EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection

Yaqing Wang[1], Fenglong Ma[1], Zhiwei Jin[2], Ye Yuan[3], Guangxu Xun[1], Kishlay Jha[1], Lu Su[1], Jing Gao[1]
[1] SUNY at Buffalo, [2]University of Chinese Academy of Sciences, [3]Beijing University of Technology.
University at Buffalo
The State University of New York

Motivation

What is Fake News ?

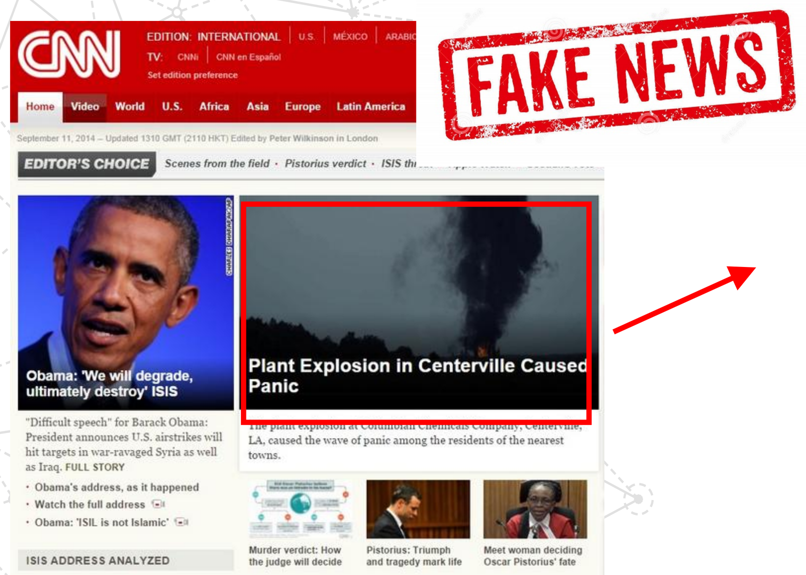
- **"Fake news** is a deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media. This false information is mainly distributed by social media."
---Wikipedia

Global concern brought by Fake News

- Within the final three months of the 2016 U.S. presidential election, the fake news generated to favor either of the two nominees was shared by more than 37 million times on Facebook.

Challenges of Fake News Detection

- Fake news is often generated on newly emerged (time- critical) events and is hard to verify.
- Fake news takes advantage of multimedia contents to mislead readers and gets rapid dissemination.



The Columbian Chemicals plant explosion was reported to have involved "dozens of fake accounts that posted hundreds of tweets for hours, targeting a list of figures precisely chosen to generate maximum attention. "

Proposed Solution

- Extract common multi-modal features (i.e. remove event-specific features) across different events, because the common features are also shared by and are effective on newly emerged events.

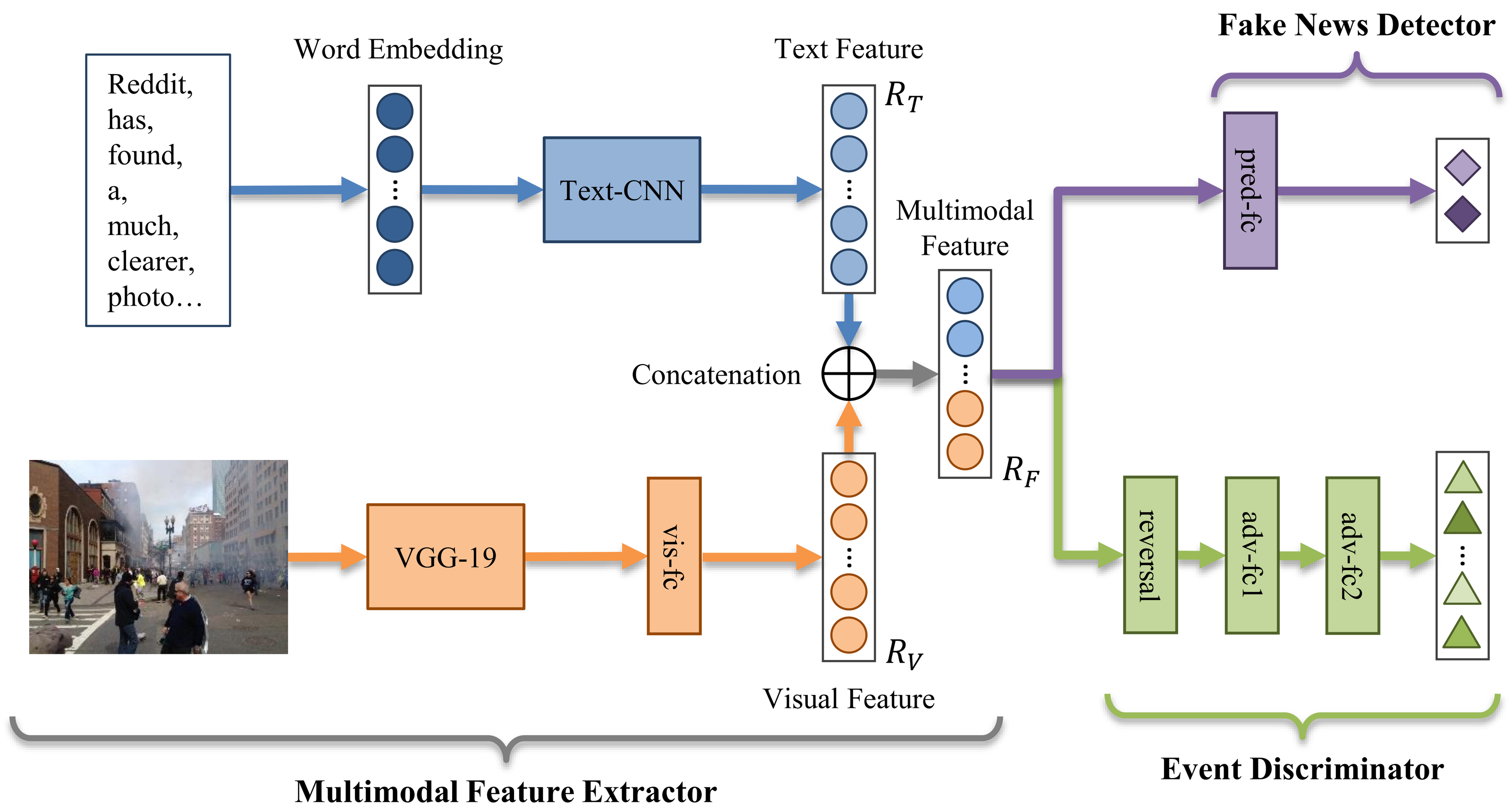
How to remove event-specific features?

- Employ Adversarial Mechanism to find event-specific features and remove them.

EANN Model

Model Overview

- The multi-modal feature extractor cooperates with the fake news detector to identify fake news.
- **Adversarial Mechanism**: The multi-modal feature extractor fools the event discriminator to learn the common features across different events.



- **The fake news detector θ_d** aims to cooperate with the multi-modal feature extractor θ_f to minimize the fake news detection loss L_d.

$$(\hat{\theta}_f, \hat{\theta}_d) = arg \min_{\theta_f, \theta_d} L_d(\theta_f, \theta_d)$$

- **The event discriminator θ_e** aims to correctly classify the post into one of the events (i.e. minizine the event discrimination loss L_e) based on multi-modal features.

$$\hat{\theta}_e = arg \min_{\theta_e} L_e(\theta_f, \theta_e)$$

- **The multi-modal feature extractor θ_f** aims to achieve two goals:
  1. **Detect fake news:** cooperate with the fake news detector θ_d to minimize the fake news detection loss L_d.
  2. **Remove event-specific features:** fool the event detector θ_e to maximize the event discrimination loss L_e.

$$\hat{\theta}_f = arg \min_{\theta_f} L_d(\theta_f, \theta_d) - \boxed{\lambda} L_e(\theta_f, \theta_e)$$

The λ controls the trade-off between losses L_d and L_e

Experiments

Datasets

- **Twitter** and **Weibo** are both popular multimedia social media websites. The datasets collected from them contain the text posts and the corresponding attached images.

	Twitter	Weibo
# of fake News	7898	4749
# of real News	6026	4779
# of images	514	9528

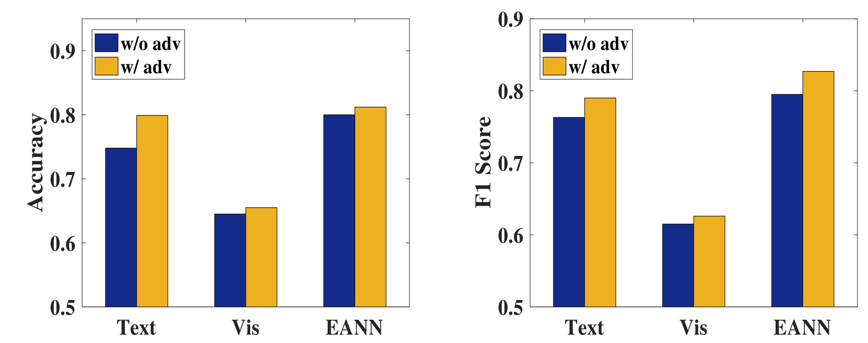
Performance Validation

- Compared with the state-of-the-art fake news detection models, EANN achieves the best performance on two datasets overall.

Dataset	Method	Accuracy	Precision	Recall	F_1
Twitter	Text	0.532	0.598	0.541	0.568
	Vis	0.596	0.695	0.518	0.593
	VQA	0.631	0.765	0.509	0.611
	NeuralTalk	0.610	0.728	0.504	0.595
	att-RNN	0.664	0.749	0.615	0.676
	EANN−	0.648	0.810	0.498	0.617
	EANN	**0.715**	**0.822**	**0.638**	**0.719**
Weibo	Text	0.763	0.827	0.683	0.748
	Vis	0.615	0.615	0.677	0.645
	VQA	0.773	0.780	0.782	0.781
	NeuralTalk	0.717	0.683	**0.843**	0.754
	att-RNN	0.779	0.778	0.799	0.789
	EANN−	0.795	0.806	0.795	0.800
	EANN	**0.827**	**0.847**	0.812	**0.829**

Importance of Adversarial Mechanism

- **Adversarial Mechanism** helps improve the performance of single-modal and multi-modal models respectively on both accuracy and F1 score by removing event-specific features.



The performance comparison for the models w/ and w/o event discriminator.

Importance of multi-modal features for fake news detection

- Fake news missed by single text modality model but detected by EANN.

(a) Five headed snake

(b) Photo: Lenticular clouds over Mount Fuji, Japan. #amazing #earth #clouds #mountains

- Fake news missed by single image modality model but detected by EANN.

(a) Want to help these unfortunates? New, Iphones, laptops, jewelry and designer clothing could aid them through this!

(b) Meet The Woman Who Has Given Birth To 14 Children From 14 Different Fathers!