SPAM DETECTOR

NLP course, May 2019

Shir and Johana

The idea

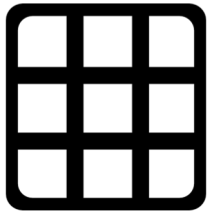
Using text messages data to learn classifying whether it is a spam

What? limited size and imbalanced data

13%
Spam

87%
Ham

How? Feature construction and models



1. TF-IDF

3 algorithms: logistic regression, naïve Bayes, random forest

2. Word embeddings

- Trained word2vec VS. pre-trained (Glove). Vector size 50.
- NN with following layers: embedding, convolutional (128 filters), pooling, hidden layer (size 10), output layer (size 1 binary). > relu for hidden, sigmoid for output layer.

Performance: on test set (after validation set)

	Baseline	Logistic regression (tfidf)	Naive Bayes (tfidf)	Random Forest (tfidf)	NN (trained word2vec)	NN (pre-trained)
Accuracy	76	97	80	97	98	98
Precision (0l1)	87 l 12	98 l 96	97 l 39	98 l 99	99 l 97	98 l 97
Recall (0l1)	87 l 12	99 l 84	80 l 85	100 l 83	100 l 93	100 l 88
F1 (0l1)	87 l 12	99 l 90	88 l 54	99 l 90	99 l 95	99 l 93
ROC AUC	49	91	82	91	96	93

Why? Top features in RF make sense

1+1 sale!

Call us on XXX!

	Word	Importance		Word	Importance
1	number	0.14783	**6**	repli	0.02708
2	call	0.03395	**7**	mobil	0.02657
3	numberp	0.03308	**8**	free	0.02517
4	txt	0.03200	**9**	tone	0.02234
5	claim	0.02776	**10**	urgent	0.02059

Txt back your vote!

Want to get free cookie? Reply to XXX


KEEP
CLAM
AND
AVOID SPAM