User Preference-aware Fake News Detection

Yingtong Dou[1], Kai Shu[2], Congying Xia[1], Philip S. Yu[1], Lichao Sun[3]
[1]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, USA
[2]Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA
[3]Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA
{ydou5,cxia8,psyu}@uic.edu,kshu@iit.edu,james.lichao.sun@gmail.com

Accepted by SIGIR'21

210603 Chia-Chun Ho

Outline

Introduction

Approach

Experiments

Conclusions

Comment

Introduction

Among Fake News Detection Techniques

- Fact-Checking is most straightforward approach
  - Usually labor-intensive to acquire evidence from domain experts
- In addition, computational approaches using feature engineering or deep learning have shown many promising result:
  - SAFE, FakeBERT used TextCNN and BERT to encode news textual information
  - GCNFN, GNN-CL leveraged the GCN to encode the news propagation patterns on social media
- These methods focus on modeling news content and its user exogenous context and ignore the user endogenous preferences

Introduction

Correlation between user preferences and their online news consumption behaviors

- *Naive Realism* indicates that
  - consumers tend to believe the their perceptions of reality are the only accurate views
  - others who disagree are regarded as uniformed, irrational, or biased.
- *Confirmation Bias* theory reveals that consumers prefer to receive information the confirms their existing views.
  - A user believes the election fraud would probably share similar news with a supportive stance, and the news asserting election is stolen would attract users with similar beliefs.

Introduction

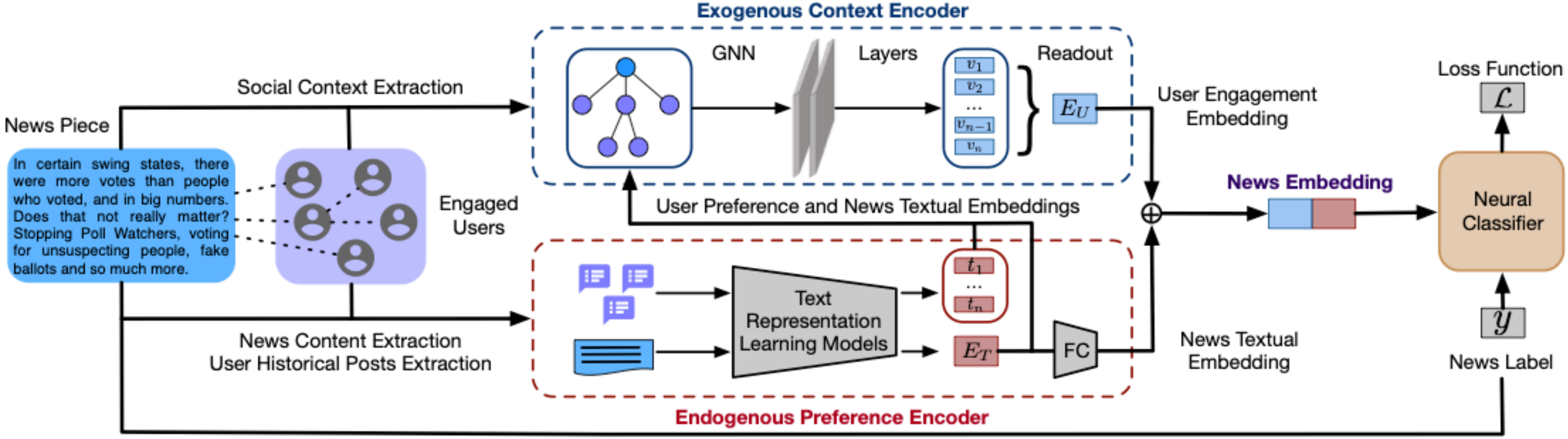
Correlation between user preferences and their online news consumption behaviors

- To model user endogenous preferences,
  - existing works have attempted to **utilize historical posts as a proxy**
  - have shown promising performance to detect sarcasm, hate speech, and fake news spreaders on social media.
- In this paper, authors consider the historical posts of social media users as their endogenous preference in news consumption.

Introduction

User Preference-aware Fake Detection (UPFD)

- Proposed an end-to-end fake news detection framework to model endogenous preference and exogenous context jointly.

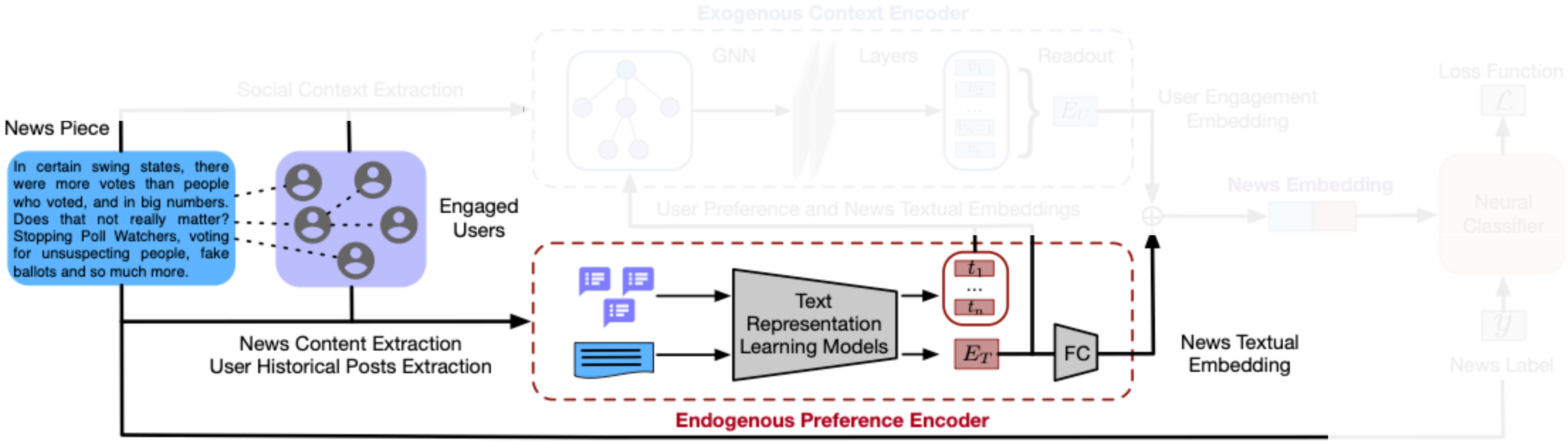


Introduction

User Preference-aware Fake Detection (UPFD)

(1) To model the user endogenous preference

- encode news content and user historical posts using various text representation learning approaches

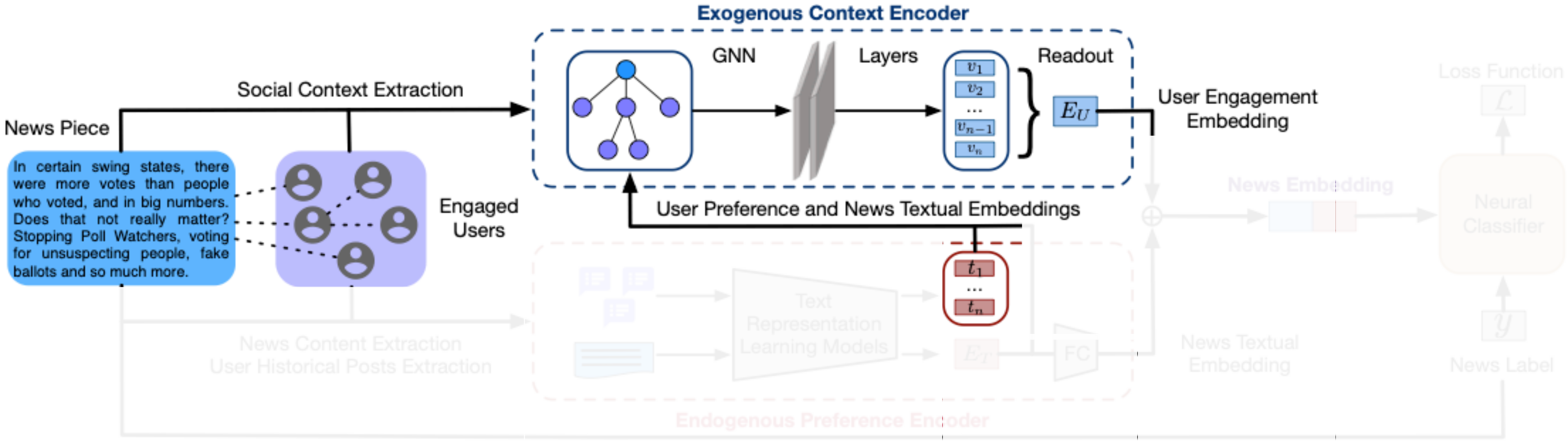


Introduction

User Preference-aware Fake Detection (UPFD)

(2) To obtain the user exogenous context

- build a tree-structured propagation graph for each news based on its sharing cascading on social media.

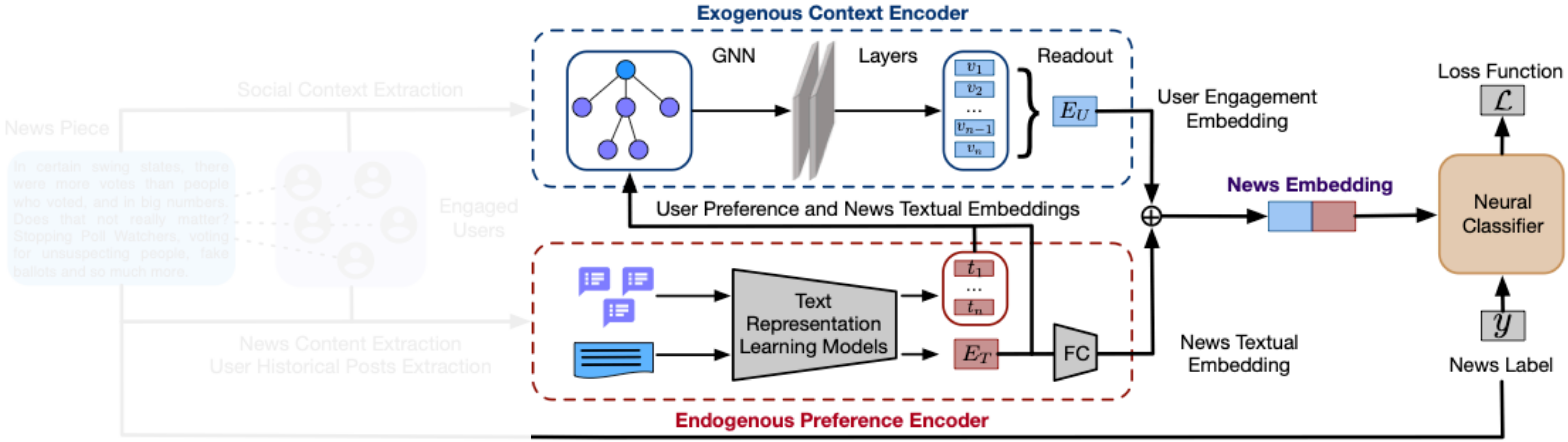


Introduction

<u>U</u>ser <u>P</u>reference-aware <u>F</u>ake <u>D</u>etection (UPFD)

(3) To integrate the endogenous and exogenous information

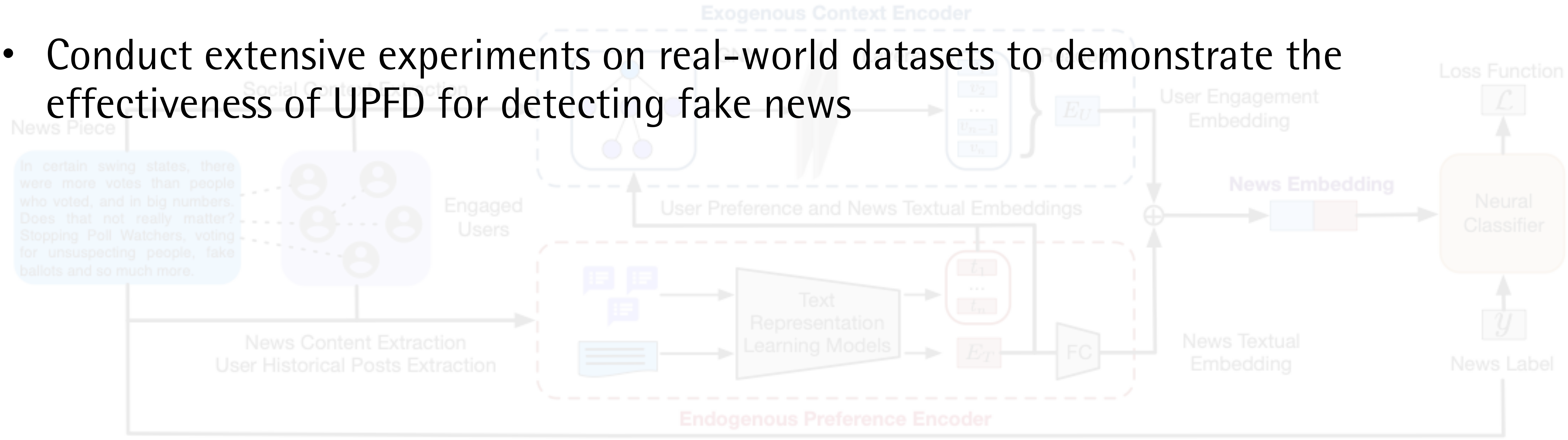
- The user engagement embedding and news textual embedding are used to train a neural classifier to detect fake news



Introduction

Major Contribution of UPFD

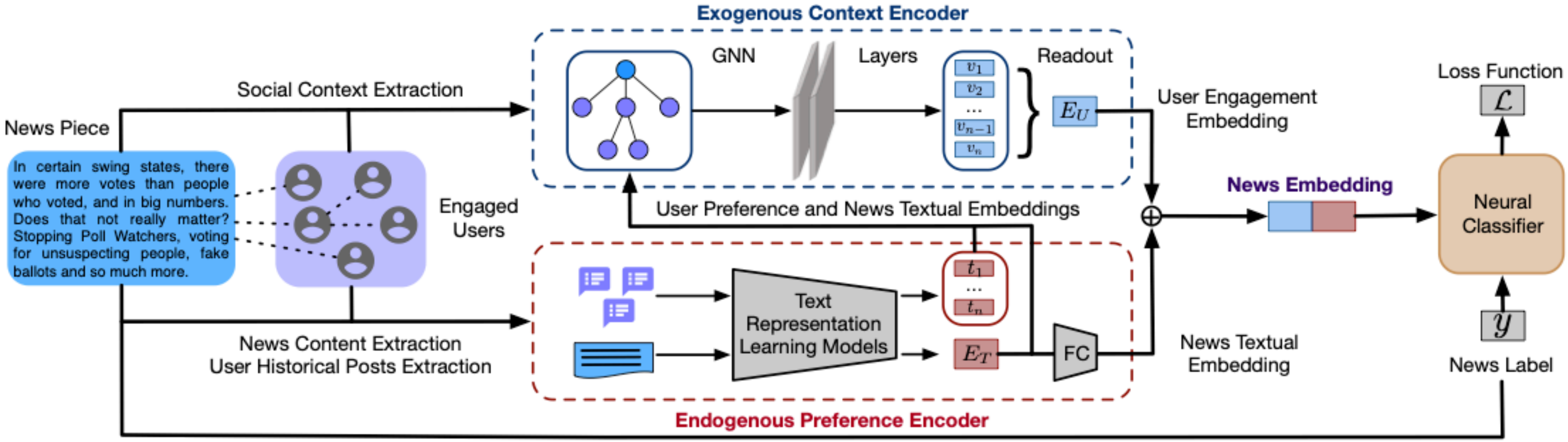
- Study a novel problem of user-preference-aware fake news detection on social media
- Propose a principled way to exploit both endogenous preference and exogenous context jointly to detect fake news
- Conduct extensive experiments on real-world datasets to demonstrate the effectiveness of UPFD for detecting fake news

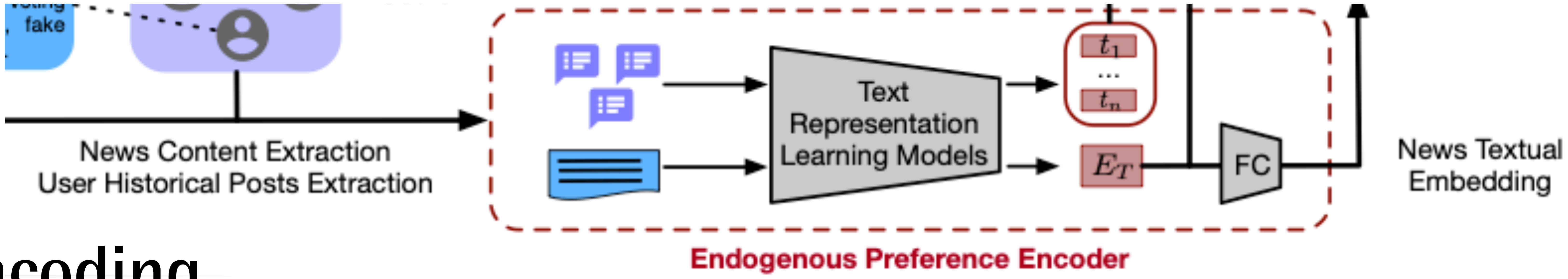


Approach

Major components of UPFD

- Endogenous Preference Encoding
- Exogenous Context Extraction
- Information Fusion

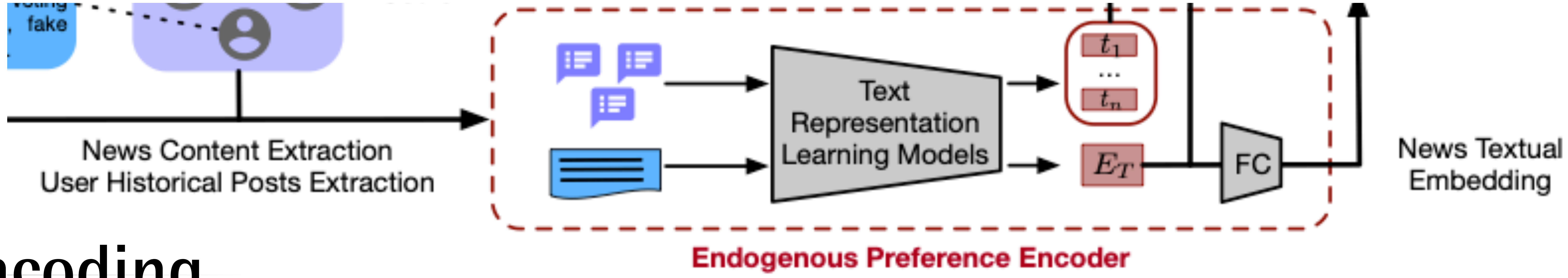




Approach

Endogenous Preference Encoding

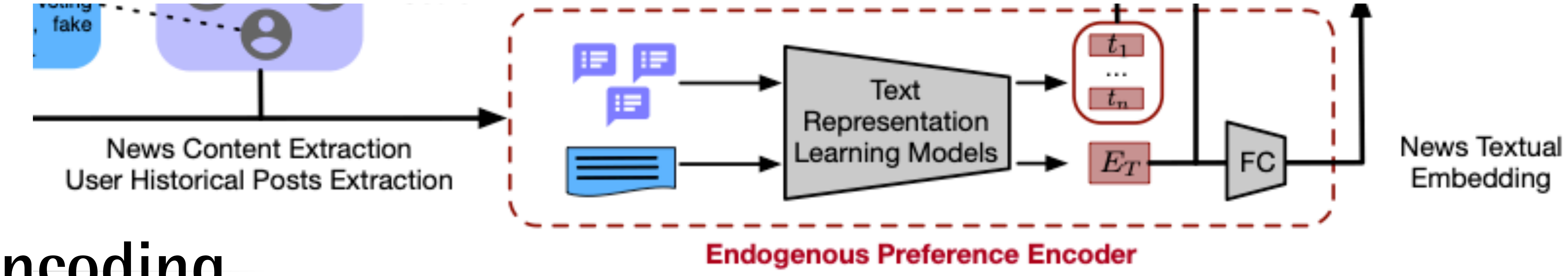
- Model the users' personality, sentiment and stance using their historical posts
  - Leverage the historical posts of a user to encode his/her preference implicitly.
- In this paper, authors select the FakeNewsNet dataset which contains news content and its social engagement information on Twitter.
  - Then use the Twitter Developer API to crawl historical tweets of all accounts that retweeted the news in FakeNewsNet
  - Crawl the recent 200 tweets for each account, so as to near 20 million tweets being crawled in total.



Approach

Endogenous Preference Encoding

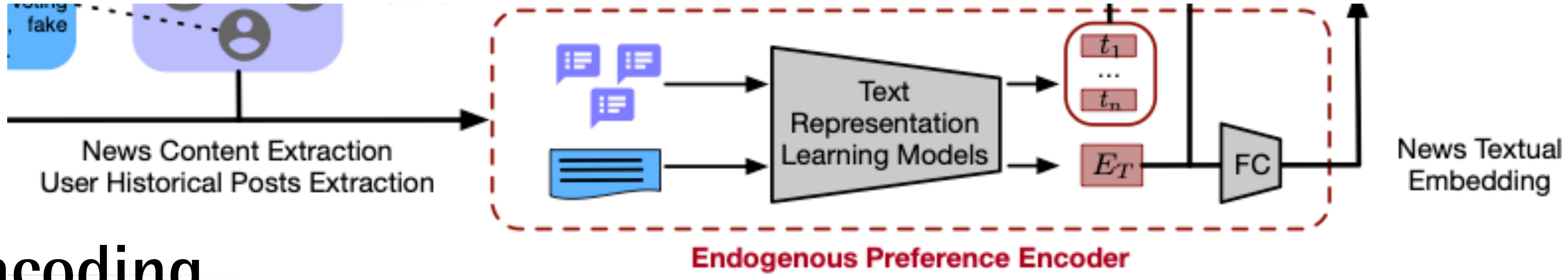
- For inaccessible users (suspended or deleted account), use randomly sampled tweets from accessible users engaging the same news as its corresponding historical posts.
  - Because deleting the inaccessible user will break the intact news propagation cascading and result in a less effective exogenous context encoder.
- Also remove the special characters (e.g. @) and urls, before applying text representation learning methods.



Approach

Endogenous Preference Encoding

- To encode the news textual information and user preferences, employ two types of text representation learning approaches based on language pertaining.
    - word2vec: choose the 680k 300 dimensional vectors pretrained by spaCy
    - BERT: employ pretrained embeddings (BERT-large) using bert-as-a-service
- Instead of training on the local corpus, the word embedding pretrained on large corpus are supposed to encode more semantic similarities between different words and sentences.



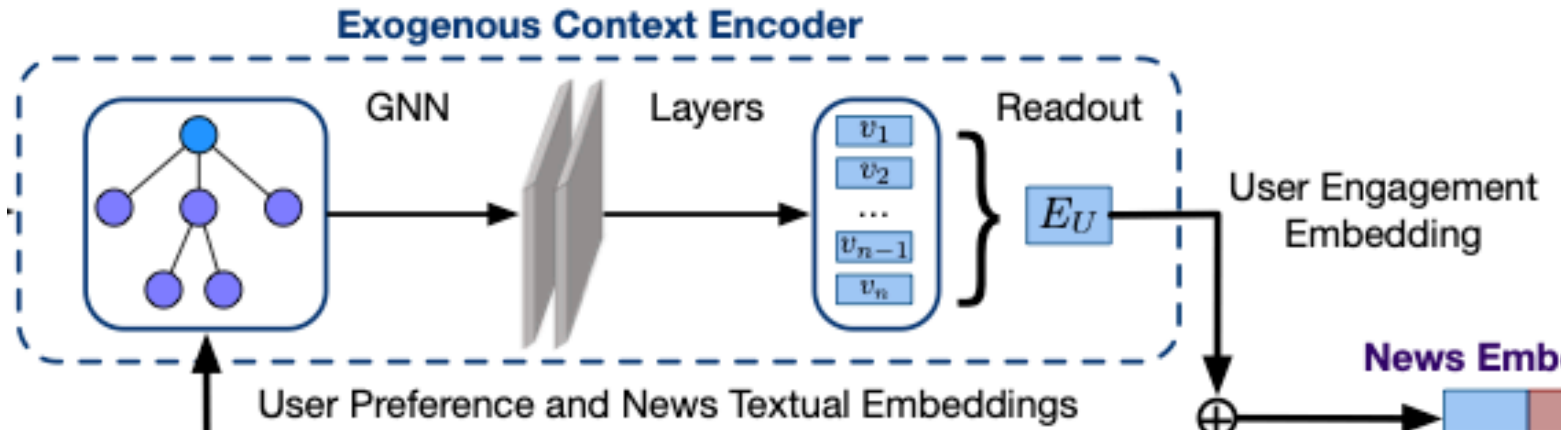
Approach

Endogenous Preference Encoding

- word2vec (spaCy)
  - Average the vectors of existing words in combined recent 200 tweets to get user preference representation.
  - The news textual embedding is obtained similarly.
- BERT (BERT-large)
  - Due to BERT's input sequence length limitation (512 tokens), couldn't use BERT to encode 200 tweets as one sequence, so authors resort to encode each tweet separately and average them afterward to obtain a user's preference representation.
  - Generally, the tweet text is way shorter than the news text, authors empirically set the max input sequence length of BERT as 16 tokens to accelerate the tweets encoding time

Approach

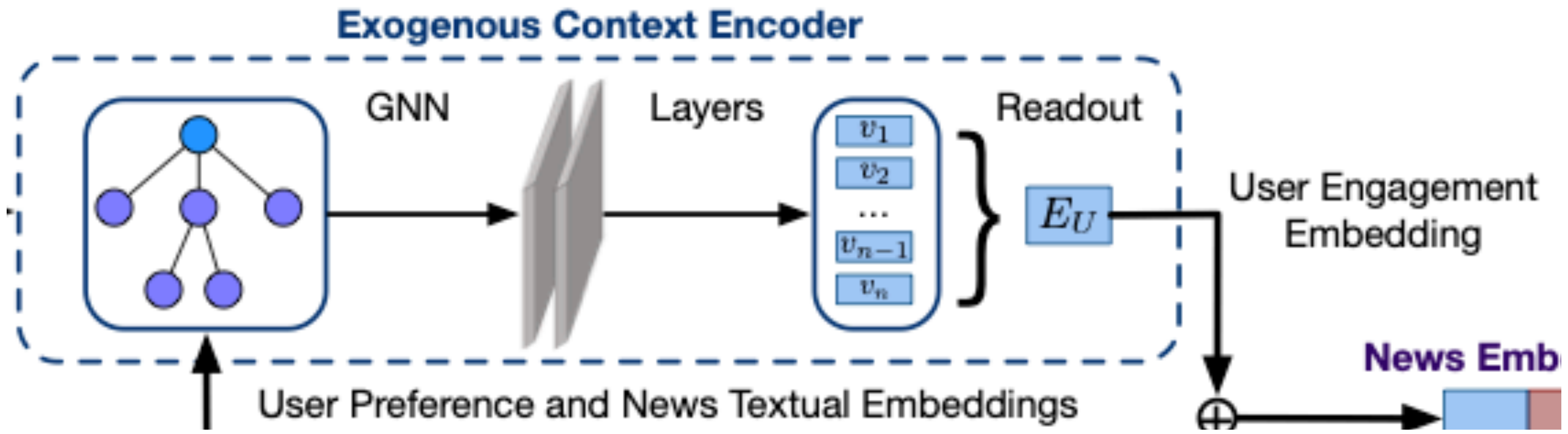
Exogenous Context Extraction



- Given a news piece on social media, the user exogenous context is composed of all users that engaged with the news.
- Utilize the retweet information of news pieces to build a news propagation graph.
- The root node represents the news pieces, and other nodes represent users who share the root news.
- Define a new piece as v_1, and $\{v_2, \ldots, v_n\}$ as a list of users that retweeted v_1 ordered by time.

Approach

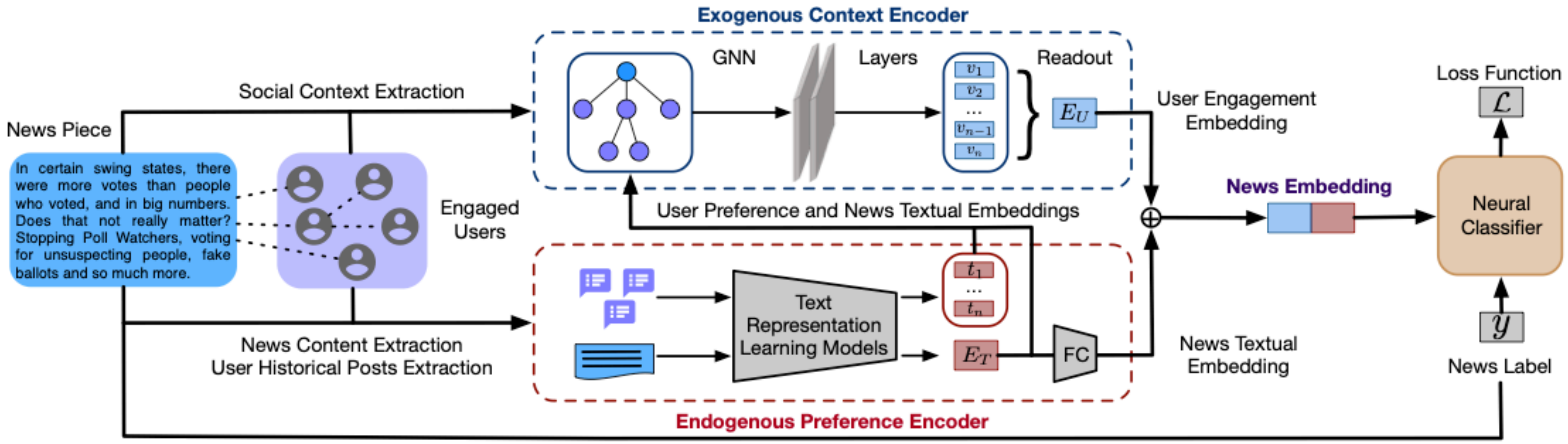
Exogenous Context Extraction



- Define two following rules to determine the news propagation path:
  - For any account v_i , if v_i retweets the same news later than at least one following account in $\{v_1, \ldots, v_n\}$, estimate the news spreads from the account with the latest timestamp to account v_i.
  - If account v_i doesn't follow any accounts in the $\{v_1, \ldots, v_n\}$, conservatively estimate the news spreads from the accounts with the most number of followers.
- Based on the above rules, can build the news propagation graphs on Twitter. Note that this approach can be applied to other social media platforms like Facebook as well.

Approach

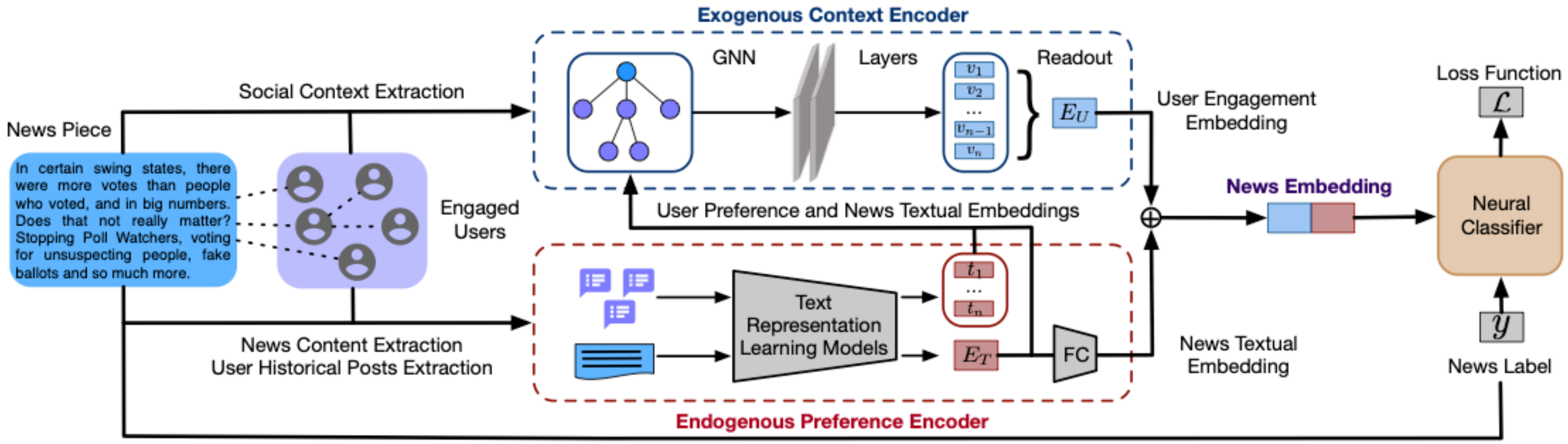
Information Fusion



- Previous works have demonstrated that fusing the user features with a news propagation graph could boost the fake news detection performance.
- Propose a hierarchical information fusion approach, first fuse the endogenous and exogenous information using the GNN, the news textual embedding and user preference embedding can be taken as node features.
- Apply a <u>**readout function**</u> over all node embeddings to obtain the embedding of a news propagation graph. The readout function makes the <u>mean pooling</u> operation over all node embeddings to get the graph embedding (i.e., user engagement embedding).

Approach

Information Fusion



- Since the news content usually contains more explicit signals regarding the news' credibility, so fuse the news textual embedding and user engagement embedding by concatenation as the ultimate news embedding to enrich the news embedding information.

- The fused news embedding is finally fed into a 2-layer MLP with two output neurons representing the predicted probabilities for fake and real news.

- The model is trained using binary cross-entropy loss function and is updated with SGD.

Experiments

Setup: Dataset

- To investigate both the user preference and propagation pattern of fake news, authors choose the FakeNewsNet datasets.
- It contains fake and real news information from two fact-checking websites and the related social engagement from Twitter.

Table 1: Dataset and graph statistics.

Dataset	#Graphs (#Fake)	#Total Nodes	#Total Edges	#Avg. Nodes per Graph
Politifact (POL)	314 (157)	41,054	40,740	131
Gossipcop (GOS)	5464 (2732)	314,262	308,798	58

緋聞警察

Experiments

Setup: Baselines

- Implement the baselines only with the parts for encoding the news content, user comments, and news propagation graph.
- CSI: employs an LSTM to encode the news content information to detect fake news.
- SAFE: uses TextCNN to encode the news textual information
- GCNFN: the first fake news detection framework to encode the news propagation graph using GCN
- GNN-CL: encodes the news propagation graph using DiffPool (a GNN designed for graph classification)
- Authors also add two baselines that apply MLP directly on news textual embeddings encoded by word2vec and BERT

Experiments

Setup: Settings

- Implement all models using PyTorch
  - all GNN models are implemented with PyTorch-Geometric package.
- Unified graph embedding size (128)
- Batch size (128)
- Optimizer (Adam)
- L2 regularization weight (0.001)
- Train-Val-Test (2:1:7) for all model
- Results are averaged over 5 different runnings

Experiments

Research Questions

- **RQ1:** How are the performances of the proposed UPFD framework compared to previous works?
- **RQ2:** What are the contributions of endogenous/exogenous information and other variants of the proposed framework?

Experiments
RQ1: Performance Evaluation

	Model	POL		GOS	
		ACC	F1	ACC	F1
News Only	SAFE [36]	73.30	72.87	77.37	77.19
	CSI [23]	76.02	75.99	75.20	75.01
	BERT+MLP	71.04	71.03	85.76	85.75
	word2vec+MLP	76.47	76.36	84.61	84.59
News + User	GNN-CL [8]	62.90	62.25	95.11	95.09
	GCNFN [17]	83.16	83.56	96.38	96.36
	UPFD (ours)	**84.62***	**84.65***	**97.23****	**97.22*****

- Observe that UPFD has the best performance comparing to all baselines.
- UPFD outperforms the best baseline GCNFN around 1% on both datasets with statistical significance
- Result of UPFD and GCNFN demonstrate that user comments (used by GCNFN) are also beneficial to fake news detection
- The user endogenous preference could impose additional information when user comment information is limited
- Since all baselines either encode the news content or user comments without considering the historical posts, prove the historical posts as user endogenous preferences could improve the fake news detection performance.

Stars denote statistically significant under the t-test ($*$ $p \leq 0.05$, $**$ $p \leq 0.01$, $***$ $p \leq 0.001$)

(Ours): Note that the UPFD with the best performance on the both datasets uses BERT as the text encoder and GraphSAGE as the graph encoder.

Experiments
Research Questions

- **RQ1:** How are the performances of the proposed UPFD framework compared to previous works?

- **RQ2:** What are the contributions of endogenous/exogenous information and other variants of the proposed framework?

Experiments

RQ2: Ablation Study: Encoder Variants

Feature	POL				GOS			
	GraphSAGE		GCNFN		GraphSAGE		GCNFN	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Profile	77.38	77.12	76.94	76.72	92.19	92.16	89.00	88.96
word2vec	80.54	80.41	80.54	80.41	<u>96.81</u>	<u>96.80</u>	94.97	94.95
BERT	**84.62**	**84.53**	<u>83.26</u>	<u>83.14</u>	**97.23**	**97.22**	96.18	96.17

- Table show performance of 2 GNN variants using 3 different node features.
- GraphSAGE: a GNN to learn node embeddings via aggregating neighbor nodes information
- GCNFN: a GNN-based fake news detection model which leverages two GCN layers to encode the news propagation graph
- Endogenous features (word2vec & BERT) are consistently better than the profile feature (only encodes the user profile information)
- Also observe that GraphSAGE+BERT have the average best performance among others

Note that the BERT performance could be further improved via fine-tuning, and authors leave it as future work.

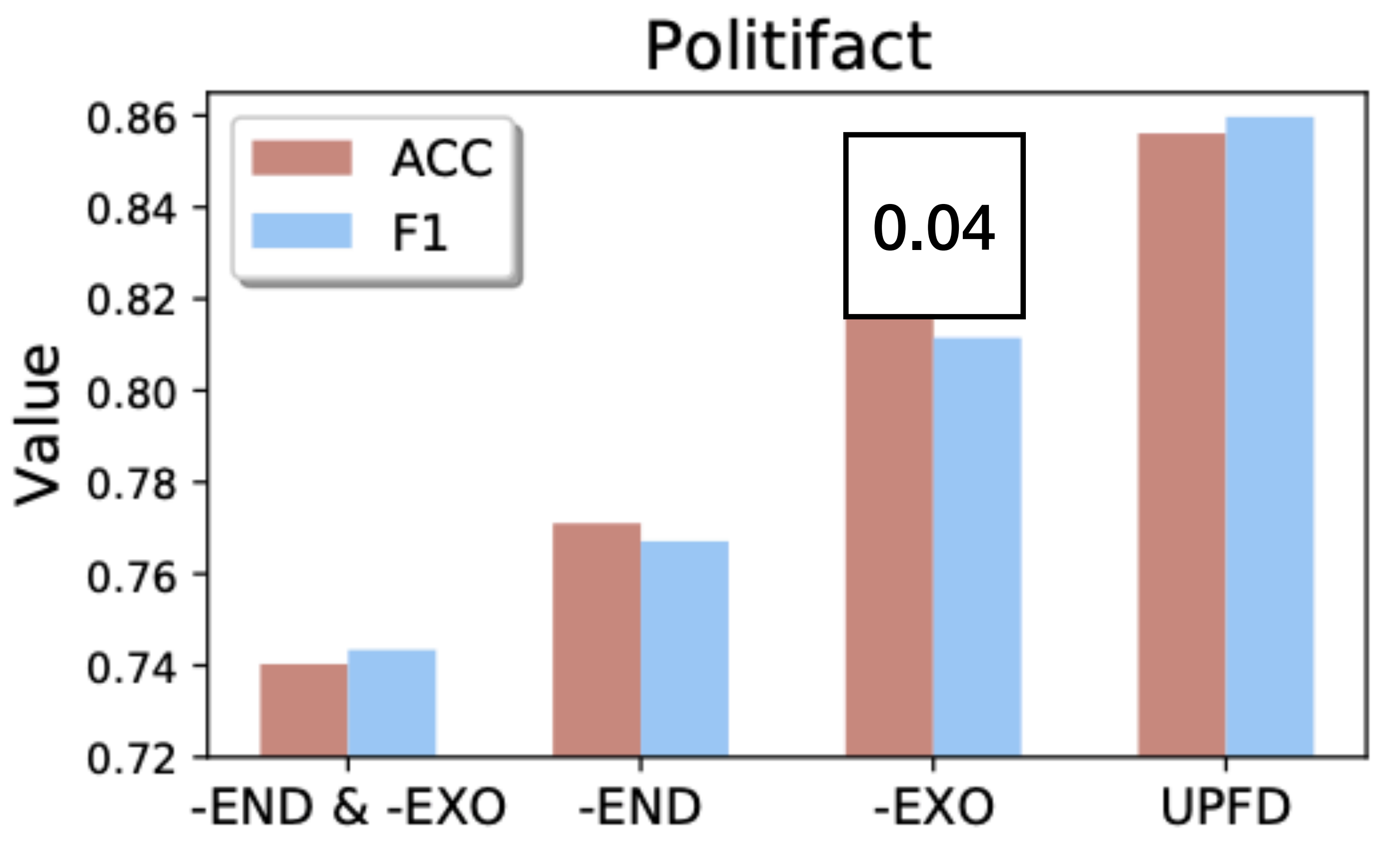
Experiments

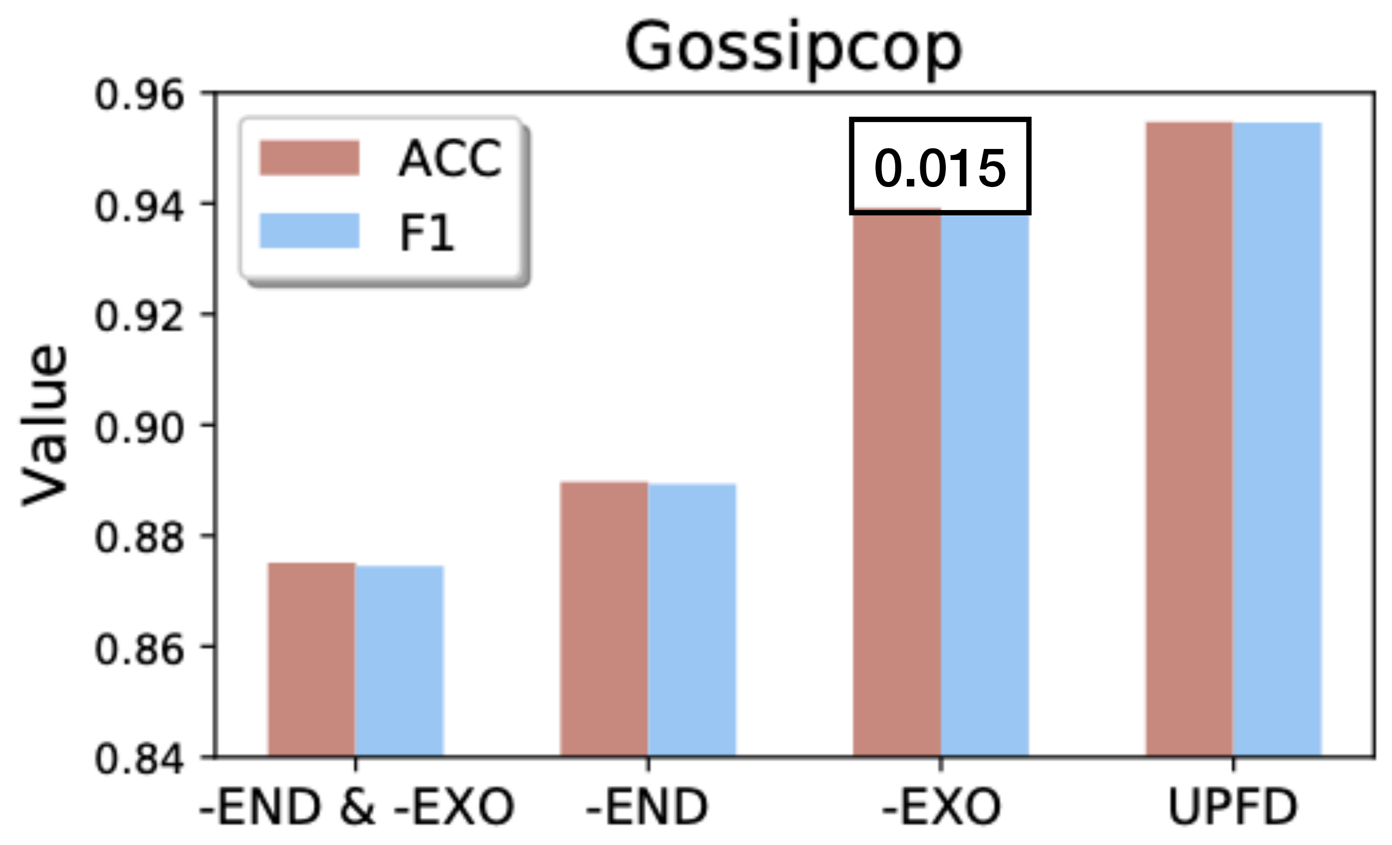
RQ2: Ablation Study:

Framework Variants

- Design 3 UPFD variants that remove the endogenous info, exogenous info or both of them.
- Employ the GCNFN (word2vec) as the graph (text) encoder for both datasets, and remove news concatenation to ensure a fair comparison
- $-EXO$ is implemented by removing all edges in the news propagation graph, thus encodes the news embedding solely based on node features without exchanging information between nodes
- $-END$ takes the user profile as node features and doesn't contain user endogenous preference information.
- $-EXO$ & $-END$ replaces the node features of the $-EXO$ with user profile features

Experiments
RQ2: Ablation Study: Framework Variants





- Find that removing either component from the UPFD will reduce its performance
- Indicates that exogenous information (i.e., news propagation graph) is more informative on Politifact since removing it results in a larger performance drop (0.04)
- Obvious that endogenous information contributes more to performance gain than exogenous information
- This observation further verifies the necessity of modeling user endogenous preferences.

	POL		GOS	
	ACC	F1	ACC	F1
-EXO	81.63	81.15	93.92	93.81
UPFD	85.61	85.97	95.47	95.46

Conclusion

- Argues that **user endogenous news consumption preference** plays a vital role in the fake news detection problem.
- Collect the user historical posts to implicitly model the user preference and leverage the news propagation graph on social media as the exogenous social context of users.
- UPFD is proposed to fuse the endogenous and exogenous information and predict the news' credibility on social media.
- Experimental results demonstrate the advantage of modeling the user endogenous preference

Comments
of User Preference-aware Fake Detection

- User endogenous news consumption preference
- Preprocessing inaccessible account historical tweets
- News propagation graph rules
- Readout function?