Fake News Detection

Group members: Haoru Xiao, Hayden Chen, Jianqiu Bai, Weihan Song, Yao Wang, Zi Li

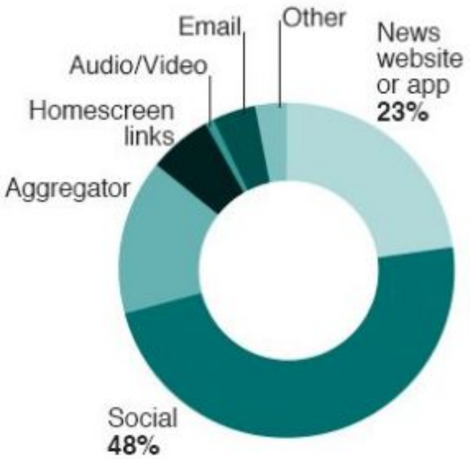
Contents

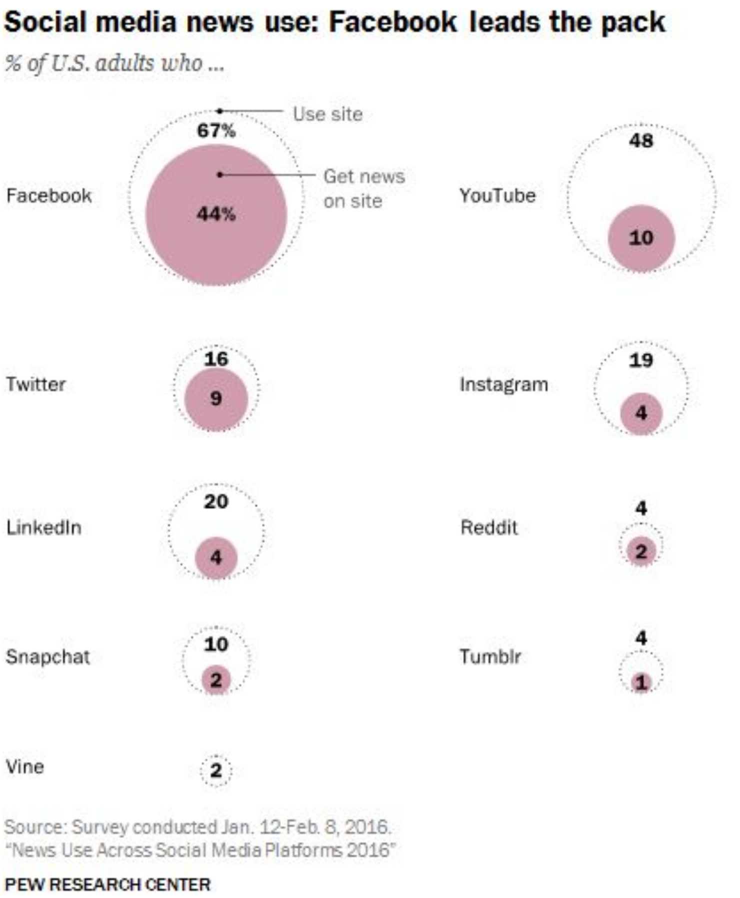
News consumption trend

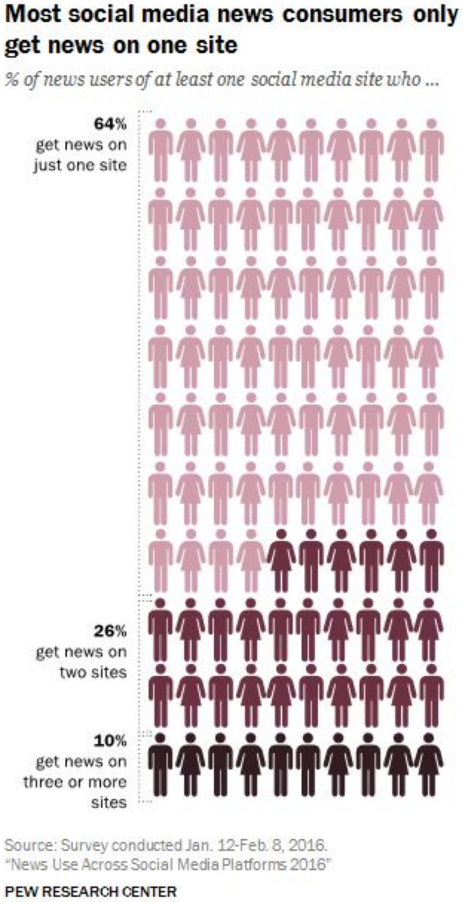
Compare to traditional news organizations, consuming news from social media is

- More timely.
- less expensive.
- Easier to share, comment and discuss with friends and other readers.

First news source when using a smartphone in US







Sources https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/
https://www.bbc.com/news/uk-36528256

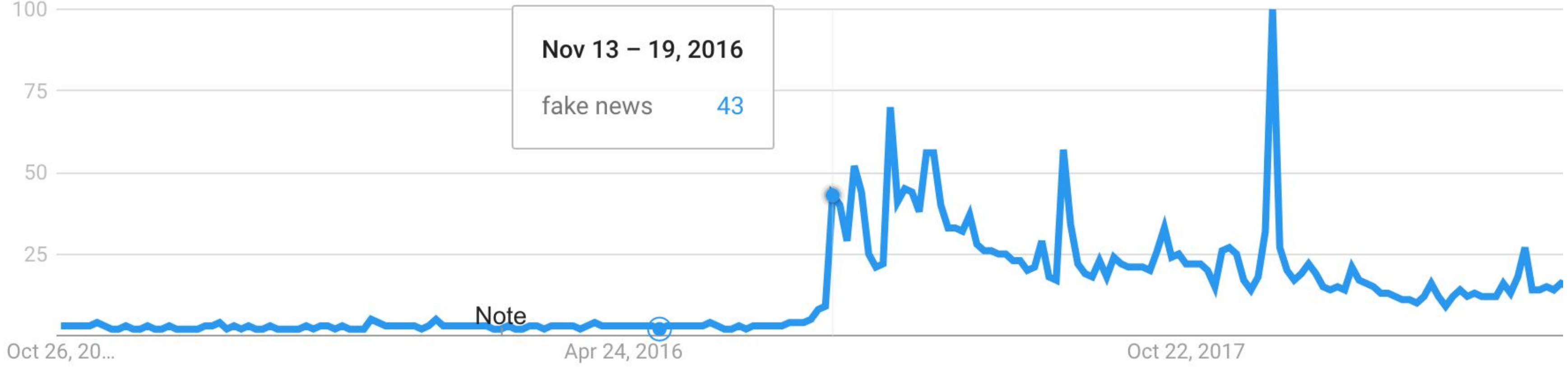
Post-truth

"relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief."

Word of the year 2016

Source https://languages.oup.com/word-of-the-year/word-of-the-year-2016

Google trend for the keyword “fake news”



Source: https://trends.google.com/trends/explore?date=today%205-y&geo=US&q=fake%20news

Top 5 Fake Election Stories by Facebook Engagement
(three months before election)

"Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement" *(960,000, Ending the Fed)*

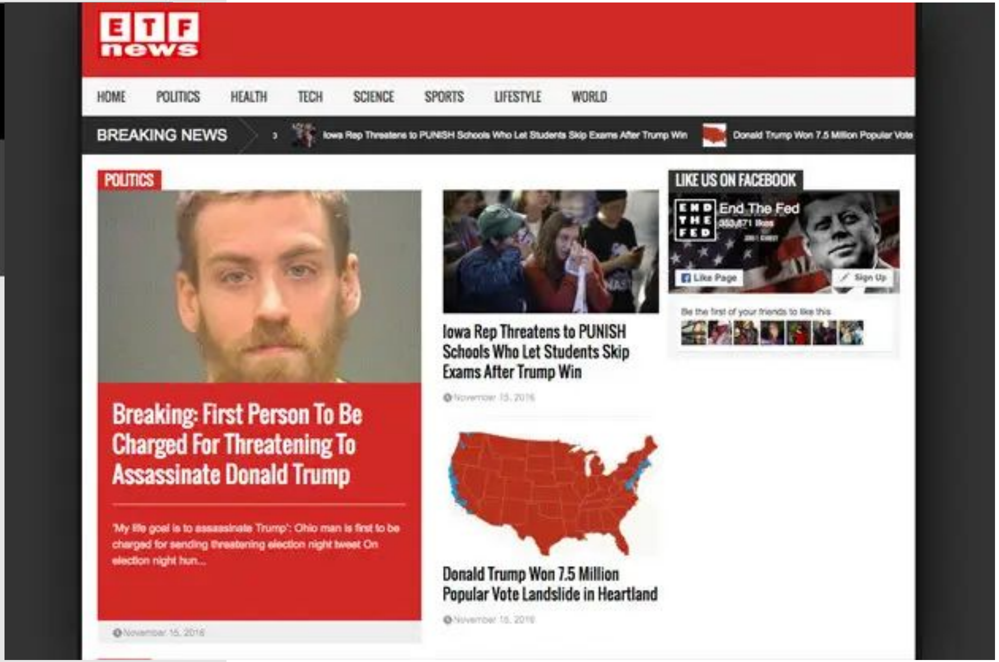
"WikiLeaks CONFIRMS Hillary Sold Weapons to ISIS... Then Drops Another BOMBSHELL! Breaking News" *(789,000, The Political Insider)*

"IT'S OVER: Hillary's ISIS Email Just Leaked & It's Worse Than Anyone Could Have Imagined" *(754,000, Ending the Fed)*

"Just Read the Law: Hillary Is Disqualified From Holding Any Federal Office" *(701,000, Ending the Fed)*

"FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide" *(567,000, Denver Guardian)*

ENGAGEMENT REFERS TO THE TOTAL NUMBER OF SHARES, REACTIONS, AND COMMENTS FOR A PIECE OF CONTENT ON FACEBOOK SOURCE: FACEBOOK DATA VIA BUZZSUMO



Source: https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

WikiLeaks CONFIRMS Hillary Sold Weapons to ISIS... Then Drops Another BOMBSHELL! Breaking News

Kosar
Featured Contributor

James — August 9, 2016 at 12:59 pm

The only thing that bothers me is that in exposing what Hillary has done too many people forget Obama is her boss and is as guilty as she is and they both should reside in Gitmo.

Lewis — August 9, 2016 at 9:23 am

Not surprising as she has been corrupt for the last 30 years

Source http://archive.is/Avh0w

Impact on the society

- Break the authenticity balance of the news ecosystem.
- Intentionally persuade consumers to accept biased or false beliefs.
- Trigger distrust and confuse people, impeding their abilities distinguish truth from fallacy.

Difficulty in detecting fake news

- The content diverse in topics, style and media platforms.
- Distort truth with diverse linguistic styles, like citing true evidence within the incorrect context to support false claim.
- Need auxiliary information such as knowledge base and user social engagements.
- Related to time-critical events, which existing knowledge bases are unable to verify properly.
- User social engagements data is big, incomplete, unstructured and noisy.

What is fake news?

- ***Authenticity***
  - Information that can be verified as false
- ***Intent***
  - Dishonest intention to mislead consumers

Only when both of these attributes are met, news is considered as fake news.

Examples of NOT fake news

- Satire news with proper context, which has **no intent** to mislead or deceive consumers and is unlikely to be mis-perceived as factual
- Rumors that did not originate from **news events**
- Conspiracy theories, which are **difficult to verify** as true or false
- Misinformation that is created **unintentionally**
- Hoaxes that are only motivated by fun or to scam targeted individuals

Two-player game

Publisher

- Short-term utility: maximize the number of consumers reached.
- Long-term utility: reputation in terms of news authenticity.

Consumer

- Information utility: get true and unbiased information (extra investment cost).
- Psychology utility: satisfy prior opinions and social needs.

Both players try to maximize their overall utilities.

Fake news happens when Short-term utility dominates publishers' overall utility and Psychology utility dominates consumers'

Fake News Characterization

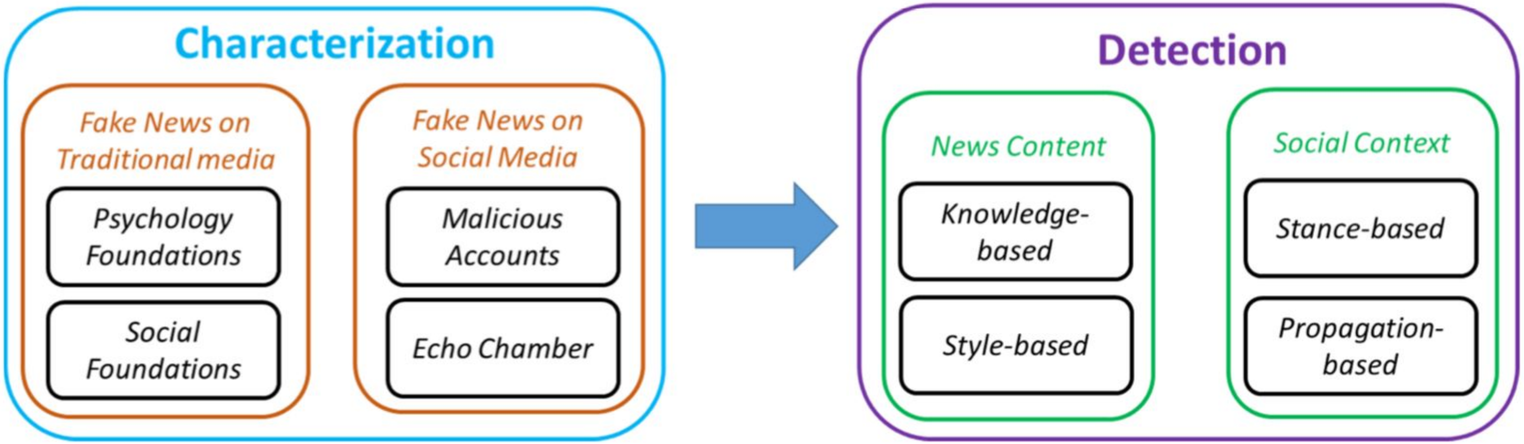


Figure 1: Fake news on social media: from characterization to detection.

Fake News on Traditional News Media

Psychological Foundations of Fake News

individual vulnerabilities

- Naive Realism: people only believe what they think is right, others' opinions are irrational and biased
- Confirmation Bias: people prefer to seek and accept information which is same with their own views

Fake News on Traditional News Media

Social Foundations of the Fake News Ecosystem

People may choose “socially safe” options following the norms established in the community.

Fake News on Social Media

Malicious Accounts on Social Media

low cost -> malicious user accounts (social bot, cyborg users, and trolls)

Fake News on Social Media

Echo Chamber Effect

psychological challenges are enhanced due to the way news appearing on social media (eg. like-minded people)

- social credibility
- frequency heuristic

FAKE NEWS DETECTION

1 Problem Definition

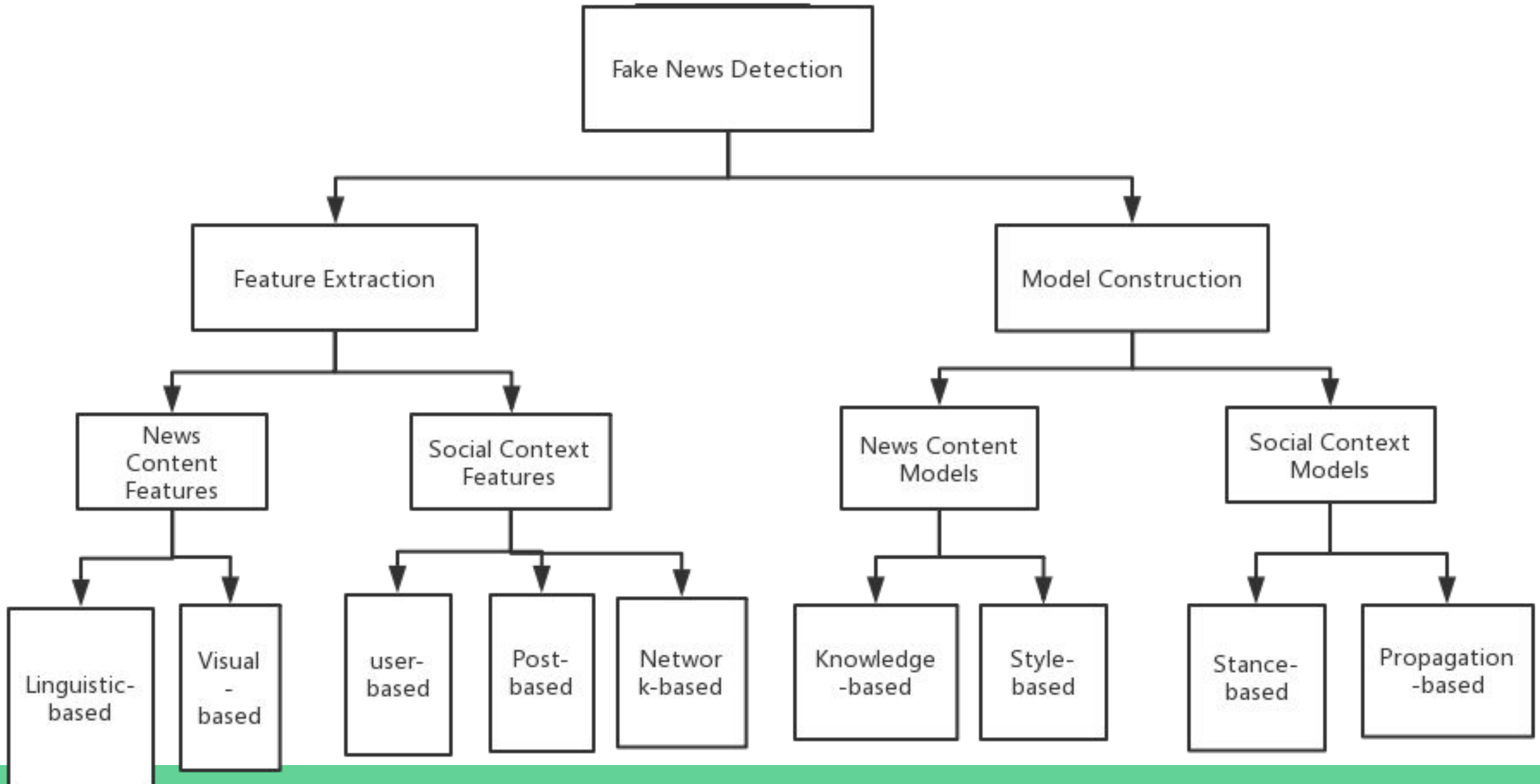
$$\mathcal{F}(a) = \begin{cases} 1, & \text{if } a \text{ is a piece of fake news,} \\ 0, & \text{otherwise.} \end{cases}$$

F: the prediction function

a: a News Article

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from http://arxiv.org/abs/1708.01967

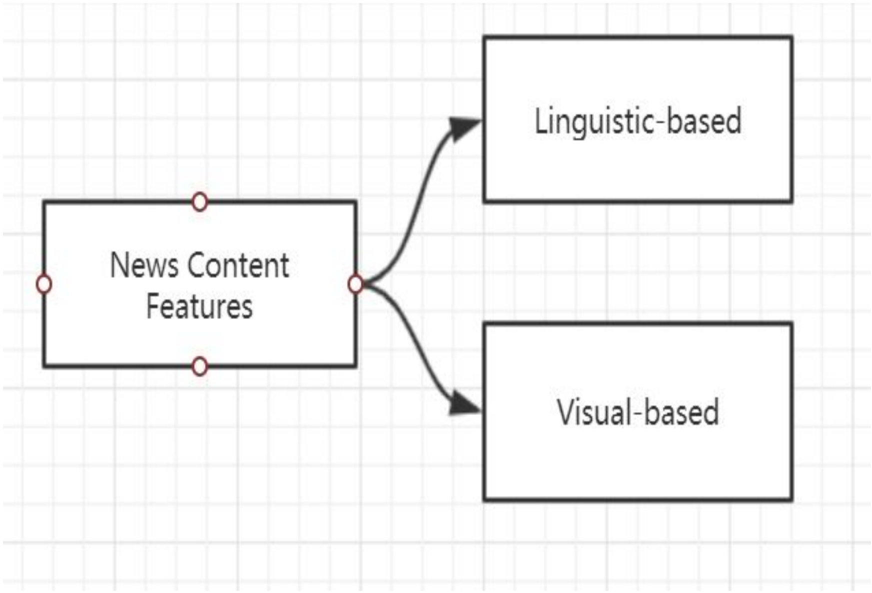
a data mining framework for fake news detection



2 Feature Extraction

2.1 News content features

- source
- headline
- body text
- image/video



Linguistic-based

Fake news are created for

some financial or political

purpose, they always have

inflammatory language

that can attract you

as soon as possible.

震惊！99.99%的人都不知道的死法！

震惊！美国总统看到后都惊呆了！

震惊！男人看了会沉默，女人看了会流泪！不转不是中国人！

震惊！父亲居然当着女儿的面做这种动作！

震惊！女子在医院对护士这样说话！

震惊！此老人竟然凭借此方法不老！

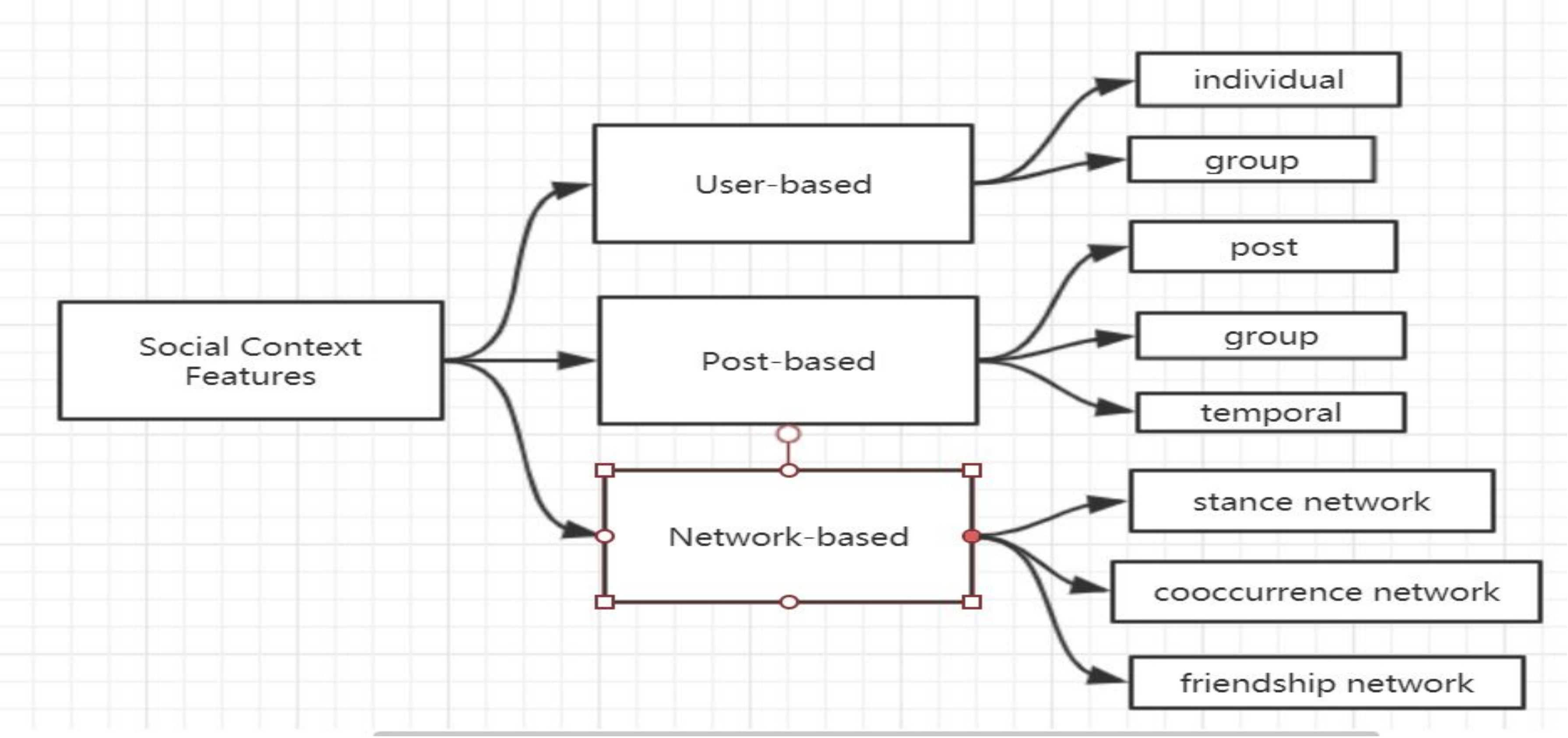
Reference link: http://www.ittime.com.cn/news/news_18857.shtml

Visual-based

Using fake images to provoke anger or other emotional response.

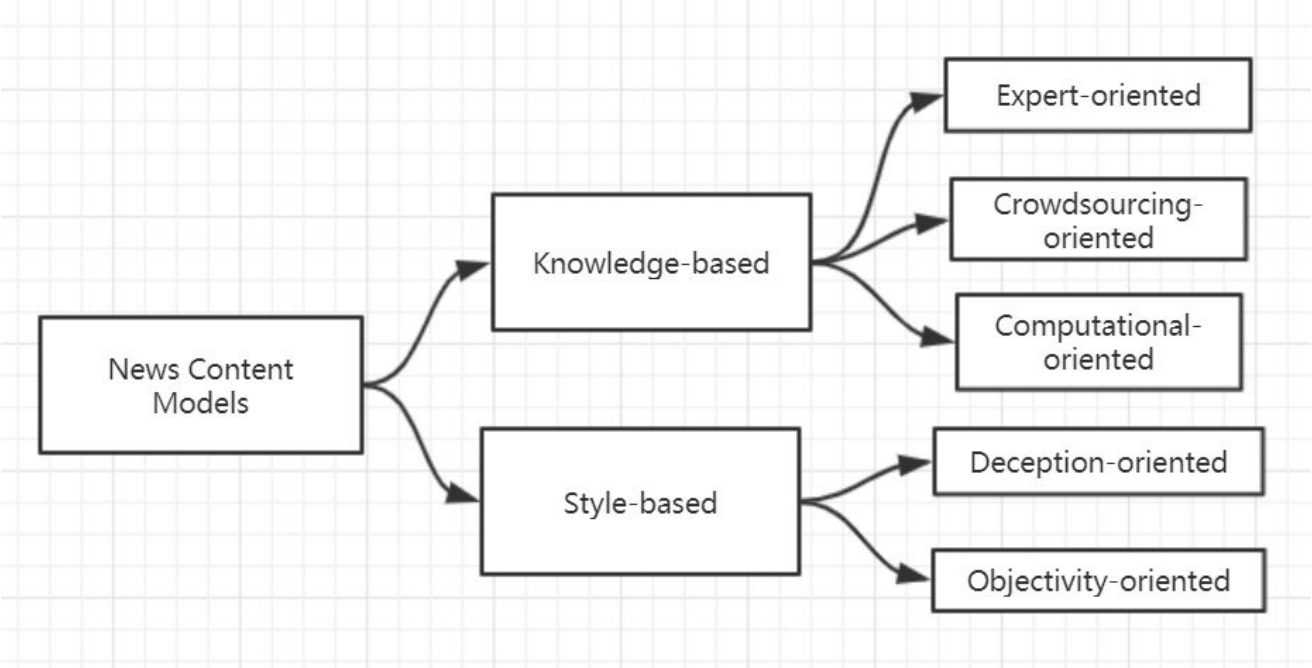
Reference link:
https://cn.gijn.org/2018/03/08/%E4%BA%8B%E5%AE%9E%E6%A0%B8%E6%9F%A5%E9%AB%98%E6%89%8B%E4%B8%89%E6%AC%BE%E5%BC%BA%E5%8A%9B%E5%B7%A5%E5%85%B7-%E8%BE%A8%E5%88%AB%E7%BD%91%E7%BB%9C%E5%9B%BE%E7%89%87%E7%9C%9F%E5%81%87/

2.2 Social Context Features

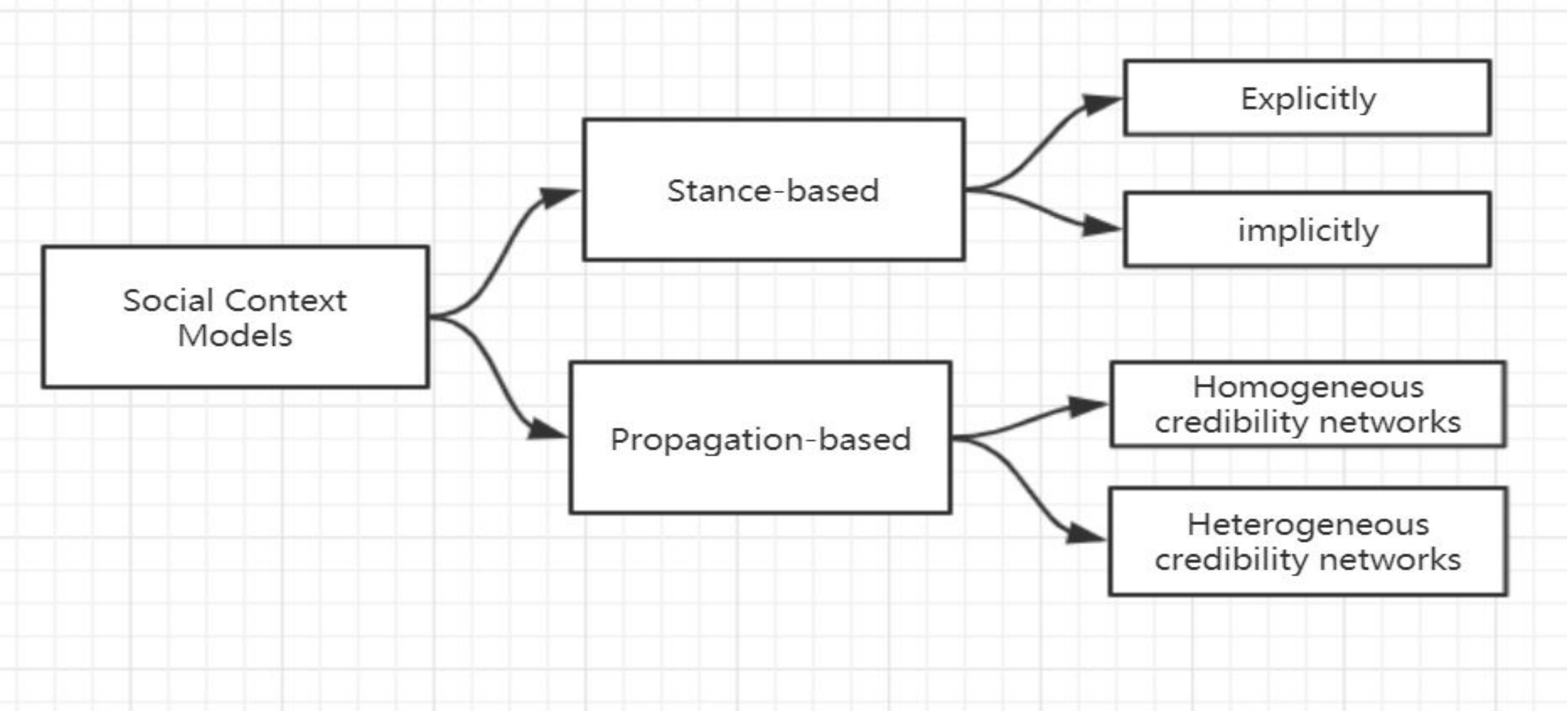


3 Model Construction

3.1 News Content Model



3.2 Social Context Models



Datasets

Dataset \ Features	News Content		Social Context		
	Linguistic	Visual	User	Post	Network
BuzzFeedNews	✓				
LIAR	✓				
BS Detector	✓				
CREDBANK	✓		✓	✓	✓

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://www.kdd.org/exploration_files/19-1-Article2.pdf.

Datasets

Dataset \ Features	News Content		Social Context		
	Linguistic	Visual	User	Post	Network
BuzzFeedNews	✓				
LIAR	✓				
BS Detector	✓				
CREDBANK	✓		✓	✓	✓

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://www.kdd.org/exploration_files/19-1-Article2.pdf.

Datasets

BuzzFeedNews

- Source : Facebook
- Author: 9 news agencies
- Time: Sep. 19 - 23 and Sep. 26 - 27
- Content: 2016 U.S. Election
- Veracity: Fact-checked claim-by-claim by 5 BuzzFeed journalists

Source: https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data.

Datasets

BuzzFeedNews

- Source : Facebook
- Author: 9 news agencies
- Time: Sep.19 - 23 and Sep. 26 - 27
- Content: 2016 U.S. Election
- Veracity: Fact-checked claim-by-claim by 5 BuzzFeed journalists

Source: https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data.

Datasets

BuzzFeedNews

	A	B	C	D	E	F	G	H	I	J	K	L
1	account_id	post_id	Category	Page	Post URL	Date Publi	Post Type	Rating	Debate	share_	reaction	comment_
2	1.841E+14	1.03506E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	video	no factual content			146	15
3	1.841E+14	1.03527E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		1	33	34
4	1.841E+14	1.03531E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		34	63	27
5	1.841E+14	1.03532E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		35	170	86
6	1.841E+14	1.03535E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	video	mostly true		568	3188	2815
7	1.841E+14	1.03537E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		23	28	21
8	1.841E+14	1.03541E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	video	mostly true		46	409	105
9	1.841E+14	1.03543E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		7	62	64
10	1.841E+14	1.03545E+15	mainstream	ABC News Politics	https://www.facebook	9/19/2016	link	mostly true		7	39	6

Source: https://github.com/BuzzFeedNews/2016-10-facebook-fact-check/tree/master/data.

Datasets

LIAR

- Source : PolitiFact fact-checking website API
- Content: 12836 human-labeled short statements
- Veracity: fine-grained multiple classes

  (pants-fire, false, barely-true, mostly true, true)

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://www.kdd.org/exploration_files/19-1-Article2.pdf.

Datasets

LIAR

- Source : PolitiFact fact-checking website API
- Content: 12836 human-labeled short statements
- Veracity: fine-grained multiple classes

  (pants-fire, false, barely-true, mostly true, true)

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://www.kdd.org/exploration_files/19-1-Article2.pdf.

Datasets

BS Detector

- Source : A web browser extension
- Veracity: Searches all links for references to unreliable sources

Source: https://www.kaggle.com/mrisdal/fake-news#fake.csv.

Datasets

BS Detector

	A	C	D	E	F	G	H	I	J	L	M	N	O	P	Q	R	S	T
1	uuid	author	published	title	text	language	crawled	site_url	coui	thread_title	spam	main_img	replies	participan	likes	comment	shares	type
2	6a175f46b	Barracuda	2016-10-26	Muslims E	Print	english	2016-10-2	100percer	US	Muslims BUS	0	http://bb4	0	1	0	0	0	bias
3	2bdc29d1	reasoning	2016-10-29	Re: Why D	Why Did	english	2016-10-2	100percer	US	Re: Why Did	0	http://bb4	0	1	0	0	0	bias
4	c70e149fd	Barracuda	2016-10-31	BREAKING	Red State	english	2016-10-3	100percer	US	BREAKING: V	0	http://bb4	0	1	0	0	0	bias
5	7cf7c1573	Fed Up	2016-11-01	PIN DROP	Email Kayl	english	2016-11-0	100percer	US	PIN DROP SF	0.07	http://10C	0	0	0	0	0	bias
6	0206b5471	Fed Up	2016-11-01	FANTASTI	Email	english	2016-11-0	100percer	US	FANTASTIC!	0.87	http://10C	0	0	0	0	0	bias
7	8f30f5ea1	Barracuda	2016-11-02	Hillary Go	Print	english	2016-11-0	100percer	US	Hillary Goes	0	http://bb4	0	1	0	0	0	bias
8	d3cc0fe38	Fed Up	2016-11-04	BREAKING	BREAKIN	english	2016-11-0	100percer	US	BREAKING! N	0.7	http://10C	0	0	0	0	0	bias
9	b4bbf8b5	Fed Up	2016-11-05	WOW! WH	BREAKIN	english	2016-11-0	100percer	US	WOW! WHIS	0.19	http://10C	0	0	0	0	0	bias
10	a19aabaa	Fed Up	2016-11-06	BREAKING		english	2016-11-0	100percer	US	BREAKING: C	0.14	http://10C	0	0	0	0	0	bias

Source: https://www.kaggle.com/mrisdal/fake-news#fake.csv.

Datasets

CREADBANK

- Source : Approximately 60 million tweets
- Time: Cover 96 days starting from Oct. 25
- Content: Related to over 1000 news content
- Veracity: Credit by 30 annotators from Amazon Mechanical Turk

http://compsocial.github.io/CREDBANK-data/.

Datasets

CREADBANK

- Source : Approximately 60 million tweets
- Time: Cover 96 days starting from Oct. 25
- Content: Related to over 1000 news event
- Veracity: Credit by 30 annotators from Amazon Mechanical Turk for each event

http://compsocial.github.io/CREDBANK-data/.

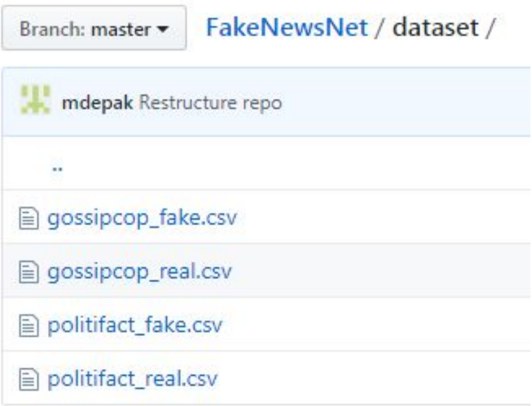
Datasets

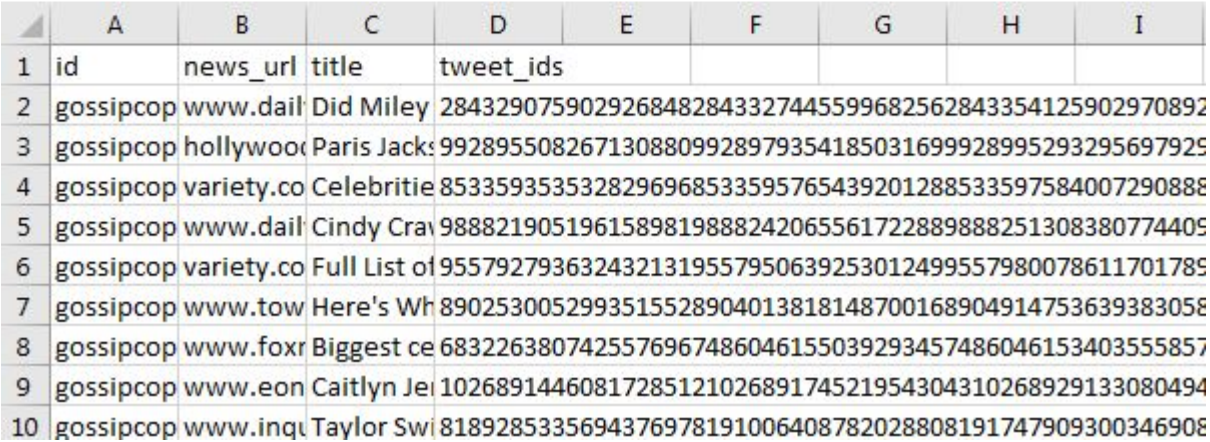
Features / Dataset	News Content		Social Context		
	Linguistic	Visual	User	Post	Network
BuzzFeedNews	✓				
LIAR	✓				
BS Detector	✓				
CREDBANK	✓		✓	✓	✓

Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from https://www.kdd.org/exploration_files/19-1-Article2.pdf.

Datasets

BuzzFeedNews20

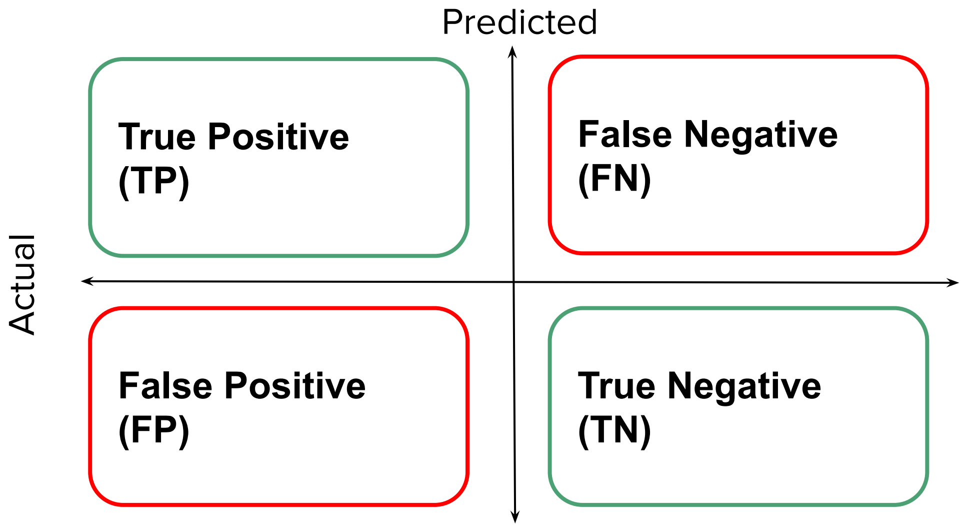




	A	B	C	D	E	F	G	H	I
1	id	news_url	title	tweet_ids					
2	gossipcop	www.dail	Did Miley	284329075902926848284332744559968256284335412590297089					
3	gossipcop	hollywooc	Paris Jacks	992895508267130880992897935418503169992899529329569792					
4	gossipcop	variety.co	Celebritie	853359353532829696853359576543920128853359758400729088					
5	gossipcop	www.dail	Cindy Crav	988821905196158981988824206556172288988825130838077440					
6	gossipcop	variety.co	Full List of	955792793632432131955795063925301249955798007861170178					
7	gossipcop	www.tow	Here's Wh	890253005299351552890401381814870016890491475363938305					
8	gossipcop	www.foxr	Biggest ce	683226380742557696748604615503929345748604615340355585					
9	gossipcop	www.eon	Caitlyn Jei	1026891446081728512102689174521954304310268929133080494					
10	gossipcop	www.inqu	Taylor Swi	818928533569437697819100640878202880819174790930034690					

https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset

Evaluation Metrics

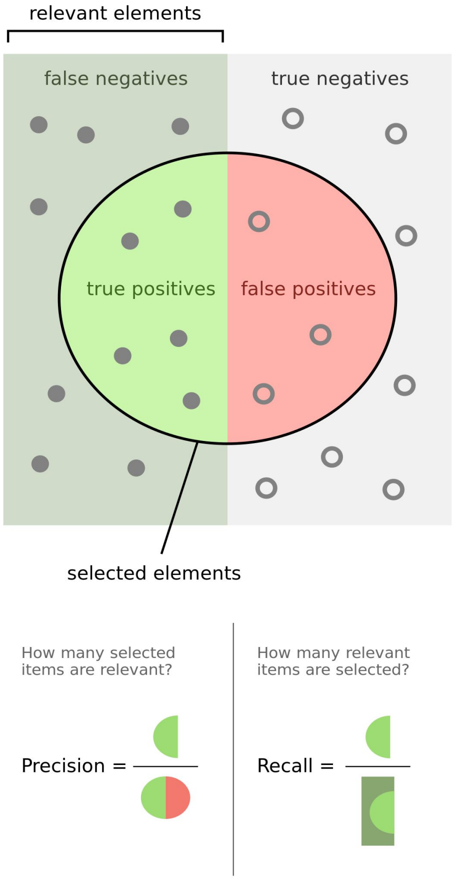


$$Precision = \frac{|TP|}{|TP| + |FP|}$$

$$Recall = \frac{|TP|}{|TP| + |FN|}$$

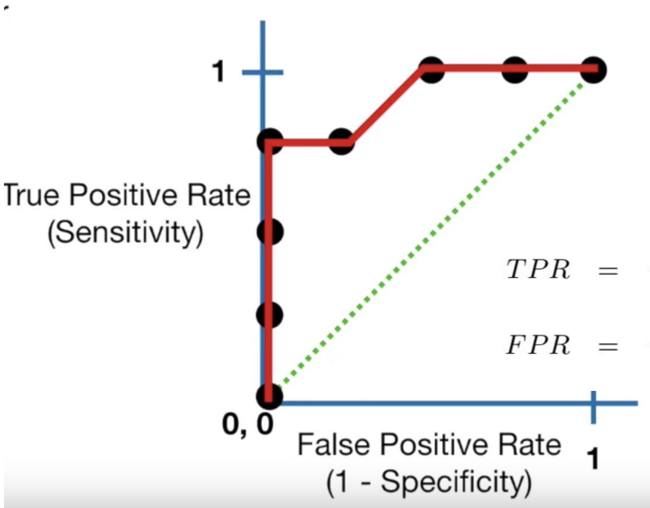
$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$



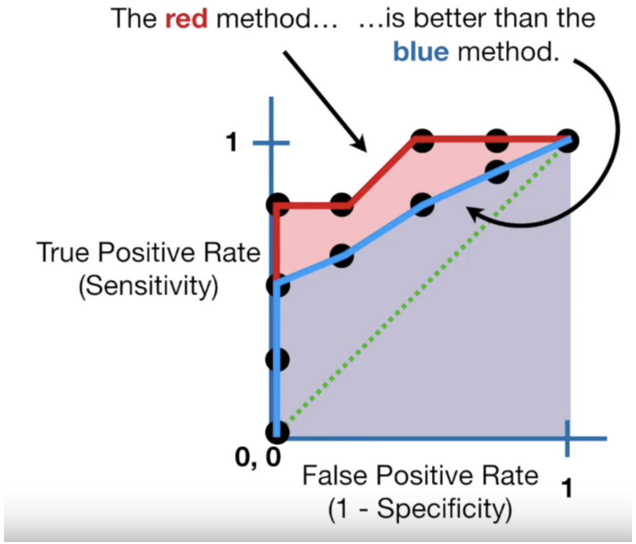
Evaluation Metrics

Receiver Operating Characteristics (ROC) curve: graphical tools for evaluating classifiers



$$TPR = \frac{|TP|}{|TP| + |FN|}$$

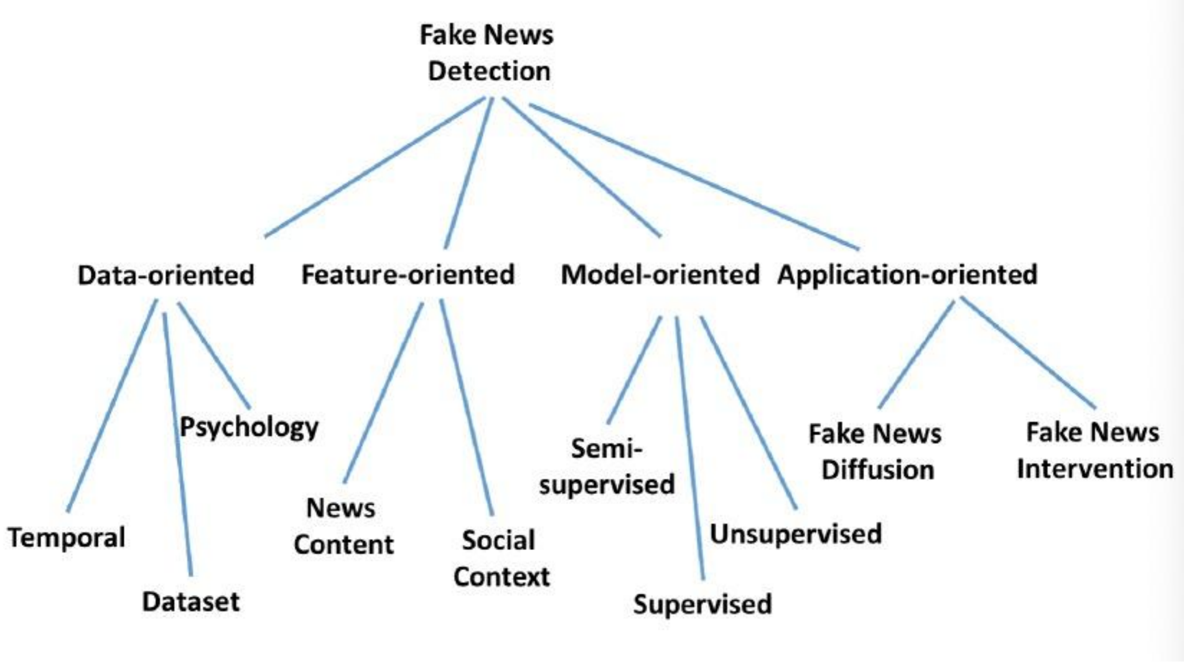
$$FPR = \frac{|FP|}{|FP| + |TN|}$$



ROC curve: makes it easier to identify the best threshold for making decision

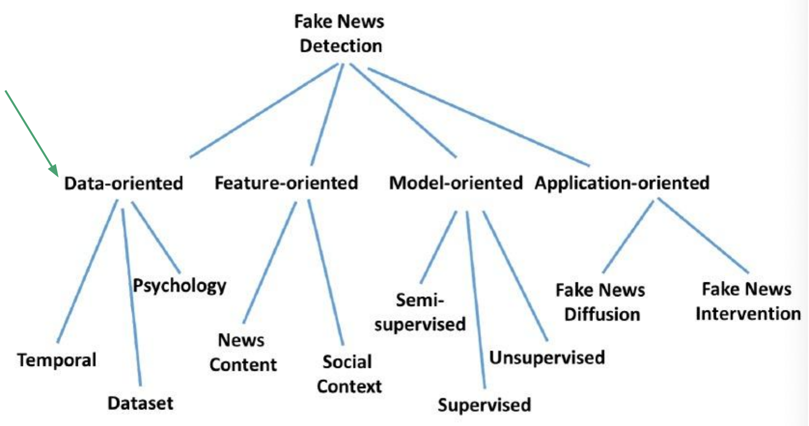
Area Under the Curve (AUC) value: helps decide which categorization is better

Open Issues and Future Research



Source: Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. Retrieved from http://arxiv.org/abs/1708.01967

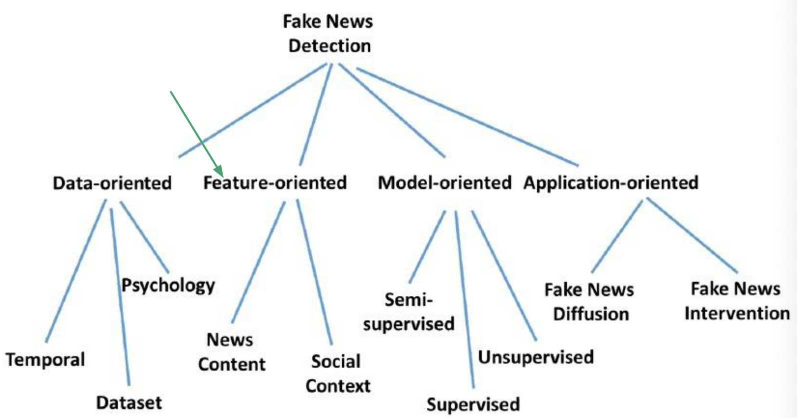
Data-oriented



Data Characteristics:

- Dataset: Large-scale fake news benchmark
- Temporal: Early fake news detection.
- Psychological: Quantitative studies, Intention studies

Feature-oriented

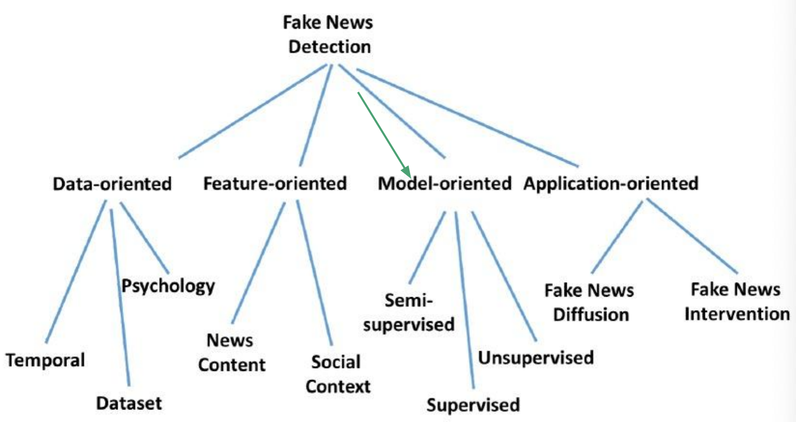


Effective features on fake news:

- News Content:
    - Linguistic-based techniques (text)
    - Visual-based techniques (image and video)

- Social Context:
    - User-based (user specific features)
    - Postbased (Convolutional Neural Networks)
    - Network-based (Different relationships among users, network representations)

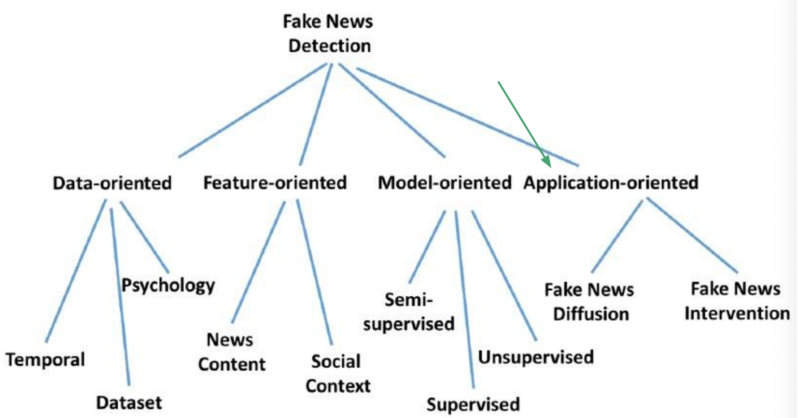
Model-oriented



Effective and practical models:

- Supervised Learning (mostly used today):
  - Naive Bayes, decision tree, logistic regression, KNN, SVM, etc.
- Promising research directions:
  - Aggregation methods
  - Probabilistic methods
- Semi-supervised or unsupervised learning method

Application-oriented



Research beyond fake news detection:

- Fake news diffusion
  - Social dimensions (Heterogeneity and weak dependency of social connections)
  - Life cycle (Time phase where people's attention and reactions switch to different stages)
  - Spreader identification (Identifying the key spreaders as clarifier or persuaders)
- Fake news intervention
  - Proactive intervention
  - Reactive intervention

Thank You For Listening!