Rumor Detection of COVID-19 Pandemic on Online Social Networks

Anqi Shi, Zheng Qu, Qingyao Jia, Chen Lyu

Shanghai University of Finance and Economics, Shanghai, China

presenter：Zheng Qu

2020-10-10

CAMERA

Outline

- Motivation
- Data Collection and System Design
- Evaluation and Results
- Conclusion

CAMERA

Outline

- **Motivation**
- Data Collection and System Design
- Evaluation and Results
- Conclusion

CAMERA

Motivation

- **Social networks** have penetrated into all aspects of people's daily lives.
- The new coronavirus epidemic (**COVID-19**) 

causes the deep health crisis across the world.
- Messages about the COVID-19 are now largely emerging on social platforms, including some false information or **rumors**.



CAMERA

Motivation

- Rumors are regarded as some kinds of **public opinion viruses**.
- After being widely spread, rumors can spread panic and pose a potential threat to social stability.
- Therefore, it is an essential task to **detect rumors of the COVID-19** on social networks.

CAMERA

Our Work

- We extract four types of features: **text features, user-related features, interaction-based features,** and **emotion-based features** to characterize the information related to the COVID-19.

- We design a novel rumor detection model by using **an ensemble learning approach** for the COVID-19 related messages.

- On the Weibo platform, we perform extensive experiments to validate our model.
  - ✓The experiment results show that our model **has a high detection accuracy** for rumor detection.

CAMERA

Outline

- Motivation
- **Data Collection and System Design**
- Evaluation and Results
- Conclusion

Data Collection

- Based on the real data collected from Weibo:

  - User's personal information:
    - including gender, profile, residence, membership level, microblog authentication, number of following, number of fans, number of Weibo or microblogs, etc.

  - Content information of microblog/Weibo:
    - including publish time, publish channel, blog content, number of reposts, number of comments, number of likes, etc.

CAMERA

System Design

- Construct four types of features:
  - Text features:
    - collected from the content of a microblog or message of COVID-19.
  - User-related features:
    - the basic features of Weibo users.
  - Interaction-based features:
    - reflect the degree of attention during the propagation process.
  - Emotion-based features:
    - reflect the personal emotion of the user and the attitude of the released message.

CAMERA

System Design

- The XGBoost model:
  - One of the ensemble learning modes
  - Prevent overfitting
  - Deal with missing data effectively
  - Improve the training speed

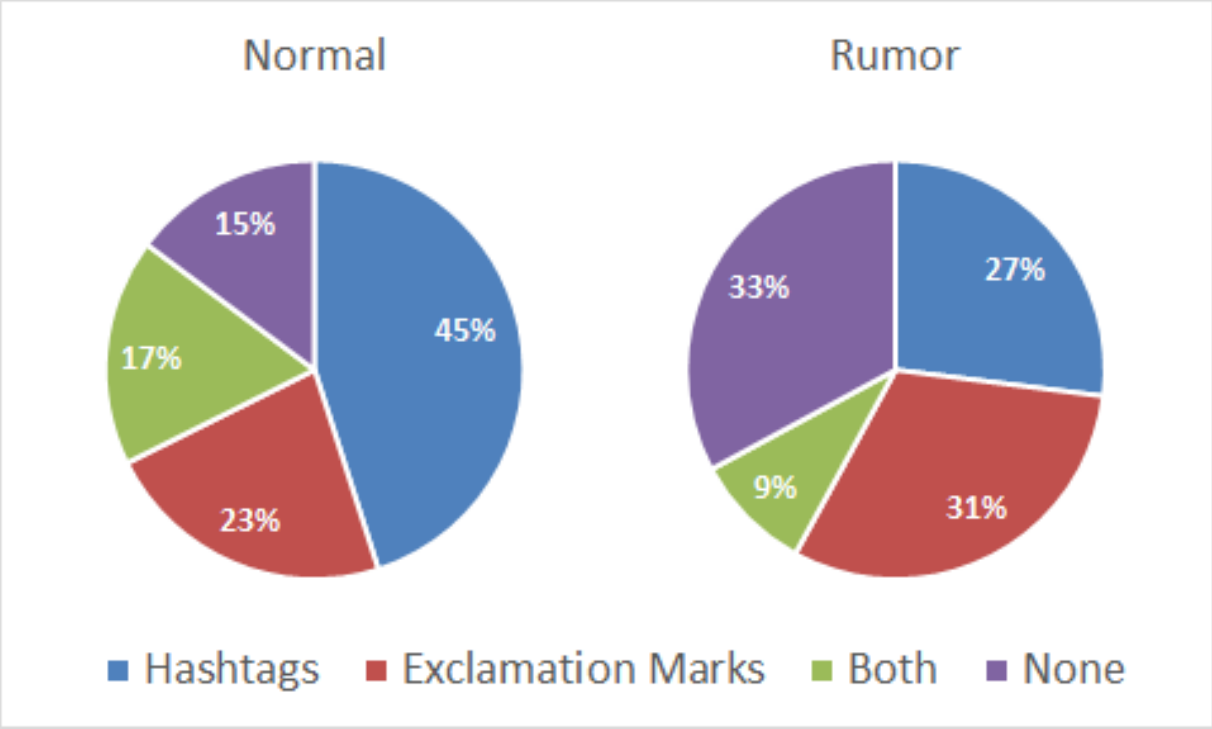
Outline

- Motivation
- Data Collection and System Design
- **Evaluation and Results**
- Conclusion

CAMERA

Data Analysis

- Text features:



	Pictures	Videos	None
Normal	44%	40%	16%
Rumor	43%	34%	23%

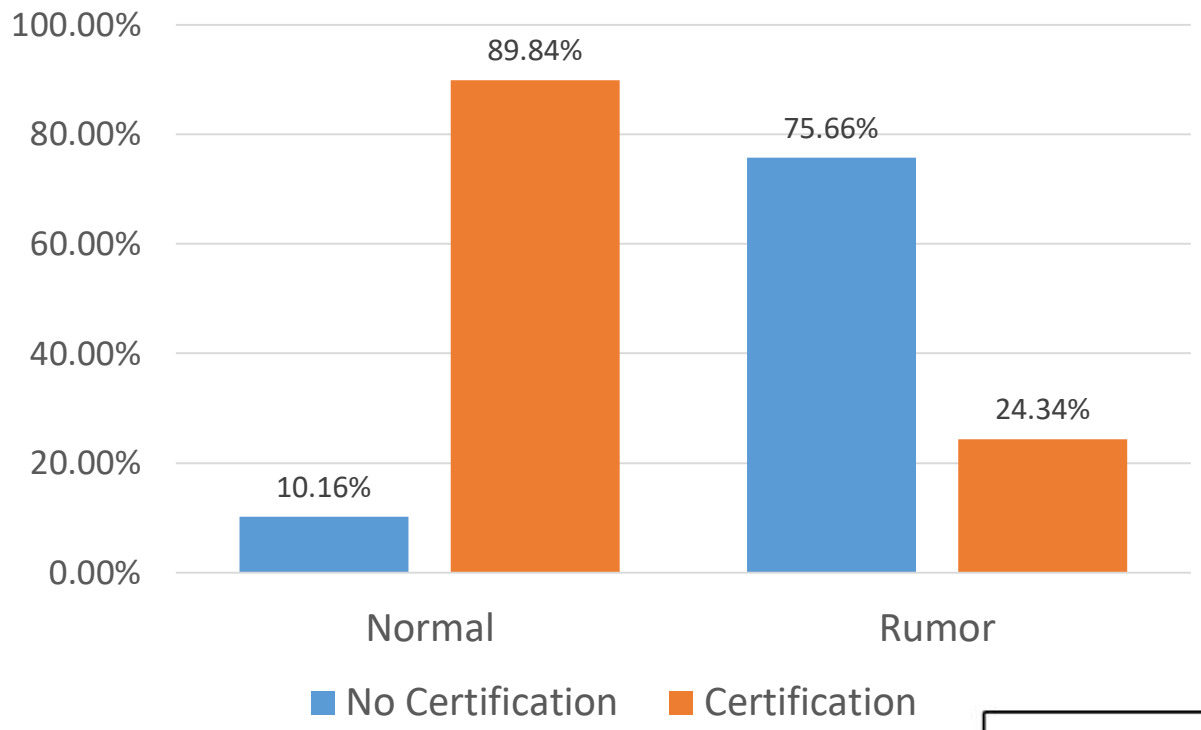
CAMERA

Data Analysis

- User-related Features:

		Normal	Rumor
Gender	Female	25.89%	30.73%
	Male	74.11%	69.27%
Weibo	Mean	23164	14903
	Median	12477	4412
Member	0	308	272
	1	3	0
	2	31	14
	3	50	27
	4	65	32
	5	98	57
	6	224	162
	7	117	77
Reputation	Mean	0.96	0.77
	Median	0.99	0.97

$$\text{Reputation} = \frac{number\ of\ fans}{number\ of\ followers + number\ of\ fans}$$

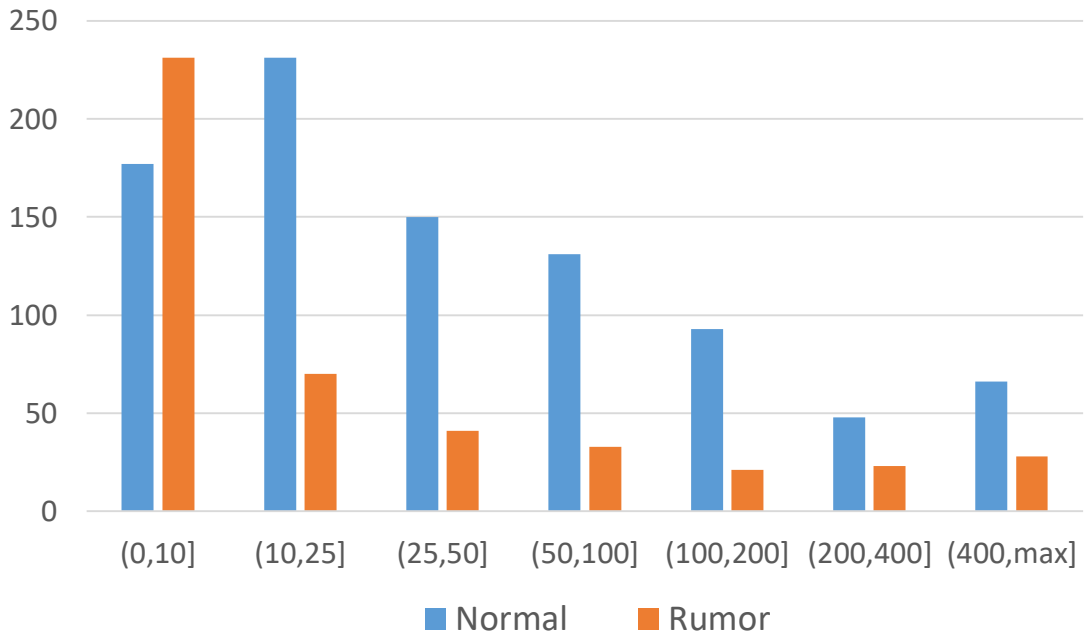


CAMERA

Data Analysis

- Interaction-based Features:

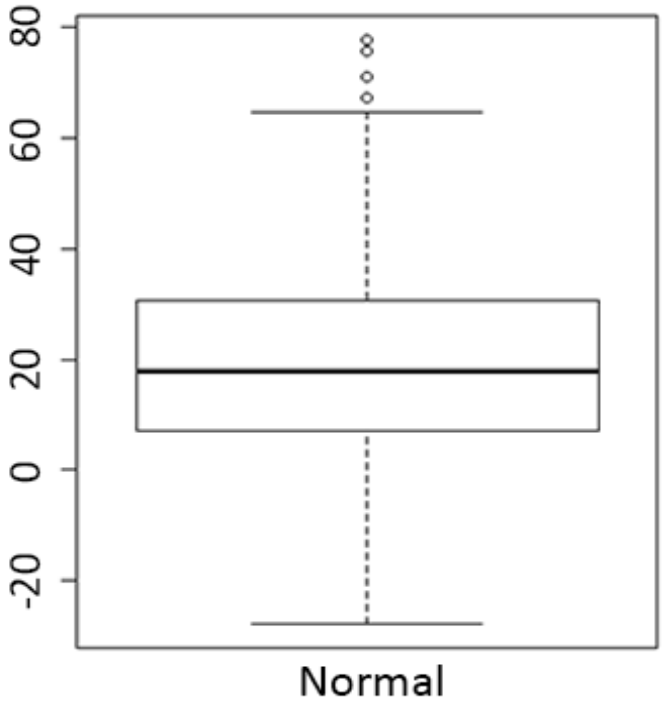
		Normal	Rumor
Reposts	Mean	365	162
	Median	30	2
Comments	Mean	188	117
	Median	29	12
Likes	Mean	1573	1131
	Median	154	19

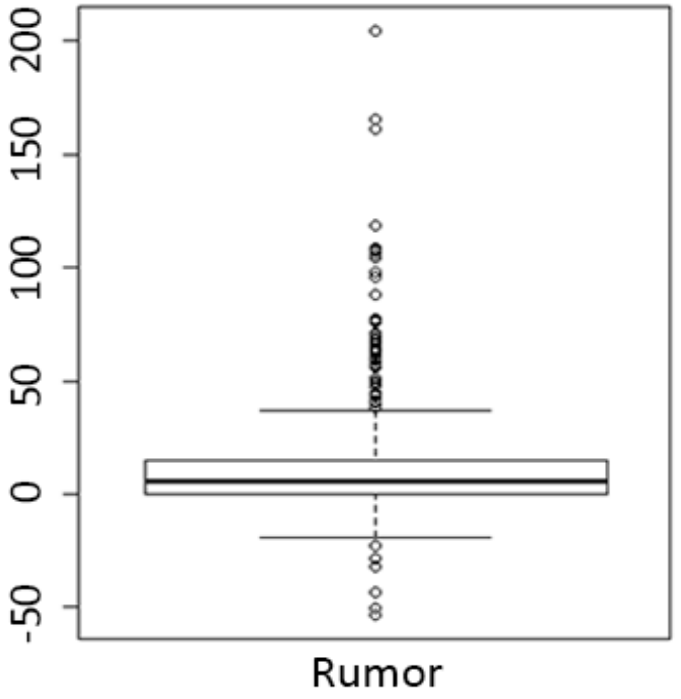


CAMERA

Data Analysis

- Emotion-based Features:





CAMERA

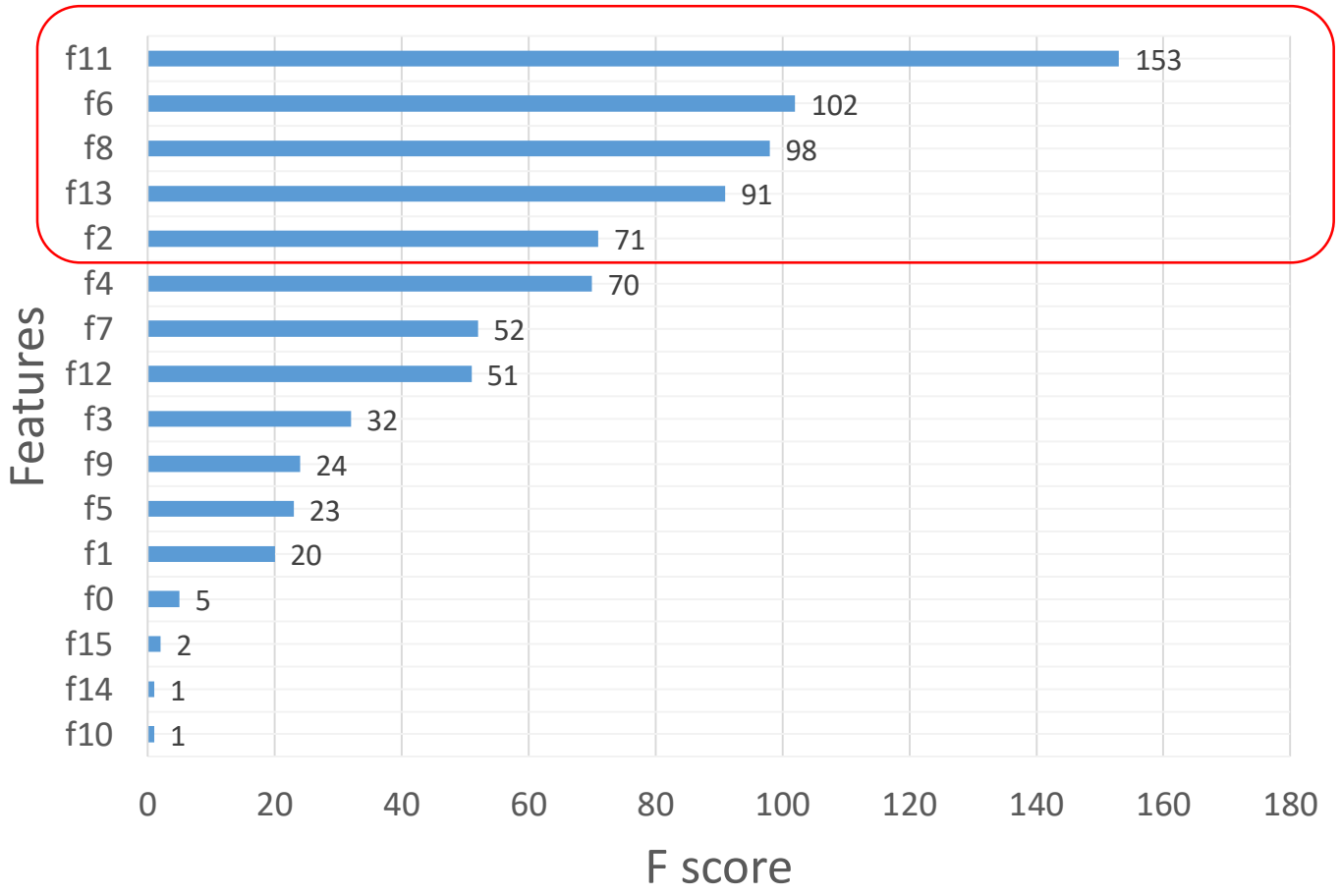
Performance Metric

- The most common evaluation indicators for classification models are accuracy, precision, recall, F1 value, and AUC value:

Model	Accuracy	Precision	Recall	F1	AUC
Text Characteristics	0.70	0.67	0.54	0.60	0.78
User-related Features	0.87	0.85	0.83	0.84	0.91
Interaction-based Features	0.80	0.94	0.55	0.69	0.84
Emotion-based Features	0.69	0.61	0.64	0.63	0.71
Our Detection Model	**0.91**	**0.94**	**0.85**	**0.89**	**0.96**

CAMERA

Feature Importance



f0:	Gender	f1:	Member
f2:	**Certification**	f3:	Following
f4:	Fans	f5:	Weibo
f6:	**Reputation**	f7:	Time
f8:	**Emotion**	f9:	Hashtags
f10:	Exclamation Marks	**f11:**	**Reposts**
f12:	Comments	**f13:**	**Likes**
f14:	Pictures	f15:	Videos

CAMERA

Outline

- Motivation
- Data Collection and System Design
- Evaluation and Results
- **Conclusion**

CAMERA

Conclusion

- We design a rumor detection model for COVID-19 related messages on social networks：
  - We deploy the XGBoost ensemble learning algorithm to construct a new rumor detection model;
  - **Our model combining 16 features in four dimensions can achieve 91% accuracy**.

- We find that the top five features that have the greatest impact on the accuracy of rumor detection are the number of reposts, reputation, emotion, number of likes, and Weibo certification.

CAMERA

Thank You !