DeepTriage: Exploring the Effectiveness of Deep Learning for Bug Triaging

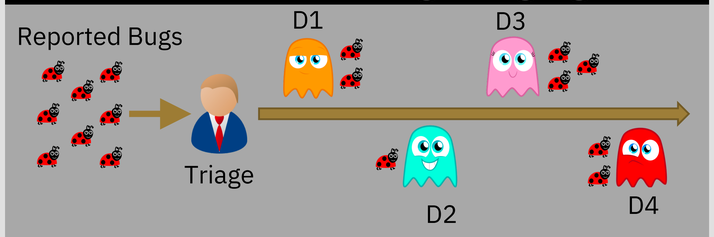
Senthil Mani, Anush Sankaran, Rahul Aralikatte

IBM Research AI

Contact email: anussank@in.ibm.com

CoDS-COMAD 2019

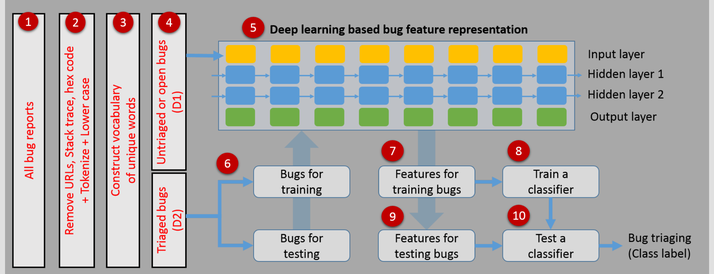
Introduction: Bug Triaging



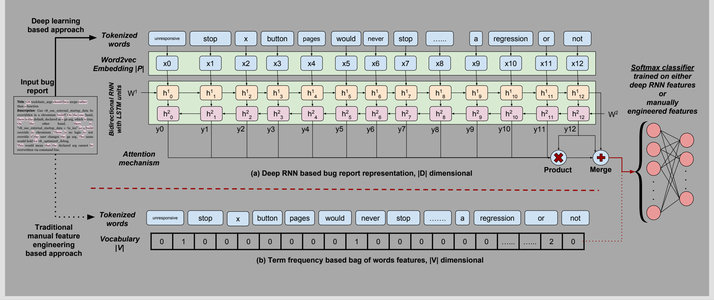
Sample Bug from Chromium:

DeepTriage: Proposed Approach

Overall Pipeline:



Deep Bidirectional Recurrent Neural Network with Attention:



Dataset: Bug Reports

Property	Chromium	Core	Firefox
Total Bugs	383,104	314,388	162,307
Unlabeled Bugs	263,936	186,173	138,093
Bugs for classifier	118,643	128,215	24,214
Vocabulary size	71,575	122,578	57,922

- 10 fold cross validation with chronological splits
- Vocabulary is built with min. word frequency as 5
- Tokenization is performed using Stanford's NLTK
- Removed URLs, hex codes, and stack traces

Other Applications

Using the deep features extracted using DBRNN-A, not only automated bug triaging could be performed, but also could be extended for other potential applications:

- Which bug will get fixed: To predict which bug will get fixed.
- Bug-fix time prediction model: To learn the time required to fix a bug.
- Reopen analysis: To study which bugs get reopened during its lifecycle.
- Bug priority estimation: Priority of the bug is to be estimated before triaging happens.

Experimental Results

Rank #10 accuracy on Chromium dataset

Sampl.	Classifier	CV#1	CV#10	Average
>= 0	BOW + MNB	21.9	33.3	26.0 ± 3.0
	BOW + Cosine	18.4	21.5	20.2 ± 1.2
	BOW + SVM	11.2	10.8	10.1 ± 0.6
	BOW + Softmax	12.5	08.7	09.1 ± 1.1
	DBRNN-A + Softmax	34.9	39.7	**37.9 ± 1.9**
>= 5	BOW + MNB	22.2	33.6	26.2 ± 3.1
	BOW + Cosine	18.6	22.0	20.4 ± 1.3
	BOW + SVM	11.3	09.0	09.2 ± 1.0
	BOW + Softmax	12.8	11.4	10.8 ± 0.9
	DBRNN-A + Softmax	32.2	38.2	**36.8 ± 2.2**
>= 10	BOW + MNB	22.4	34.3	26.6 ± 3.3
	BOW + Cosine	18.8	21.0	20.6 ± 1.3
	BOW + SVM	12.2	11.9	11.7 ± 0.4
	BOW + Softmax	11.9	11.5	11.3 ± 0.2
	DBRNN-A + Softmax	36.2	46.0	**41.8 ± 3.1**
>= 20	BOW + MNB	22.9	36.0	27.8 ± 3.7
	BOW + Cosine	19.3	23.0	21.5 ± 1.4
	BOW + SVM	12.2	11.9	11.7 ± 0.3
	BOW + Softmax	11.9	11.7	11.5 ± 0.3
	DBRNN-A + Softmax	36.7	47.0	**42.7 ± 3.5**

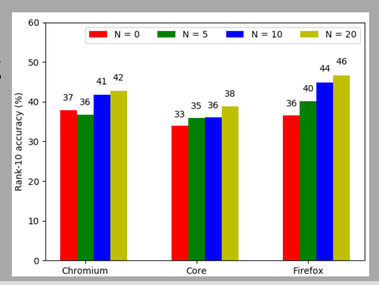
Rank #10 accuracy on Core dataset

Sampl.	Classifier	CV#1	CV#10	Average
>= 0	BOW + MNB	21.6	32.1	29.5 ± 3.6
	BOW + Cosine	16.3	29.1	22.6 ± 3.9
	BOW + SVM	13.6	14.1	13.6 ± 1.0
	BOW + Softmax	14.3	10.8	10.8 ± 1.4
	DBRNN-A + Softmax	30.1	35.1	**33.9 ± 1.7**
>= 5	BOW + MNB	20.7	36.2	31.5 ± 5.2
	BOW + Cosine	15.7	29.9	23.5 ± 4.6
	BOW + SVM	16.4	13.1	12.9 ± 1.5
	BOW + Softmax	14.9	14.0	12.7 ± 1.2
	DBRNN-A + Softmax	33.8	38.0	**35.9 ± 2.1**
>= 10	BOW + MNB	18.4	42.5	34.5 ± 7.7
	BOW + Cosine	16.0	35.5	25.1 ± 6.2
	BOW + SVM	17.5	16.2	16.7 ± 0.6
	BOW + Softmax	15.6	14.1	14.3 ± 0.6
	DBRNN-A + Softmax	32.5	39.6	**36.1 ± 2.1**
>= 20	BOW + MNB	21.3	41.8	35.1 ± 7.0
	BOW + Cosine	16.8	38.9	28.9 ± 8.2
	BOW + SVM	14.6	16.4	15.5 ± 0.9
	BOW + Softmax	18.8	15.3	14.0 ± 2.4
	DBRNN-A + Softmax	33.3	43.3	**38.8 ± 3.2**

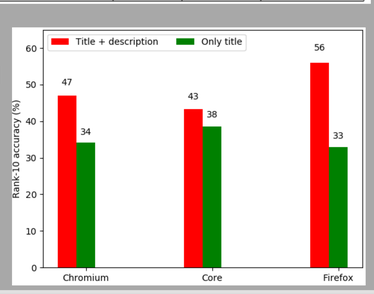
Rank #10 accuracy on Firefox dataset

Sampl.	Classifier	CV#1	CV#10	Average
>= 0	BOW + MNB	19.1	35.55	27.4 ± 5.2
	BOW + Cosine	17.3	30.1	25.7 ± 4.1
	BOW + SVM	13.4	14.6	14.1 ± 1.0
	BOW + Softmax	11.9	13.6	14.6 ± 1.9
	DBRNN-A + Softmax	33.6	38.1	**36.5 ± 1.7**
>= 5	BOW + MNB	21.1	36.5	33.1 ± 5.1
	BOW + Cosine	20.8	35.2	28.5 ± 4.8
	BOW + SVM	14.4	15.2	16.5 ± 1.1
	BOW + Softmax	18.2	13.7	14.8 ± 1.8
	DBRNN-A + Softmax	27.6	44.5	**40.1 ± 5.3**
>= 10	BOW + MNB	21.7	38.5	33.1 ± 4.8
	BOW + Cosine	18.1	36.6	28.7 ± 5.8
	BOW + SVM	09.9	12.8	11.9 ± 1.1
	BOW + Softmax	14.3	12.7	12.1 ± 1.8
	DBRNN-A + Softmax	35.1	51.4	**44.8 ± 5.6**
>= 20	BOW + MNB	22.0	38.4	30.4 ± 6.2
	BOW + Cosine	18.4	38.3	29.8 ± 6.3
	BOW + SVM	18.7	21.9	19.6 ± 2.2
	BOW + Softmax	16.5	12.9	13.1 ± 1.3
	DBRNN-A + Softmax	38.9	55.8	**46.6 ± 6.4**

Effect of no. of training samples:



Effect of Using title and description:



Conclusion

- Learnt a paragraph level feature representation of a bug report preserving the semantic and the syntactic relation over a longer context.

[1] J. Anvik, L. Hiew, and G. C. Murphy, "Who should fix this bug?" in International Conference on Software Engineering, 2006, pp. 361–370.

Code & dataset made public @ http://bugtriage.mybluemix.net/