Botection · 博文强识

Progress Report I

Shangbin Feng, Herun Wan, Ningnan Wang
Xi'an Jiaotong University
{wind_binteng,wanherun,mrwangyou}@stu.xjtu.edu.cn

Report Outline

I. Introduction

II. Overall Architecture

III. Data Collection & Preprocessing

IV. Bi-LSTM Textual Network

V. Random Forest Classifier

VI. Result Analysis

VII. Deployment

VIII. Conclusion & Future Work

I. Introduction

Account-level

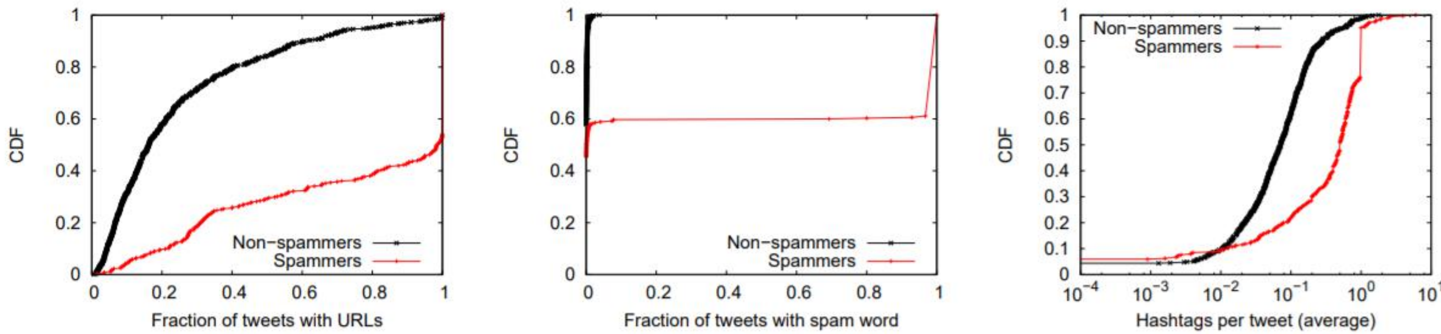
Unsupervised

Supervised

Tweet-level

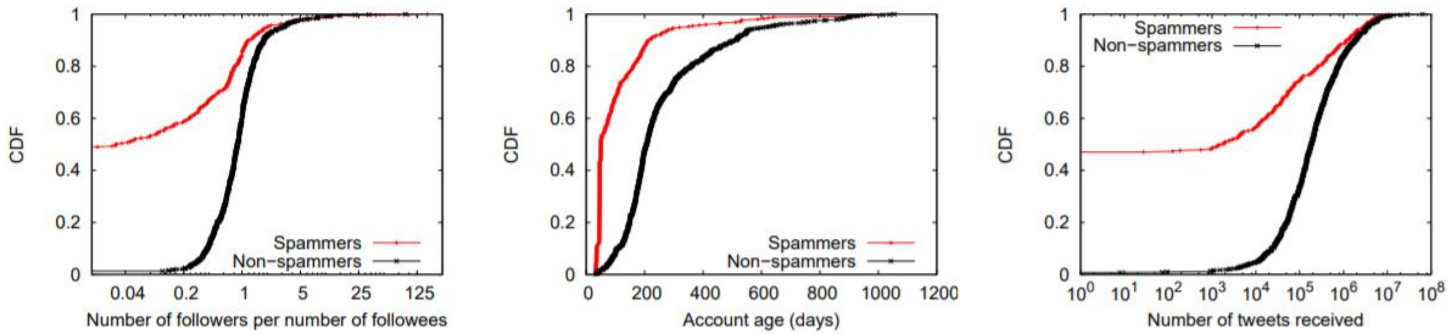
I. Introduction

Supervised & Account-level[1]



(a) Fraction of tweets containing URLs (b) Fraction of tweets with spam words (c) Average number of hashtags per tweet

Figure 3: Cumulative distribution functions of three content attributes



(a) Fraction of followers per followees (b) Age of the user account (c) Number of tweets received

Figure 4: Cumulative distribution functions of three user behavior attributes

I. Introduction

Supervised & Tweet-level[2]

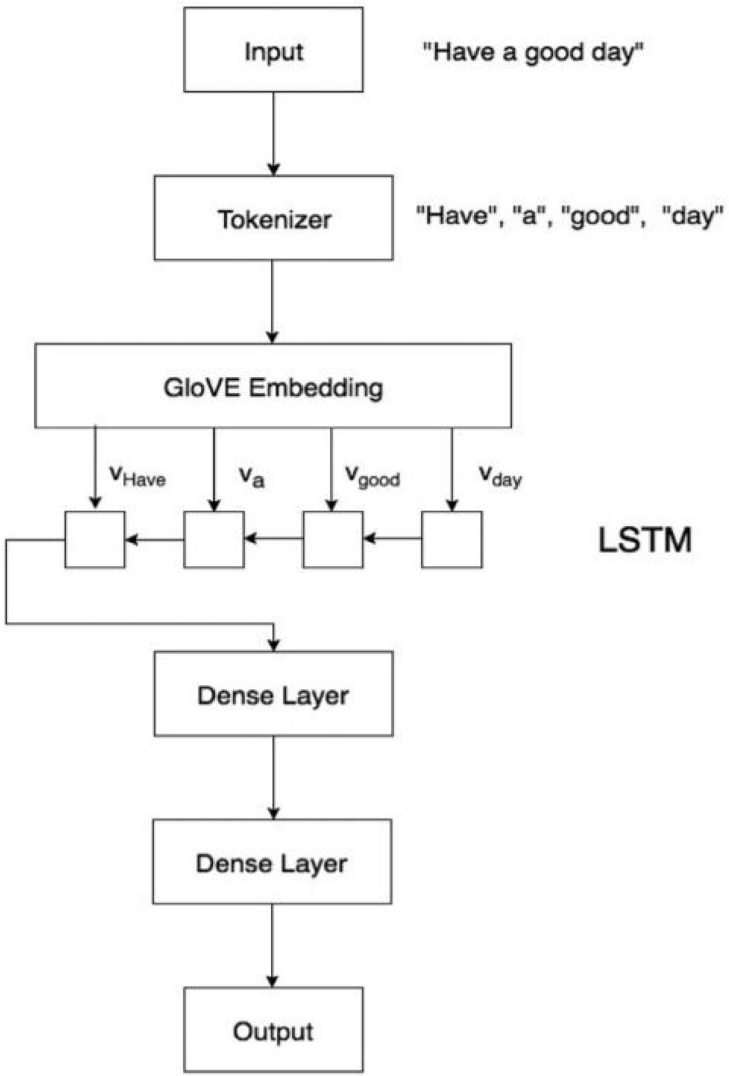
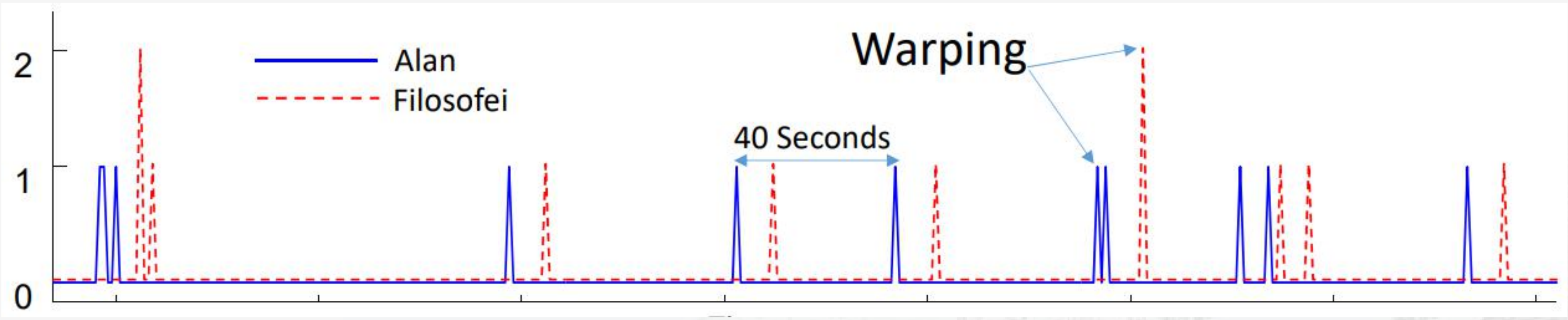


Fig. 1. Architecture of model for tweet-level bot detection that takes only the tweet content as its input.

I. Introduction

Unsupervised & Account-level[3]



I. Introduction

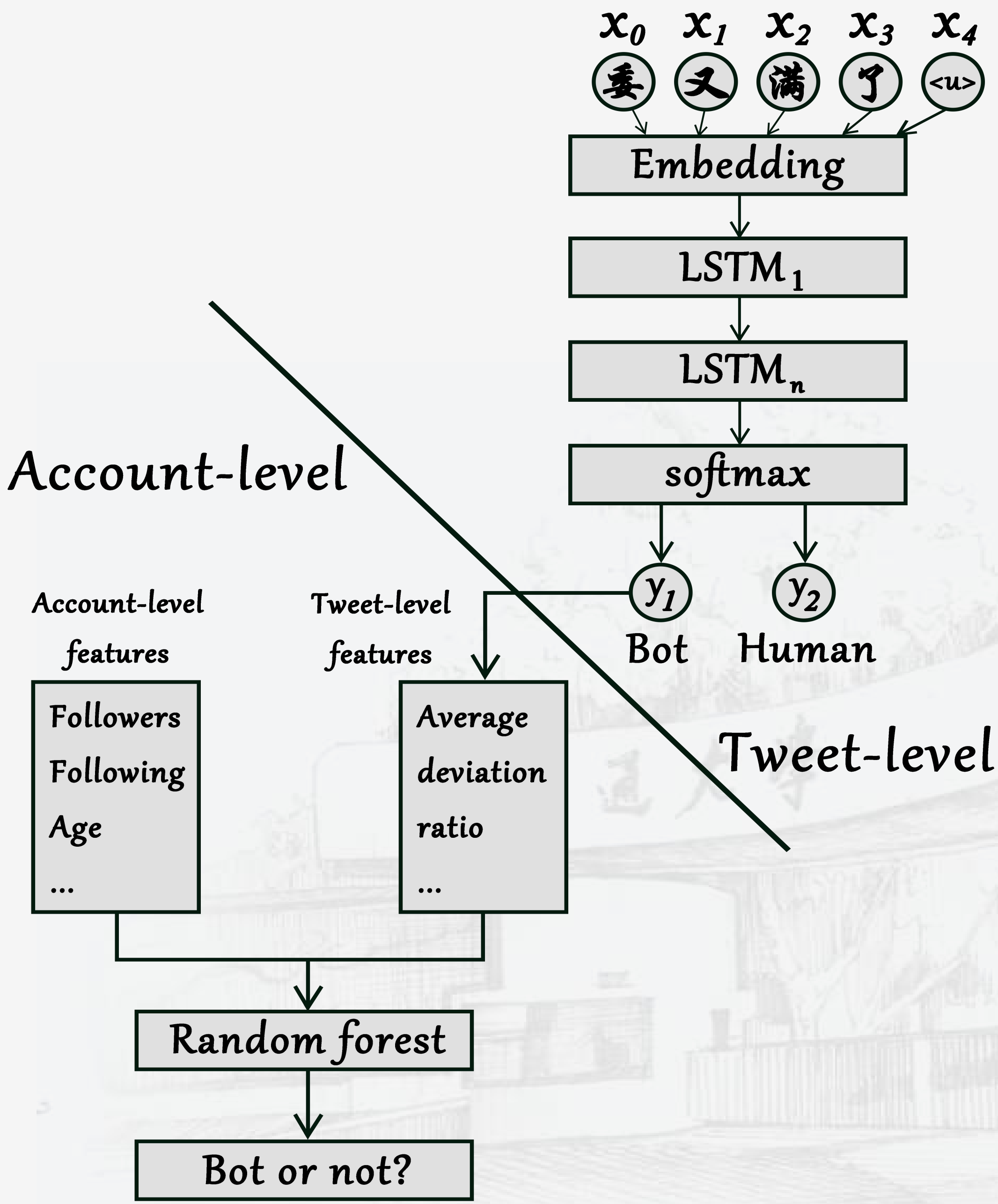
Contribution

-Be the first to integrate account-level & tweet-level detection(to the best of our knowledge)

-Collect Weibo data, label and preprocess into a dataset which would be publicized to the research community

-Our proposed model achieves competitive performance compared with existing state-of-the-art bot detection systems

II. Overall Architecture: Botection



III. Data Collection & Preprocessing

Data collection

Data preprocessing

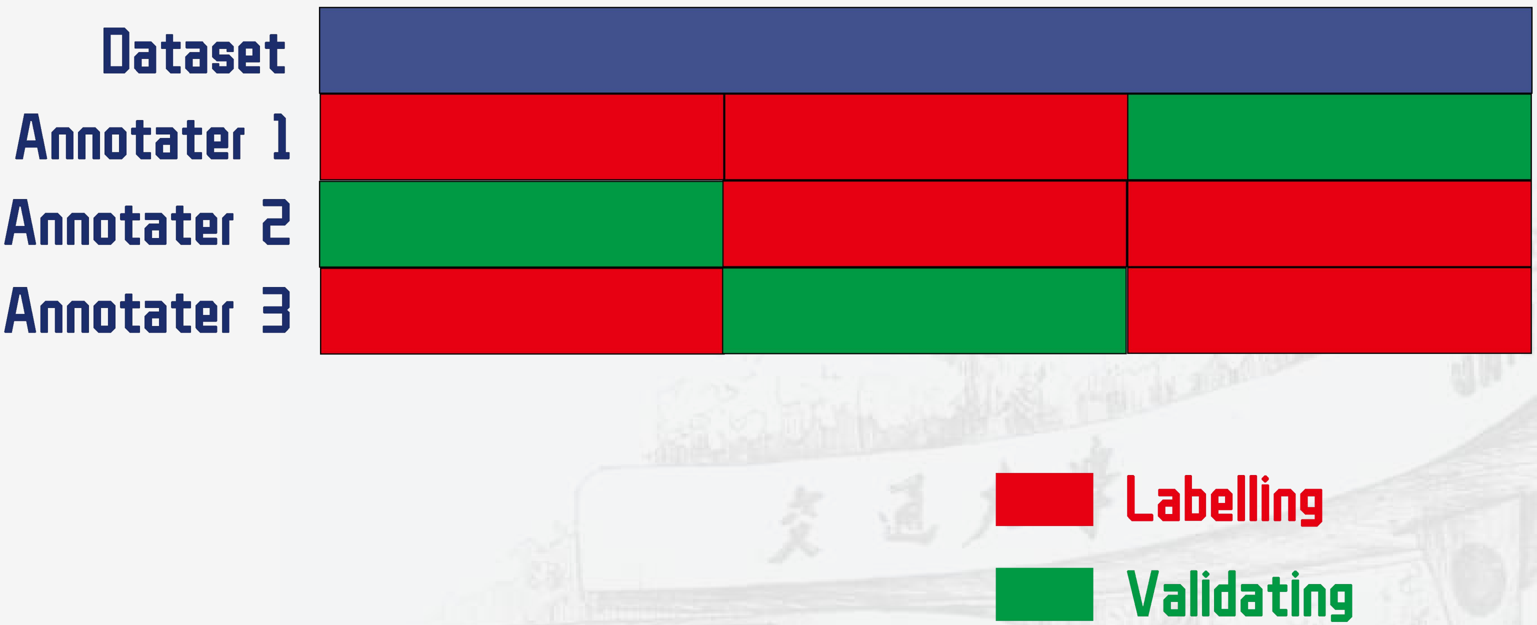
-≥4kB

-@ → ‘ ttttt ’
→ ‘ ggggg ’
<url> → ‘ uuuuu ’
☺ & emoji → ‘ eeeee ’
out-of-vocabulary words → ‘ ooooo ’

-lol//@fsb: hhhhh//@whr: Repost → lol

III. Data Collection & Preprocessing

Labelling



III. Data Collection & Preprocessing

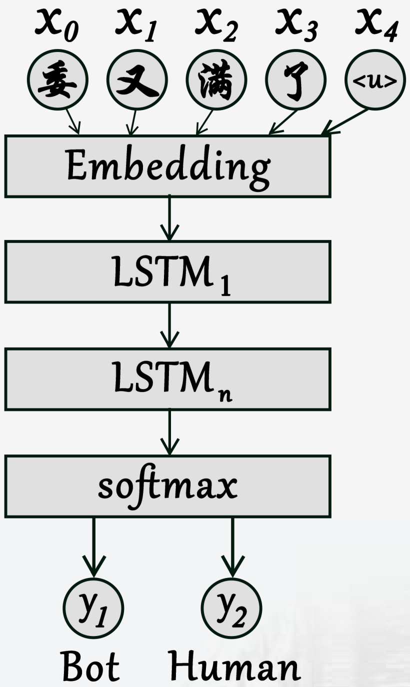
Final dataset composition

-1154 accounts with ≥4kB textfile

-985 accounts available

-95385 posts in total

IV. Bi-LSTM Textual Network



tokenizer:
PyNLPIR by [4]
Chinese Academy of Sciences

embedding:
pretrained on Weibo dataset by [5]
Word2vec / Skip-Gram with Negative Sampling
with a vocabulary of size 195203, vector dimension of 300
randn for special characters/zeros for oov

LSTMn:
Bi-LSTM layers
n = 3(hyperparameter)

softmax:
converts the output of the last LSTMn layer into two numbers y1 & y2,
representing the probability of being bot/human.

IV. Bi-LSTM Textual Network

Training:

HIDDEN_DIM = 100
NUM_LAYER = 3

BATCH_SIZE = 32(with regard to a total of 95385 tweets)

EPOCH = 100(with no_up settings for validation set)

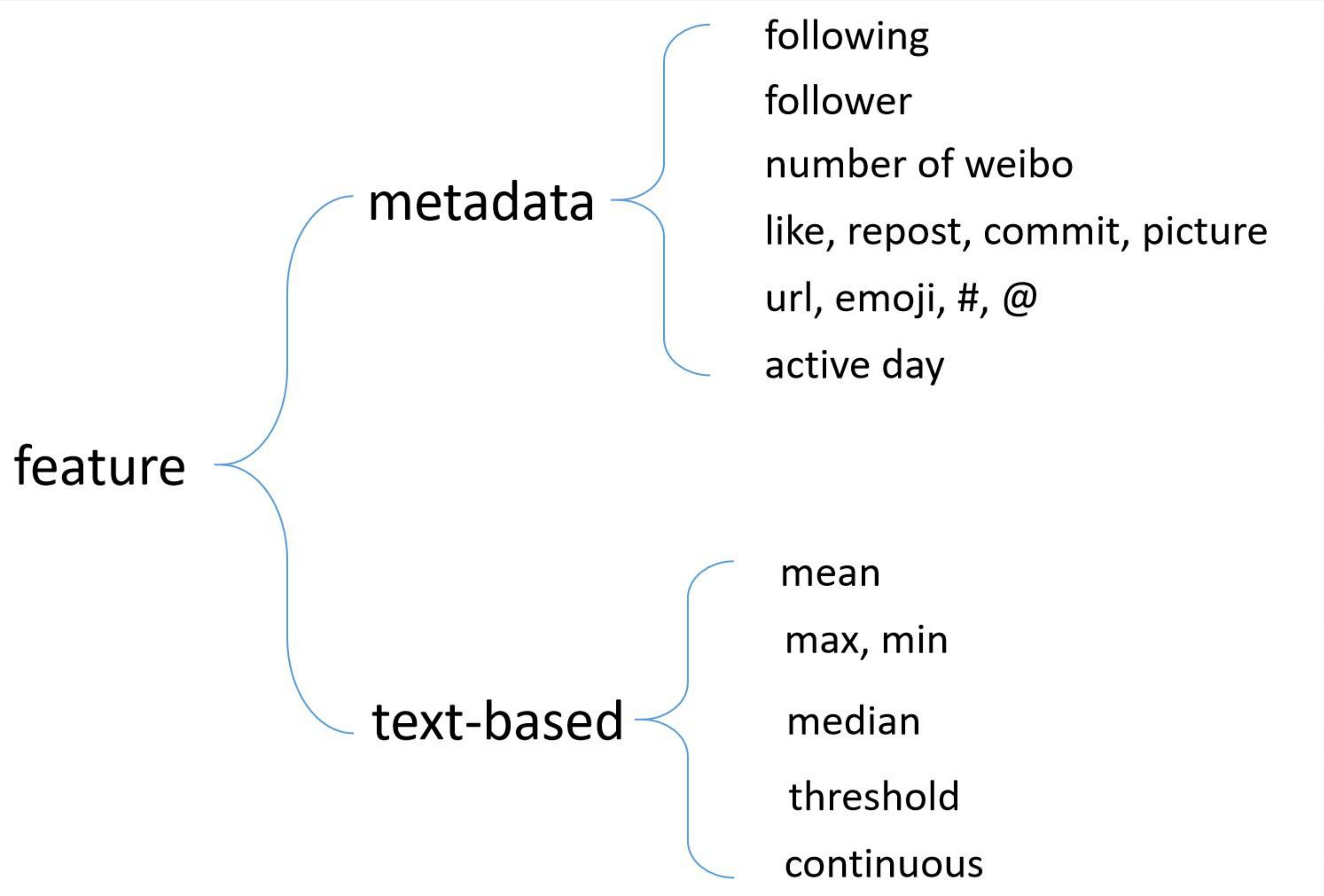
DROPOUT = 0.5, DROPOUT_SCHEDULE = 0.95
WEIGHT_DECAY = 5e-4

OPTIMIZER = torch.adam
LEARNING_RATE = 1e-3

thx for the server!! (although CUDA drive version is antiquated...)

V. Random Forest Classifier

Feature selection:



V. Random Forest Classifier

hyperparameters

Number of decision trees *n*
Max depth of each decision tree *d*

These hyperparameters are tuned using the predetermined validation set.

The result is:
n = 25
d = 15

VI. Result Analysis

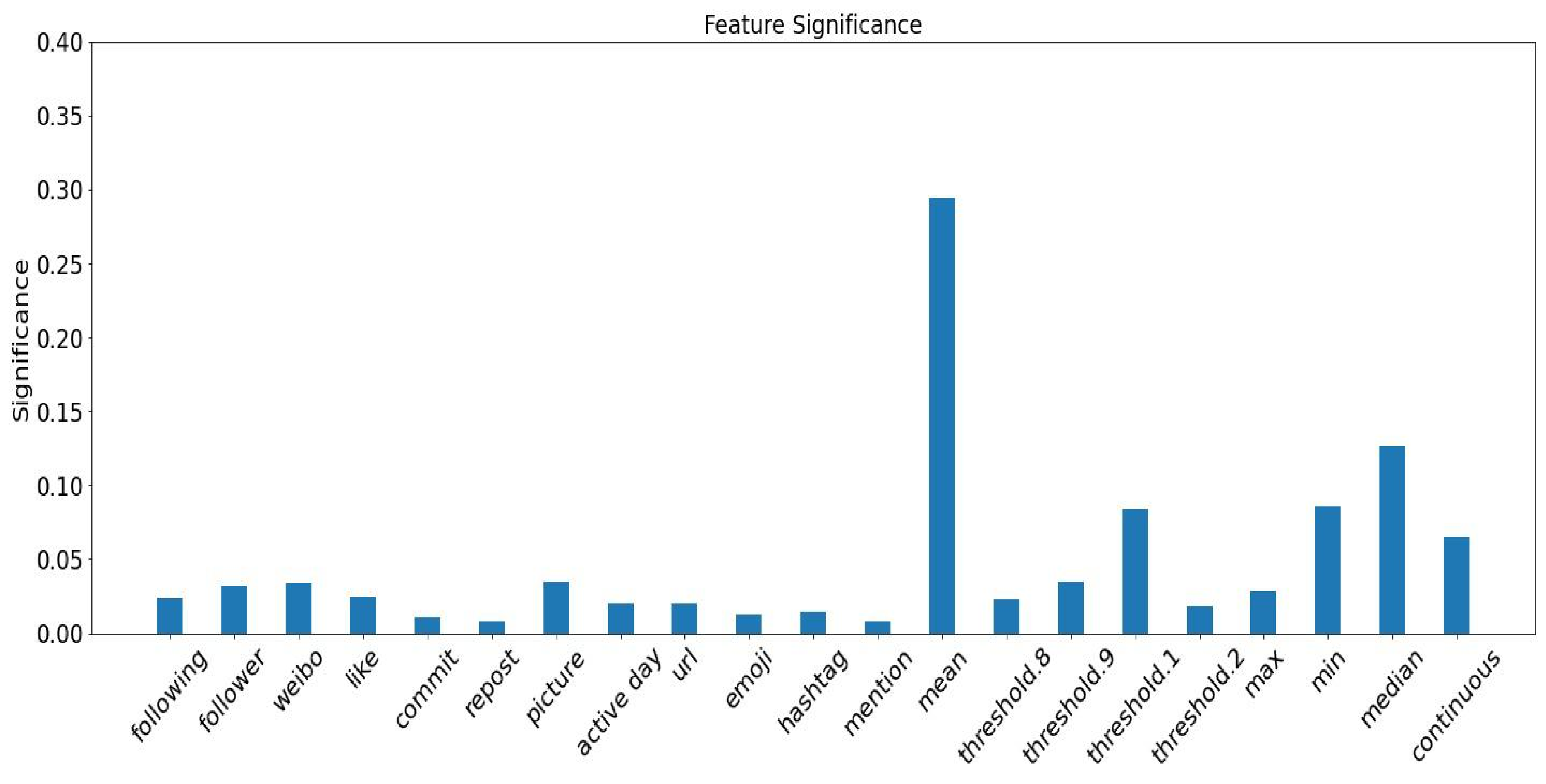
feature	precision	recall	specificity	accuracy	f-measure	MCC
with text-based	**1.000**	**0.974**	**1.000**	**0.987**	**0.987**	**0.974**
without text-based	0.866	0.829	0.866	0.847	0.847	0.694

Text-based features significantly improve the overall performance of Botection

VI. Result Analysis

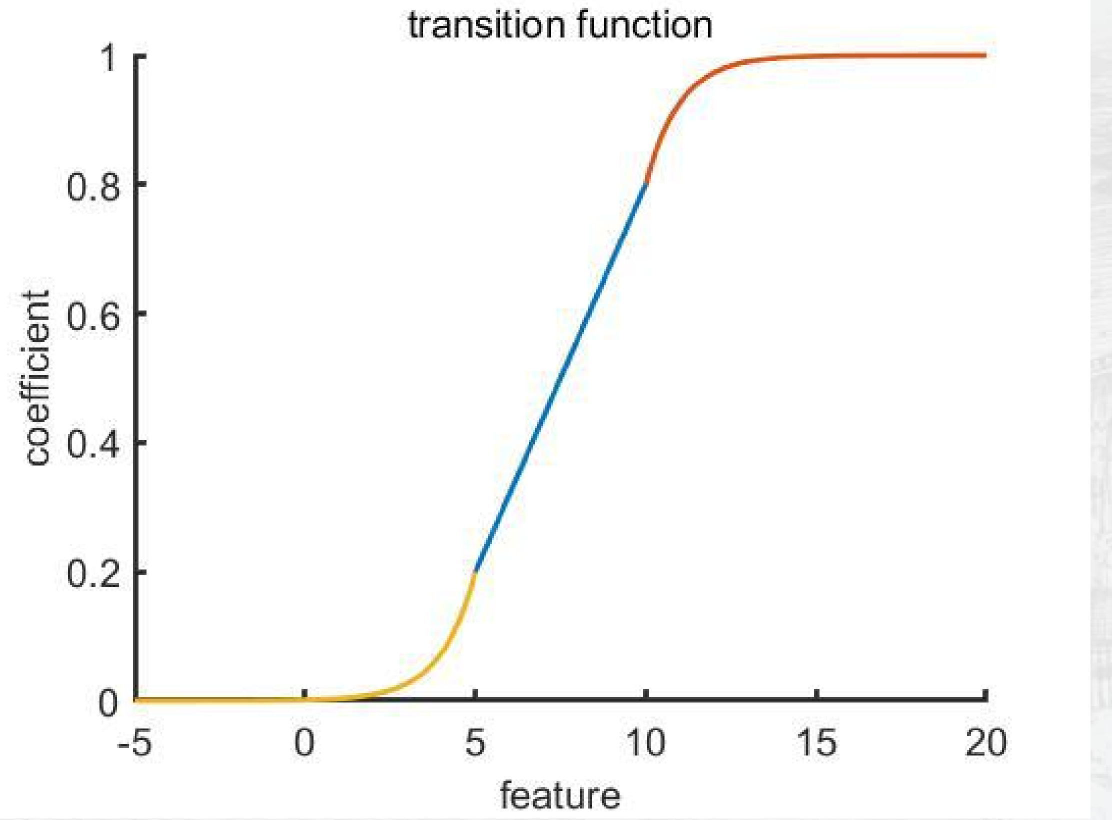
feature	precision	recall	specificity	accuracy	f-measure	MCC
BotOrNot[6]	0.471	0.208	0.918	0.734	0.288	0.174
C. Yang et al.[7]	0.563	0.170	0.860	0.506	0.261	0.043
Miller et al.[8]	0.555	0.358	0.698	0.526	0.435	0.059
W. Feng et al.[9]	0.940	**0.976**	0.935	0.961	0.963	0.920
Ahmed et al.[10]	0.945	0.944	0.945	0.943	0.944	0.886
Cresci et al.[11]	0.982	0.972	0.981	0.976	0.977	0.952
Ours	**1.000**	0.974	**1.000**	**0.987**	**0.987**	**0.974**

VI. Result Analysis



VI. Result Analysis

interpretability:
1) probability
2) metadata coefficient
3) text coefficient



VII. Deployment

Front end & back end

- -Vue.js
- -ThinkPHP 5

VII. Deployment

请输入微博用户名

@人民日报

人民日报是微博机器人的概率为_____

被关注数得分：______

关注数得分：______

活跃时间得分：______

……

A demo of the website is available for trail

VIII. Conclusion & Future Work

Summing up briefly:

-Collected Weibo data, labelled & preprocessed into a dataset which could be publicized to the research community

-Proposed **Botection**, which:

combines textual info and metadata

integrates tweet-level & account-level detection

achieves significant(surprisingly) performance on real-world data

-Analyzed feature selection and its effect on the performance

-Proposed interpretable values for user reference

This project is at 90% progress currently.

VIII. Conclusion & Future Work

Future work:

For the 10% remaining:

-Deployment as planned

-Github repository

For paper publication:

-Dataset acquiring[12]

-Reading group

-Idea Practice

temporal pattern of tweet metadata

group anomaly behaviour

source of tweet pics

advanced feature design(graph, neighbor, ...)

comprehensive feature evaluation(effective+feasible+robust)

...

-Demo-track paper?

Reference

[1]Z. Alom, B. Carminati and E. Ferrari, "Detecting Spam Accounts on Twitter," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, pp. 1191-1198, doi: 10.1109/ASONAM.2018.8508495.

[2]Kudugunta, S., Ferrara, E., 2018. Deep neural networks for bot detection. Information Sciences.. doi:10.1016/j.ins.2018.08.019

[3]Chavoshi, Nikan & Hamooni, Hossein & Mueen, Abdullah. (2016). DeBot: Twitter Bot Detection via Warped Correlation. 10.1109/ICDM.2016.0096.

[4]https://github.com/tsroten/pynlpir

[5]https://github.com/Embedding/Chinese-Word-Vectors

[6]C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, “Botornot: A system to evaluate social bots,” in Proc. 25th Int. Conf. Companion on World Wide Web, 2016.

Reference

[7]C. Yang, R. Harkreader, and G. Gu, “Empirical evaluation and new design for fighting evolving twitter spammers,” IEEE Trans. Information Forensics Security, vol. 8, no. 8, pp. 1280–1293, 2013.

[8]F. Wei and U. T. Nguyen, "Twitter Bot Detection Using Bidirectional Long Short-Term Memory Neural Networks and Word Embeddings," 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), Los Angeles, CA, USA, 2019, pp. 101-109, doi: 10.1109/TPS-ISA48467.2019.00021.

[9]Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, “Twitter spammer detection using data stream clustering,” Information Sciences, vol. 260, pp. 64–73, 2014.

[10]F. Ahmed and M. Abulaish, “A generic statistical approach for spam detection in online social networks,” Computer Communications, vol. 36, no. 10-11, pp. 1120–1129, 2013.

Reference

[11]S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, “Dna-inspired online behavioral modeling and its application to spambot detection,” IEEE Intelligent Systems, vol. 31, no. 5, pp. 58–64, 2016.

[12]S.Cresci,“Mib datasets,” http://mib.projects.iit.cnr.it/dataset.html, 2017.

西安交通大學
XI'AN JIAOTONG UNIVERSITY

Thank you!

Shangbin Feng, Herun Wan, Ningnan Wang
Xi'an Jiaotong University
{wind_binteng,wanherun,mrwangyou}@stu.xjtu.edu.cn