D-FEND: A Diffusion-Based Fake News Detection Framework for News Articles Related to COVID-19

ACM SAC, 2022

Soeun Han,* Yunyong Ko,* Yushim Kim,† Seong Soo Oh,* Heejin Park,* Sang-Wook Kim*

* Hanyang University, Seoul, Republic of Korea

† Arizona State University, Arizona, United States

Hanyang University

Soeun Han

April 25-29, 2022

Fake News

- A news article created intentionally with false information
- Social confusion caused by COVID-19 related fake news

Hundreds die in Iran over false belief drinking methanol cures coronavirus

Posted Tue 28 Apr 2020 at 3:14pm

Iran is among the countries hit hardest by coronavirus in the Middle East. (AP: Vahid Salemi)

'Hundreds dead' because of Covid-19 misinformation

By Alistair Coleman
BBC Monitoring

12 August 2020

Figure 1. Social confusion caused by fake news articles.

Fake News Detection

Hanyang University

- **Content-based detection**
  - Using the difference between linguistic characteristics in the content of true news and fake news
  - Examples: semantic, writing style, syntactic, frequency of words, …
  - Limitations
    - Easy to manipulate by publishers
    - Easy to imitate true news very closely
    - Dependent on the language with which the article is written

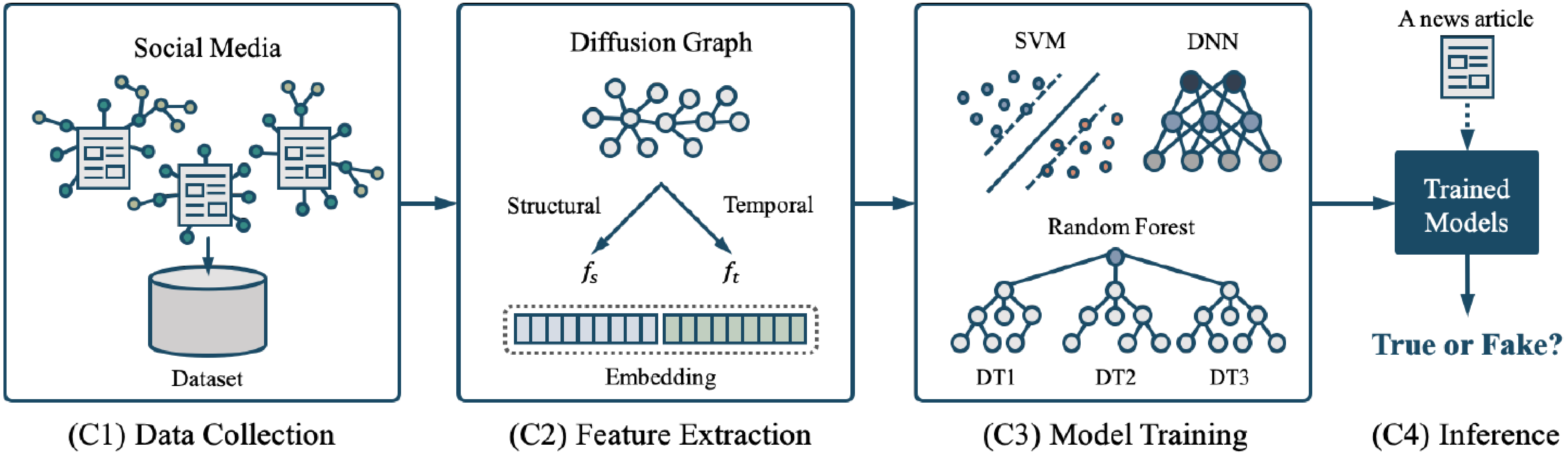
Fake News Detection

- **Social context-based detection**
  - Using the information from users who consume news on social media and various user engagement information
  - Examples: user profile, user relationship network, and user behavior information (like, retweet, share), …
  - Diffusion-based detection
    - To detect fake news by analyzing the difference in diffusion patterns of news on social media
    - Advantages
      - Not easy to manipulate by publishers
      - Independent of the language with which the news is written

Motivation

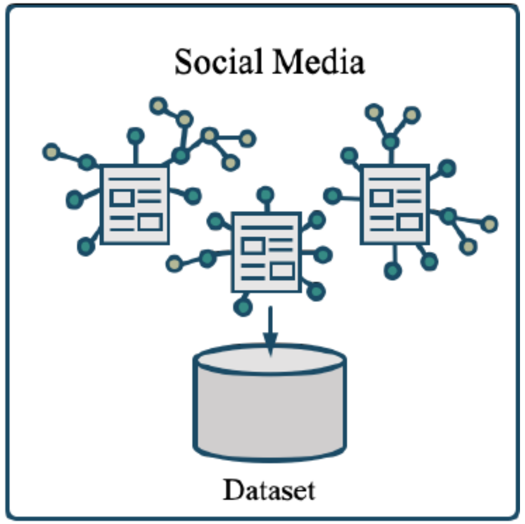
- **The deficiency of the diffusion information about COVID-19 related news articles**
  - **Challenges**
    - Lack of social media diffusion information
    - Diffusion feature extraction for detection of COVID-19 related fake news
    - Lack of comparisons with existing models

Proposed Framework (D-FEDN)



(C1) Diffusion data collection

- **To collect news data and diffusion information on social media**
  - Collecting true and fake news data related to COVID-19
  - Collecting diffusion information of news on social media



(C1) Data Collection


News | Coronavirus pandemic

US researchers share COVID-19 vaccine with the world

Using traditional technology that can be scaled widely and cheaply, Corbevax offers a potential solution to vaccine inequity.

News

Social Media

(C1) Diffusion data collection

- **To collect diffusion information based on the CoAID**
  - **CoAID (Covid-19 heAlthcare mIsinformation Dataset)***
    - Collected news information related to COVID-19 and some social-context information
    - Classified news into true news and fake news
    - Consisted of a total of 3,921 news and 150,002 initial tweets
  - **CoAID⁺ Collection**
    - To collect additional information about diffusion (retweet) through Twitter API based on CoAID

Table 2. Descriptive statistics of CoAID⁺

Feature Name	Fake	True
# of news	157	2,606
# of tweets	9,745	140,257
# of retweets	3,528	45,287
# of nodes	85.51	70.69
Max. depth	1.80	1.57

* Limeng et al. 2020. Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885.

(C2) Feature extraction

- **To extract features by analyzing the diffusion patterns of news**
  - Comparative analysis of the diffusion patterns of true and fake news
  - Effective feature extraction for fake news detection

- **Analysis features**
  - Structural features
    - Analysis of structural patterns of connections between nodes
    - Examples: maximum depth, number of nodes, …
  - Temporal features
    - Analysis of temporal patterns of connections between nodes
    - Examples: time difference between the first tweet and the last retweets, …

(C2) Feature extraction

- **Structural features**

  (S1) Maximum depth

  (S2) Number of nodes

  (S3) Maximum width at a certain hop

  (S4) Average distance of all node pairs

  (S5) Maximum out-degree

  (S6) Number of tweets that first posted the news article

  (S7) Depth from the news article to the influential posting

  (S8) Number of tweets with retweets

  (S9) Fraction of tweets with retweets

(C2) Feature extraction

- **Temporal features**

(T1) Average time difference between the adjacent retweet nodes

(T2) Time difference between the first tweet and the last retweets

(T3) Time difference between the first tweet and the tweet with maximum out-degree

(T4) Time difference between the tweet and its last retweet

(T5) Average time difference between the adjacent retweets in the deepest path

(T6) Time difference between the first and last ‘tweets’ posting the news article

(T7) Average time among tweets posting the news article

(T8) Time difference between the first tweet and its first retweet

(T9) Average time difference between tweets and their first retweet

(C2) Feature extraction

Hanyang University

- **Preliminary analysis of the structural and temporal features on diffusion**
  - Results
    - Fake news spreads farther over more users than real news
    - Fake news spreads faster and has a shorter life span (T2)

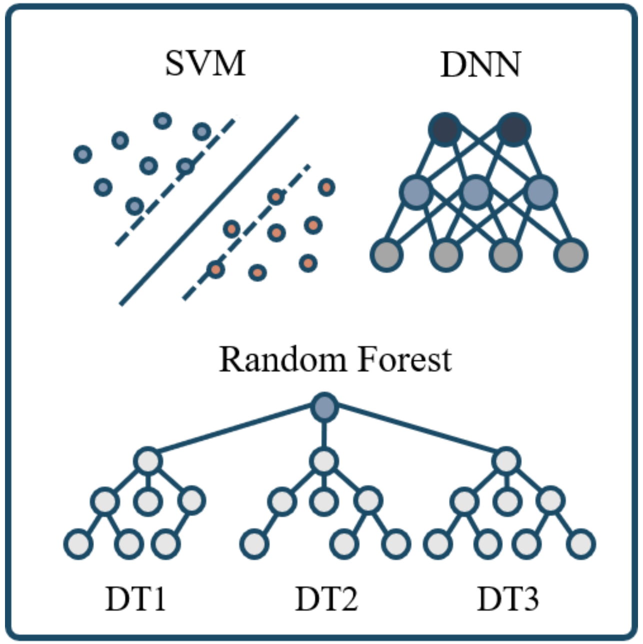
Features	Fake		True		Features	Fake		True	
	Mean	Median	Mean	Median		Mean	Median	Mean	Median
S1	2.59	2	2.47	2	T1	25,427	None	63,877	None
S2	164.53	63	159.23	53	T2	3,274,014	888,347	3,271,199	1,026,472
S3	127.77	58	126.21	45	T3	450,698	102,811	907,908	133,686
S4	2.25	2.18	2.20	2.12	T4	2,714,204	1,728,996	2,729,019	2,050,785
S5	18.33	3	19.44	2	T5	534,308	23,223	825,426	22,469
S6	118.94	48	117.57	43	T6	36,442	None	51,541	None
S7	1.04	1	1.04	1	T7	88,948	38,605	106,606	43,454
S8	10.28	3	8.35	2	T8	258,692	31,962	560,844	66,466
S9	0.17	0.10	0.10	0.06	T9	187,146	9,334	119,822	9,550

Table 3. The extracted structural and temporal features of news articles in CoAID+

(C3) Model training

- **To learn classification models by using the features extracted in (C2)**

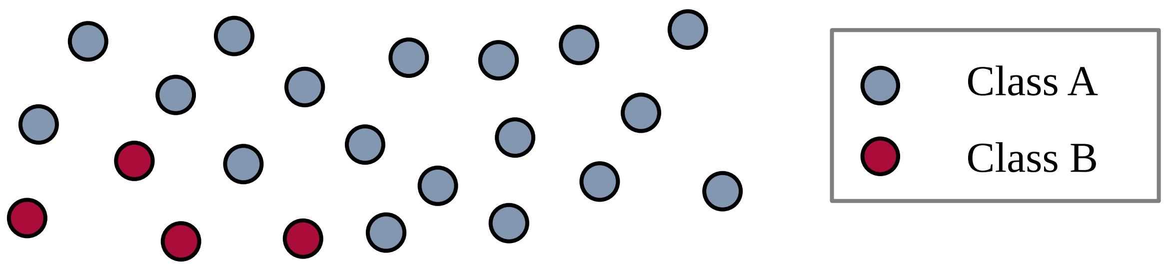
- Classification models
  - Decision tree (DT)
  - Random forest (RF)
  - Support vector machine (SVM)
  - Deep neural network (DNN)



(C3) Model Training

(C3) Model training

- **Class imbalance problem**
  - The problem of the unbalanced class proportions
  - Over-fitting for a specific class

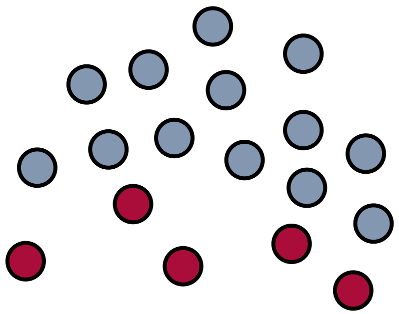


- CoAID^{+} has an unbalanced ratio of true news and fake news
  - True news : fake news = 93 : 7

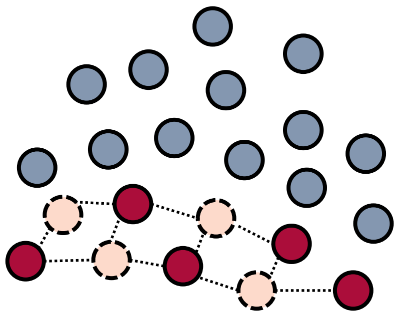
(C3) Model training

- **Solution**
  - To adjust the proportions of classes evenly

- In our case, we adjust the proportions of true and fake news through over-sampling
  - SMOTE (Synthetic Minority Oversampling Technique)
    - Synthetic data is generated by considering the distance between neighbors of minor class data



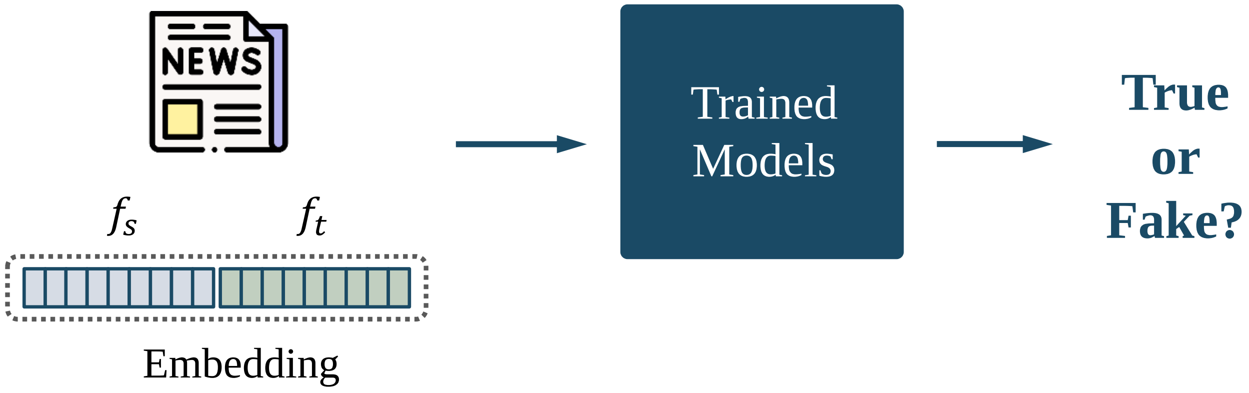
Original Data



SMOTE

(C4) Inference

- **To identify fake news by the trained model (C3)**



Experimental Setup

- **Data set: CoAID[+]**
  - Consists of 3,921 news and 198,817 tweets.
- **Validation**
  - Leave-One-Out Cross Validation (LOOCV)
- **Accuracy metrics**
  - $Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$
  - $Precision = \frac{TP}{TP + FP}$
  - $Recall = \frac{TP}{TP + FN}$
  - $F1\text{-}score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$

Evaluation Questions

- EQ1. How accurately does D-FEND detect fake news articles?
- EQ2. Which type of features (structural/temporal) is more effective in fake news detection?
- EQ3. How sensitive are the accuracies of SVM and DNN models in D-FEND to their hyperparameters?

EQ1. Model Accuracy

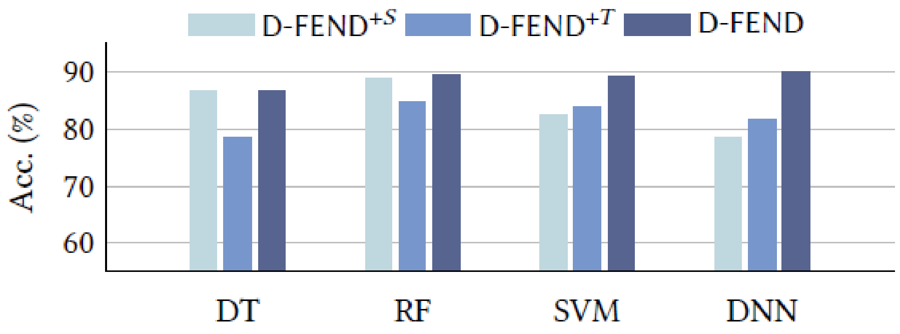
- Comparison of fake news detection accuracy related to COVID-19
  - More than 86% accuracy in all models
  - In particular, the DNN model shows the highest accuracy of about 90%

Model	Class	Acc.	Prec.	Recall	F1-score
DT	True	0.8663	0.8606	0.8741	0.8673
	Fake		0.8721	0.8584	0.8652
RF	True	0.8963	0.8967	**0.8958**	**0.8962**
	Fake		**0.8959**	0.8968	0.8963
SVM	True	0.8933	0.9283	0.8525	0.8888
	Fake		0.8636	0.9341	0.8975
DNN	True	**0.8997**	**0.9414**	0.8525	0.8947
	Fake		0.8652	**0.9469**	**0.9042**

  - Structural and temporal diffusion features are effective in detecting fake news related to COVID-19

EQ2. Effectiveness of Features

- Comparison of fake news detection effect of structural and temporal features
  - D-FEND^{+S}: using only 9 structural features
  - D-FEND^{+T}: using only 9 temporal features
  - D-FEND: using all 18 structural and temporal features



  - The structural and temporal features are effective in detecting fake news
  - The Structural and temporal features are complementary to each other

EQ3. Hyperparameter Sensitivity

- Accuracy comparison for hyperparameters of SVM and DNN models
  - DNN models are relatively insensitive to hyperparameters
  - All DNN models show 89% or better accuracy

SVM		$\gamma = 0.1$				$\gamma = 1$				$\gamma = 10$			
		Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
$C = 0.1$	True	0.7616	0.7486	0.7876	0.7676	0.6514	0.7810	0.4208	0.5470	0.5688	0.9321	0.1485	0.2561
	Fake		0.7759	0.7355	0.7552		0.6036	0.8820	0.7167		0.5374	0.9892	0.6964
$C = 1$	True	0.8063	0.8093	0.8014	0.8053	0.8746	0.9079	0.8338	0.8693	0.8240	0.7873	0.8879	0.8346
	Fake		0.8033	0.8112	0.8072		0.8464	0.9154	0.8795		0.8715	0.7601	0.8120
$C = 10$	True	0.8402	0.8604	0.8122	0.8356	**0.8933**	0.9283	0.8525	0.8888	0.8225	0.7847	0.8889	0.8336
	Fake		0.8222	0.8682	0.8446		0.8636	0.9341	0.8975		0.8719	0.7561	0.8099

DNN		Small				Medium				Large			
		Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
L3	True	0.8948	0.9267	0.8574	0.8907	0.8948	0.9222	0.8623	0.8913	**0.8977**	0.9354	0.8545	0.8931
	Fake		0.8673	0.9322	0.8986		0.8707	0.9272	0.8981		0.8661	0.941	0.902
L5	True	0.8904	0.9214	0.8535	0.8862	**0.8997**	0.9414	0.8525	0.8947	0.8958	0.9333	0.8525	0.8911
	Fake		0.8636	0.9272	0.8943		0.8652	0.9469	0.9042		0.8643	0.939	0.9001

Conclusions

Hanyang University

- To construct a new diffusion dataset, named CoAID$^+$, and providing CoAID$^+$ publicly to vitalize the study on diffusion based fake news detection

- To propose a comprehensive framework for effectively detecting fake news related to COVID-19, named D-FEND based on the diffusion information of news articles

- To validate the effectiveness of D-FEND in fake news detection, successfully detecting fake news articles with 88.89% accuracy on average

Thank you