

Como parâmetro de comparação, levar em consideração a v1 submetida para o CBIC 2017

- Melhor resultado todas as mensagens: F1 **0.5054** (modelo B)

Evoluções v2

- Constraint na criação e em cada geração de indivíduos
 - Mais de uma massive function: fitness zero
 - Não ter if_then_else ou polaritySum* na raiz – penalização de 20% do fitness
- Pesos para os dicionários

Dicionários/Pesos						
W1	W2	W3	W4	W5	W6	W7
LIU	Sentiwordnet	AFFIN	Vader	Slang	Effect	Semeval2015

- Parâmetros principais [1]
 - Gerações: 51
 - População: 500

Testes 1

Usando valores discretos como terminais (0.0, 0.5, 1.0, 1.5, 2.0) para possíveis pesos e um range de valor real [0,2]

#	Indivíduo
1	polaritySumAVGUsingWeights(replaceNegatingWords(removeAllPunctuation(replaceBoosterWords(removeStopWords(replaceNegatingWords(replaceBoosterWords(boostUpperCase(x))))))), 0.5, 0.0, 0.31063506398117546, 0.0, 0.0, 0.0, 0.0)
2	polaritySumAVGUsingWeights(removeAllPunctuation(replaceBoosterWords(removeStopWords(replaceNegatingWords(x)))), 0.0, 0.0, 1.6561909157376327, positiveWordsQuantity(replaceNegatingWords(removeAllPunctuation(x))), negativeWordsQuantity(x), 0.0, 0.0)
3	polaritySumAVGUsingWeights(replaceBoosterWords(removeAllPunctuation(removeStopWords(replaceNegatingWords((x))))), 0.5, 0.0, 0.5, positiveWordsQuantity(removeStopWords(replaceBoosterWords(removeLinks(removeAllPunctuation(replaceBoosterWords(replaceBoosterWords(removeLinks(removeAllPunctuation(replaceNegatingWords(removeAllPunctuation(replaceBoosterWords(removeAllPunctuation(removeAllPunctuation(replaceNegatingWords(x)))))))))))))), mul(mul(0.5, negativeWordsQuantity(x)), mul(0.5, 0.5)), 0.0, 0.0)

Resultados

3 models evaluated - 7 dictionaries

AVG All F1 SemEval: 0.6011

Best All F1 value: 0.6101

Desvio padrão: 0.64

Algumas observações:

- Geração em que foi obtido o melhor indivíduo
 - Modelo 1: geração 40
 - Modelo 2: geração 38
 - Modelo 3: geração 44

Testes 2

Usando somente o valor discreto 0.0 e mantendo os valores reais [0,2]

#	Indivíduo
1	polaritySumAVGUsingWeights(removeStopWords(removeAllPunctuation(replaceNegatingWords(replaceBoosterWords(x))))), sub(1.4917314878762928, 1.0666968653301865), if_then_else(hasURLs(removeAllPunctuation(removeStopWords(x))), 1.695040482181927, 0.0), if_then_else(hasURLs(removeLinks(replaceNegatingWords(removeAllPunctuation(x)))), 0.5791933567949965, 0.252301850185894), if_then_else(hasURLs(removeLinks(boostUpperCase(removeAllPunctuation(replaceNegatingWords(x))))), mul(1.4917314878762928, 1.5050939233561567), sub(1.4917314878762928, 1.4917314878762928)), 0.0, 0.0, 0.0)
2	polaritySumAVGUsingWeights(removeAllPunctuation(replaceNegatingWords(boostUpperCase(removeStopWords(replaceBoosterWords(removeLinks(x))))))), 0.4405453203256887, 0.0, 0.2494501229744468, 0.0, 0.0, 0.0, 0.0)
3	if_then_else(hasURLs(x), add(sub(sub(0.0, add(negativeWordsQuantity(boostUpperCase(removeStopWords(removeStopWords(removeAllPunctuation(x))))), mul(0.0, 0.0)), add(negativeWordsQuantity(removeStopWords(boostUpperCase(removeStopWords(removeStopWords(removeAllPunctuation(x))))), mul(0.0, 0.0))), positiveWordsQuantity(replaceNegatingWords(x))), add(polaritySumAVGUsingWeights(replaceNegatingWords(removeLinks(removeAllPunctuation(removeStopWords(x))))), 0.32170186496512987, 0.0, 1.7518934561906048, 0.32170186496512987, 0.0, 0.0, 0.0), add(sub(add(0.0, mul(0.0, 0.0)), add(negativeWordsQuantity(boostUpperCase(removeStopWords(removeAllPunctuation(x))))), positiveWordsQuantity(x))), positiveWordsQuantity(x)))

3 models evaluated - 7 dictionaries

AVG All F1 SemEval: 0.608

Best All F1 value: 0.6105

Desvio padrão: 0.26

Algumas observações:

- Geração em que foi obtido o melhor indivíduo
 - Modelo 1: geração 45
 - Modelo 2: geração 33
 - Modelo 3: geração 38

Próximos passos

- Mutação diferenciada para os valores dos pesos dos dicionários com probabilidade diferente
 - Manter as duas mutações
 - Aumentar a probabilidade de mutação dos valores, uma vez que está fortemente ligado com a criação do indivíduo
- Modificação dos parâmetros de população e gerações
 - Bons indivíduos estão sendo obtidos em gerações muito próximas do limite, o que sugere que ainda pode haver espaço para melhorias nos indivíduos se houverem mais gerações

- Manter 25 mil ciclos (50 * 500)

[1] Sean Luke, Gabriel Catalin Balan, and Liviu Panait. Population Implosion in Genetic Programming - Department of Computer Science, George Mason University 4400 University Drive MSN 4A5, Fairfax, VA 22030, USA