

Dec 2 - Dec 9 Progress

Reproduction

- Can now reproduce Lin Tan's results on PROMISE data. We have his PROMISE and AST baselines, but not his semantic features (the code to generate those are protected by a patent).
- Difference was in how `bug > 1` samples were being treated; they are now set to 1 instead of being purged.

Experiments

- Running models with the new pre-processing on PROMISE data, they're doing better.

Results so far:

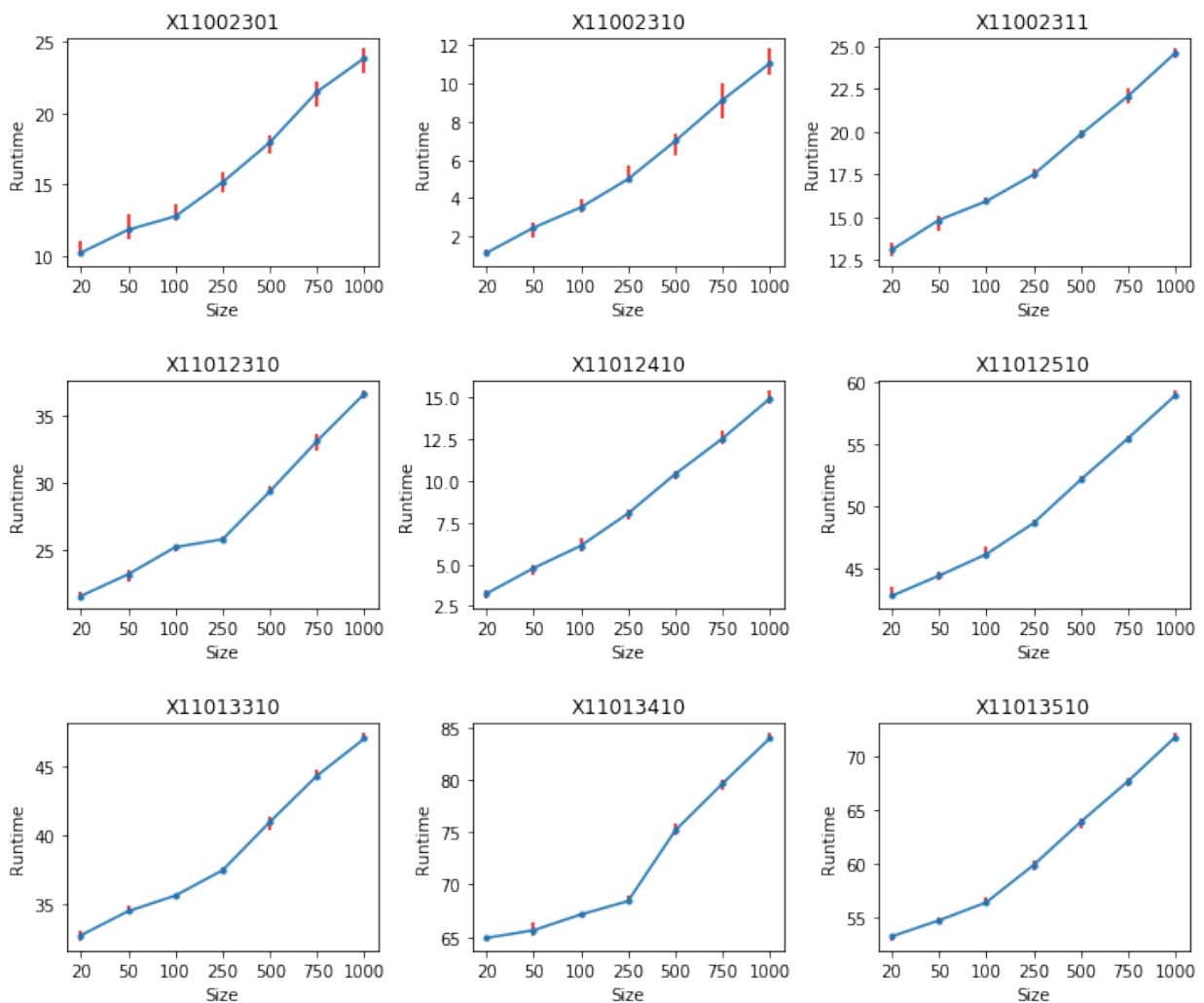
Dataset	Our method				MSR			Lin Tan		
	P	R	F	Runtime	P	R	F	P	R	F
ant 1.5 - 1.6	51.9 (3.7)	72.8 (10.9)	59.9 (1.9)	20.6 (0.4)	33	80	47	44.8	51.1	47.7
ant 1.6 - 1.7	45.6 (2.7)	67.5 (3.0)	54.6 (0.7)	37.7 (0.3)	21	98	35	41.8	77.1	54.2
camel 1.2 - 1.4	25.6 (1.3)	55.2 (4.8)	34.6 (1.4)	8.6 (0.8)	20	82	32	24.8	75.2	37.3
camel 1.4 - 1.6	28.6 (1.1)	48.9 (5.9)	35.9 (1.6)	31.5 (0.2)	28	68	40	28.3	63.7	39.1
xerces 1.2 - 1.3	16.1 (1.0)	75.4 (17.4)	31.3 (3.0)	43.9 (0.3)	23	28	26	16	46.4	23.8

- Where we are not SOTA yet, I will run more hyper-parameter search. The above are done with only the top-10 models, rather than top-28.

Research Questions

Old

- *How far can we stray from DL literature and still do well?*
 - Hamming distance 3
- *Can these models outperform standard DL models?*
 - Yes.
- *Can these models outperform SOTA SE methods?*
 - In general, yes
- *Are these models transferable to other datasets?*
 - Where transfer does NOT mean cross-project defect prediction (CPDP), generally, yes.
 - For CPDP, experiments are required
- *Are these models scalable?*
 - Yes.

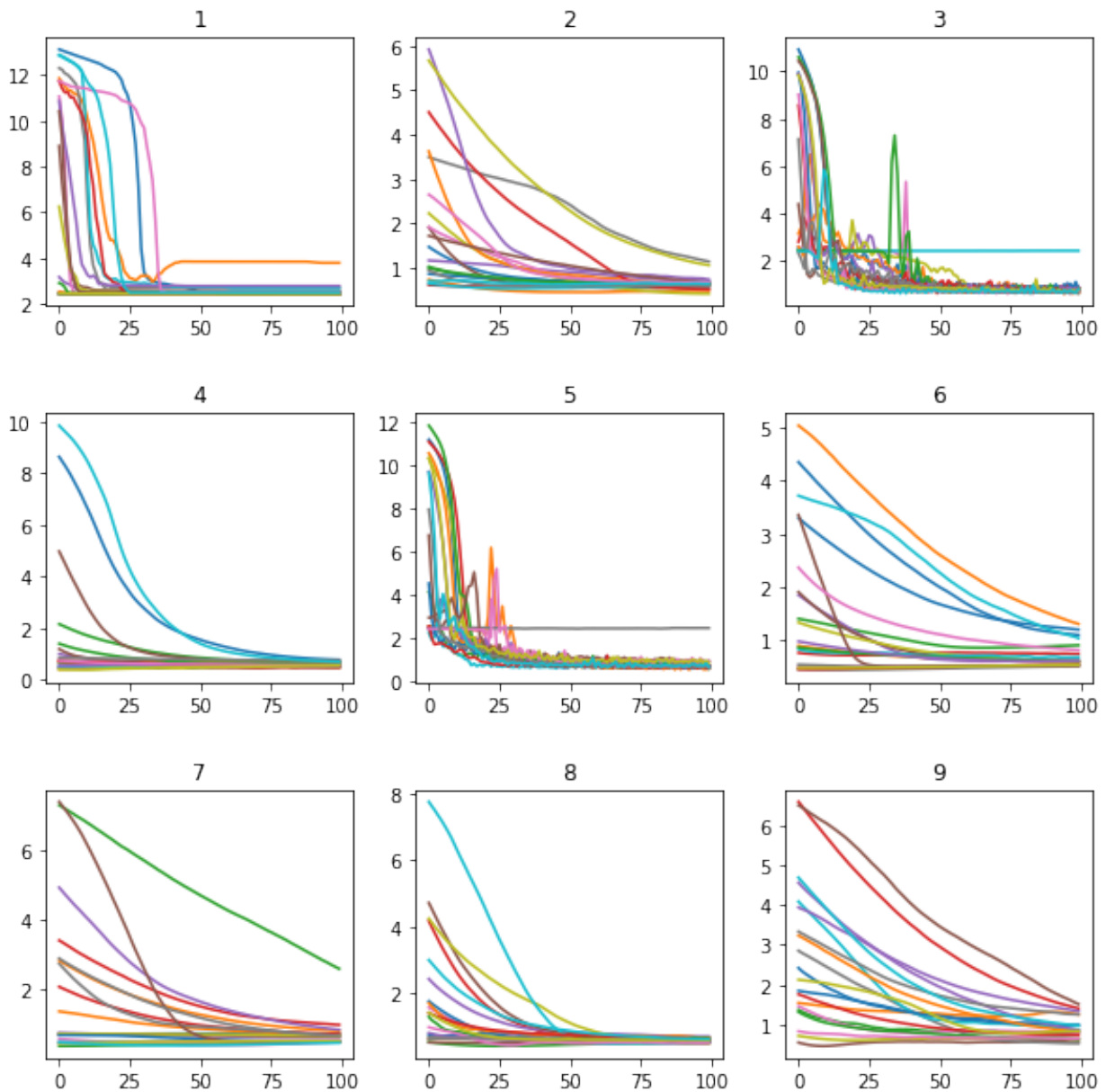


New

- *Why does oversampling help so much, such that the majority of models choose it?*
 - Because oversampling causes the minority class samples to contribute more to the loss, and

therefore, the model cannot ignore them.

- The performance boost from using these data points (<5%) comes from oversampling rather than simply using them.
- **How many epochs are really needed to train these models?**
 - About 65-75 seems to work for most models.



Work for next week

- Finish experiments
- Deep learners on AST features

Work for later

- Run cross-project defect prediction experiments
- Start writing paper (?)

Other Discussion

- RA Paperwork?
- Journal to aim for?