DarkData Scoring Function

Anirudh Prabhu, Stephan Zednik

June 22, 2016

1 Scoring Algorithm

Current Scoring Function :

$$\sum_{0}^{c}\left(\frac{\sum_{0}^{n_c}(f_{n_c}\cdot w_{n_c})}{\sqrt{n_c}}\right)$$

New function :

$$Score = \log_{10}\left(\sum_{i=0}^{c}\left(\frac{\sum_{i=0}^{n_c}(f_{n_c})}{max\,(f_{n_c})}\cdot w_c\right)\times\left(\frac{na_c}{na_{max}}\right)\right)$$

- $c \longrightarrow$ Compatibility (Strong, Slight, Some, Negative , Indifferent)
- 0 to $n_c \longrightarrow$ All compatibility assertions
- $na_c \longrightarrow$ Number of Assertions for a specific 'c'.
- $f_{n_c} \longrightarrow$ Confidence for a specific assertion for a specific compatibility.
- $w_c \longrightarrow$ Weight for a specific compatibility.
- $na_{max} \longrightarrow$ Maximum value from all na_c values.

Example :

Candidate 1

	c_{slight}	c_{some}	c_{strong}	$c_{indifferent}$	$c_{negative}$
f	0.5	0.7	0.5	1	0.5
w	1	3	5	0	-1
na	1	1	1	1	1

Applying the formula,

$$Score = (0.5) + (0.7 \times 3) + (0.5 \times 5) + (0) + (0.5 \times -1) = 4.6$$

Once normalized , $\log_{10}(Score) = 0.6627$
Candidate 2

	c_{slight}	c_{some_1}	c_{some_2}	c_{some_3}	c_{strong}
f	0.5	0.7	0.5	1	0.5
w	1	3	3	3	5
na	1	3	3	3	1

Applying the formula,

$$Score = (0.5 \times \frac{1}{3}) + ((0.7 + 0.5 + 1) \times 3) + (0.5 \times 5 \times \frac{1}{3}) = 7.5983$$

Once normalized , $\log_{10}(Score) = 0.8807$
Candidate 3

	c_{strong_1}	c_{strong_2}	c_{strong_3}	c_{strong_4}	c_{strong_5}
f	0.5	0.7	0.5	1	0.5
w	5	5	5	5	5
na	5	5	5	5	5

Applying the formula,

$$Score = (0.5 + 0.7 + 0.5 + 1 + 0.5) \times 5 = 16$$

Once normalized , $\log_{10}(Score) = 1.2041$
Candidate 4

	c_{slight_1}	c_{slight_2}	c_{slight_3}	c_{slight_4}	c_{slight_5}
f	0.5	0.7	0.5	1	0.5
w	1	1	1	1	1
na	5	5	5	5	5

Applying the formula,

$$Score = (0.5 + 0.7 + 0.5 + 1 + 0.5) \times 1 = 3.2$$

Once normalized , $\log_{10}(Score) = 0.5051$

Notes : Initial Attempt

$$\sum_{i=0}^{c} \left(\frac{\sum_{i=0}^{n_c} (f_{n_c} \cdot w_c) \times \left(\frac{na_c}{na_{max}} \right)}{max\,(f_{n_c} \cdot w_c)} \right)$$

First Iteration of the formula uses the $f_{n_c} \cdot w_c$ combination to normalize, but this makes the weights assigned loose its importance and the final result only give priority to results where candidates have the same type of compatibility. (Irrespective of whether the value of the compatibility assertion is strong or weak.)

The second formula, solves this issue by using *log* to normalize the results instead.

2 User Scenario

Scenario 1

Let us consider candidate 4 as an example for this scenario and see how the score changes when assertions are added, changed or removed.

Original Case
Candidate 4

	c_{slight_1}	c_{slight_2}	c_{slight_3}	c_{slight_4}	c_{slight_5}
f	0.5	0.7	0.5	1	0.5
w	1	1	1	1	1
na	5	5	5	5	5

Applying the formula,

$$Score = (0.5 + 0.7 + 0.5 + 1 + 0.5) \times 1 = 3.2$$

Once normalized , $\log_{10}(Score) = 0.5051$

Replace one of the slight compatibility assertions with a "Strong" assertion

	c_{strong_1}	c_{slight_1}	c_{slight_2}	c_{slight_3}	c_{slight_4}
f	0.5	0.5	0.7	0.5	1
w	5	1	1	1	1
na	1	4	4	4	4

Applying the formula,

$$Score = ((0.5 + 0.7 + 0.5 + 1) \times 1) + (\frac{0.5}{1} \times 5 \times \frac{1}{4}) = 3.325$$

Once normalized , $\log_{10}(Score) = 0.5217$

Replace one of the slight compatibility assertions with a "Some" assertion

	c_{strong_1}	c_{slight_1}	c_{slight_2}	c_{slight_3}	c_{some_1}
f	0.5	0.5	0.7	0.5	1
w	5	1	1	1	3
na	1	3	3	3	1

Applying the formula,

$$Score = ((0.5+0.7+0.5) \times 1) + (\frac{0.5}{1} \times 5 \times \frac{1}{3}) + ((\frac{1}{1} \times 3 \times \frac{1}{3})) = 3.5333$$

Once normalized , $\log_{10}(Score) = 0.5481$

Replace 2 of the slight compatibility assertions with a "Negative" assertion

	c_{strong_1}	c_{slight_1}	$c_{negative_1}$	$c_{negative_2}$	c_{some_1}
f	0.5	0.5	0.7	0.5	1
w	5	1	-1	-1	3
na	1	1	2	2	1

Applying the formula,

$$Score = (0.5 \times \frac{1}{2}) + (\frac{0.5}{1} \times 5 \times \frac{1}{2}) + ((\frac{1}{1} \times 3 \times \frac{1}{2})) + (\frac{(0.7+0.5)}{1} \times -1 \times \frac{2}{2}) = 1.8$$

Once normalized , $\log_{10}(Score) = 0.2552$

Add 2 strong compatibility assertions to the original example

	c_{strong_1}	c_{strong_2}	c_{slight_1}	c_{slight_2}	c_{slight_3}	c_{slight_4}	c_{slight_4}
f	0.5	0.8	0.5	0.7	0.5	1	0.5
w	5	5	1	1	1	1	1
na	2	2	5	5	5	5	5

Applying the formula,

$$Score = ((0.5 + 0.7 + 0.5 + 1 + 0.5) \times 1) + (\frac{0.5 + 0.8}{1} \times 5 \times \frac{2}{5}) = 5.8$$

Once normalized , $\log_{10}(Score) = 0.7634$

Remove 2 slight compatibility assertions from the original example

	c_{slight_1}	c_{slight_2}	c_{slight_3}
f	0.5	0.7	0.5
w	1	1	1
na	3	3	3

Applying the formula,

$$Score = ((0.5 + 0.7 + 0.5) \times 1) = 1.7$$

Once normalized , $\log_{10}(Score) = 0.2304$