Evaluating Boosting Algorithms to □lassi□ar□□lass□□ □om□arison an□ □m□ro□m□nts

Authors: □oshi□ □umar □ Agar□al

□rs□nt□□ □□ tr□□ a□

□ontri□utions o□ this □a□□

- □□o algorithms ar□ introdu□□□ □
  - □arBoost □
  - □arBoost □
- □□v□ral □ lass□s o□ □ oosting algorithms ar□ □om□ar□ □m□iri□□□ all□
  - Ada□oost
  - □□□□E□
  - Ada□ost an□ variants

□hat ar□ □ar□ □lass□s□

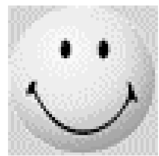
- □lass□s that ar□ *in□r□□□ unt□ll□* □□□□□ r□lativ□ to th□ □ntir□ atast□□
- □□am□l□s□
  - *□ □ommr□□□□ur□has□s vs □ rosing□*
  - *n□t□r□ s□urit□ □ha□s vs l□git logins□*
  - *□rau□□□ tion □□□git □ isa □ur□has□□□*
  - *o□□um□nt □lassi□□ation*

□□tin□ ar□ □lass□

- □h□□ ar□ using an □algorithm to □ri□□ t a rar□□ lass **C** th□r□ ar□ □□ □ossi□l□ out□om□s□

	□ri□□ t□□as **C**	□ri□□ t□□as **C**
□s A□tuall□ **C**	□ru□ □ ositiv□ □ □	□als□ □ □□gativ□ □ □
□s A□tuall□ **C**	□als□ □ ositiv□ □ □	□ru□ □□ gativ□ □ □

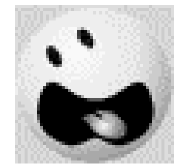
□ru□□ositiv□



□ri□□ t□□

A□ tuall□is □

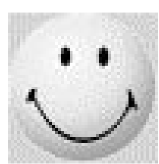
□als□□ositiv□



□ri□□ t□□

is a□ tuall□□ □

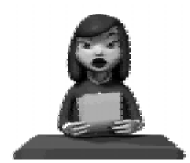
□ru□□ □gativ□

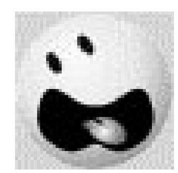


□ri□□ t□□ □

A□tuall□is □ □

□als□□ □gativ□





□ri□□ t□□ □

is a□tuall□□

h□Boosting□

- i□tifying rar□□ lass□□**C**□rom th□ rst o th□ atast□ **C** □is har□
- □an imrov□th□r□orman□ o a n□ a□□ □□lassi□ation algorithm
- thor□ti□ al □usti□ation
- s□ms to or□ □ll in ra□ ti□□
  - ho□□□ oosting or□ s□ns havil□ on th□ training □ata

□o to lassi□ar□lass□

- *a□□ura□□* is not a goo□m□tri□ or i□nti□ ation o rar□□ lass□
- □sir high *□ all*
  - *□an istinguish □ □rom □□*
- □sir high *r□ision*
  - *□an istinguish □ □rom □□*
- *□ all* an□ *r□ision* usuall□ onli□t

ru□E□ aml□

- □uos□ hav□a atast hr□□□ □ o □ata is in **C** an□□□ □is in **C**
- i□ naivl□□ lassi□ □rthing as **C**
  - a□□ura□is □ □□□ maningl□ss□
  - lo r□ all
- i□ naivl□□ lassi□ □rthing as **C**
  - lo r□□ ision

hat is Boosting□

- goal high r□ all A□□ high r□□ ision
- a ma□hin □arning t□ hni□u□ to itrativl□ imrov□an □ isting mo□□ using ights
- itrativl□□ hang ights using training □ata
- or th□ mo□l in n□ t itration□ lal□ grat r ight on th□ in□ orr□t □lassi□ations at th□ ns o □al□ ing l□ss r ight on th□□ orr□t □lassi□ation□
- vr□ mo□□ ontri□uts to th□ inal r□□ tion □ nsm□ l□□as □voting□

□hr□t□s o□ Boosting Algorithms

- Boosting Algorithms that □on□t A□stain
  - v□r□ mo□□ i□ us□
- Boosting Algorithms that A□stain
  - onl□ □goo□ mo□□s ar□ us□
- □ost□s□nsitiv□ Boosting Algorithms
  - i□□i□ats □ □ □□ □ □□□□□□□□□

A□aBoost

- □oosting algorithm that □o□s□□ t a□stain
- □o□ular algorithm
- tr□at □ □□□□□□□□ □all □ □uall□
- at □v□r□ it□ration th□ training mo□□ al□a□s r□turns som□thing
- □irst tr□at all □ata □ □uall□ □giv□ □□ual □□ight to all □ata □oints□□ □□□□ □□ □ □n□

A□aBoost □□tails

- □iv□n training □ata □ □□□□□□□ □ith r□s□ons□s □ □ □□□□□
- □or it□ration t□ □□□□□
  - □□□arn a □ mo□□ h_t using □□ights □ $_t$
  - □□□om□ut□ im□ortan□□ □□ight A_t
    - r_t □ sum□□□□□□□□ $_t$ □i□ h_t □□□□□□ □
    - A_t □ □□□□□ln□□□□ r_t □□ □ln□□□□ r_t □□
  - □□□om□ut□ n□□ □□ights □ $_t$□□ □i□ □or a□□ h □
    - □ $_t$□□ □i□□□□□ $_t$ □i□□□ □ □ A_t □ h_t □□□□□□□□ $_t$
    - □h□r□ $_t$ □hos□n s□t□ sum□□□ $_t$□□ □i□ □ □ □
- □inal o□□□
  - □□□□□ sign □sum□□□□□□□ □ A_t h_t □□□□□□

□□ar□Boost □ □

- im□ro□□m□nt o□□r A□aBoost
- similar to A□aBoost□□ □ut i□□i□□ats □□th□ □ositi□□ an□ n□gati□□ □r□□□ tions
- insta□□ o□ A_t □or all □r□□□ tions□ no□ us□ $A_t^{□}$ an□ A_t^{n} □or □ositi□□ an□ n□gati□□ □r□□□ tions r□s□□ ti□□l□

arBoost□tails □

- □□arn a□□ mo□h $_t$ using ights □ $_t$
- a□□omut□mortan□ ight A $_t$ □
  - □□$_t$□sum□□□ □□i□h$_t$□□□
  - □□$_t$□sum□□□ □□i□h$_t$□□□
  - A$_t$□ □ □□□□n□□□$_t$□□n□□□$_t$□□
- □□□□omut□mortan□ ight A $_t^n$
  - □□$_t$□sum□□□ □□i□h$_t$□□□
  - □□$_t$□sum□□□ □□i□h$_t$□□□
  - A$_t^n$ □ □□□□n□□□$_t$□□n□□□$_t$□□

arBoost□tails □

- □a□omut□□ $_t$□□i□or ositiv□□□ tions
  - □$_t$□□i□□□□□□i□□ □ □A$_t$□□h$_t$□□□□□□
- □□□□omut□ $_i$ □$_t$□□i□or ngativ□□□ tions
  - □$_t$□□i□□□□□□i□□ □ □A$_t^n$ □h$_t$□□□□□□
  - hr□□ $_t$ □hosn□s □t□sum□□$_t$□□i□□ □ □
- □inal□□□
  - □□□□□sign □ sum□□□□A$_t$□h$_t$□□□□
  - □ sum□□□□A$_t^n$ h$_t$□□□□ □

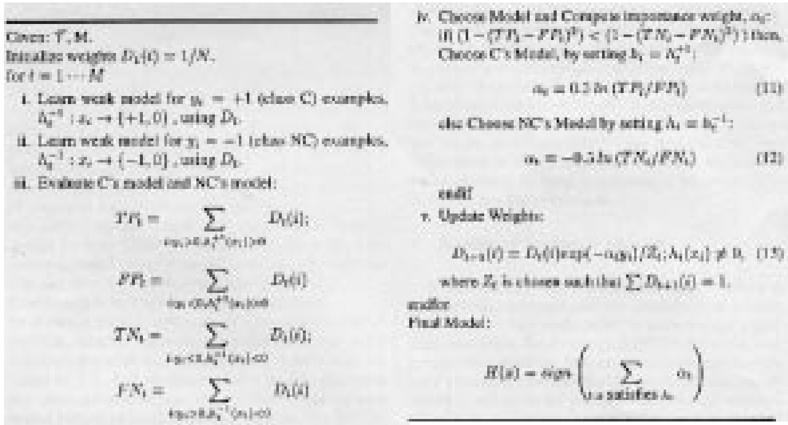
□□□□E□

- a □oosting algorithm that □s not a□stain□unli□□AaBoost □arBoost□
- i□th□□ urr□t mo□is not □ goo□□it uss□a a□ult mo□
- th□a□ult mo□naiv□□□ lassi□s□ v□rthing as □□ goo□r□ all□□or r□□ ision□

arBoost□

- □s□n□ t us□a a□ult mo□
- □uils□□ mo□s □ □an□□ at v□r□□ itr□tion using ights □ $_t$ □an□on□o□ ths□□s □ hosn□at v□□itration

arBoost□ □□tails



Given: $\mathcal{T}, M$.
Initialize weights $D_1(i) = 1/N$.
for $t = 1 \cdots M$

i. Learn weak model for $y_i = +1$ (class C) examples. $h_t^{+1} : x_i \to \{+1, 0\}$, using D_t.
ii. Learn weak model for $y_i = -1$ (class NC) examples. $h_t^{-1} : x_i \to \{-1, 0\}$, using D_t.
iii. Evaluate C's model and NC's model:

$$TP_t = \sum_{i:y_i \geq 0, h_t^{+1}(x_i) \neq 0} D_t(i);$$

$$FP_t = \sum_{i:y_i < 0, h_t^{+1}(x_i) \neq 0} D_t(i)$$

$$TN_t = \sum_{i:y_i < 0, h_t^{-1}(x_i) \neq 0} D_t(i);$$

$$FN_t = \sum_{i:y_i > 0, h_t^{-1}(x_i) \neq 0} D_t(i)$$

iv. Choose Model and Compute importance weight, α_t:
if $(1-(TP_t - FP_t)^2) < (1-(TN_t - FN_t)^2)$ then,
Choose C's Model, by setting $h_t = h_t^{+1}$:

$$\alpha_t = 0.5 \ln (TP_t / FP_t) \quad (11)$$

else Choose NC's Model by setting $h_t = h_t^{-1}$:

$$\alpha_t = -0.5 \ln (TN_t / FN_t) \quad (12)$$

endif
v. Update Weights:

$$D_{t+1}(i) = D_t(i)\exp(-\alpha_t y_i)/Z_t; h_t(x_i) \neq 0, \quad (13)$$

where Z_t is chosen such that $\sum D_{t+1}(i) = 1$.
endfor
Final Model:

$$H(x) = sign\left(\sum_{t:x \text{ satisfies } h_t} \alpha_t\right)$$

ost□□□nsitiv□Boosting Algorithms

- AaBoost □ sam□ight or tru□an□ als□ri□□ tions
- arBoost□ □□ □irht ights or ositiv□□ □□□□□□an□ngativ□ ri□□ tions □□□□□□□
- irht ights or □ □□□□□□□□□□□
- □algorithms: Aaost□□ B□□□ B□

ost□sn□sitiv□Algorithms

- Aaost tst□
  - as s□□ ri□lin □ an t□al□□ □
- □B□an□□ B□
  - as s□□ ri□lin □ ing□□□

□ Boi□□□ ations

- □B□
  - A_t □□or all t
  - mols ar mo□ rati□all vots ar□□ ual□
- □B□
  - A_t □omut□□ ust li□AaBoost
  - that is mor a□ □urat mols hav grat□r imortan□ □
- aramtr□□□ □ost□□□□□ost□□□□

ight ating is mortant

- ho hang $_t$ at a h itration is important to l oosting or s
- E sum o ights or all t
- all algorithms tr to in ras ight o an thr oring ight o an an in r all as a rsult
- ovri or ight o an too mu h th r ision is lost

E rimnts

- r t at
- all s nthti ata
- a h mo uss attri uts a h attri ut h r t h thr mo is in or
- atast attri uts ar in nt
- atast attri uts ar orrlat

atast tu

- tst all th algorithms
  - AaBoost arBoost E Aaost B
  - us st aramtrs o a h algorithm
- vrsions o atast us
  - sh sh sh
  - a h gts rogrssivl ouir harr to istinguish th an

atast sults

DSet		ABst	RB-1	SLIP	RB-2	CSB1	CSB2	ACst
snc1	R	77.04	77.04	81.13	74.93	77.04	77.04	77.04
	P	94.19	94.19	95.94	98.61	94.19	94.19	94.19
	F	84.76	84.76	87.92	85.16	84.76	84.76	84.76
snc2	R	76.78	76.78	69.39	76.78	76.78	76.78	77.31
	P	95.41	95.41	85.53	95.41	95.41	95.41	95.75
	F	85.09	85.09	76.62	85.09	85.09	85.09	85.55
snc3	R	41.03	41.16	35.09	49.21	43.14	43.14	79.02
	P	59.58	65.00	59.11	62.37	67.84	67.84	51.73
	F	48.59	51.10	44.04	55.01	52.74	52.74	62.53
snc4	R	54.75	58.71	47.63	56.99	52.64	38.13	69.53
	P	79.35	90.45	94.75	82.60	55.65	51.89	76.38
	F	64.79	71.20	63.39	67.45	54.10	43.95	72.79
snc5	R	47.23	51.58	43.14	53.96	56.46	55.67	63.19
	P	76.50	87.87	98.14	89.50	50.53	77.86	92.47
	F	58.40	65.00	57.93	67.33	53.33	64.92	75.08

□atast□□ □tu□

- tst□all th□algorithms
- E□□ atast□□ attri□uts□ar□□ orrlat□
- □vrsions o□□ atast□□ us□
  - irt□t rortions o□□ r□□ □us in th□raining ata
  - g□ r□□ □ □□□mans that in th□training ata thr ar□□ □or vr□□ □□ ataoints□

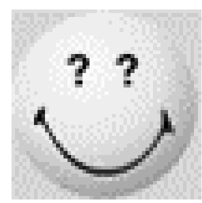
□atast□□ sults

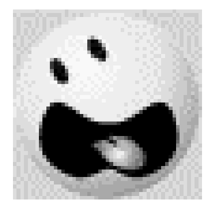
Cfrc		ABst	RB-1	SLIP	RB-2	CSB1	CSB2	ACst
2.9%	R	50.93	57.20	64.80	56.27	57.07	84.00	67.20
	P	71.27	81.25	57.51	82.10	60.03	44.74	68.39
	F	59.41	67.14	60.94	66.77	58.51	58.39	67.79
5.7%	R	63.20	68.67	63.60	73.20	67.07	84.13	76.00
	P	80.61	83.74	71.19	81.70	72.90	46.43	72.15
	F	70.85	75.46	67.18	77.22	69.86	59.84	74.03
13.1%	R	76.53	77.47	70.00	79.73	77.07	88.67	74.80
	P	80.62	83.36	78.59	86.04	72.70	58.18	85.65
	F	78.52	80.30	74.05	82.77	74.82	70.26	79.86
23.1%	R	78.27	84.40	80.80	80.80	86.13	96.13	85.07
	P	83.74	85.43	82.56	86.45	76.81	68.60	84.39
	F	80.91	84.91	81.67	83.53	81.21	80.07	84.73

□ummar□

- igt □ ating is imortant
  - r□all vs r□□ ision
- □urris□
  - arBoost □ □ttr than AaBoost
  - arBoost □ □ttr than □ □□E□
  - arBoost □ □□ o□ ttr hn r□□ in□□
  - an □ □□ ttr than □ ost snsitiv□ algorithms li□AaBost

□han□s□□□ustions□





- smil□a□□ s ar□rom smil□□ tionar□□om