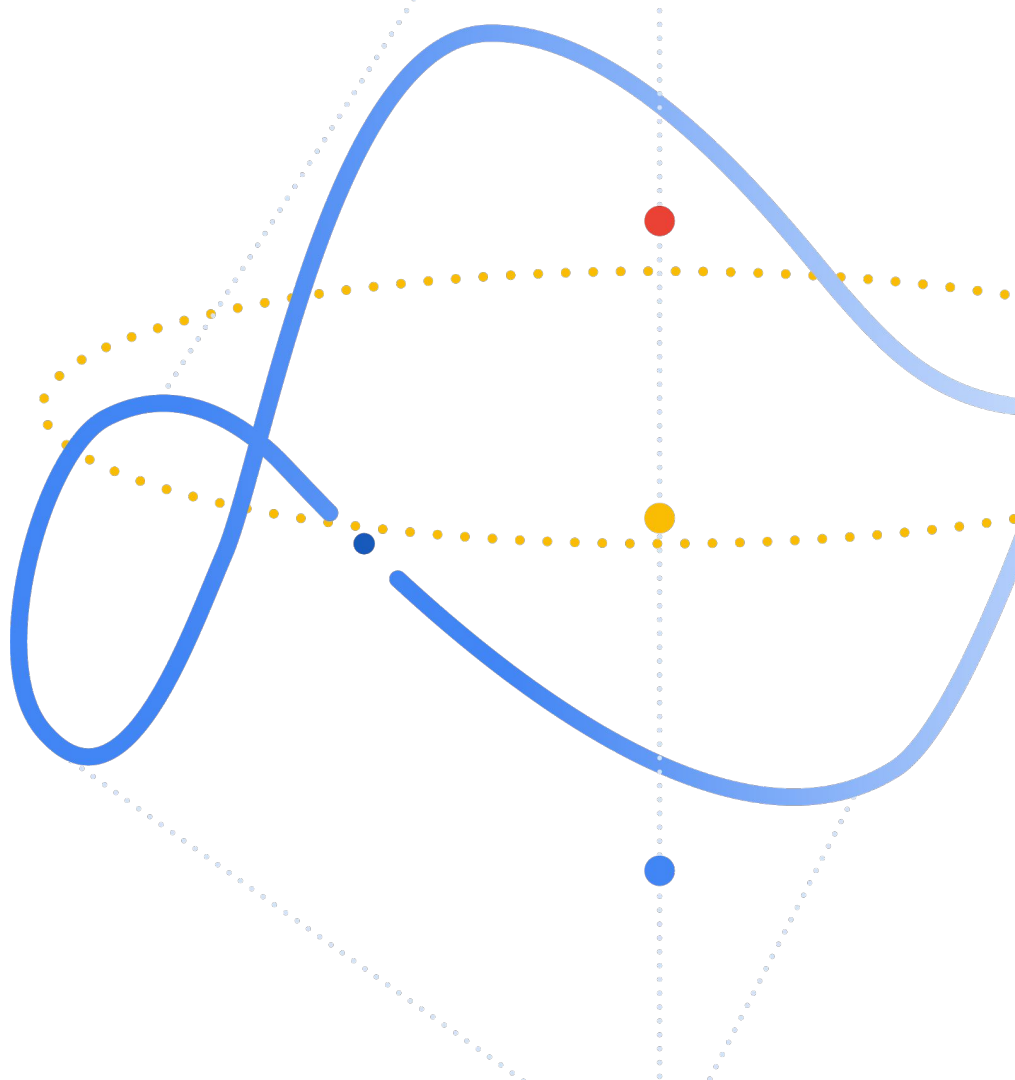


Deep Dive in NLP with Tensorflow 2.x

<https://bit.ly/tfug-presentation>





Abhishek Kumar

Lead Data Science Team @
Publicis Sapient, India

Masters – University of California,
Berkeley

Google Developers Expert –
Machine Learning

Pluralsight Author

Speaker – Strata Data, O'Reilly AI
Conf , ODSC



Menti.com and use code **43 99 65**



Workshop - Plan



What is NLP ?

NLP is an approach to **process**, **analyze** and **understand** large amount of text data.



Top Applications of NLP

- Machine translation
- Speech Recognition
- Question Answering
- Text summarization
- Text Classification



\$26.4B

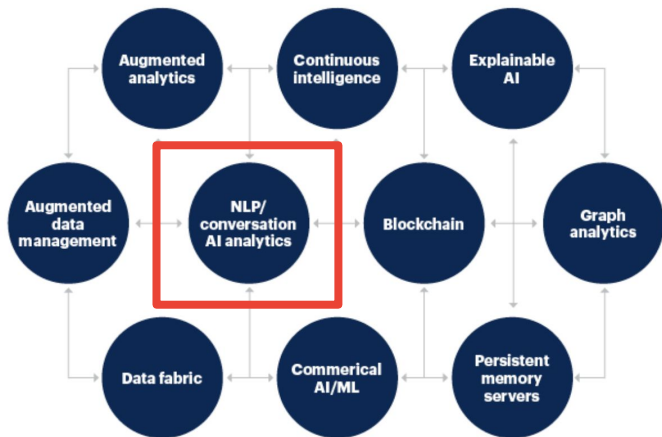
**Global NLP Market by
2024**

80%

**Of global data by 2025 will
be unstructured
- IDC**

<https://openarticlesubmission.com/natural-language-processing-market-projected-to-grow-26-4-billion-by-2024/>
<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Top 10 technology trends in data and analytics



gartner.com/SmarterWithGartner

Source: Gartner
© 2019 Gartner, Inc. All rights reserved. PR 781030

Gartner®



Text Modeling

IMDB Review

A big disappointment for what was touted as an incredible film. Incredibly bad. Very pretentious. It would be nice if just once someone would create a high profile role



Bag-of-words

Individual words are important

IMDB Review

A **big disappointment** for what was touted as an **incredible** film.
Incredibly bad. Very **pretentious**. It would be **nice if** just once someone would create a high profile role

Word	Count	TF-IDF
Big
Disappointment
Incredible
Bad
Pretentious
Nice

Missing Context



Word Embedding

Use corpus context to create dense word representation

IMDB Review

A **big disappointment** for what was touted as an **incredible** film.
Incredibly bad. Very **pretentious**. It would be **nice if** just once someone would create a high profile role

Word	Word2Vec	FastText
Big
Disappointment
Incredible
Bad
Pretentious
Nice

Only Word Level .. But how to model beyond “Word Meaning”



Text as Sequence Modeling through Deep Learning

Modeling spatial relationship

IMDB Review

A big disappointment for what was touted as an incredible film.
Incredibly bad. Very pretentious. It would be nice if just once someone would create a high profile role

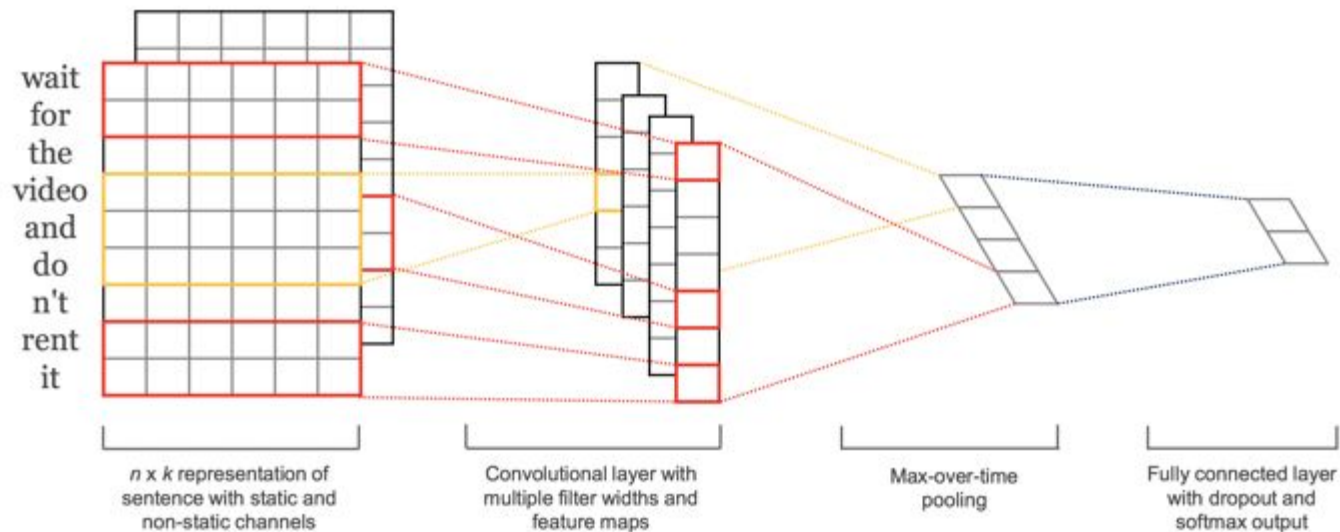
Text as sequence modeling through deep learning

Use spatial patterns

Approaches : CNN, LSTM, Transformers etc.



Convolutional Neural Networks

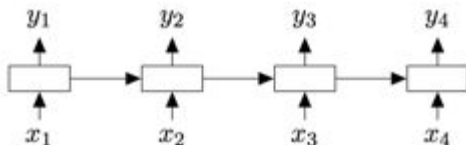


Convolutional neural networks for sentence classification, Yoon Kim (2014)

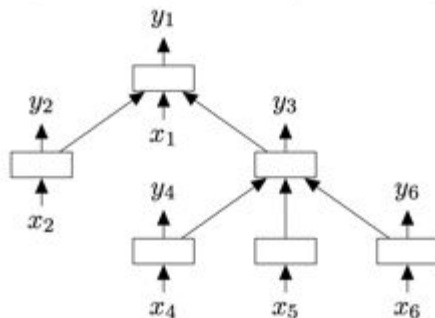


LSTM (Long Short Term Memory)

Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks, Tai et.al 2015



Chain-structured LSTM



Tree-structured LSTM

<https://arxiv.org/pdf/1503.00075.pdf>



O'REILLY[®]

Artificial Intelligence Conference

PRESENTED WITH  intel AI

Industrialized capsule networks for text analytics

Vijay Agneeswaran (Walmart Labs), Abhishek Kumar (Publicis Sapient)

<http://bit.ly/aiconf2019>



And the NLP world exploded

Transformers

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

BERT, RoBERTa, DistilBERT ...

BERT: Pre-training of Deep Bidirectional Transformers for
Language Understanding

Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language
{jacobdevlin, mingweichang, kentonl, kristout}@google.com

Open AI : GPT, GPT-2, GPT-3

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}



HUGGING FACE

On a mission to solve NLP,
one commit at a time.



Workshop - Plan



Demo : Basic NLP with TF2.x

<https://bit.ly/tfug-01>

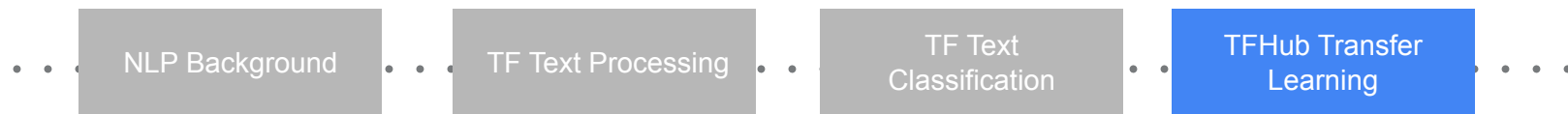


Demo : Text Classification with TF2.x

<https://bit.ly/tfug-02>



Workshop - Plan



Tensorflow Hub

<https://tfhub.dev/>

TensorFlow Hub

Search for models, collections & publishers

Send feedback

Filters

Clear all

Problem domain

Text embedding

Model format

TFjs

TFLite

Coral

TF Version

TF1

TF2

Fine tunable

Architecture

Publisher

Dataset

Language

Text embedding

universal-sentence-encoder-large

Published by: Google Updated: 04/22/2020

Encoder of greater-than-word length text trained on a variety of data.

Transformer

Text embedding

universal-sentence-encoder-multilingual

Published by: Google Updated: 04/20/2020

16 languages (Arabic, Chinese-simplified, Chinese-traditional, English, French, German, Italian, Japanese, Korean, Dutch, Polish, Portuguese, Spanish, Thai, Turkish, Russian) text encoder.

CNN

Text embedding

bert_en_uncased_L-12_H-768_A-12

Published by: TensorFlow Updated: 04/23/2020

Bidirectional Encoder Representations from Transformers (BERT).

Transformer | Wikipedia and BooksCorpus

Text embedding

universal-sentence-encoder

Published by: Google Updated: 06/22/2020

Encoder of greater-than-word length text trained on a variety of data.

DAN

Text embedding

tf2-preview/nnlm-en-dim128

Published by: Google Updated: 04/29/2020

Token based text embedding trained on English Google News 200B corpus.

NNLM | Google News

Text embedding

universal-sentence-encoder-multilingual-...

Published by: Google Updated: 04/20/2020

16 languages (Arabic, Chinese-simplified, Chinese-traditional, English, French, German, Italian, Japanese, Korean, Dutch, Polish, Portuguese, Spanish, Thai, Turkish, Russian) text encoder.

Text embedding

tf2-preview/gnews-swivel-20dim

Published by: Google Updated: 04/29/2020

Token based text embedding trained on English Google News 130GB corpus.

Text embedding

tf2-preview/nnlm-en-dim50

Published by: Google Updated: 04/29/2020

Token based text embedding trained on English Google News 7B corpus.



Demo : NLP with TFHub

<https://bit.ly/tfug-03>





Resources

Browse > Data Science > Machine Learning

This course is part of the **TensorFlow in Practice Specialization**

Natural Language Processing in TensorFlow


★★★★★ 4.6 3,412 ratings | 👍 91%  Share


 Laurence Moroney

Browse > Data Science > Machine Learning

Natural Language Processing Specialization

Break into the NLP space. Master cutting-edge NLP techniques through four hands-on courses!

★★★★★ 4.5 69 ratings  Share

 Younes Bensouda Mourri [+2 more instructors](#)

Text

Word embeddings


Text classification with an RNN

Text generation with an RNN

Neural machine translation with attention

Image captioning

Transformer model for language understanding

Fine tuning BERT 

https://www.tensorflow.org/tutorials/text/word_embeddings



Thank You

