

Exploring deep learning for Sentiment Analysis of user generated reviews

CSML1010 Individual Assignment

GOURI KULKARNI

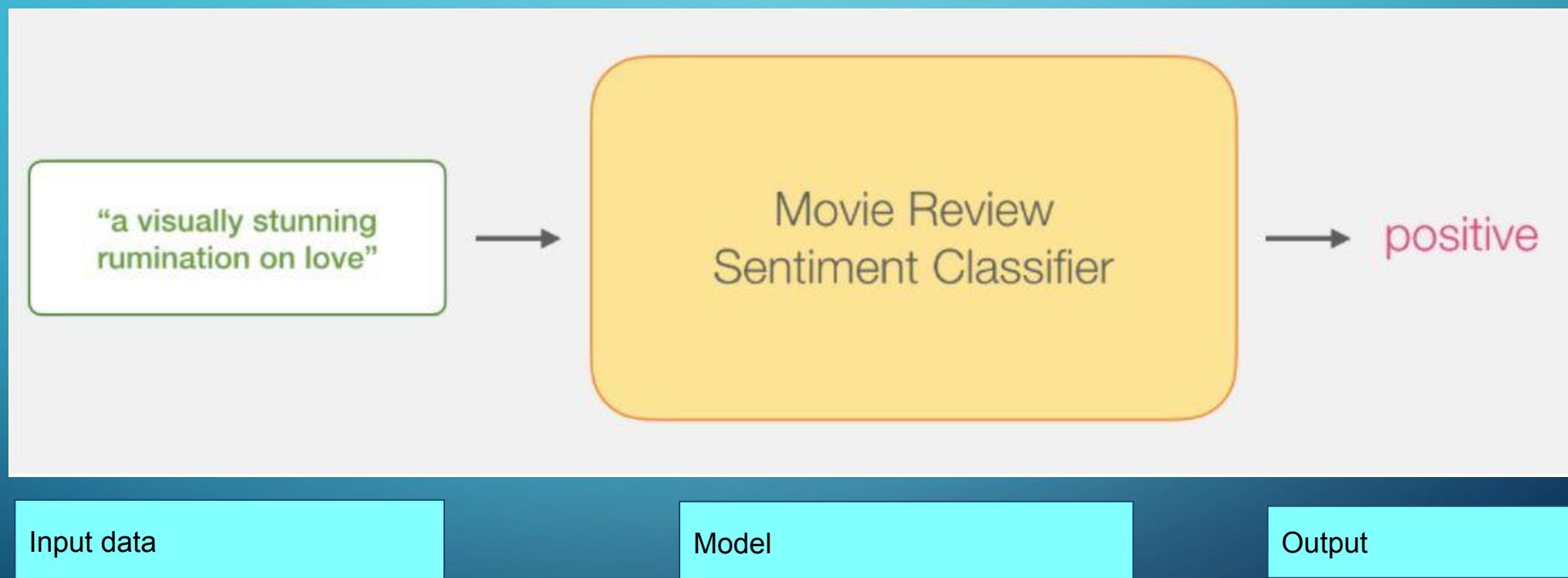
From using traditional bag of words and scikit-learn to using Transfer Learning to solve the problem of classifying user generated reviews

PROBLEM STATEMENT

MOVIE REVIEWS SENTIMENT CLASSIFICATION

- **Example of a movie review**
“A visually stunning rumination on love”
- **To classify the sentiment of Movie Reviews**
- **Take a sentence and produce a 0 or 1**
- **0 being a bad review , 1 being a good review**

Visualizing the Goal



Challenges posed by traditional methods

- Words that have different meaning under different contexts
- Requirement for a large amount of data for training purposes
- Uni directional approach can lead to wrong predictions
- Task of optimizing feature selection is time consuming
- Task of optimizing n-grams task is time consuming
- Cleaning the data can lead to loss of valuable information
- Fine tuning is tedious

Misclassification example : “no good” - a bad review classified as a good review

Deep learning techniques to solve NLP tasks

Taking our simple task of classifying reviews as “ good “ or “bad” reviews

Two step process :

- use a language model trained on a large unlabelled corpus
- fine tune the model for NLP tasks

Two approaches :

- BERT DistilBERT pre-trained model
- CNN deeplearning

BERT for NLP

- BERT stands for Bidirectional Encoder Representations from Transformers
- Developed by Google
- BERT is a pre-trained model which has already learnt from a massive corpus
- BERT is bidirectional - learns information from the token's left and right during training phase
- Fine-tuning can be achieved quickly by adding layers

BERT for NLP - evolution

Pretraining models on large unlabelled data started with Word2Vec and GloVE

They captured contextual representations between words like “bank”

Using embeddings gave better predictions but the models were shallow

ELMo solved the problem of Polysemy - same word having different meanings

ULMfit trained models that could be fine tuned to provide excellent results with less data

BERT = Transfer Learning = Pre-Training + Fine Tuning

BERT for NLP - libraries, models

Released in 2018

PyTorch or TensorFlow libraries

Pretrained models developed

and open sourced by the team at HuggingFace

Bertweet

BertGeneration

Blenderbot

Blenderbot Small

CamemBERT

CTRL

DeBERTa

DialoGPT

DistilBERT

DPR

ELECTRA

Encoder Decoder Models

FlauBERT

FSMT

Funnel Transformer

herBERT

LayoutLM

LED

Longformer

LXMERT

MarianMT

MBart

MobileBERT

MPNet

MT5

OpenAI GPT

OpenAI GPT2

Pegasus

PhoBERT

BertConfig

BertTokenizer

BertTokenizerFast

Bert specific outputs

BertModel

BertForPreTraining

BertModelLMHeadModel

BertForMaskedLM

BertForNextSentencePrediction

BertForSequenceClassification

BertForMultipleChoice

BertForTokenClassification

BertForQuestionAnswering

TFBertModel

TFBertForPreTraining

TFBertModelLMHeadModel

TFBertForMaskedLM

TFBertForNextSentencePrediction

TFBertForSequenceClassification

TFBertForMultipleChoice

TFBertForTokenClassification

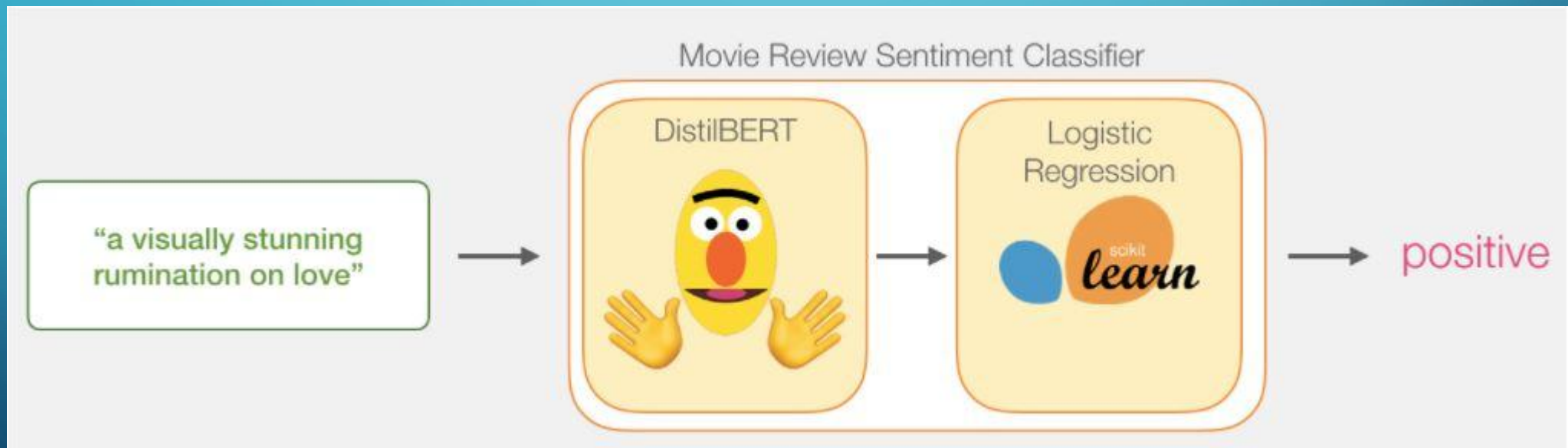
TFBertForQuestionAnswering

FlaxBertModel

FlaxBertForMaskedLM

BERT for NLP

Utilizing DistilBERT and sklearn Classifier/s to classify reviews



TensorFlow for NLP

TensorFlowHub

Hello. Welcome to TensorFlow Hub.

The TensorFlow Hub lets you search and discover hundreds of trained, ready-to-deploy machine learning models in one place.

TensorFlow Hub is a repository for machine learning models.

From image classification, text embeddings, audio, and video action recognition, TensorFlow Hub is a space where you can browse trained models and datasets from across the TensorFlow ecosystem. Use it to:

1

Find trained models for transfer learning to save time on training

2

Publish your own models

3

Deploy models on device and in the browser

Evolution of TensorFlow

The Tensorflow library uses Keras as the high level library

Focuses on deep learning neural networks unlike scikit learn which focuses on traditional methods

Developed by Google for internal use

Released in 2015

Demonstration :

Overcoming challenges posed by BOW approach using Transfer learning

- 1: BERT Huggingface , scikit-learn model
- 2: CNN model , BERT Tensorflow

<https://colab.research.google.com/drive/1K4NzJVtwjlsOfOPode-VGTg5hD9RFjQH>

THANK YOU