Sentiment Analysis with large movie review data

Fawad Kirmani

Background:

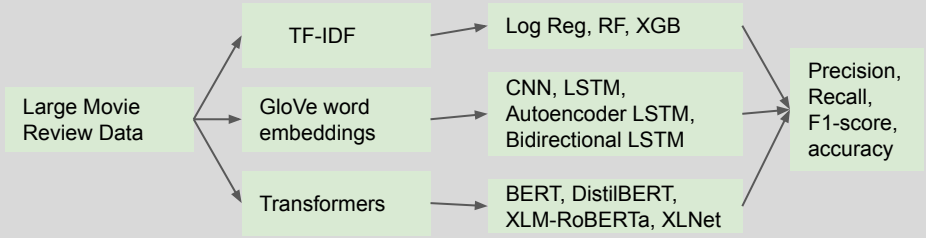
As more and more products are getting sold online. Product reviews are an essential part of the purchase process. The project aim is to learn the sentiment behind the reviews of the products using natural language processing. The main goal of the project was to:

- Learn and explore traditional and state of the art sentiment analysis techniques.
- Achieve equal or better results than Andrew Maas et al. (2011) with traditional and state of the art methods.

In the original work on same dataset, Andrew Maas et al. (2011) achieved accuracy of 89%

Problem: How to do sentiment analysis/classification of movie reviews data with different approaches? How well different approaches perform compared to Andrew Maas et al. (2011) ?

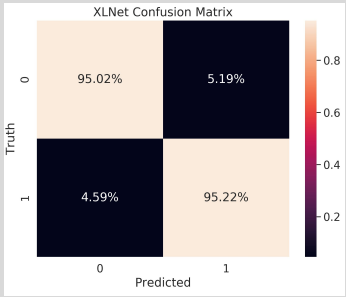
Methods	Models	Precision	Recall	F1-score	Accuracy
TF-IDF	LogReg*	83%	83%	83%	83%
	RF	84%	84%	84%	84%
	XGB	82%	82%	82%	82%
GloVe word embeddings	CNN	90%	90%	90%	90%
	LSTM	89%	89%	89%	89%
	Autoencoder LSTM	89%	89%	89%	89%
	Bi-directional LSTM	89%	89%	89%	89%
Transformers	BERT	93%	93%	93%	93%
	DitilBERT	94%	94%	94%	94%
	XLM-RoBERTa	94%	94%	94%	94%
	XLNet	95%	95%	95%	95%



Approach:

- Machine Learning models with TF-IDF tokenizer
- Deep Learning models with GloVe Word-Embeddings
- Classification with Transformers based Pretrained language models

Best Result



Evaluation:

- The goal to learn and implement traditional and state of the art sentiment classification methods have been achieved.
- Achieved equal or better results compared to Andrew Maas et al. (2011) with GloVe word-embeddings and transformers based pretrained models.
- Didn't achieved good results with TF-IDF vectorization.

Discussion:

- **Achieved:** Best accuracy of 95% with XLNet. Original XLNet paper achieved accuracy of 96%
- **Challenges faced:** Hyperparameter tuning for transformer based models because of GPU resources and time.
- **Future work:**
  - Exploring more on hyperparameters of transformers.
  - Exploring state of the art methods not based on transformers

Repo: https://github.com/fkirmani/csce771/tree/main/project-home