Can Statistical Language Models be used to improve Spectrum Based Fault Localization Rankings? (Additional Material)

Anonymous Author(s)

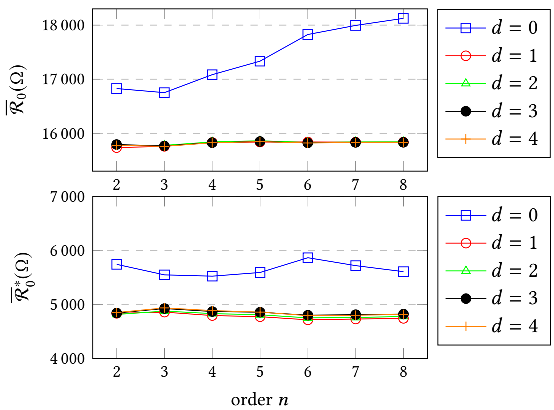

Figure 1: Plots of $\overline{\mathcal{R}}_0(\Omega)$ (top) and $\overline{\mathcal{R}}^*_0(\Omega)$ (bottom) for different orders n and different token abstraction levels d.

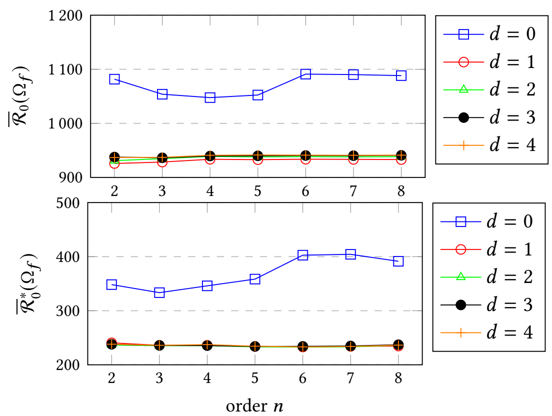

Figure 2: Plots of $\overline{\mathcal{R}}_0(\Omega_f)$ (top) and $\overline{\mathcal{R}}^*_0(\Omega_f)$ (bottom) for different orders n and different token abstraction levels d.

ISSTA, Santa Barbara
2016. 978-x-xxxx-xxxx-x/YY/MM...$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

SBFL ranking metric	formula
Ample	$\left\lvert \frac{e_f}{e_f+n_f} - \frac{e_p}{e_p+n_p} \right\rvert$
Anderberg	$\frac{e_f}{e_f+2(n_f+e_p)}$
Arithmetic Mean	$\frac{2e_f n_p - 2n_f e_p}{(e_f+e_p)\cdot(n_p+n_f)+(e_f+n_f)\cdot(e_p+n_p)}$
Cohen	$\frac{2e_f n_p - 2n_f e_p}{(e_f+e_p)\cdot(n_p+e_p)+(e_f+n_f)\cdot(n_f+n_p)}$
Dice	$\frac{2e_f}{e_f+n_f+e_p}$
Euclid	$\sqrt{e_f + n_p}$
Fleiss	$\frac{4e_f n_p - 4n_f e_p - (n_f-e_p)^2}{2e_f n_f e_p + 2n_p n_f e_p}$
Geometric Mean	$\frac{e_f n_p - n_f e_p}{\sqrt{(e_f+e_p)\cdot(n_p+n_f)\cdot(e_f+n_f)\cdot(e_p+n_p)}}$
Goodman	$\frac{2e_f - n_f - e_p}{2e_f+n_f+e_p}$
GP13	$e_f \left(1 + \frac{1}{2e_p+e_f}\right)$
Hamann	$\frac{e_f+n_p-n_f-e_p}{e_f+n_f+e_p+n_p}$
Hamming etc.	$e_f + n_p$
Harmonic Mean	$\frac{(e_f n_p - n_f e_p)((e_f+e_p)(n_p+n_f)+(e_f+n_f)(e_p+n_p))}{(e_f+e_p)\cdot(n_p+n_f)\cdot(e_f+n_f)\cdot(e_p+n_p)}$
Jaccard	$\frac{e_f}{e_f+n_f+e_p}$
Kulczynski1	$\frac{e_f}{n_f+e_p}$
Kulczynski2	$\frac{1}{2}\left(\frac{e_f}{e_f+n_f} + \frac{e_f}{e_f+e_p}\right)$
M1	$\frac{e_f+n_p}{n_f+e_p}$
M2	$\frac{e_f}{e_f+n_p+2(n_f+e_p)}$
Naish2 (Op2)	$e_f - \frac{e_p}{e_p+n_p+1}$
Ochiai	$\frac{e_f}{\sqrt{(e_f+n_f)\cdot(e_f+e_p)}}$
Ochiai2	$\frac{e_f n_p}{\sqrt{(e_f+e_p)\cdot(n_p+n_f)\cdot(e_f+n_f)\cdot(e_p+n_p)}}$
Overlap	$\frac{e_f}{\min(e_f . n_f . e_p)}$
Rogers & Tanimoto	$\frac{e_f+n_p}{e_f+n_p+2(n_f+e_p)}$
Rogot1	$\frac{1}{2}\left(\frac{e_f}{2e_f+n_f+e_p} + \frac{n_p}{2n_p+n_f+e_p}\right)$
Rogot2	$\frac{1}{4}\left(\frac{e_f}{e_f+e_p} + \frac{e_f}{e_f+n_f} + \frac{n_p}{n_p+e_p} + \frac{n_p}{n_p+n_f}\right)$
Russell & Rao	$\frac{e_f}{e_f+n_f+e_p+n_p}$
Scott	$\frac{4e_f n_p - 4n_f e_p - (n_f-e_p)^2}{(2e_f+n_f+e_p)\cdot(2n_p+n_f+e_p)}$
Simple Matching	$\frac{e_f+n_p}{e_f+n_f+e_p+n_p}$
Sokal	$\frac{2(e_f+n_p)}{2(e_f+n_p)+n_f+e_p}$
Sørensen-Dice	$\frac{2e_f}{2e_f+n_f+e_p}$
Tarantula	$\frac{\frac{e_f}{e_f+n_f}}{\frac{e_f}{e_f+n_f}+\frac{e_p}{e_p+n_p}}$
Wong1	e_f
Wong2	$e_f - e_p$
Wong3	$\begin{cases} e_f - e_p & \text{if } e_p \leq 2 \\ e_f - \left(2 + \frac{1}{10}(e_p - 2)\right) & \text{if } 2 < e_p \leq 10 \\ e_f - \left(2.8 + \frac{1}{1000}(e_p - 10)\right) & \text{otherwise} \end{cases}$
Zoltar	$\frac{e_f}{e_f+n_f+e_p+\frac{10000 n_f e_p}{e_f}}$

Table 1: Overview of all examined SBFL metrics.