CIFellows 2020-2021

Computing Innovation Fellows

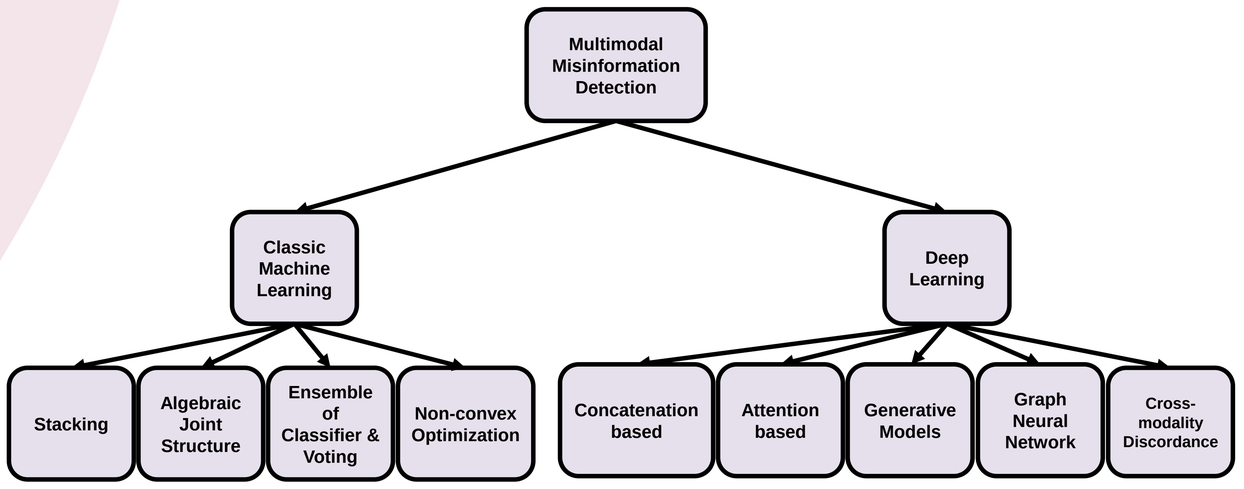
Multi-class Multi-modal Misinformation Detection with Concatenation-based Architecture and Variational Autoencoder

Sara Abdali-Georgia Institute of Technology sabdali3@gatech.edu

Introduction

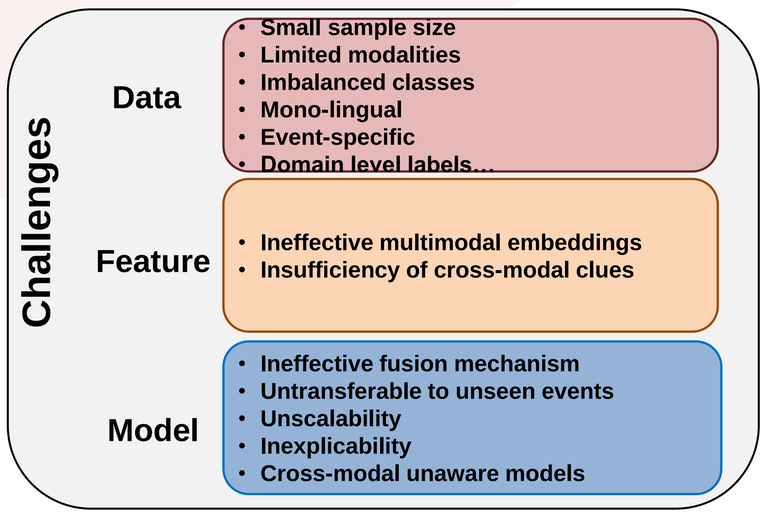
- Social media platforms are evolving from text-based forums into multi-modal environments and consequently the nature of misinformation in social media is changing accordingly.
- Misinformation spreaders have recently targeted contextual correlations between modalities e.g., text and image. Thus, many research efforts have been put into development of automatic techniques for detecting possible cross-modal discordances in web-based media. However, there are still limitations in detecting multimodal misinformation.

What has been done so far?



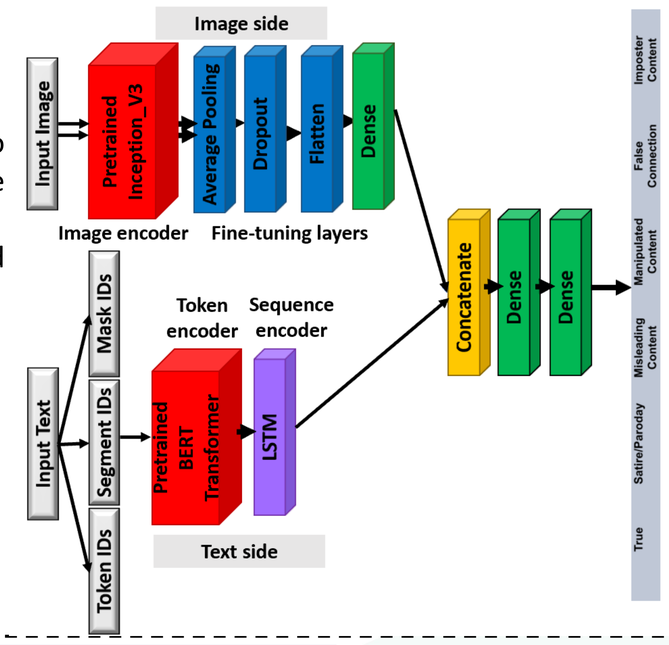
Limitations

- Recent studies on multi-modal learning have made significant contributions to the field of multi-modal fake news detection.
- There are still weaknesses which recognizing them opens the door to new opportunities.



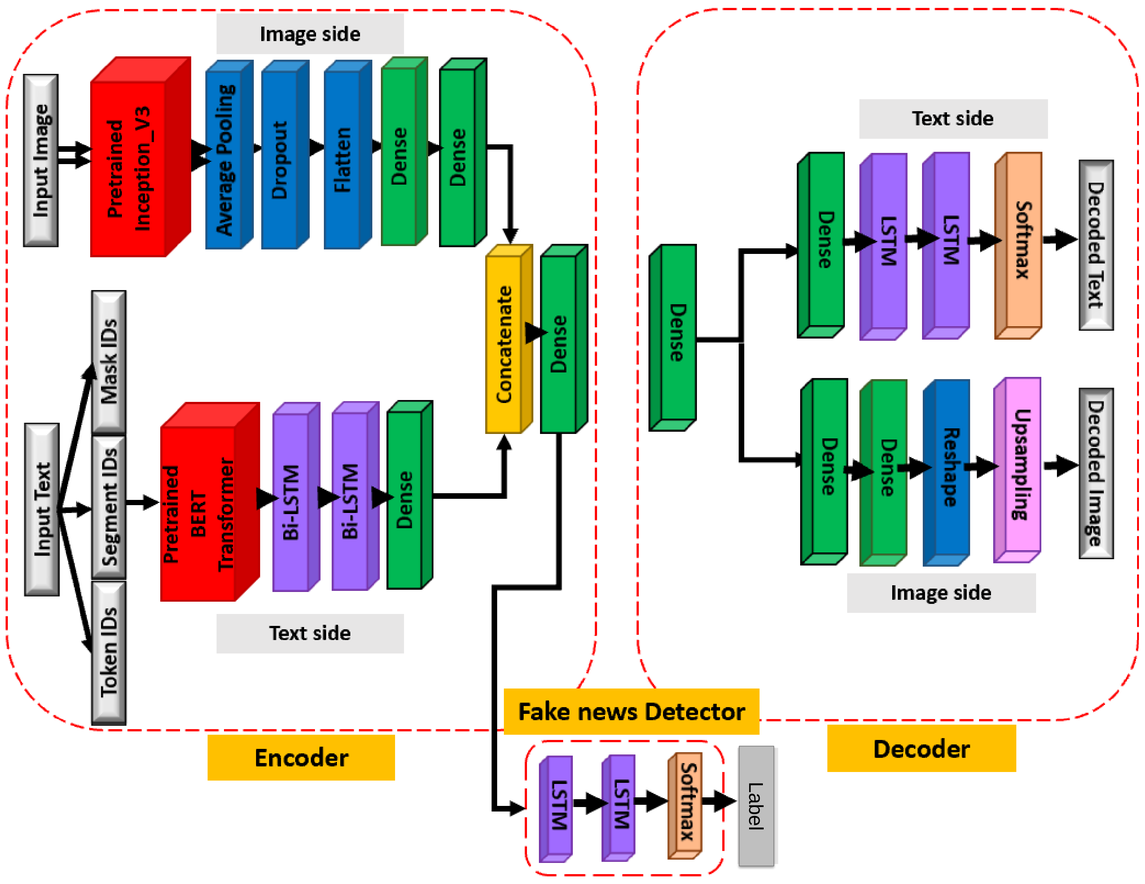
Our concatenation-based architecture

- Concatenation-based models are easy to implement and result in admissible accuracy.
- we use BERT for encoding text and Inception network for encoding image.



Our VAE based architecture

- We use a variational autoencoder as our generative model.
- we use BERT for encoding texts and Inception version3 for encoding images.



Dataset & Examples

- **Fakeddit** is a dataset collected from Reddit, a social news and discussion website. Fakeddit consists of over 1 million submissions from 22 different subreddits posted from 3/19/2008 to10/24/2019 by over 300,000 users.

Image is taken from: Nakamura, K., Levy, S., & Wang, W.Y. (2020). Fakeddit: A New Multimodal Benchmark Dataset for Fine-grained Fake News Detection. *LREC*.

Preliminary Results

Model	Training size	Test Size	Accuracy
Concatenation-based	57689	49388	**0.85**
Multimodal VAE	57689	49388	0.70

our survey on multimodal misinformation detection.

This project is supported by National Science Foundation under Grant number 2127309 to the Computing Research Associate for the CIFellows. This work used the Hive cluster, which is supported by the National Science Foundation under grant number 1828187. This research was supported in part through research cyber infrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA.