

Multi-modal Identification of State-Sponsored Propaganda on Social Media

Xiaobo Guo, Soroush Vosoughi
{xiaobo.guo.gr, soroush}@dartmouth.edu

Minds, Machines and Society Group,
Department of Computer Science, Dartmouth
Twitter: MMS_Dartmouth

December 10, 2020

Table of Contents

1 Motivation

- Background
- Aim

2 Methodology

- Dataset
- Features
- Model

3 Results

- Performance
- Attention Analysis

Table of Contents

1 Motivation

- Background
- Aim

2 Methodology

- Dataset
- Features
- Model

3 Results

- Performance
- Attention Analysis

Characters

- Propaganda
 - Purpose
 - Veracity
 - Parties
- Social Media
 - Anonymity
 - Astroturfing

Examples

- Russian interference in the 2016 US presidential election.
- Spreading content about the Catalan Referendum.
- Amplifying messaging supportive of the Saudi government

Our Model

- Early Detection
 - Limit Influence
- Content-based
 - User Independent
- Veracity-agnostic
 - Fake news or not

Table of Contents

1 Motivation

- Background
- Aim

2 Methodology

- Dataset
- Features
- Model

3 Results

- Performance
- Attention Analysis

Dataset details

Organization	Data Type	# of tweets	Start date	End date
IRA	training	4,102	2015-04-01	2016-01-19
	validation	642	2016-01-20	2016-01-31
	continuous test	896	2016-02-01	2016-02-28
	delay test	216	2016-10-01	2017-06-30
Russian	training	5,194	2015-04-01	2015-12-14
	validation	742	2015-12-15	2015-12-31
	continuous test	1,534	2016-01-01	2016-02-28
	delay test	6,544	2016-10-01	2017-06-30
Iranian	training	5,452	2015-04-01	2015-12-14
	validation	300	2015-12-15	2015-12-31
	continuous test	1,620	2016-01-01	2016-02-28
	delay test	6,642	2016-10-01	2017-06-30

Table: Dataset details including organizations, number of tweets and time period.

Visual Features



Figure: Original picture



Figure: Structural picture

Textual Features

Text Type	Textual Content
Original	#Putin's 1st New Year's "achievement" in #Syria URL
Tag	TAG 1st New Year's "achievement" in TAG URL
Miss	1st New Year's "achievement" in URL
Structure	T W W W W W T U

Table: Example of four different modifications of textual content.

The structure of our model

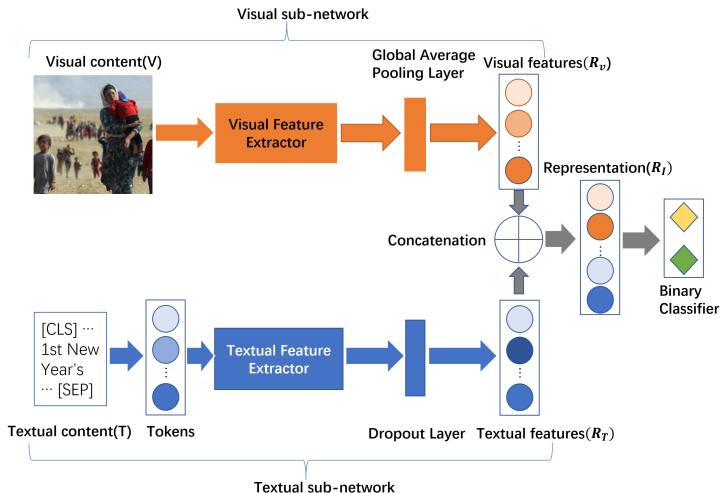


Figure: Structure of multi-model network

Table of Contents

1 Motivation

- Background
- Aim

2 Methodology

- Dataset
- Features
- Model

3 Results

- Performance
- Attention Analysis

F1 Score with different features

	C	D
Resnet-50	0.714	0.644
Inception-v3	0.706	0.639
VGG-19	0.681	0.497
Style	0.477	0.424
Content	0.684	0.519
Texture-Content	0.664	0.510
Image-Structure	0.604	0.446

Table: Visual features

	C	D
Original	0.854	0.643
Tag	0.803	0.649
Miss	0.788	0.614
Structure	0.715	0.548

Table: Textual features

*C means test on continuous test data.

*D means test on delay test data.

ROC and AUC with different models

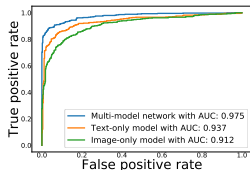


Figure: IRA (C)

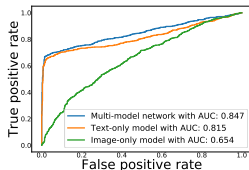


Figure: Russian (C)

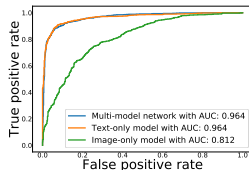


Figure: Iranian (C)

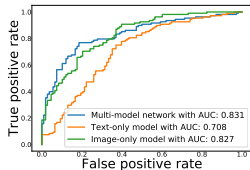


Figure: IRA (D)

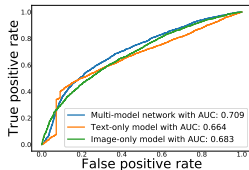


Figure: Russian (D)

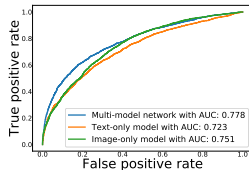


Figure: Iranian (D)

*C means test on continuous test data.

*D means test on delay test data.

Attention Analysis



Figure: False Positive

Figure: False Negative

Thank You!