

Paper Title : MVAE: Multimodal Variational Autoencoder for Fake News Detection, *Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, Vasudeva Varma, The Web Conference-2019 (The Web Conference-2019 2019)*

MVAE: Multimodal Variational Autoencoder for Fake News Detection

In the paper “**MVAE: Multimodal Variational Autoencoder for Fake News Detection**” the authors have proposed a more robust model to detect fake news. The author proposed a multimodal variational autoencoder which learns shared representations (both visual and textual). Basically it consists of a **encoder, decoder and a fake news classifier** module. The MVAE outperforms the current baseline models (VQA, EANN, att-RNN, mainly single modality) because of discovering correlation between the modalities. The paper mainly concentrates on dataset taken from Twitter and Weibo. For textual data it is pre-processed (using Word2vec) then fed into **LSTM** and for the visual data it is fed into VCGNet architecture. Then finally with the help of fake news detector it is classified.

We can extend the paper in the following ways:

1. We can use **attention based multimodal feature extractor** which is a **self-attention-based encoder** to get deeper representations of text and image, and a co-attention module to fully extract cross-modal complementary information. The Bi-LSTM layer mentioned in the paper **can be combined with Attention**s to improve model's interpretability and accuracy as Attention ignores irrelevant text or images and focus on the contextual thing only.
2. The conclusion may fail to address the heterogeneity gap that occurs in multimodal data. To overcome this we can use **Graph based models** which bridges the gap between multimodalities (un-weighted and bidirectional graphs)
 - The entities in the news and images can be represented as nodes and the relationship between them as edges. This graph way model can interpret the relation and fakeness in a more intuitive way than a traditional vector based model..
3. If the Autoencoder is **fine-tuned with Reinforcement learning** then its performance can be improved in addition to its accuracy. An additional feedback signal can be fed into the training process. In this the Autoencoder can be treated as the **agent** and the signal can be provided by a human classifier.
4. The MVAE can be trained to perform other tasks like **sentiment analysis** in addition to fake news detection to make it more comprehensive.
5. Attention mechanisms can be inculcated to improve its **explainability** as they allow the model to focus on specific parts (text, image) which causes the decision making. (Explainable AI). Using various model **explainability algorithms**.
6. **Adversarial training** can be incorporated to make it more resilient and robust. Adversarial training is basically feeding in misleading examples to prevent manipulation of input data. Also Images are very prone to adversarial attacks hence the training will improve its recognizability and robustness. Some methods like:-
 - **FGSM** (Fast gradient sign method) which adds small perturbations in the direction of the model's gradient.
 - **PGD** (Projected gradient Descent).
 - **Using Adversarial autoencoders** to improve robustness.
7. **Including Audio/Video** can help the users comprehend the fakeness easily. Audio speech analysis can be used to detect the signs of deception from the video. Other Computer vision techniques like **object detection, image segmentation** can be used to infer findings.

8.As a matter of future usability of the model,we can integrate the MVAE model with **real-time systems** to detect fakeness in social media,news on a real time basis.With a feedback loop the model keeps fine-tuning and getting updated.