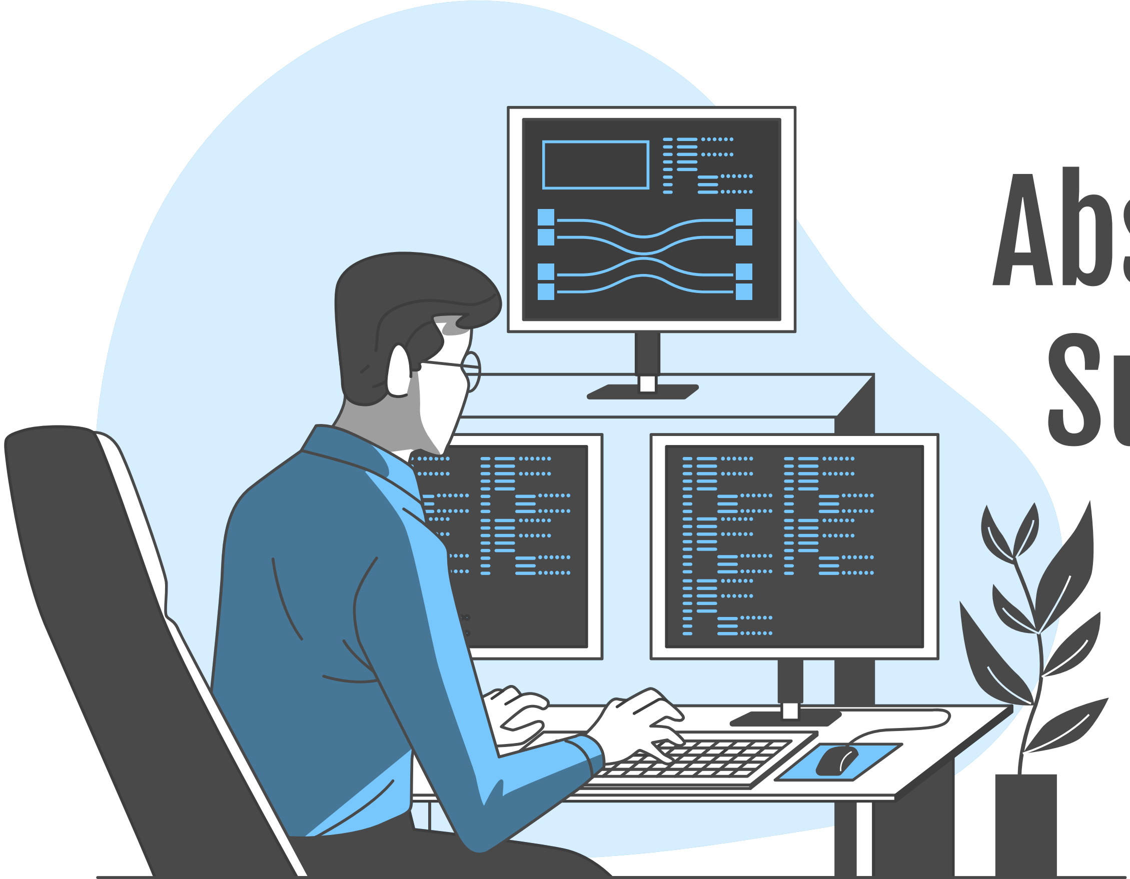
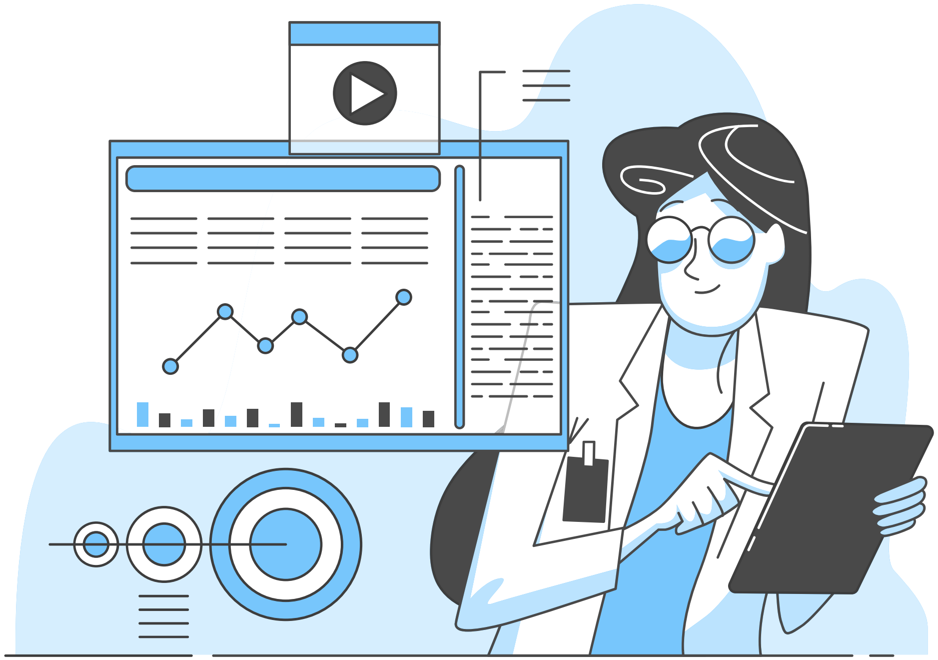


Abstractive Text Summarization

Dareen Hussein
Israa Fahmy

Table of Contents



01

Proposal summary

- Problem statement
- Challenges
- Our application

Problem statement

In the big data era, there has been an explosion in the amount of text data from a variety of sources. This volume of text needs to be effectively summarized to be useful. There are important applications for text summarization in various NLP related tasks.

Challenges

Hard Abstractive

It has complex capabilities incorporating real-world knowledge.

Compared to Human

Human performance is far better than the text summarizer when compared side-by-side.

Fine Tuning

Highly unstable process, with many hyperparameters settings

Our application

Customer reviews can often be long and descriptive. Analyzing these reviews manually, is really time-consuming. This is where we can apply Natural Language Processing to generate a summary for long reviews.

02

Related work

- Previous solutions
- Comparative analysis

Previous Solution

TextRank

Extractive and unsupervised text summarization technique.

...

BERT Extractive Summarizer

Utilizes text embeddings and KMeans clustering to identify sentences.

...

Generating Headlines with RNN

Encoder-decoder RNN with LSTM units and attention to generating headlines.

...

Comparative Analysis

TextRank	BERT Extractive Summarizer	Generating Headlines with RNN
Extractive Summarizer	Extractive Summarizer	Abstractive Summarizer
Uses Cosine Similarity to compute the similarity between sentences.	Calculate ELBOW to determine the optimal cluster.	Used the training and holdout loss. And the BLEU performance.
GloVe word embeddings as vector representation of words.	Utilizes the BERT model for text embeddings.	An embedding layer that transforms the word into a distributed representation

03

Our progress

- Proposed solution
- Model Overview
- Dataset
- Progress Report
- Issues encountered

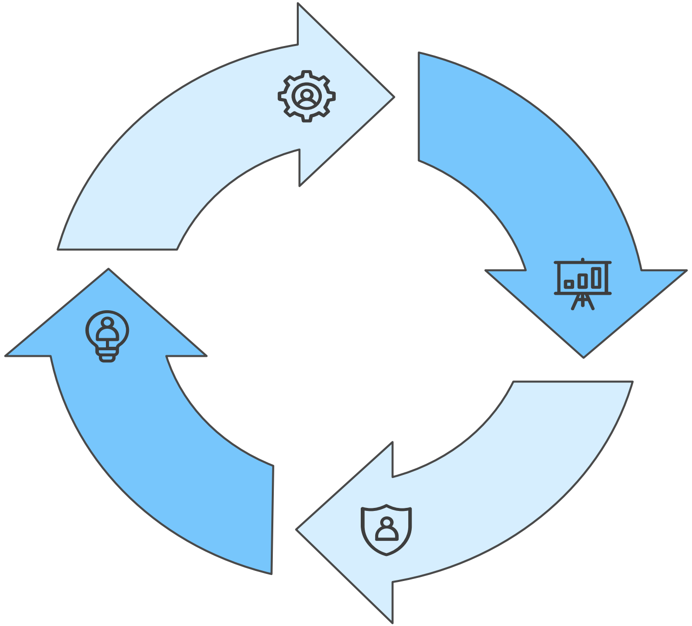
Proposed Solution

Abstractive Summarizer

Generate new sentences from the original text.

Seq2Seq

The input is a sequence of words and the output is a sequence of tags for every word in the input sequence



Encoder-Decoder Architecture

An Encoder LSTM reads the entire input sequence. The decoder is trained to predict the next word in the sequence given the previous word.

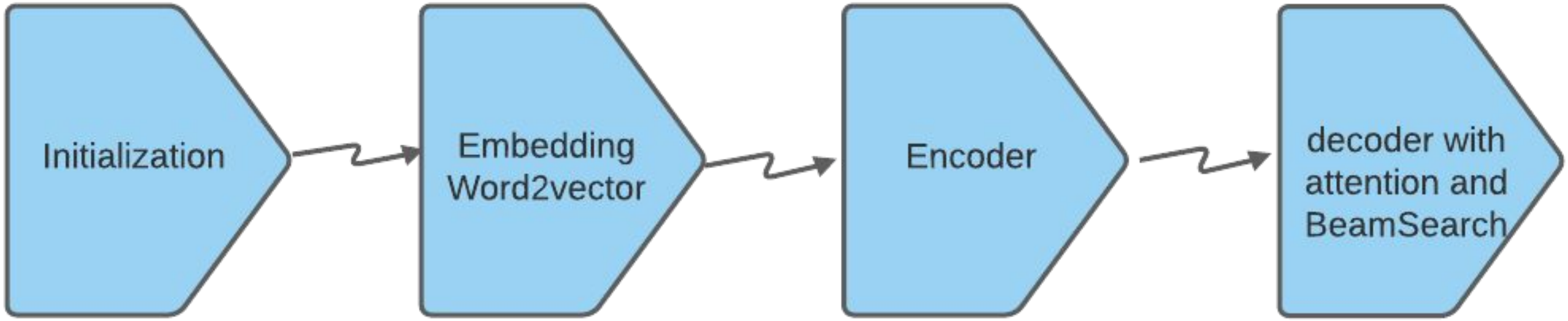
Attention Mechanism

It aims to predict a word by looking at a few specific parts of the sequence only, rather than the entire sequence.

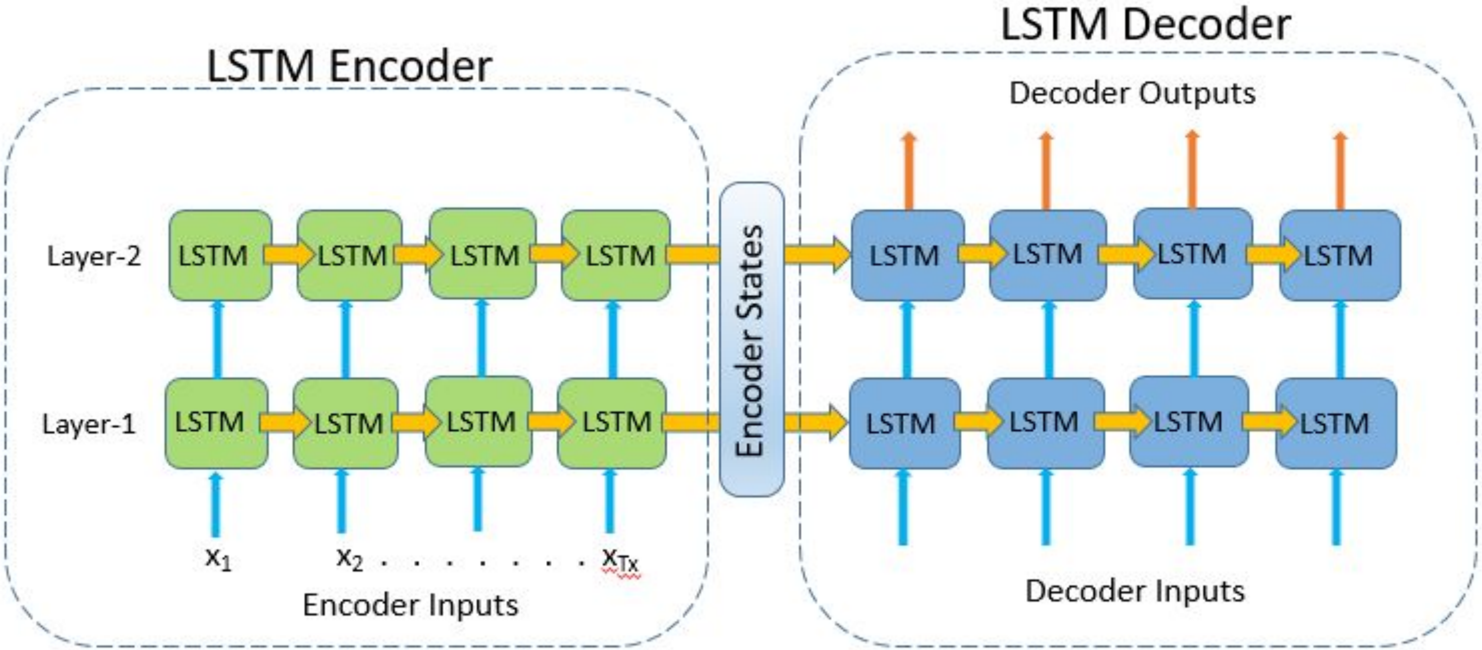
Greedy Vs Beam Search

Beam search saves computational powers and performs well..

Model Architecture



Encoder-decoder



Dataset

Name: Amazon Fine Food Reviews

Description: This dataset consists of reviews of fine foods from amazon.

Size: 642.49 MB

Data includes: 568,454 reviews

Progress report

1. Compared all different previous solutions and decided to deploy best 2

2. We setted up the environment needed for each and solved the bugs found

3. Changed parameters and compared accuracies

4. Added additional layers to increase accuracy and attached attention layer.

Issues encountered

01 Encode-decoder limitations

Solved by importing attention.

02 Development environment

Keras does not officially support attention layer

03 Training time

We couldn’t use the entire training set.
We added early stopping.

Issues encountered (Cont.)

04

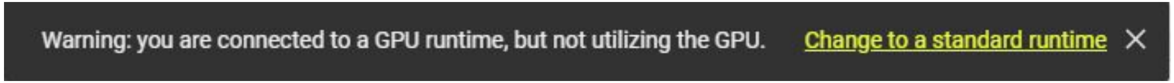
RAM Utilization

Faced a problem loading the entire dataset into the RAM. Besides the lack of GPU usage.

05

Technical problems

Disconnection during session. Beside the need to attach the attention layer and dataset every time.



04

Results

- Experiments summary
- Output samples

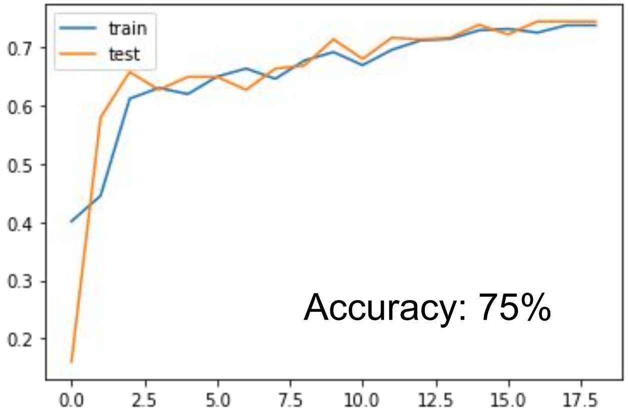
Hyperparameters Experiments

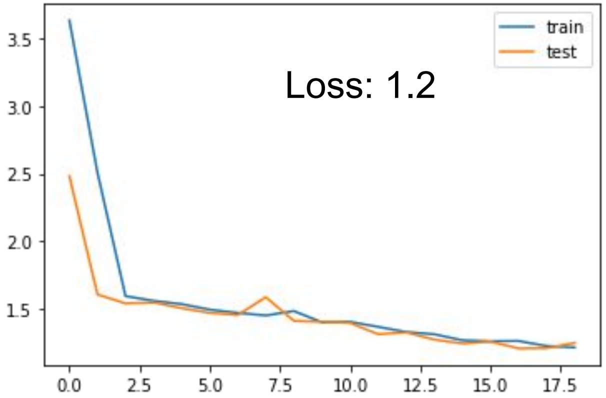
Parameter	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Dropout	0.4	0.4	0.4	0.3
Activation	softmax	softmax	softmax	softmax
L1	1e-5	1e-5	1e-2	1e-2
L2	1e-4	1e-4	1e-2	1e-2
Optimizer	RmsProp	Adam	Adam	rmsprop
Loss function	Sparse categorical cross entropy	Sparse categorical cross entropy	Sparse categorical cross entropy	Sparse categorical cross entropy

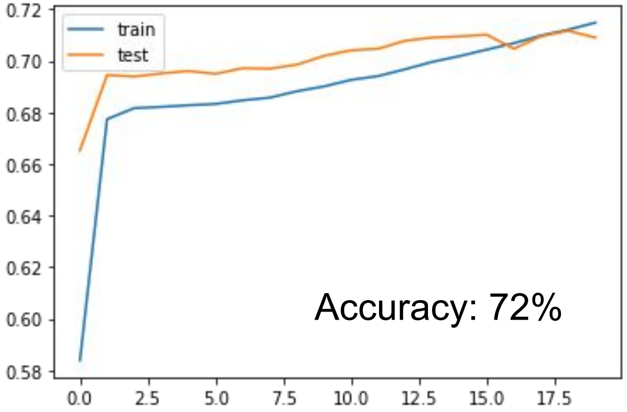
Hyperparameters Experiments (Cont.)

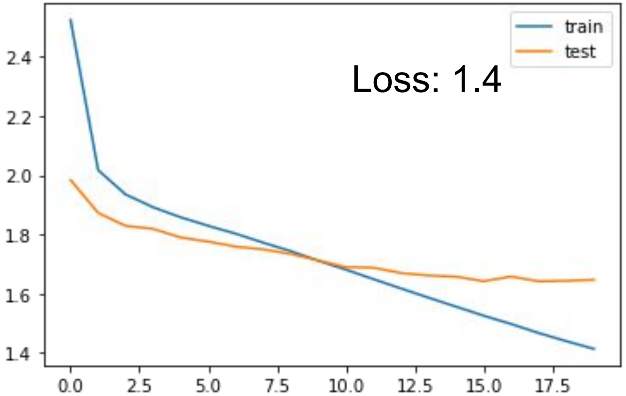
Parameter	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Stacked LSTM layers	1	3	3	5
Loss	2.79	1.99	1.6	1.2
Accuray	0.43	0.57	0.7.09	0.745

Best Result Graphs









Sample Output

Review: delicious coffee ever mess stomach amazing tastes wonderful everyone work loves
Original summary: delicious
Predicted summary: great coffee

Review: purchased larger size love size perfect keep purse snack especially times others dessert snack cannot eat must gluten free spouse touch diet food loves
Original summary: cannot get enough
Predicted summary: great snack

Review: many kit wines cost three four times made many kits find fine table wine recommend adding water five gallon mark flavor
Original summary: good wine
Predicted summary: great product

Review: give little background ordered great deal box single serving chips thought would give try variety flavors package great enjoy chips leave lot less mess average bags potato chips crumbs bag prepared like normal chips result differently chip personal taste whether like taste texture would highly recommend trying
Original summary: excellent chips
Predicted summary: love these chips

Future work & Timeline

MS1

Compiled/trained/ tested our base models code and get initial results.

Try different **word embeddings** (bert) & Implement pointer generator network & hopefully try reinforecemnt learning methods.

MS2

- Use the **beam search strategy** for decoding the test sequence instead of using the greedy approach (argmax)

- Evaluate the model performance using **BLUE** and **ROUGE** scores.

Final

- Try **different customized attention methods** instead of importing third party layer.

Members Contribution

01

Israa

- Added attention layer to base code.
- Deployed base model and changed parameters.
- Changed stacked LSTM layers.

02

Dareen

- Loaded and checked different base line codes.
- Deployed base model and changed parameters.
- Downloaded Dataset.

Resources

- Ilya Sutskever, Oriol Vinyals, and Quoc Le. 2014. Sequence to sequence learning with neural networks. In *NIPS*.
- Vishal Gupta and Gurpreet Singh Lehal. 2010. A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence* , 2(3):258–268.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. In *arXiv preprint arXiv:1609.08144* .
- Ian Good fellow, Aaron Courville ,and Yoshua Bengio. Deep learning. Book in preparation for MIT Press, 2015.
- Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. CoRR, abs/1506.03099, 2015.
- William Chan, Navdeep Jaitly, Quoc V. Le, and Oriol Vinyals. Listen, attend and spell. CoRR, abs/1508.01211, 2015.

Resources (Cont.)

- Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. CoRR, abs/1409.2329, 2014.
- Andrej Karpathy and Fei-Fei Li. Deep visual-semantic alignments for generating image descriptions. CoRR, abs/1412.2306, 2014.
- Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5 – rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL ’02, pages 311–318, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- Aravind PaiAravind is a sports fanatic. His passion lies in developing data-driven products for the sports domain. He strongly believes that analytics in sports can be a game-changer. (2020, May 10). Text summarization: Text summarization using deep learning. Retrieved March 30, 2021, from https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/


Thank you!
Dareen & Israa