Cross domain sentiment analysis with LSTM using Keras in Tensorflow

Guide - Prof. Wencen Wu

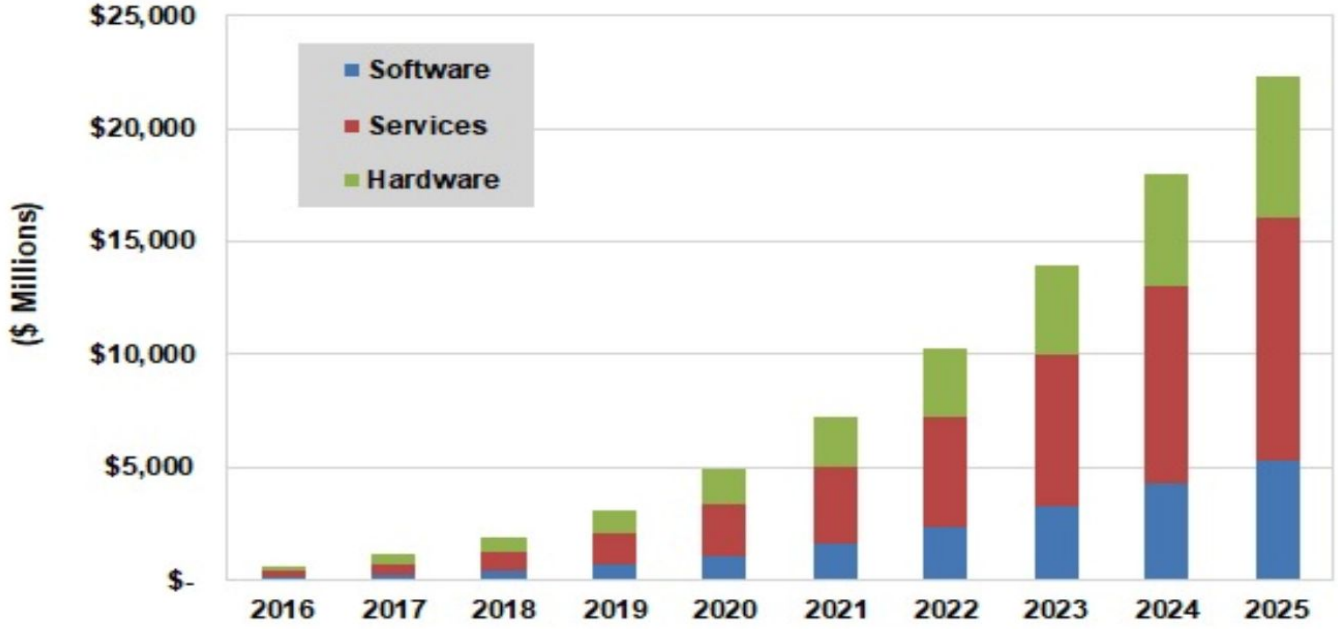
Team - Ameya Nawale, Anurag Shinde, Hanisha Thirtham, Venkatesh Devale

MOTIVATION

- Today's popular sentiment analysis systems are more inclined towards a specific domain for example, only for movies or play store or app store or food and so on.
- To develop a generalized system that works for all the domains will need huge infrastructure and domain experts which requires a lot of investment.

Tractica

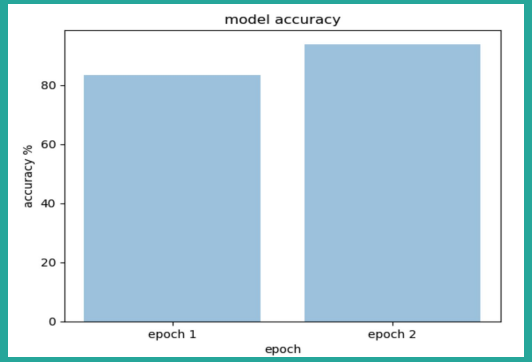
Natural Language Processing Total Revenue by Segment, World Markets: 2016-2025



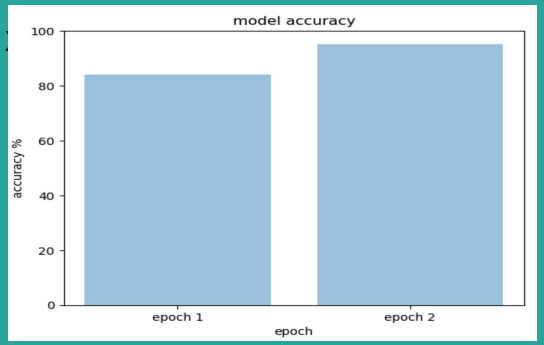
Source: Tractica

Models

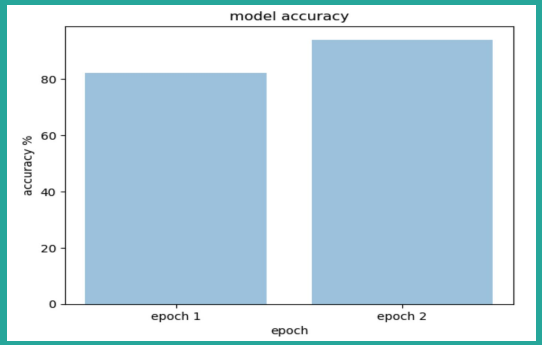
- LSTM + Dense
- Dropout + LSTM + Dropout + Dense
- LSTM + Recurrent Dropout + Dense
- Convolution 1D + Max Poll + LSTM + Dense
- Convolution 1D + Max Poll + Flatten + Dense + Dense



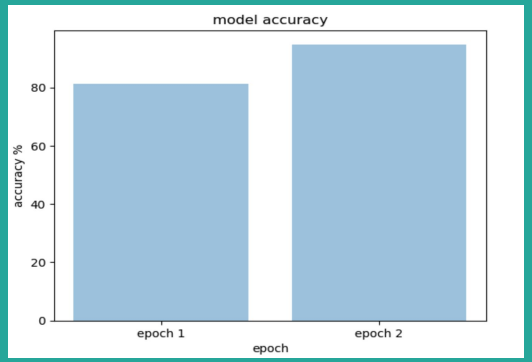
LSTM + Dense (Accuracy = 93.85%, Compute Time= 256 sec)



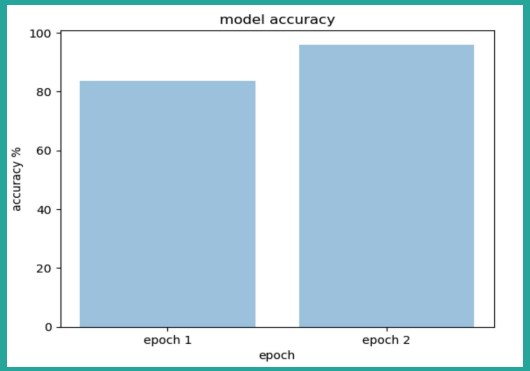
Dropout + LSTM + Dropout + Dense (Accuracy = 95.23%, Compute Time= 260 sec)



LSTM + Recurrent Dropout + Dense (Accuracy = 94.04 % Compute Time= 55ms)



Convolution 1D + Max Poll + LSTM + Dense(Accuracy = 94.75%, Compute Time = 131 sec)

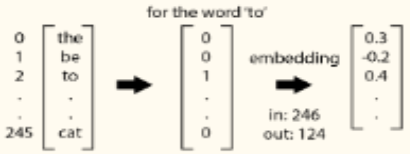


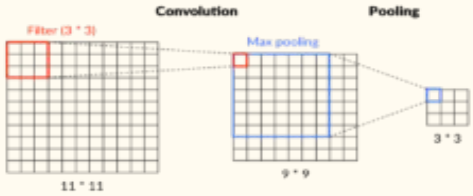
Convolution 1D + Max Poll + Dense + Dense (Accuracy = 95.95%, Compute Time = 36 sec)

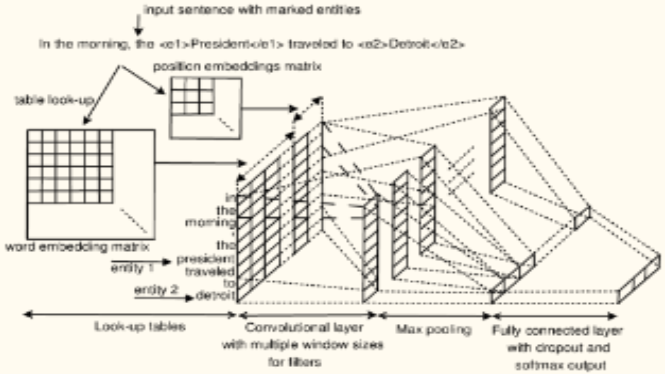
Technologies used

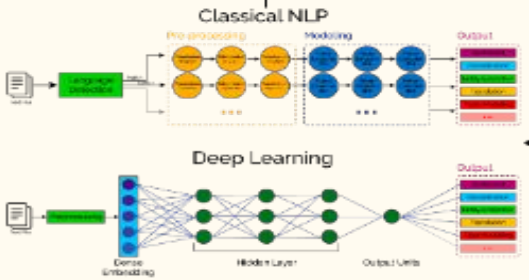
- Python
- Keras
- Tensorflow
- Python Flask
- React

Application Architecture



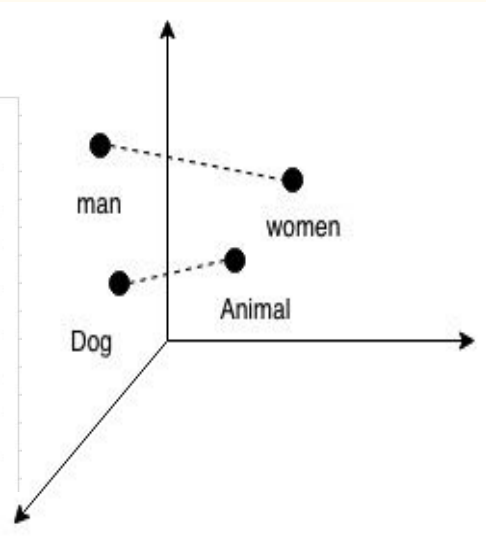


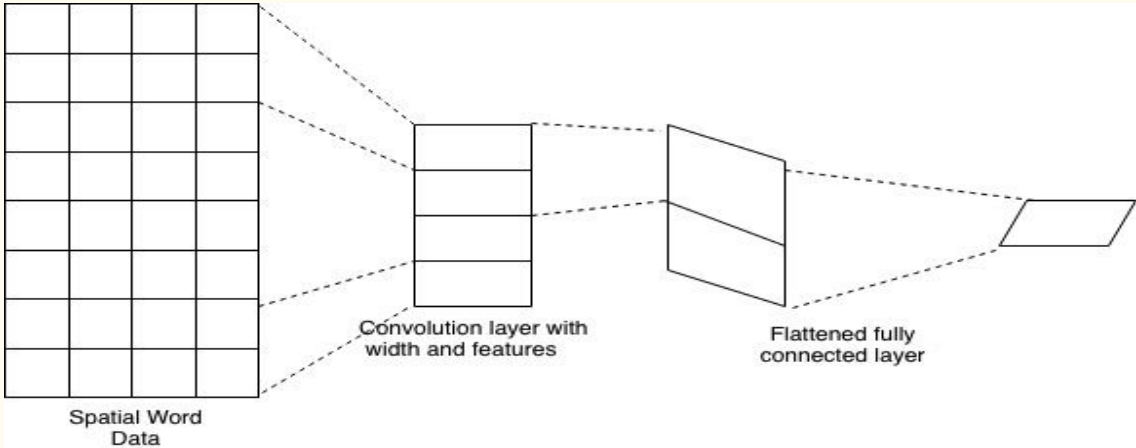




Steps

1. Data collection and preprocessing
2. Forming tokenizer using tensor flow.
3. Forming Word2Vector
4. Forming the model using a combination of the previously mentioned models. (Convolution Neural Network, LSTM)





Demo and the results

```
"list":["This is not a good book",
        "Mr. X is not that bad as a employee.",
        "Nicely written code, found very few bugs.",
        "His articles are difficult to understand and many find it boring.But the concepts are
            important and I will highly recommend this professor",
        "This is not the best book i have read. I will say it is not bad either"
        ]
"final_output": [
    "negative sentiment",
    "Neutral with more positive sense",
    "positive sentiment",
    "positive sentiment",
    "Neutral"
]
```

Conclusion

Thus startups of small scale to medium scale can use this generalized model to give sentiment insights for different domains such as restaurants, health, movie reviews, etc.,. This will have great impact on small organisations by reducing the infrastructure needs and maintenance costs and hence helping them in expanding their business.

Major learning points:

- We need a lot of data and computational power to build a great machine learning model.
- Data preprocessing, features extraction and deciding architecture is comparatively difficult than selecting the machine learning model itself.

References

[1] Jiajun Cheng, Pei Li, Zhaoyun Ding, Sheng Zhang, Hui Wang, "Sentiment Classification of Chinese Microblogging Texts with Global RNN" in First International Conference on Data Science in Cyberspace, IEEE, 2016 2.

[2] Wang, J.; Yu, L.C.; Lai, K.R.; Zhang, X. Dimensional Sentiment Analysis Using a Regional CNN LSTM Model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; pp. 225–230 3.

[3] Sak, Haşim, Andrew Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." Fifteenth annual conference of the international speech communication association. 2014

Team Contribution

Ameya Nawale
- Frontend, Restaurant Data preprocessing
- Model - Dropout + LSTM + Dropout + Dense

Anurag Shinde
- Backend, Amazon Data preprocessing
- Model - LSTM + Recurrent Dropout + Dense

Hanisha Thirtham
- Frontend, Google App Data Preprocessing
- Model - Convolution 1D + Max Poll + LSTM + Dense

Venkatesh Devale
- Backend, Creating tokenizer after concatenating the processed data
- Model- Convolution 1D + Max Poll + Flatten + Dense + Dense