is more commonly used than E.

Some people use α as a parameter of F.

$$F_\alpha = \frac{1}{\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}} \qquad (0 \leq \alpha \leq 1).$$

There is nothing wrong with this definition of F but use of this definition might cause an unnecessary confusion because $F_{\alpha=0.5}$=$F_{\beta=1}$. An attention is needed that the commonly used notation F_1 means $F_{\beta=1}$, not $F_{\alpha=1}$.

3 Further investigation in β

Still, some of you are not sure why β^2 is used instead of β in $\alpha = \frac{1}{\beta^2+1}$. The best way to understand this is to read Chapter 7 of van Rijsbergen's masterpiece[van Rijsbergen, 1979]. However, let me try to explanation the reason.

β is the parameter that controls the weighting between P and R. Formally, β is defined as follows:

$$\beta = R/P, \quad where \ \frac{\partial E}{\partial P} = \frac{\partial E}{\partial R}.$$

The motivation behind this condition is that at the point where the gradients of E w.r.t. P and R are equal, the ratio of R against P should be a desired ratio β.

Please recall that E is defined as follows:

$$\begin{aligned} E &= 1 - \frac{1}{\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}}, \\ &= 1 - \frac{PR}{\alpha R + (1-\alpha)P}. \end{aligned}$$

Now we calculate $\frac{\partial E}{\partial P}$ and $\frac{\partial E}{\partial R}$. By the quotient rule on the derivative of a composite function, $(f/g)' = (f'g - fg')/g^2$. For conciseness, let $g = \alpha R + (1-\alpha)P$.

$$\begin{aligned} \frac{\partial E}{\partial P} &= -\frac{R(\alpha R + (1-\alpha)P) - PR(1-\alpha)}{g^2}. \\ \frac{\partial E}{\partial R} &= -\frac{P(\alpha R + (1-\alpha)P) - PR\alpha}{g^2}. \end{aligned}$$

Then, $\frac{\partial E}{\partial P} = \frac{\partial E}{\partial R}$ is equivalent to:

$$R(\alpha R + (1-\alpha)P) - PR(1-\alpha) = P(\alpha R + (1-\alpha)P) - PR\alpha,$$

which can be simplified to:

$$\alpha R^2 = (1-\alpha)P^2.$$

As $\beta = R/P$, we can replace R with βP.[2]

$$\begin{aligned}
& \alpha\beta^2 P^2 = (1-\alpha)P^2. \\
\Rightarrow\quad & \alpha\beta^2 = 1-\alpha. \\
\Rightarrow\quad & \alpha(\beta^2+1) = 1. \\
\Rightarrow\quad & \alpha = \frac{1}{\beta^2+1}. \qquad (1)
\end{aligned}$$

4 End Note

There is one thing that remains unsolved, which is why the F-measure is called F. A personal communication with David D. Lewis several years ago revealed that when the F-measure was introduced to MUC-4, the name was accidentally selected by the consequence of regarding a different F function in van Rijsbergen's book as the definition of the "F-measure".

Finally, if you have any comments, please contact me by email.

References

[Chinchor, 1992] Nancy Chinchor, MUC-4 Evaluation Metrics, in Proc. of the Fourth Message Understanding Conference, pp. 22–29, 1992. http://www.aclweb.org/anthology-new/M/M92/M92-1002.pdf

[van Rijsbergen, 1979] C. J. van Rijsbergen, Information Retrieval, London: Butterworths, 1979. http://www.dcs.gla.ac.uk/Keith/Preface.html

[2]In van Rijsbergen's book, $\beta = P/R$ but I believe this is a typing error.