IRE Scope Document

GROUP-13
Mentor: Vijayasaradhi Indurthi

GROUP MEMBERS

1. Pinkesh Badjatiya (201402002)
2. Anupam Pandey (20162118)
3. Nakul Vaidya (201501108)

Problem Statement

Figuring out the weirdness score and ranking weird or odd news stories.

Applications

The weird news is one of the most catchy categories of news which tempts the user to read the content. Many Internet media and news companies use it as a tool to attract users like scoopwhoop, RVCJ etc to boost hits on their pages/sites and to increase the user engagement. This project will provide users a platform where they can decide which kind of news they want to read and help them to not to get carried away with such weird news and waste their time.

Challenges:

- **Weirdness score Metric**
  - Figuring out if a news is Weird or not is relatively easier as it involves learning a classifier, while quantifying the weirdness of a news is a more challenging task. Analyzing the problem and figuring out a way to quantize the problem and its tractability is a major challenge of this project.
- **Ranking (absolute or relative)**
  - Figuring out if a news is weird or not is comparatively easier than ranking the news articles based on the weirdness. Figuring out a way to be able to rank a given set of news is challenge. We aim at working on figuring out a way to rank news articles in this project.
- **Dataset**

- The current biggest dataset consists of ~73K positive samples along with ~75K negative samples. The dataset is good enough for the classification task but it does not contain information about the relative weirdness of a news. To be able to test the metric or a proposed ranking of the news based on weirdness, we need to measure its accuracy based on a gold standard dataset.

2nd Deliverables Details:

Weird News Classification:

- Classification of the news i.e. whether it is weird or not.
- Comparison with existing Machine Learning techniques to provide a survey of the performance of various models for this task.
- Providing a novel Deep Learning architecture for this classification task.

3rd Deliverables Details:

The Ranking Task:

- Ranking of the weird news as per its weirdness score.
- Providing a way to rank a set of news based on their weirdness scores.
- Provide a metric that can be used to quantify the weirdness of a news.

Dataset for ranking evaluation:

- The current datasets do not provide a way to evaluate the ranking proposed by any algorithm. We plan to develop a good quality dataset that can be used as a gold standard for evaluation.

Tools to be used

1. Python
2. NLTK library for NLP
3. Tenserflow/Keras/pytorch for neural networks
4. Keras & sklearn for models

References

1. www.upi.com/Odd News/
2. www.oditycentral.com/
3. https://github.com/IREL-IIITH/IREL-Reading-Group