Spam Review Detection with Graph Convolutional Networks.

Ao Li, Zhou Qin, Runshi Liu, Yiqun Yang and Dong Li

José Antonio Álvarez Ocete

March 16, 2022

Máster en Ciencia de Datos
Escuela Politécnica Superior

Framework

Context

Focused on **Xianyu**, largest second-hand goods app in China.

Context

Focused on **Xianyu**, largest second-hand goods app in China.

In Xyanyu, reviews are **comments**.

Context

Focused on **Xianyu**, largest second-hand goods app in China.

In Xyanyu, reviews are **comments**.

Main challenges:

- **Scalability** - over 1 billion goods and over 10 million users.
- **Adversarial actions** taken by spammers.

Adversarial actions

Camouflage:

- Dial this number for a part-time job.
- Want to earn more money in your spare time? Contact me.

Adversarial actions

Camouflage:

- Dial this number for a part-time job.
- Want to earn more money in your spare time? Contact me.

Deforming the comments:

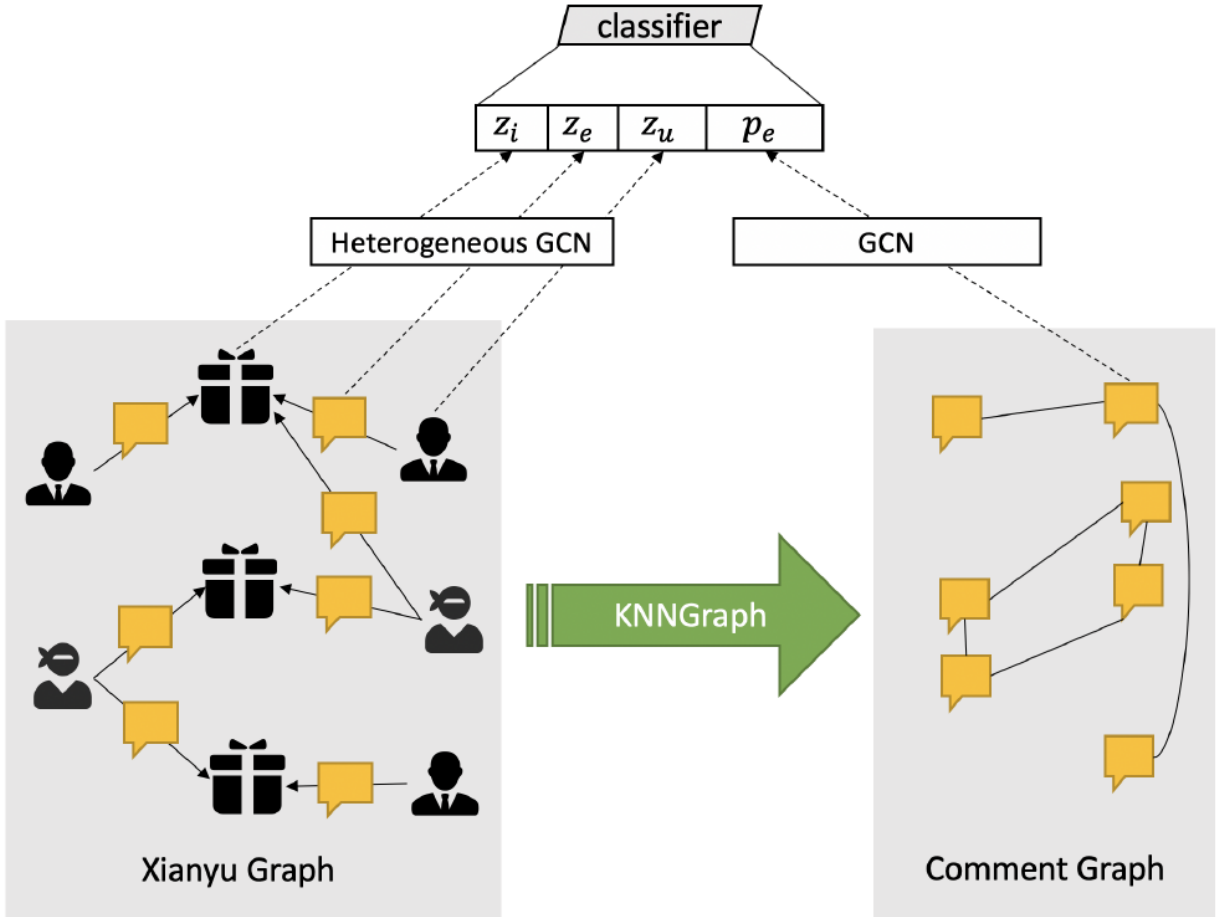
- Add my vx.
- Add my v.
- Add my wx.
- Add my WeChat.

The GAS model

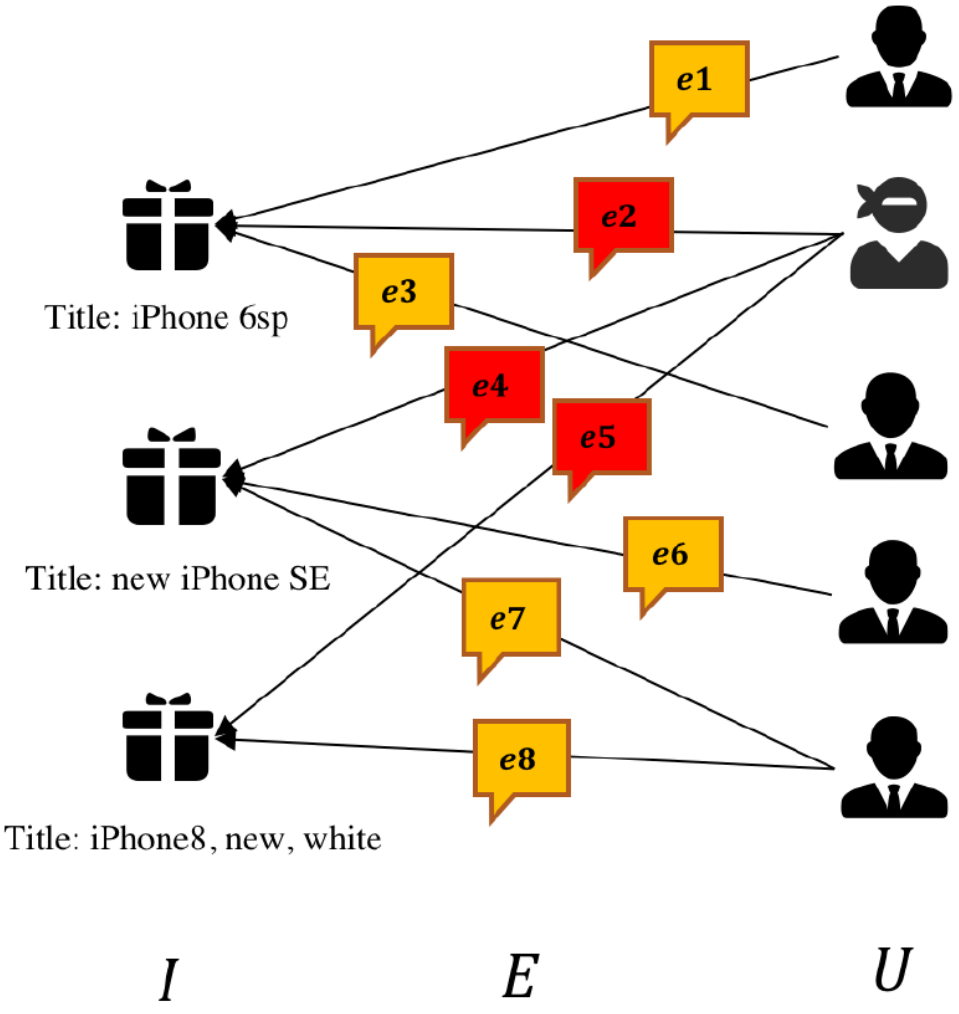
GCN-based Anti-Spam (GAS) method

Using:

- Local context.
- Global context.



Local context - Xianyu Graph



e1: more discount?

e2: get an iPhone 8plus,$130, contaxt me: #1

e3: hello, I want it

e4: get an iPhone 8plus,$130, contaxt me: #1

e5: get an iPhone 8plus,$130, contaxt me: #1

e6: how alrge is the storage space?

e7: hi, is it still on sale?

e8: is it a complete new one?

Graph Convolutional Networks (GCN)

For each leayer i, aggregation and combination can be written as:

$$h^i_{N(v)} = \sigma\left(W^i \cdot AGG\left(\left\{h^{i-1}_v, \forall v' \in N(v)\right\}\right)\right)$$

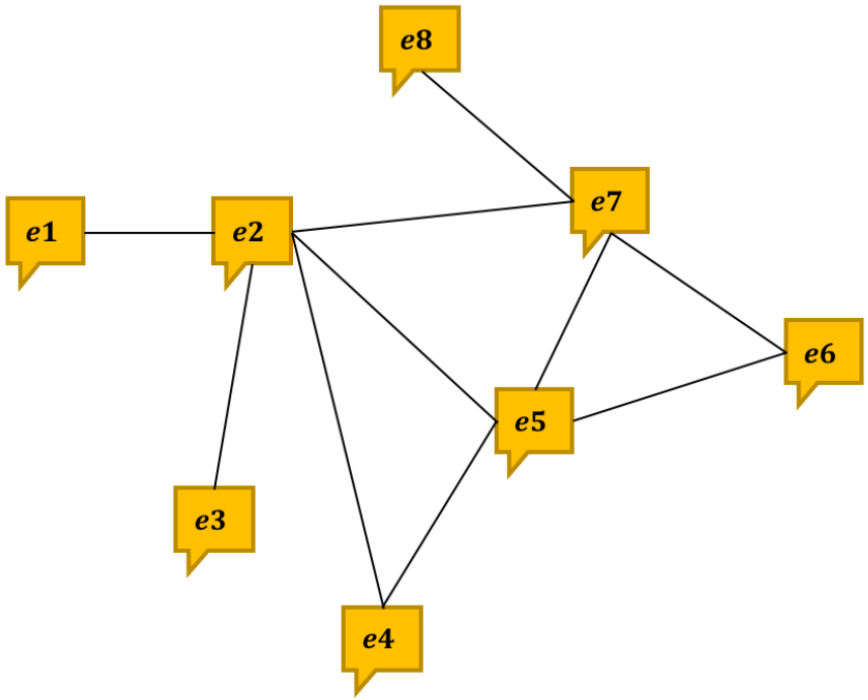
$$h^i_v = COMBINE\left(h^{i-1}_v, h^i_{N(v)}\right)$$

The GCN is fed with:

- Item features.
- User features.
- Comment features: Embeddings by **word2vec** are the input of **TextCNN**. The output of TextCNN is used as the embedding of the comment:

$$h^0_e = \text{TextCNN}(w_0, w_1, \ldots, w_n)$$

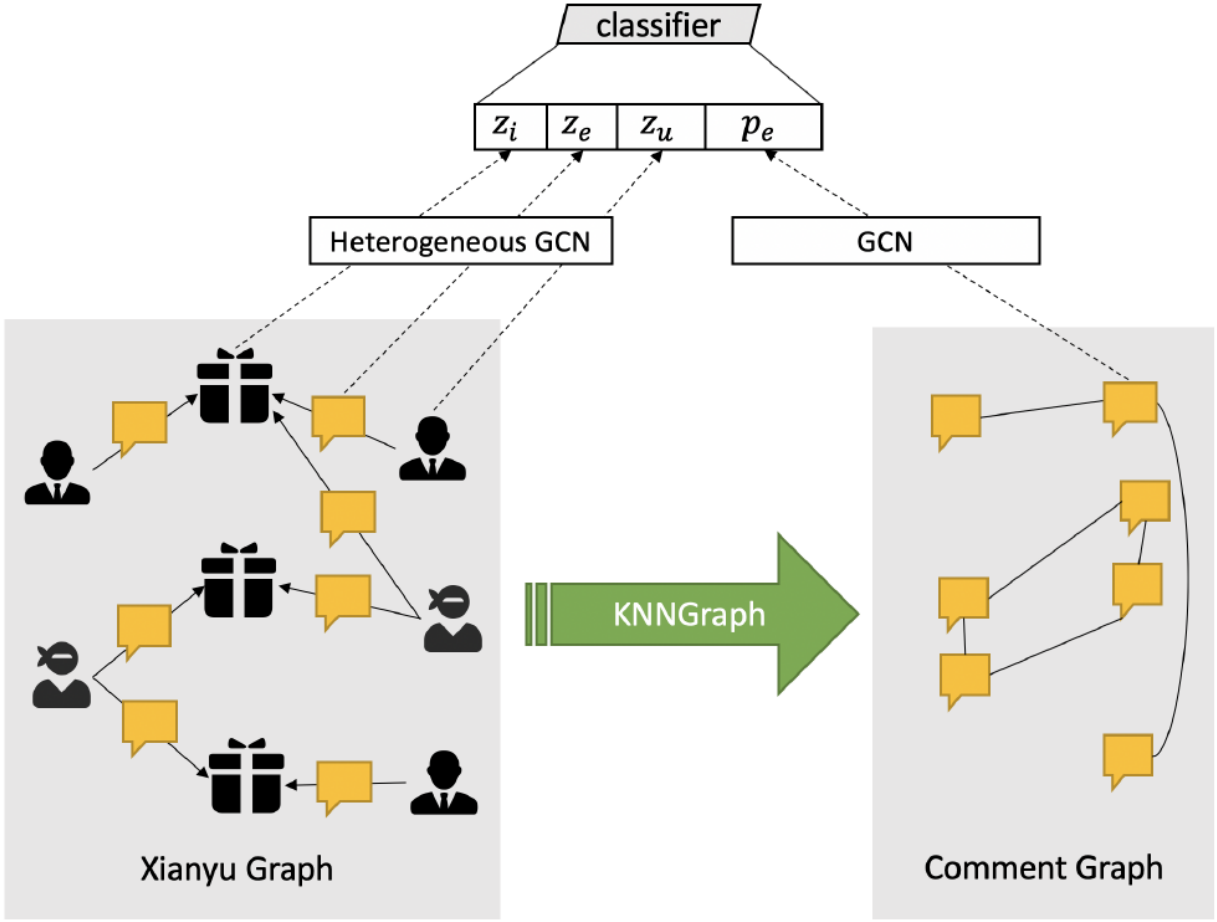
Global context - Comment graph



Processed as follows:

- Remove duplicated comments.
- Generate comments embeddings.
- Approximate KNN to obtain similar comments.
- Recomve comments by same user or under same item.

Classification step



$$y = classifier(concat(z_i, z_u, z_e, p_e))$$

Experiments and results

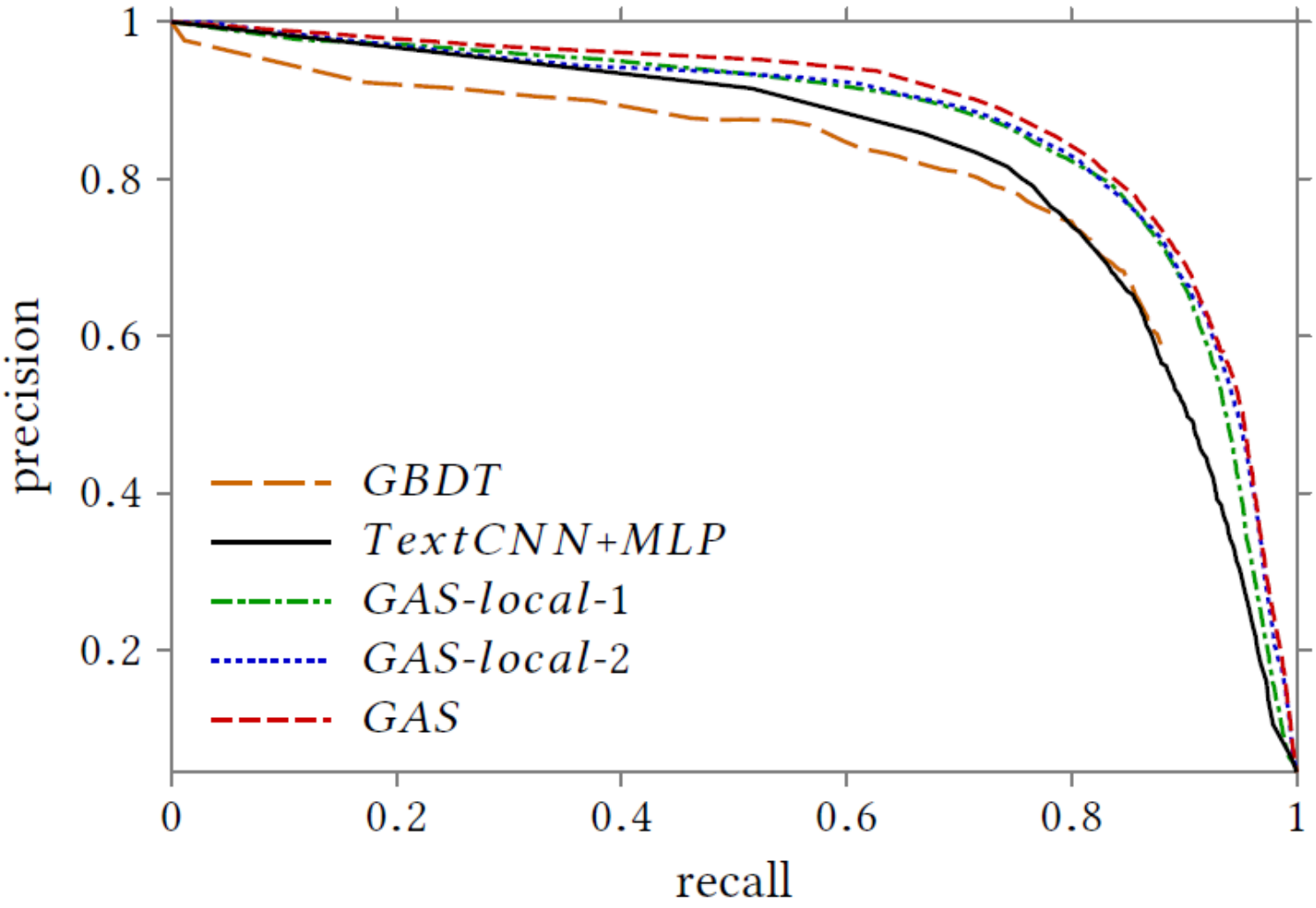
Models compared

- **GBDT**: Domain expert features with Gradient Boosting Decision Trees model.
- **TextCNN+MLP**: user features + item features with 2 layer MLP model.
- **GAS-local-1**: 1 propagation layer on Xianyu Graph.
- **GAS-local-2**: 2 propagation layer on Xianyu Graph.
- **GAS**: GAS model with 2 propagation layers on Xianyu Graph and 1 propagation layer on Comment Graph.

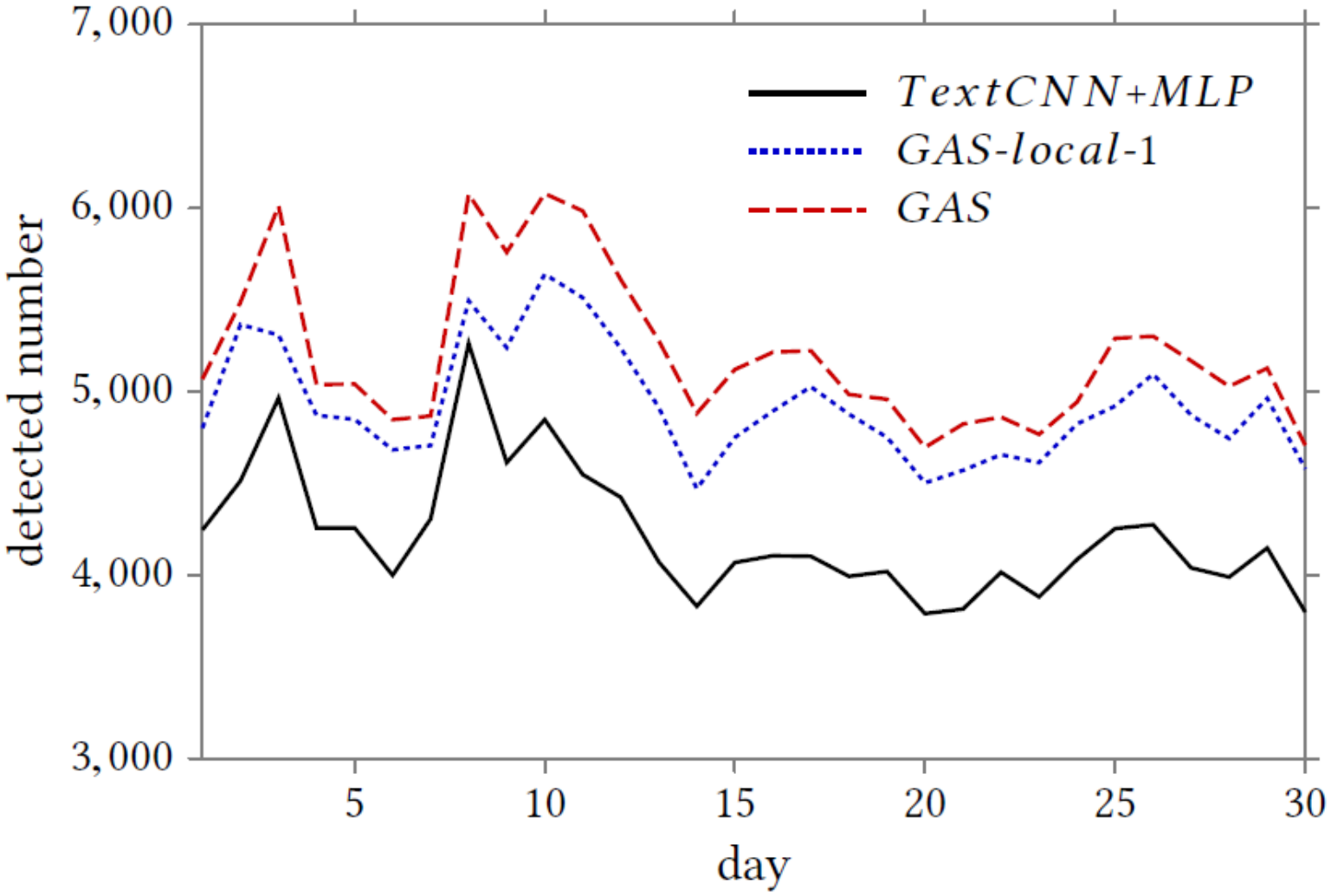
Offline evaluation (1)

Method	AUC	F1 score	recall@90% precision
GBDT	0.9649	0.7686	50.55%
TextCNN+MLP	0.9750	0.7784	54.86%
GAS-local-1	0.9806	0.8138	66.90%
GAS-local-2	0.9860	0.8143	67.02%
GAS	**0.9872**	**0.8217**	**71.02%**

Offline evaluation (2)



Online evaluation



Thank you for your attention!

References

[Li u. a. 2019] Li, Ao ; Qin, Zhou ; Liu, Runshi ; Yang, Yiqun; Li, Dong: Spam Review Detection with Graph Convolutional Networks. (2019). – URL https://arxiv.org/pdf/1908.10679v1.pdf