NC STATE UNIVERSITY

How to improve DL for SE

A case study with code smell detection

Rahul Yedida, NC State
Tim Menzies, NC State

Code smells

- *"Certain structures in the code that suggest (sometimes they scream for) the possibility of refactoring."* Beck et al. (2019)
- **Feature envy:** Class A is "envious" of class B.
- **Large class:** A do-it-all class.
- **Long method:** "What does this do??"
- **Misplaced class:** Classes that are misplaced *(obviously)*.

Beck, K., Fowler, M., & Beck, G. (1999). Bad smells in code. *Refactoring: Improving the design of existing code*, *1*(1999), 75-88.

NC STATE UNIVERSITY

Data

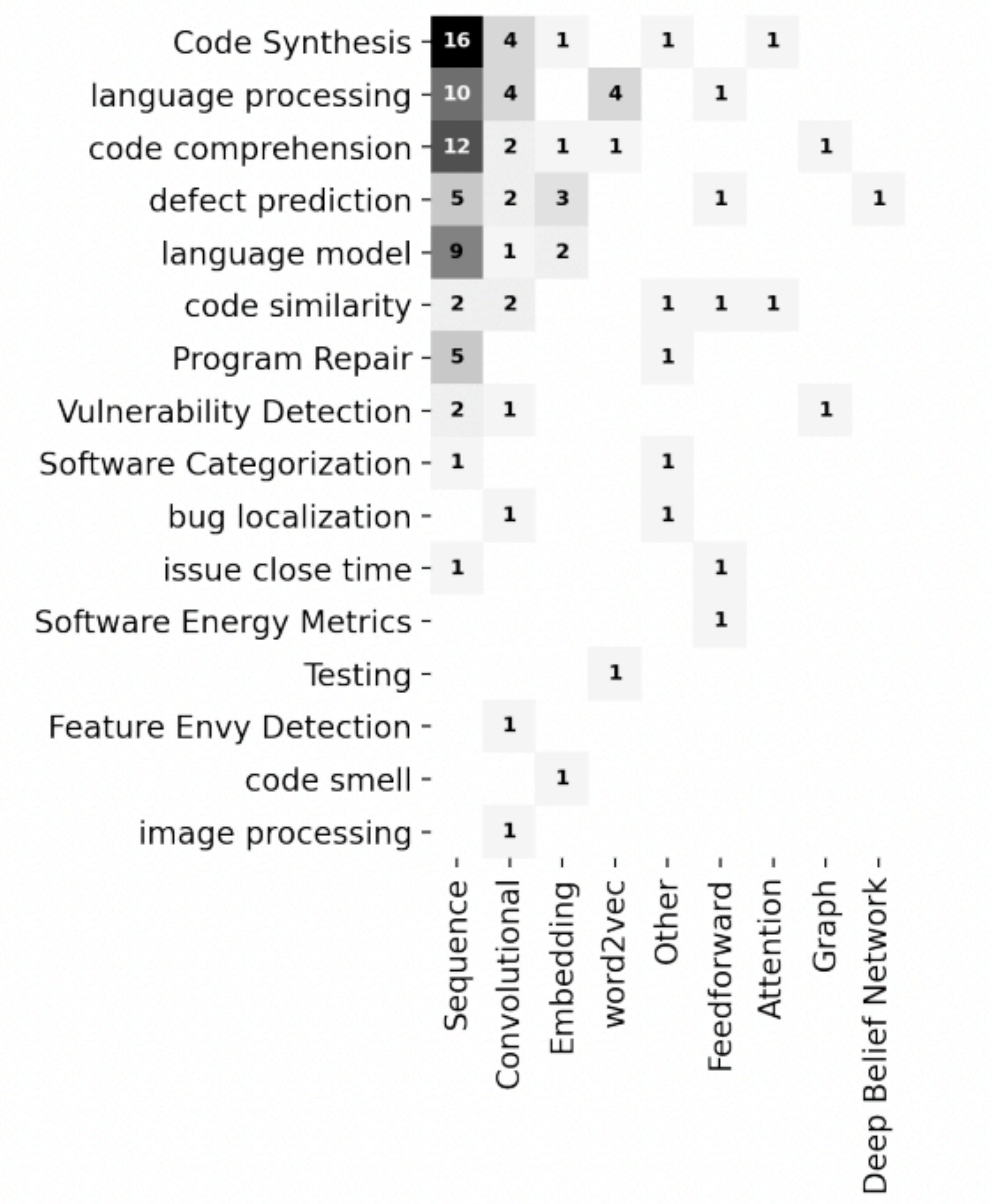
- Smell-inducing refactoring (Liu et al., 2019)
- Example: move a method from one class to another to induce *feature envy*

Liu, H., Jin, J., Xu, Z., Bu, Y., Zou, Y., & Zhang, L. (2019). Deep learning based code smell detection. *IEEE transactions on Software Engineering*.

What's wrong with prior work?

Complexity

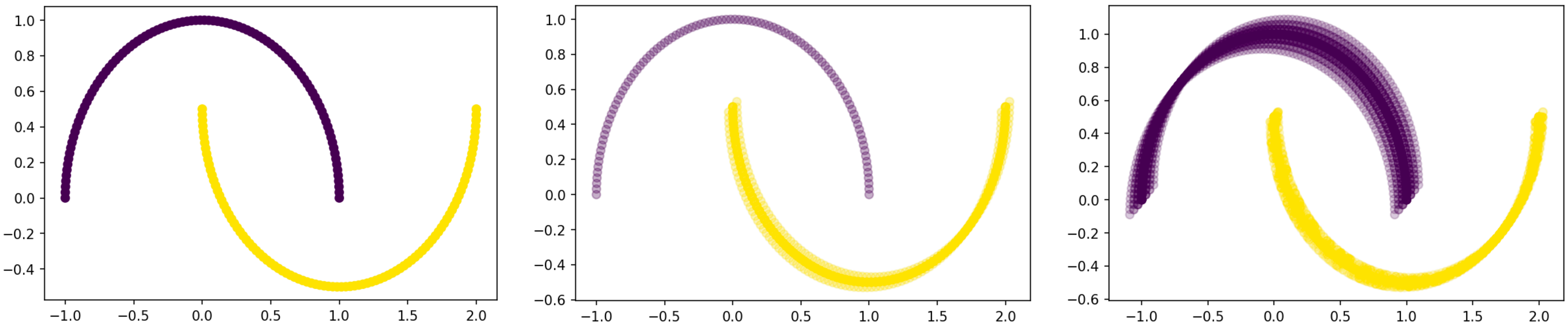
- **Seed:** Watson (2020)
- **Search:** *"deep learning AND software", "deep learning AND defect prediction", and "deep learning AND bug"*
- **Filter:** Published in top venues, highly cited, etc.
- **Backward snowballing:** Wohlin (2014)



Watson, C. A. (2020). Deep Learning in Software Engineering. The College of William and Mary.
Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1-10).

...and how do we fix it?

Fuzzy sampling



Before | **After fuzzy sampling** | **Final**

GHOST

Fuzzy sampling + ...

- Fuzzy sampling (twice)
- SMOTE
- Hyper-parameter optimization
- Weighted loss functions

Yedida, R., & Menzies, T. (2021). On the value of oversampling for deep learning in software defect prediction. *IEEE Transactions on Software Engineering*.

NC STATE UNIVERSITY

But does it work?

Yes (that's why we're here)

SOTA (reproduced)				Fuzzy sampling (with GHOST)			
Precision	Recall	F1	AUC	Precision	Recall	F1	AUC
79.5	80	79.7	88	98	98	98	99.3
79.2	79.2	79.2	87.3	97.7	97.7	97.7	99.1
80.1	79.9	80	88.3	98	98	98	99.3
80.6	80.4	80.5	88.8	98.2	98.2	98.2	99.3
80	80.1	80	88.1	97.7	97.7	97.7	99.1
78.7	78.5	78.6	87	97.8	97.8	97.8	99.2
78.5	78.2	78.3	86.3	97.7	97.7	97.7	99.2
80.3	80.2	80.2	88	97.7	97.7	97.7	99.1
79.8	79.6	79.7	87.8	97.8	97.8	97.8	99.2
77.9	77.5	77.7	85.8	97.8	97.8	97.8	99.1

Feature envy

SOTA (reproduced)				Fuzzy sampling (with GHOST)			
Precision	Recall	F1	AUC	Precision	Recall	F1	AUC
41.7	77.8	52.5	72.2	41.7	70	51.5	71.2
44.6	78.2	55.6	70.3	45.8	76.7	57	71.1
53.4	82.5	63.4	71.3	54.1	80.5	62.2	70.4
51.7	67.1	56	66.1	51.3	64.9	54.1	61.6
35.2	84.8	47.6	73.6	35.4	77.8	47.1	72.9
49.9	83.9	62.3	76.7	47.7	86.3	61.2	76.1
39.1	73.5	50.7	73.4	38.9	73.6	50.5	73.2
36.1	66.3	47.5	65.9	38	73.9	46.2	64.8
38.4	81.2	51.8	72.6	38.2	81.5	52.2	72.7
38.5	81.5	52.1	80.3	37.3	81.1	50.9	79.8

Long method

SOTA (reproduced)				Fuzzy sampling (with GHOST)			
Precision	Recall	F1	AUC	Precision	Recall	F1	AUC
33.7	78.3	47.1	50	82.9	77.5	80.1	89.5
6.1	100	11.5	50	0	0	0	68.8
61.5	100	76.2	50	76.8	81.8	79.2	75
11.9	100	21.3	50	32.2	71.4	44.4	74.6
11.4	100	20.5	50	23.8	23.6	23.7	63.1
0	0	0	50	42.9	71.4	53.6	70.7
9.1	14.3	11.1	50	27.9	28.6	28.2	65.3
0	0	0	50	0	0	0	50
30	100	46.2	50	53.8	80	64.3	75.7
15.7	100	27.1	50	57.7	73.5	64.6	80.4

Large class

SOTA (reproduced)				GHOST			
Precision	Recall	F1	AUC	Precision	Recall	F1	AUC
34.1	100	50.9	98.6	44.7	100	61.8	100
32.7	100	49.3	96.3	43.3	99.4	60.3	99.4
32.6	100	49.2	98.7	36	100	53	100
35.2	96.1	51.6	93.4	36.3	96.3	52.7	97.3
41.2	96.8	57.7	92	53.8	99.3	69.8	99.6
48.5	99.8	65.3	99.7	48.4	99.9	65.2	99.8
44.5	99.7	61.6	99.5	43.8	100	60.9	99.8
18.2	90.6	30.4	85.6	24.5	100	39.4	100
43.6	100	60.7	92.2	44	100	61.8	100
21.6	86.4	34.6	74.3	32.6	96.1	48.6	96.3

Misplaced class

Lessons learned

- Simpler learners with adequate preprocessing can outperform more complex learners, faster.
- Try simpler baselines before moving to more complex learners.
- Use GHOST as one such baseline.

Thank you!

GHOST paper	bit.ly/msr-ghost
Fuzzy sampling	bit.ly/fuzzy-sampling
These slides	bit.ly/msr-ghost-slides
Contact us	ryedida@ncsu.edu | timm@ieee.org