Privacy-Promoting Fake News Detection

Jatin Arora, Bailey Flanigan, Shilpa George (mentor)

Background

Fake news: a widespread, influential problem

"Pope Francis shocks world, endorses Donald Trump for president"

~ 2 million views

Identifying fake news requires user data

Our vision: platforms can detect fake news *without* collecting detailed user data, via...

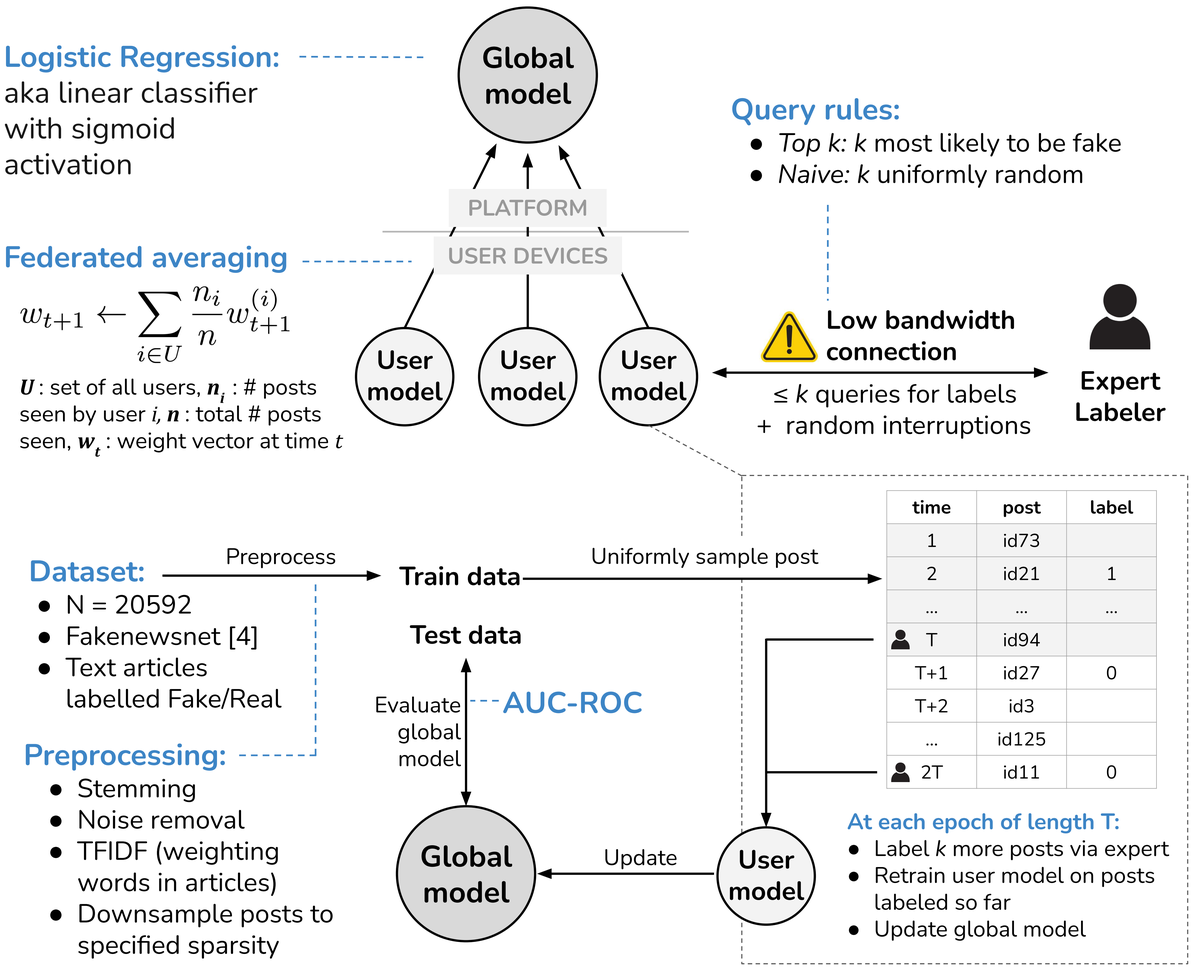
Federated learning: global model is trained by aggregating several decentralized models on edge devices.

Related work:

- Fake news detection
- Existing work on learning from streaming data with low bandwidth connection to labeler [1]
- Use of edge computing / federated learning to preserve data privacy [2,3]

Research question: Can fake news detection be done with low-bandwidth federated learning with small loss in performance?

Federated architecture and system model



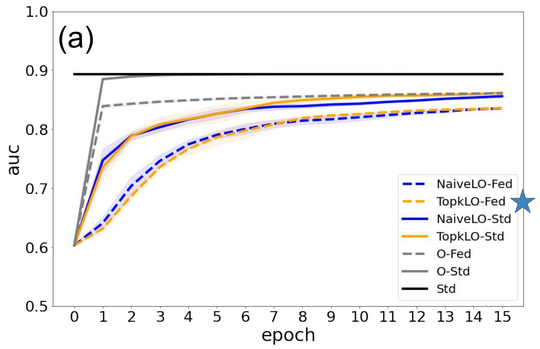
Run Parameters and Benchmarks

Run params: *# users* = 50, *query limit k* = 5, *fake:real* = 1:80, *epoch length* = 200, *base data* = 25 posts.

	Avg # new labels / epoch	Federation gap: Federated	Federation gap: Standard	Query rule
Limited bandwidth (≤*k* posts to expert)	2.5/user * #users = 125 total	– – / – –	——— / ———	Naive / Top-k
Unlimited bandwidth (*all* seen posts to expert)	50*200 = 1000 total	– – / – –	——— / ———	
Best possible (non-streaming; sees all data directly)			———	

Results in Demo Models

Model #1: *includes* proper nouns



Model #2: *excludes* proper nouns

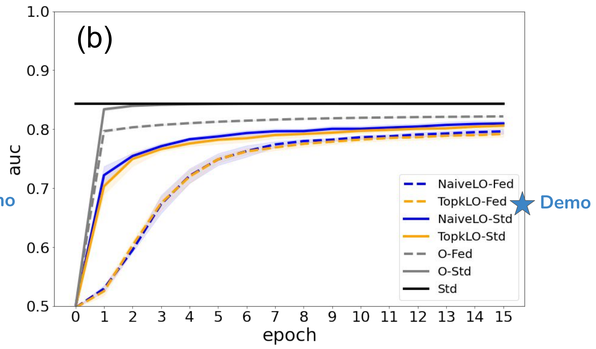
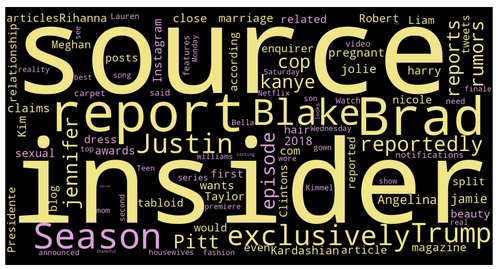


Figure 1. Model performance over 15 epochs. Standard deviations (shaded) over 5 replicates. Federation gaps at epoch 15: (a) NaiveLO: 2.1%, TopkLO: 2.6%, O: 3.2%; (b) NaiveLO: 1.3%, TopkLO: 1.4%, O: 2.1%;



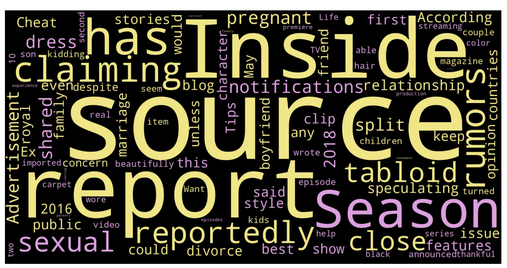


Figure 2. Wordclouds explaining decisions of respective models. Purple, yellow words are strongly associated with real, fake news respectively. Word size is proportional to importance in prediction.

Future Work

- Increase sophistication of user model
- Increase sophistication of preprocessing
- Study federation gap in other models, datasets

Sources

1. George, Shilpa, et al. "Low-bandwidth Learning From Unlabeled Data." *Under submission* (2021).
2. Satyanarayanan, Mahadev. "The emergence of edge computing." *Computer* 50.1 (2017): 30-39.
3. Sadilek, Adam, et al. "Privacy-first health research with federated learning." *NPJ digital medicine* 4.1 (2021): 1-8.
4. FakeNewsNet: Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media (2018)