Spoiler Detection Using Machine Learning

Atharv Kulkarni[1], Malhar Lohar[2], Manav Ahuja[3], Atharva Shastri[4], Prof. Yogesh Handge[5]

UG Student, Department of Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune [1,2,3,4]

Professor, Department of Computer Engineering, SCTR's Pune Institute of Computer Technology, Pune[5]

Abstract: *Over the course of the lockdown, television shows, web series, and movies have been viewed more than ever before. In order to pick a show or movie to start, we always need to browse through several reviews as it feels like a significant investment of our time, and we want to be sure that they will be genuinely entertaining and gauging. However, these reviews often contain information that reveal things about the plot that a viewer should not know prior to watch the respective show or movie. These bits of information are popularly known as spoilers and they possess the potential of greatly impacting a viewer's experience. These viewers may lose interest and in turn the production companies will to suffer from a loss of revenue. The issue at hand is that we want to read reviews before starting a movie or show, but we do not want to read a spoiler. How can we be sure that a review does not contain a spoiler? We need some warning, and that is why we are attempting to build a machine learning model that uses Natural Language Processing (NLP) to predict whether or not a particular review contains a spoiler, hence, serves as a warning. We will be using an IMDB review dataset to train and test our model.*

Keywords: Machine Learning, Neural Network, Deep Learning, LSTM, Natural Language Processing, etc.

REFERENCES

[1] Sreejita Biswas and Goutam Chakraborty. Movie reviews: To read or not to read! Spoiler detection with applied machine learning. In SAS GLOBAL FORUM. SAS, 2020.

[2] Hwanjo Yu Sungho Jeon, Sungchul Kim. Spoiler detection in tv program tweets. In Proceedings of the International AAAI Conference on Web and Social Media, pages 220–235. Information Sciences, 2016.

[3] Saarthak Sangamnerkar Allen Bao, Marshall Ho. Spoiler alert: Using natural language processing to detect spoilers in book reviews. https://arxiv.org/ pdf/2102.03882.pdf, 2021.

[4] IMDB Spoiler Dataset - https://www.kaggle.com/rmisra/imdb-spoiler-dataset

[5] GloVe: Global Vectors for Word Representation, https://nlp.stanford.edu/projects/glove/

[6] Hochreiter, J. Schmidhuber. (1997). "Long Short-Term Memory", Neural Computation, 9(8):1735-1780.

[7] Has¸im Sak, Andrew Senior, and Franc¸oise Beaufays, "Long shortterm memory based recurrent neural network architectures for large vocabulary speech recognition," arXiv preprint arXiv:1402.1128, 2014.

[8] Mike Schuster and Kuldip K Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, 1997. [9] Alex Graves, Navd.

[9] Xiangang Li and Xihong Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on. IEEE, 2015, pp. 4520–4524.

[10] Felix A Gers, Nicol N Schraudolph, and Jurgen Schmidhuber, "Learning ¨ precise timing with LSTM recurrent networks," The Journal of Machine Learning Research, vol. 3, pp. 115–143, 2003.