Effectivess comparison report

Raphael Rodrigues Campos

January 17, 2016

Caso de Estudo: Classificação Automática de Texto

Resultados e discussões

A Tabela 1 sumariza os resultados da avaliaćão empirica de algums algoritmos estado-da-arte para classificaćão de texto e os algoritmos propostos baseado na ** Extremelly Randomized Tree**.

1) ha' uma avanco na area baseada em florestas
2) temos um a abordagem combinando stacking, bagging (arvores tradicionais e LAZY) e boosting (BROOF e BERT)
3) os resulatdos de stacking so são bons qdo os novos métodos de florestas entram (COMBSOTA não eh tao bom)
4) ha' um ganho em combinar tudo mas 5) nao precisa combinar tudo praobter bons resuatdos, o stacking de florestas jah eh competitivo

% latex table generated in R 3.2.4 by xtable 1.8-0 package % Thu Apr 28 15:29:23 2016

V1	V2	20NG	4UNI	ACM	REUTERS90
SVM	microF1	**90.06 ± 0.43**	**83.48 ± 1.08**	**75.4 ± 0.66**	**68.19 ± 1.15**
	macroF1	**89.93 ± 0.43**	**73.39 ± 2.17**	**63.84 ± 0.55**	**31.95 ± 2.59**
BERT	microF1	88.93 ± 0.39	**84.61 ± 0.98**	**74.8 ± 0.59**	**67.33 ± 0.72**
	macroF1	88.59 ± 0.5	**73.61 ± 1.85**	**62.1 ± 0.99**	**29.24 ± 1.4**
BROOF	microF1	87.96 ± 0.24	**84.41 ± 1.07**	73.35 ± 0.79	**66.79 ± 0.97**
	macroF1	87.44 ± 0.28	**73.23 ± 1.1**	60.76 ± 0.8	**28.48 ± 2.17**
KNN	microF1	87.53 ± 0.69	75.63 ± 0.94	70.99 ± 0.96	**68.07 ± 1.07**
	macroF1	87.22 ± 0.66	60.34 ± 1.36	55.85 ± 0.97	**29.93 ± 2.48**
LAZY	microF1	87.96 ± 0.37	82.34 ± 0.61	74.02 ± 0.79	**66.3 ± 1.07**
	macroF1	87.39 ± 0.37	68.33 ± 1.6	59.46 ± 1.35	26.61 ± 2.12
NB	microF1	88.99 ± 0.54	62.63 ± 1.7	73.54 ± 0.71	65.32 ± 1.13
	macroF1	88.68 ± 0.55	51.38 ± 3.19	58.03 ± 0.85	**27.86 ± 0.79**
XT	microF1	85.94 ± 0.23	81.66 ± 1.03	71.94 ± 0.66	64.33 ± 0.86
	macroF1	85.57 ± 0.22	65.44 ± 2.41	57.4 ± 1.13	24.47 ± 2.22
LXT	microF1	88.39 ± 0.51	81.24 ± 0.71	69.63 ± 0.91	65.92 ± 0.82
	macroF1	88.05 ± 0.44	66.89 ± 1.23	57.33 ± 1.48	26.71 ± 2.53
RF	microF1	83.64 ± 0.29	81.52 ± 1	71.05 ± 0.31	63.92 ± 0.81
	macroF1	83.08 ± 0.35	65.44 ± 1.91	56.56 ± 0.45	24.36 ± 1.98

Table 1: Comparação entres métodos de base

% latex table generated in R 3.2.4 by xtable 1.8-0 package % Thu Apr 28 15:29:40 2016

Legenda para os métodos:

- BERT: Boosted Extremely Randomized Trees
- LXT: Lazy Extremely Randomized Trees
- RF: Random Forest com 200 árvores

V1	V2	20NG	4UNI	ACM	REUTERS90
COMBALL	microF1	**91.67 ± 0.44**	**86.74 ± 1.17**	**78.46 ± 0.72**	**80.02 ± 1.24**
	macroF1	**91.43 ± 0.42**	**79.45 ± 2.23**	**63.72 ± 1.01**	**37.84 ± 3.14**
COMB3	microF1	90.63 ± 0.57	**86.79 ± 0.86**	**77.34 ± 0.6**	**79 ± 1.14**
	macroF1	90.4 ± 0.57	**79.63 ± 1.91**	**62.91 ± 0.92**	**33.93 ± 2.97**
COMBSOTA	microF1	90.65 ± 0.45	**84.95 ± 1.15**	**77.78 ± 0.73**	74.63 ± 1
	macroF1	90.42 ± 0.44	**75.96 ± 1.78**	**63.04 ± 0.85**	27.66 ± 0.88
COMB2	microF1	90.2 ± 0.51	**86.54 ± 1.06**	76.88 ± 0.55	**78.25 ± 1.17**
	macroF1	89.95 ± 0.52	**79.41 ± 1.63**	**62.66 ± 0.81**	32.86 ± 2.23
COMB1	microF1	89.32 ± 0.42	**86.52 ± 1.18**	76.74 ± 0.73	77.22 ± 1.14
	macroF1	89.01 ± 0.44	**78.66 ± 1.9**	**62.2 ± 1.01**	31.71 ± 2.7
BERT	microF1	88.93 ± 0.39	**84.61 ± 0.98**	74.8 ± 0.59	67.33 ± 0.72
	macroF1	88.59 ± 0.5	73.61 ± 1.85	**62.1 ± 0.99**	29.24 ± 1.4
SVM-L2	microF1	90.06 ± 0.43	83.48 ± 1.08	75.4 ± 0.66	68.19 ± 1.15
	macroF1	89.93 ± 0.43	73.39 ± 2.17	**63.84 ± 0.55**	31.95 ± 2.59

Table 2: Comparação entre os métodos de stacking

- RF1000: Random Forest com 1000 árvores
- XT: Extremely Randomized Trees com 200 árvores
- XT1000: Extremely Randomized Trees com 1000 árvores
- COMB1: Stacking (Lazy + BROOF)
- COMB2: Stacking (LXT + BERT)
- COMB3: Stacking (Lazy + BROOF + LXT + BERT)
- COMBSOTA: Stacking (KNN + RF + SVM + NB)