ACL 2023

Learn over Past, Evolve for Future: Forecasting Temporal Trends for Fake News Detection

Beizhe Hu[1,2], Qiang Sheng[1,2], Juan Cao[1,2], Yongchun Zhu[1,2], Danding Wang[1], Zhengjia Wang[1,2], Zhiwei Jin[3]

[1]Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences

[2]University of Chinese Academy of Sciences

[3]ZhongKeRuijian Technology Co., Ltd

Presenter: Beizhe Hu

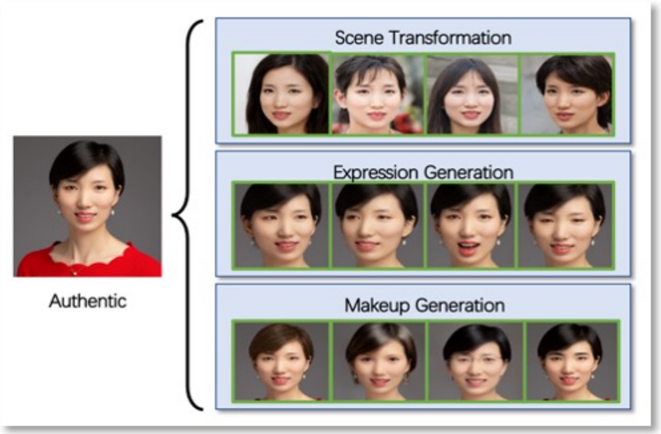
中科睿鉴

Media Synthesis & Forensics Lab

MAKE THE WORLD MORE CREDIBLE

Fake News Detection
Fact-Checking

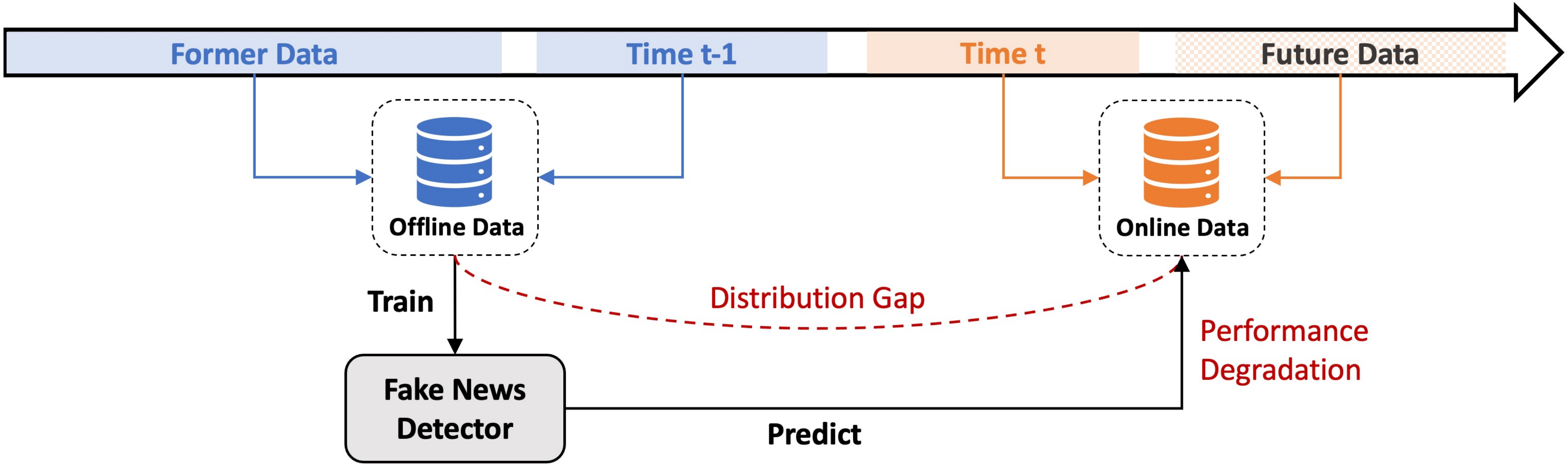
Deep Synthesized Media
Detection & Attribution



Attacking & Defense
for AI Safety

Introduction: Temporal shift in fake news detection

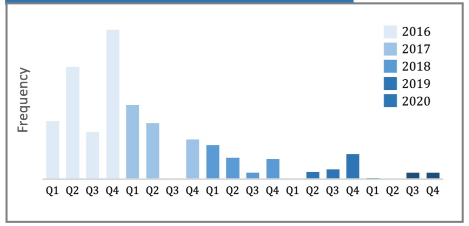
- The rapidly-evolving nature of news leads to the distributional difference between offline and online data, namely temporal shift
- Temporal shift causes significant performance degradation to the fake news detection model trained on offline data when predicting on online data



Introduction: Diverse topic-level temporal patterns

- The appearance of news events on the same topic presents diverse temporal patterns
- These temporal patterns indicate the different importance of news samples in the training set for detection in future quarters.

Topic 1: Child Trafficking



2016Q1 **Text:** Over 100 people have arrived around Funing County to steal children. More than 20 children have been lost in Funing County.

2017Q1 **Text:** Over 100 foreigners have come from Sanya and have arrived in Baoding City, Hebei Province. They specialize in children trafficking.

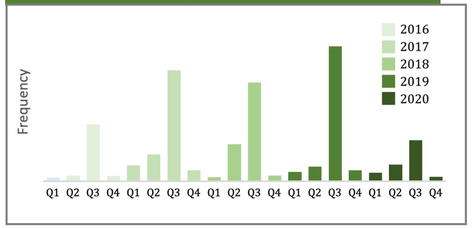
2018Q1 **Text:** Over 10,000 foreigners from Sanya and have arrived in Shulan, Jilin, Changchun, more than 2,000 children was lost.

Topic 1: Child Trafficking

Long-standing rumors, have been widely debunked

Decrease

Topic 2: College Entrance Exam



2019Q3 **Text:** Yang Guilan, whose admission letter was forgotten on the No. 818 bus, and now it is placed at the dispatching booth of Xinxi Station, please forward (from Suiyang County).

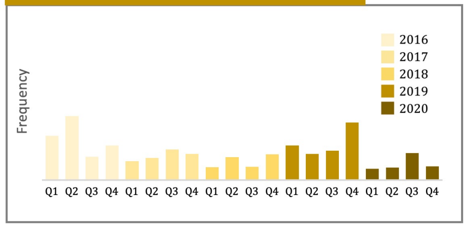
2020Q3 **Text:** Anyone in Lanzhou know Yang Guilan? Her admission ticket was lost on the No.104 bus, and is currently at the dispatching booth of Xinxi Station. (please forward if you see it)

Topic 2: College Entrance Exam

Experiencing periodic surges in a specific quarter annually

Periodicity

Topic 3: Falling Accident



2016Q2 **Text:** Netizens reported that this morning a girl fell from the apartment building of ...

2017Q4 **Text:** A girl fell from the 11th floor of a hotel because of a broken relationship.

2020Q1 **Text:** A tragedy occurred in Longgang! Yesterday, a man accidentally fell from ...

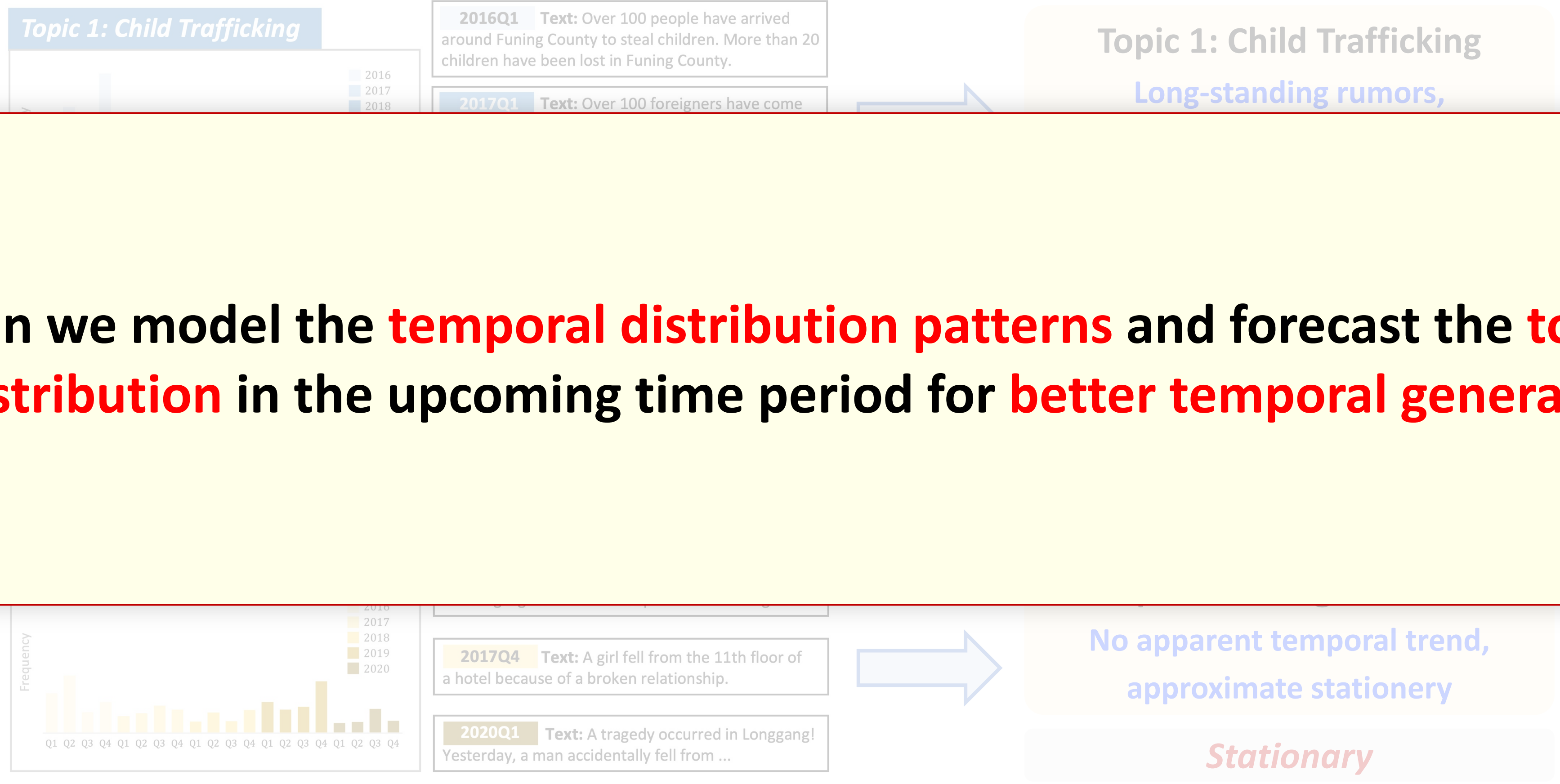
Topic 3: Falling Accident

No apparent temporal trend, approximate stationary

Stationary

Introduction: Diverse topic-level temporal patterns

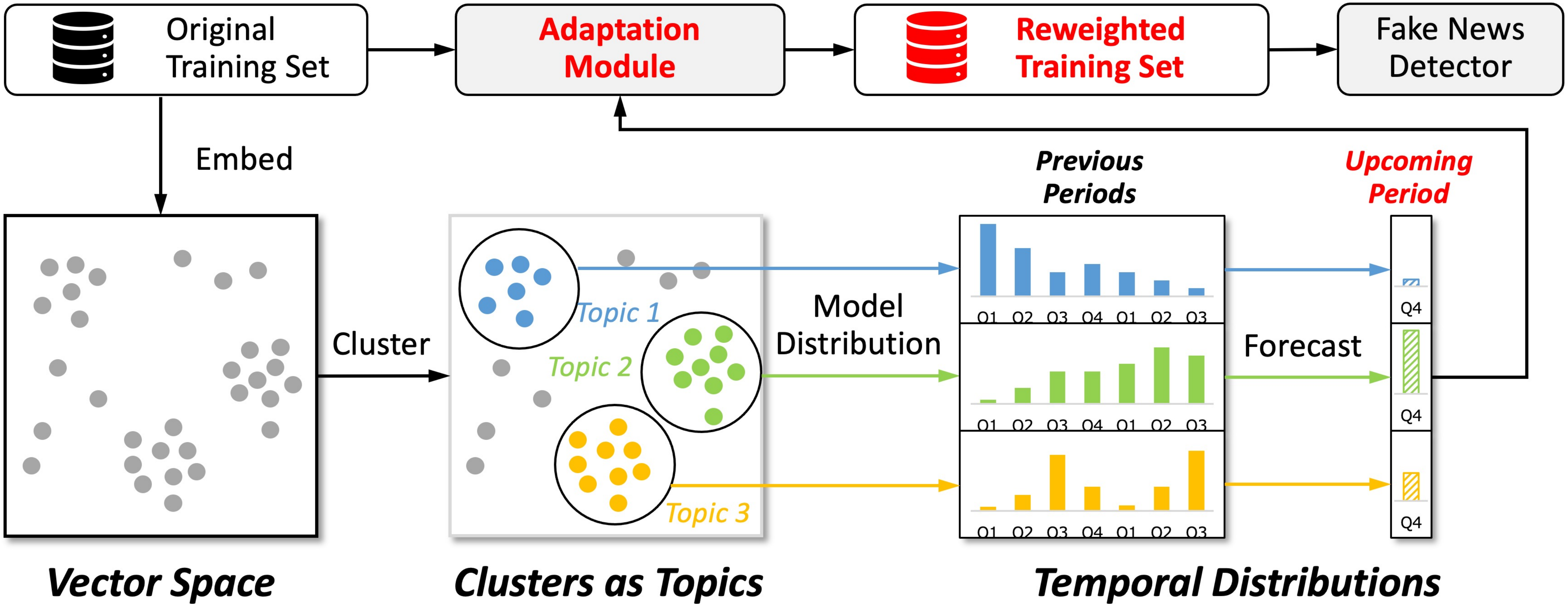
- The appearance of news events on the same topic presents diverse temporal patterns
- These temporal patterns indicate the different importance of news samples in the training set for detection in future quarters.



Can we model the temporal distribution patterns and forecast the topic-wise distribution in the upcoming time period for better temporal generalization?

Method: Forecasting Temporal Trends (FTT) Framework

- We propose a framework for **F**orecasting **T**emporal **T**rends (FTT) to guide the detector to fast adapt to future distribution.



Evaluation: Performance Comparison

2020	Metric	Baseline	EANN_T	Same Period Reweighting	Prev. Period Reweighting	Combined Reweighting	FTT (Ours)
Q1	macF1	0.8344	0.8334	0.8297	0.8355	0.8312	**0.8402**
	Accuracy	0.8348	0.8348	0.8301	0.8359	0.8315	**0.8409**
	F1_{fake}	0.8262	0.8181	0.8218	0.8274	0.8237	**0.8295**
	F1_{real}	0.8425	0.8487	0.8377	0.8435	0.8387	**0.8509**
Q2	macF1	0.8940	0.8932	0.8900	0.9004	0.8964	**0.9013**
	Accuracy	0.8942	0.8934	0.8902	0.9006	0.8966	**0.9014**
	F1_{fake}	0.8894	0.8887	0.8852	0.8953	0.8915	**0.8981**
	F1_{real}	0.8986	0.8978	0.8949	**0.9055**	0.9013	0.9046
Q3	macF1	0.8771	0.8699	0.8753	0.8734	0.8697	**0.8821**
	Accuracy	0.8776	0.8707	0.8759	0.8741	0.8707	**0.8827**
	F1_{fake}	0.8696	0.8593	0.8670	0.8640	0.8582	**0.8743**
	F1_{real}	0.8846	0.8805	0.8836	0.8829	0.8812	**0.8900**
Q4	macF1	0.8464	0.8646	0.8464	0.8429	0.8412	**0.8780**
	Accuracy	0.8476	0.8647	0.8476	0.8442	0.8425	**0.8784**
	F1_{fake}	0.8330	0.8602	0.8330	0.8286	0.8271	**0.8707**
	F1_{real}	0.8598	0.8690	0.8598	0.8571	0.8553	**0.8853**
Average	macF1	0.8630	0.8653	0.8604	0.8631	0.8596	**0.8754**
	Accuracy	0.8636	0.8659	0.8610	0.8637	0.8603	**0.8759**
	F1_{fake}	0.8546	0.8566	0.8518	0.8538	0.8501	**0.8682**
	F1_{real}	0.8714	0.8740	0.8690	0.8723	0.8691	**0.8827**

- **Observations:**
  - FTT outperforms the comparing methods across all quarters
  - The average improvement of F1 fake is larger than that of F1 real

Evaluation: Performance Comparison

Break down the performance on the testing set according to the existence of their topics.

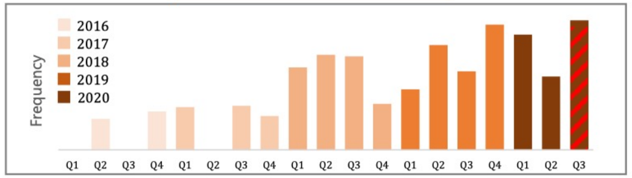
Subset of the test set	Metric	Baseline	FTT (Ours)
Existing Topics	macF1	0.8425	**0.8658**
	Accuracy	0.8589	**0.8805**
	$F1_{fake}$	0.7997	**0.8293**
	$F1_{real}$	0.8854	**0.9023**
New Topics	macF1	0.8728	**0.8846**
	Accuracy	0.8729	**0.8846**
	$F1_{fake}$	0.8730	**0.8849**
	$F1_{real}$	0.8727	**0.8843**

- **Observation:**
  - FTT achieves performance improvements on both the Existing Topics and the New Topics subsets

Evaluation: Case Study

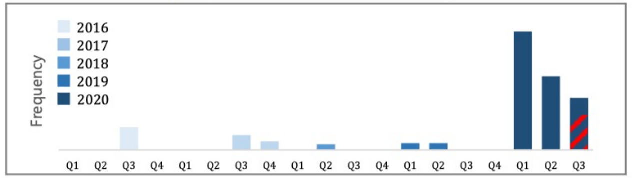
- **Objective:**
  - Evaluation of FTT Framework
- **How to do**:
  - Select topics assigned positive weights according to the forecasted results of the frequencies.
- **Results:**
  - The detector flips its previously incorrect predictions in cases after training on the reweighted set

Topic 1: Big Tech



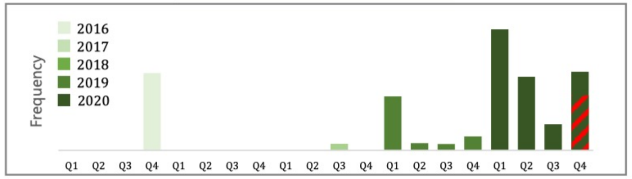
Text: Google Maps is suspected of blocking SIM cards of domestic operators. Recently, some netizens broke the news that Google Maps began to detect the SIM card of domestic operators to stop the service.	Baseline	*Real (0.49)*
	Ours	***Fake (0.58)***
	Ground Truth	*Fake*

Topic 2: Infectious Diseases



Text: Barcelona officially confirmed that Umtiti's COVID-19 test result was positive. As an asymptomatic patient he has begun home isolation. Umtiti did not follow the team to the UCL due to injury.	Baseline	*Fake (0.67)*
	Ours	***Real (0.37)***
	Ground Truth	*Real*

Topic 3: Medication Safety



Text: The Second Xiangya Hospital and Huaxi Hospital all issued documents refusing the admission of traditional Chinese medicine because of the frequent occurrence of quality problems.	Baseline	*Real (0.34)*
	Ours	***Fake (0.55)***
	Ground Truth	*Fake*

Conclusion

- **Contributions**

  - **Problem:** To the best of our knowledge, we are the ***first*** to incorporate the characteristics of topic-level temporal patterns for fake news detection.

  - **Method:** We propose a framework for Forecasting Temporal Trends (FTT) to tackle temporal generalization issue in fake news detection.

  - **Industrial Value:** We experimentally show that our FTT overall outperforms five compared methods while maintaining good compatibility with any neural network-based fake news detector.

Project

THANKS.

Our code is available at https://github.com/ICTMCG/FTT-ACL23.

Feel free to contact Beizhe Hu (hubeizhe21s@ict.ac.cn) for any questions!

Media Synthesis & Forensics Lab