Truth be Told: Fake News Detection Using User Reactions on Reddit

Vinay Setty
University of Stavanger, Norway
vsetty@acm.org

Erlend Rekve
University of Stavanger, Norway
erlend.rekve@stud.uis.no

CIKM'20

210910 Chia-Chun Ho

Outline

Introduction

Related Work

Problem Description

Proposed Model

Experiments

Conclusion and Future Work

Comments

Introduction
Automated claim verification and fake news detection

- Some approaches rely on linguistic features, source domain and speaker information.

  - Domain name, page rank and author of the claim.

- Misinformation is very difficult to detect even for critical human readers.

  - It's not hard to write high quality news articles conveying false facts.

  - Since these articles are written by humans, it's hard to come up with a specific set of features to verify their truthfulness.

- While these features boost the recall of detecting false claims, they result in poor precision since all news articles from a specific website tend to be classified as true or false.

Introduction
Other recent works

- Consider temporal patterns of the response received for news articles and model them using deep neural networks such as LSTM.

- These works do not consider the news article contents rather only focus on the textual content of the reactions.

- Models to jointly consider the textual content of the news article as well as the textual social media reactions to detect the false claims are limited in the literature.

Introduction
Contributions

- Release a large-scale dataset with social media comments for the task.

- Propose neural network models to represent textual content of news articles and social media comments jointly to detect false claims.

- The proposed solution purely depends on textual content rather than any handcrafted features or domain information.

Related Work
of fake news detection

- Some works consider manually engineered linguistic and source-based features ('16–'17).

- Most recent works proposed deep neural networks with attention mechanisms ('18–'19).

  - News article contents alone are not sufficient and none of these approaches consider additional information such as social media discussions.
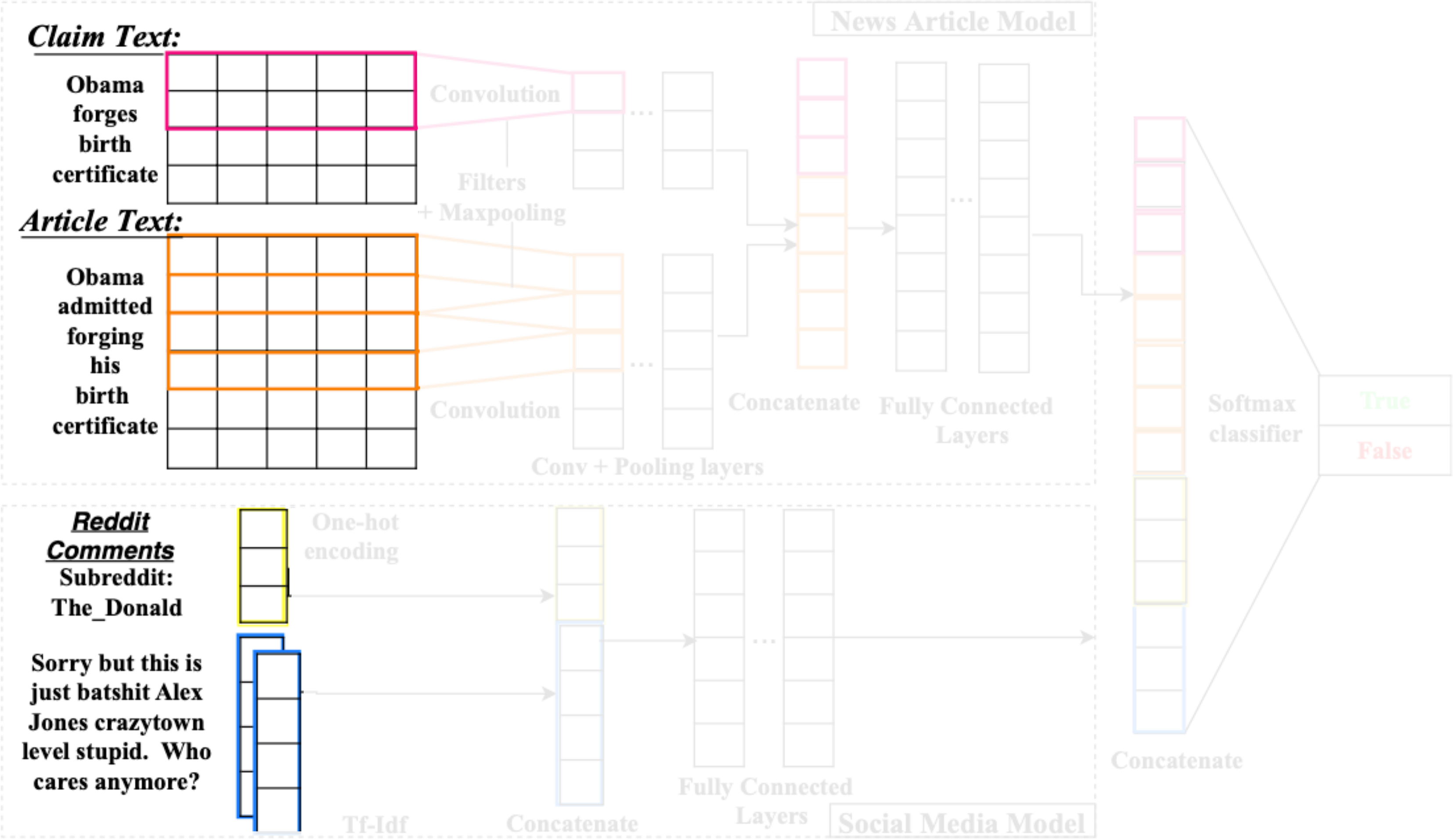
Related Work
Modeling social media

- Existing works in literature manually hand-craft features such as Facebook likes, numbers of shares, and user demographics to train classifiers ('15).

- Modeling rumor cascades in social networks as RNNs have been proposed recently ('16, '18).

- Hybrid models which consider the temporal sequence of textual content of social media comments have also been proposed ('17).

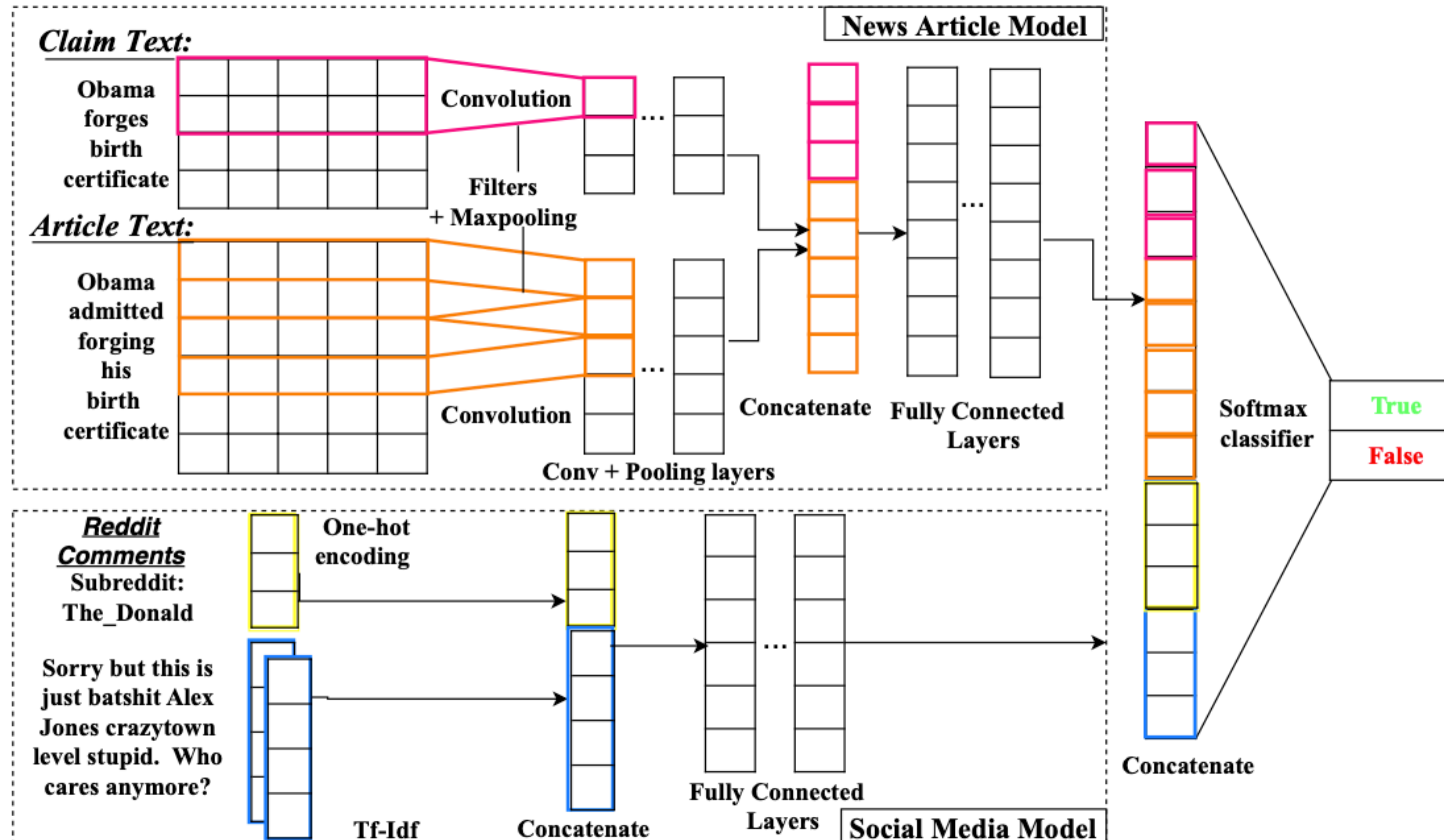 - These models do not consider the news article text and social media comments holistically.

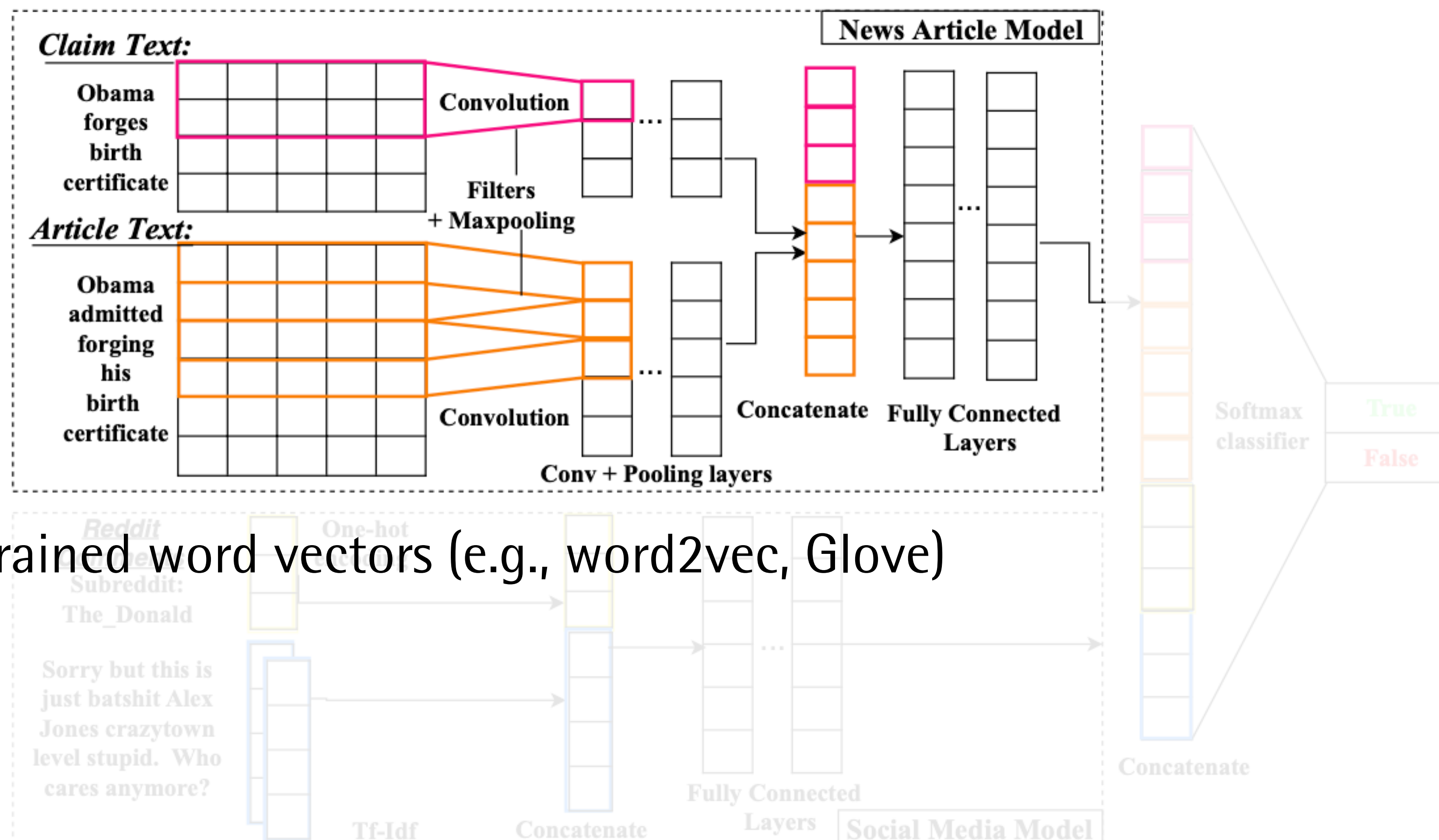# Problem Description
Example explanation

Proposed Model
Model Overview
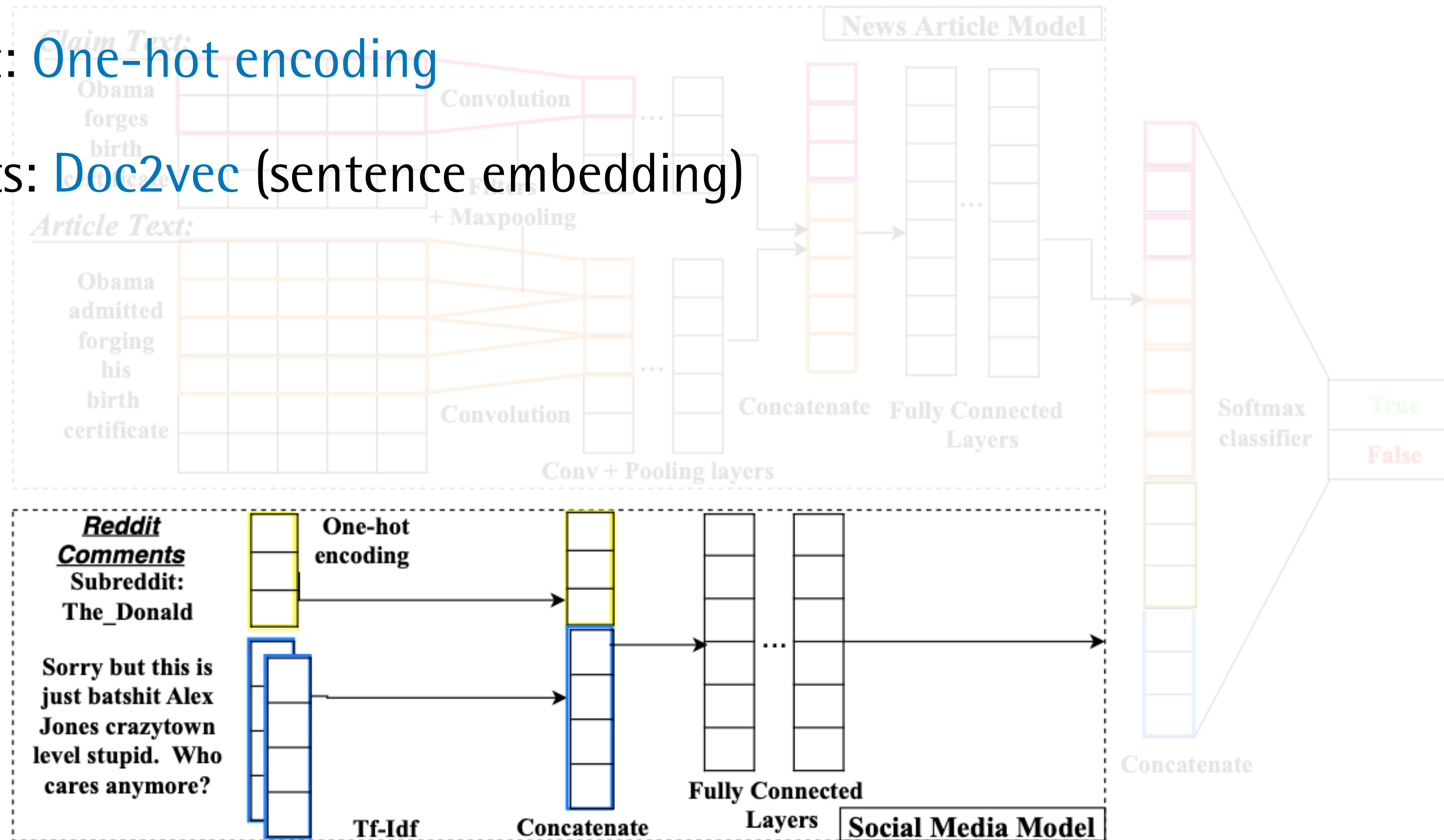
Proposed Model
News Article Model

- Use pre-trained word vectors (e.g., word2vec, Glove)

- Text-CNN

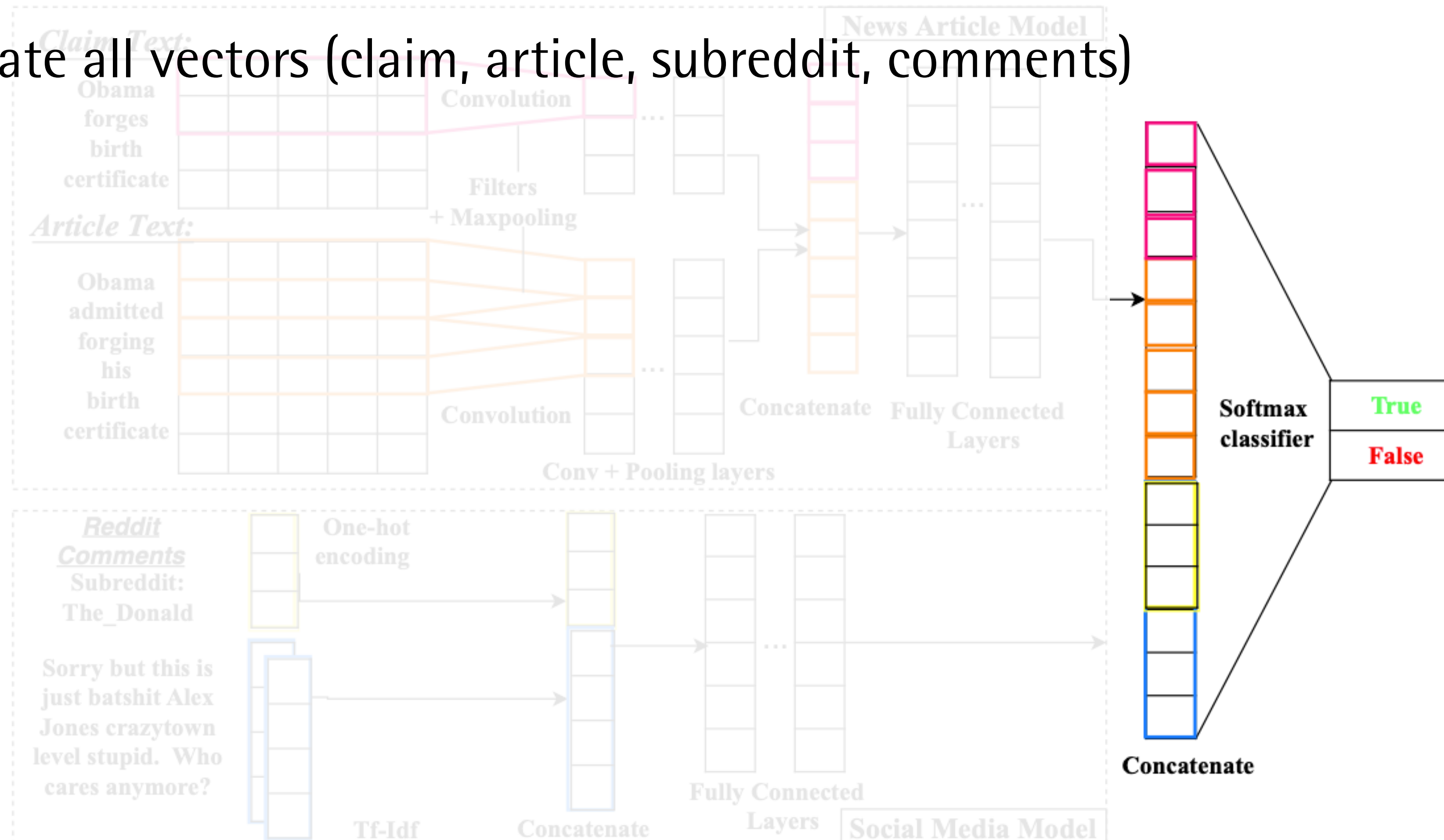# Proposed Model
Social Media Discussion Model

- Subreddit: One-hot encoding

- Comments: Doc2vec (sentence embedding)

Proposed Model
Final Classifier

- Concatenate all vectors (claim, article, subreddit, comments)

Experiments
Datasets

Labeled by	Articles	Claims	Reddit comments	False claims
Snopes	36,271	3,096	352,708	73%
Politifact	11,019	9,453	289,424	80%
Emergent	250	48	20,073	36%
Total	47,540	12,597	662,205	63 %

- Reddit comments for fact checked fake/true news

  - https://www.kaggle.com/deepnews/fakenews-reddit-comments

- Rumor data

  - Claims labeled by snopes.com, politifact.com, emergent.com

- Reddit data

  - Use Reddit API to get user comments.

  - Omit the subtree of comments which appear after the debunking comments referencing fact-checking websites.

Dataset

Reddit comments for fact checked fake/true news

Reddit comments collected for fact checked news from Snopes, Politifact etc.

deepnews • updated a year ago (Version 1)

Data   Code (1)   Discussion   Activity   Metadata

Download (903 MB)   New Notebook

Usability 4.4   License CC0: Public Domain   Tags computer science, news, deep learning

Data Explorer

902.8 MB

{i} fake_news_reddit_cikm20.js...

Summary

▸ 🗀 1 file

< **fake_news_reddit_cikm20.json** (902.8 MB)

⚠ This preview is truncated due to the large file size. The number of JSON items and individual items might be might be truncated. Create a Notebook or download this file to see the full content.

Download   Create Notebook

```
▼ "root" : [ 14 items
    ▼ 0 : { 6 items
        "label" : int 0
        ▸ "reddit_comments" : [] 0 items
        "researched_by" : string "snopes"
        "text" :
        string "analyze videos growth watch videos growth since day uploaded keep track it daily find video want track
        click follow button start following videos performance including views likes dislikes available also get watch
        videos growth since day uploaded"
        "title" : string "vidinfo"
        "url" : string "http://www.vidinfo.org/video/67155269/jeremy-meeks-bail-women-offer-to-pay-bail-for"
    }
    ▼ 1 : { 6 items
        "label" : int 1
```

⊘ This preview is truncated due to the large file size. The number of JSON items and individual items might be might be truncated. Create a Notebook or download this file to see the full content.   **Download**   **Create Notebook**

▼ "root" : [*14 items*
    ▼ 0 : { *6 items*
        "label" : int 0
        ▶ "reddit_comments" : [] *0 items*
        "researched_by" : string "snopes"
        "text" :
        string "analyze videos growth watch videos growth since day uploaded keep track it daily find video want track click follow button start following videos performance including views likes dislikes available also get watch videos growth since day uploaded"
        "title" : string "vidinfo"
        "url" : string "http://www.vidinfo.org/video/67155269/jeremy-meeks-bail-women-offer-to-pay-bail-for"
    }
    ▼ 1 : { *6 items*
        "label" : int 1
        ▶ "reddit_comments" : [] *0 items*
        "researched_by" : string "snopes"
        "text" :
        string "last week current administration missed point again joint chiefs last week current administration missed point again joint chiefs staff chairman sent angry letter washington post regarding editorial cartoon newspaper published kudos exercising freedom speech kudos washington post publishing critical letter same letter represents nothing another attempt administration shift focus away itself letter proves rather accept constructive criticism military highranking officials prefer misconstrue facts strike unwarranted offensives cartoon drawn pulitzer prizewinning tom toles features heavily bandaged quadrupleamputee soldier lying hospital bed chart reads us army line graph sloping steadily down side defense secretary donald rumsfeld playing doctor announces listing condition battle hardened bottom cartoon smaller rumsfeld drawn caption prescribing stretched thin define torture so mean toles making fun soldiers veterans way trivializing wounds american soldiers people think so cartoon clearly making strong statement statement army whole careless treatment soldiers wellbeing toles said drew cartoon response comments rumsfeld made two weeks ago pentagonsponsored study stated war iraq risks breaking army rumsfeld ignoring study called military battlehardened extremely capable force may true even battlehardened capable organization find vulnerable position more deputy communications director disabled american veterans issued statement saying offended cartoon rather that drove home point critically ill patients certainly need attended to still letter editor pronounced cartoon callous asked toles make light soldiers tremendous physical sacrifices perhaps joint chiefs staff spend time attending patients instead writing letters every citizen country toles rumsfeld pitt students mess together us want thing minimal amount harm coming soldiers trying cast toles washington post negative light insensitive soldiers struggles chiefs staff ignore unity actively try destroy it"
        "title" : string "editorial misinterpreted toon unpatriotic"
        "url" : string "http://pittnews.com/30440/archives/editorial-misinterpreted-toon-not-unpatriotic/"
    }
    ▼ 2 : { *6 items*
        "label" : int 0
        ▶ "reddit_comments" : [] *0 items*
        "researched by" : string "snopes"

Experiments
Baselines

- SVM with TF-IDF features

- LSTM

- Conditional Random Field (CRF) (WWW'17)

- Reddit comments

- News article model

Experiments
Performance

Approach	Accuracy		False Claims			True Claims			AUC
	Micro	**Macro**	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**	
SVM	0.793	0.796	0.739	0.748	0.791	0.754	0.843	0.796	0.884
LSTM	0.735	0.726	0.690	0.734	0.762	0.829	0.558	0.667	0.835
CRF [9]	0.618	0.610	0.630	0.740	0.680	0.594	0.460	0.520	0.611
Reddit Doc2Vec	0.657	0.647	0.651	0.740	0.692	0.667	0.567	0.613	0.686
News Article CNN	0.802	0.798	0.777	0.873	0.822	**0.838**	0.724	0.777	0.885
Joint (News Article + Reddit)	**0.826**	**0.825**	**0.837**	**0.895**	**0.866**	0.814	**0.817**	**0.812**	**0.909**

- There is a performance gain by concatenating the two models, a reason for it being so small is that we do not have Reddit comments from all the articles either by there not existing or limitations of the Reddit API.

- If take away all the articles not containing Reddit comments, the performance of the joint model over the news article model increases by 2.5%.

Conclusion and Future Work

- Release a large-scale dataset for fake news detection using the popular social media forum Reddit.

- Presented a simple neural network model to jointly consider the claim text, news article body and the social media comments to classify fake news.

- Will do more analysis of the Reddit dataset and exploit network features to improve fake news detection.

Comments
of Truth to Told

Labeled by	Articles	Claims	Reddit comments	False claims
Snopes	36,271	3,096	352,708	73%
Politifact	11,019	9,453	289,424	80%
Emergent	250	48	20,073	36%
Total	47,540	12,597	662,205	63 %

- New dataset with Reddit Comment

  - Unbalance

  - No other approaches use this dataset

- Use user comment to represented the opinion of reader.

- Simple text-based classifier model

  - Maybe can use user comments to generate the emotion information for the news to detect fake news.