PyData
MUMBAI

UnFound

ELMo Embeddings

Exploring the power of context.

Session flow

- Characterization Problem at hand
- Existing solutions and how they work
- Drawbacks of these solutions
- In comes ELMO
- How does it solve the problem
- Testing on our results
- Applications of ELMO

UnFound

Artificial Intelligence meets News

We want to change the way news is consumed today.

www.unfound.news

AI behind UnFound

- Summarization
- Question Answer Generation
- Quotes Generation
- News Ranking
- Timeline Detection
- Stance Detection

How to deal with Text data ?

Word Embeddings!

One-hot encoding

What if the **size of vocabulary** increases to a million words ?

Do the words **ants** and **toxic** have any resemblance ?

Word Embeddings

Hello → [0.3 0.6 0.1 0.9]

Word embeddings are distributed representations in vector space.

Understanding Embeddings

King → [0.1 0.2 0.7]

Queen → [0.2 0.3 0.5]

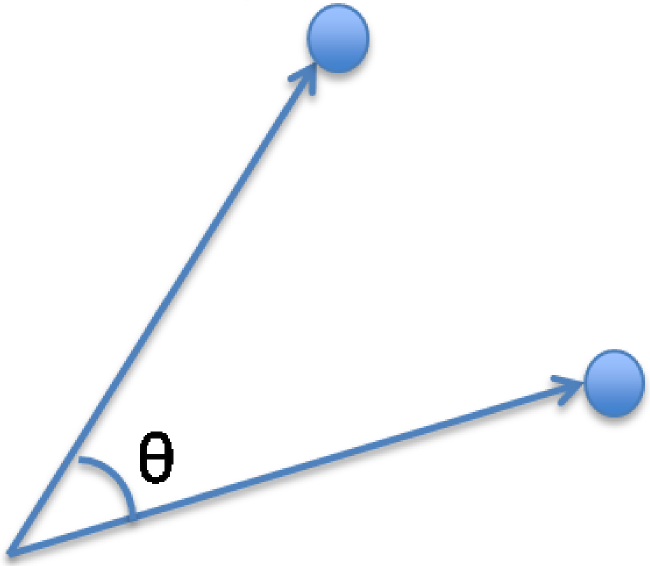
0.80

King → [0.1 0.2 0.7]

Apple → [0.9 0.5 0.6]

0.20

$$sim(A,B) = \mathbf{cos}(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$



Task- News Clustering

1. The **king** of Africa had planted this tree.

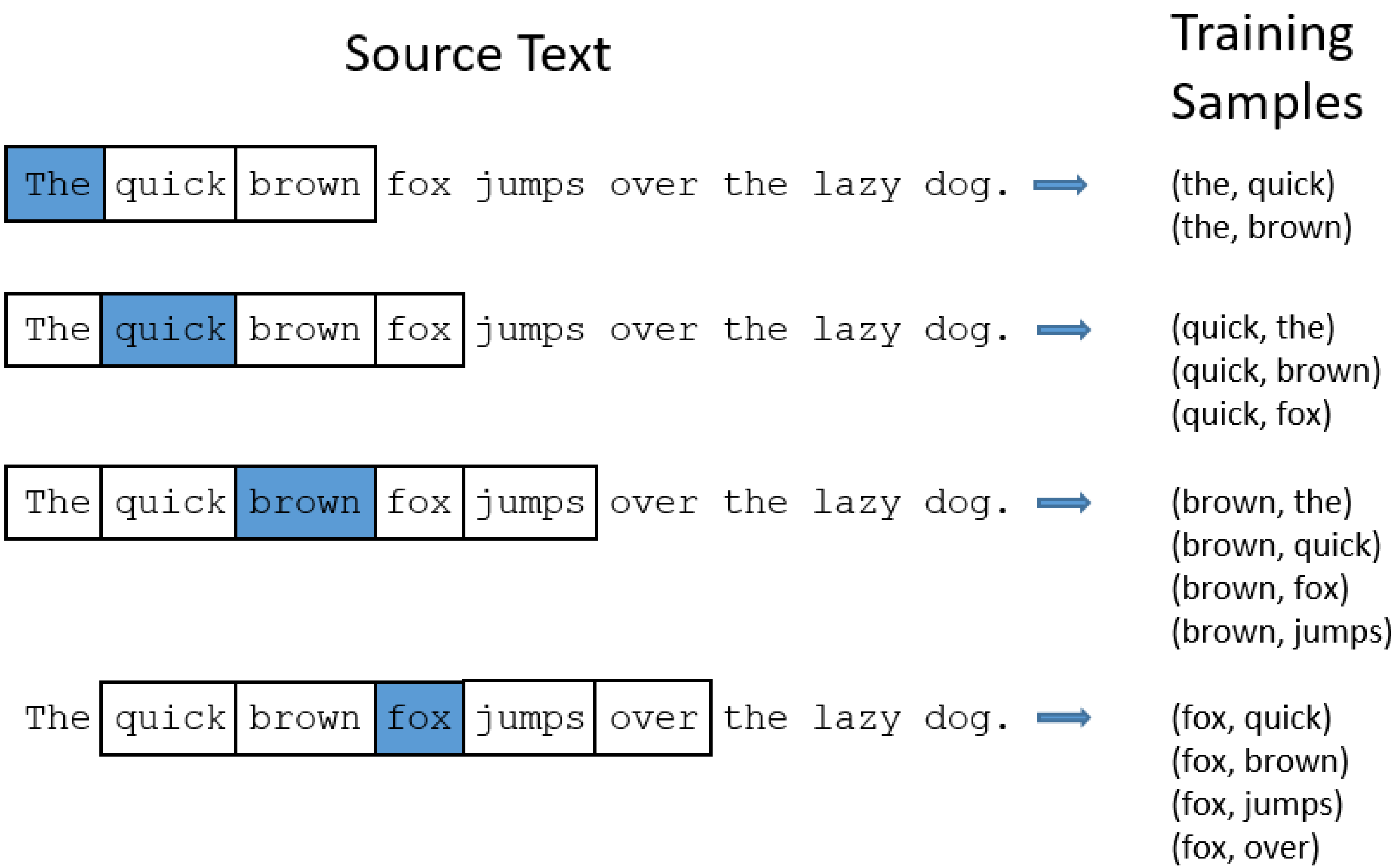
2. The **queen** of England is set to land in Africa this year .

3. This year we received a good quality of **apples**.

Popular Embedding models

- Word2Vec
- GloVe
- Doc2Vec
- fastText
- Gensim

A simple Skip-gram model



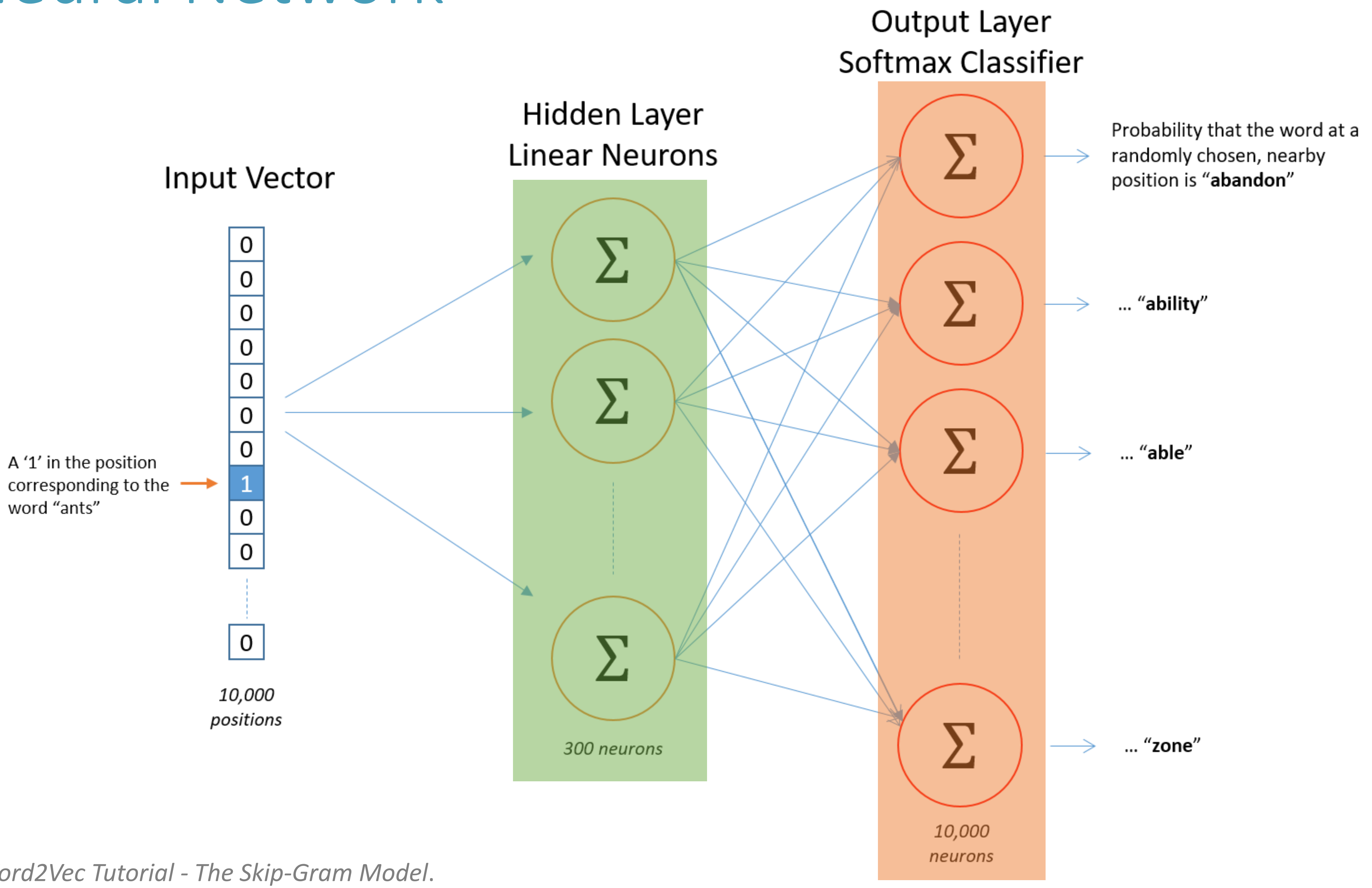
McCormick, C. (2016, April 19). *Word2Vec Tutorial - The Skip-Gram Model*.

The quick brown fox jumps over the lazy dog. The lazy dog was sleeping when the brown fox arrived. The brown fox found no other way to get on the other side. Jumping over the lazy dog was the only option that the brown fox had.

brown → fox

lazy → dog

A Shallow Neural Network



McCormick, C. (2016, April 19). *Word2Vec Tutorial - The Skip-Gram Model*.

Drawbacks of current approaches …

The play performed by the artists was very funny.

All work and no play makes everyone dull.

Polysemy

ELMo: To the Rescue!

- “Deep Contextualized Word Embeddings”
- The Paper was presented at this year’s NAACL, in June.
- Developed by AllenNLP team at the University of Washington
- Detailed info at https://allennlp.org/elmo

Elmo representations are…

1. Contextual

The representation for each word depends on the entire context in which it is used.

The play performed by the artists was very funny.

All work and no play makes everyone dull.

Elmo representations are…

2. Deep

The word representations combine all layers of a deep **pre-trained** neural network.

Elmo representations are…

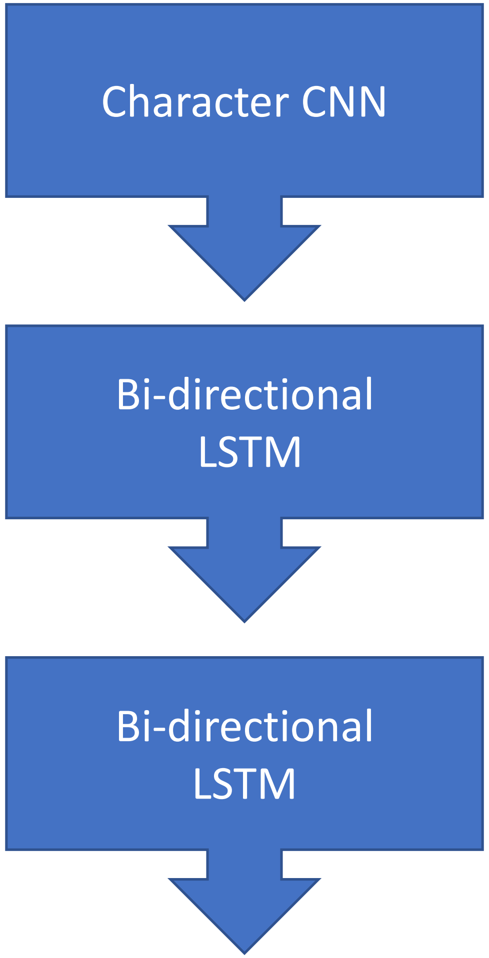
3. Character Based

ELMo representations are purely character based, allowing the network to use morphological clues to form robust representations for **out-of-vocabulary tokens** unseen in training.

US → us

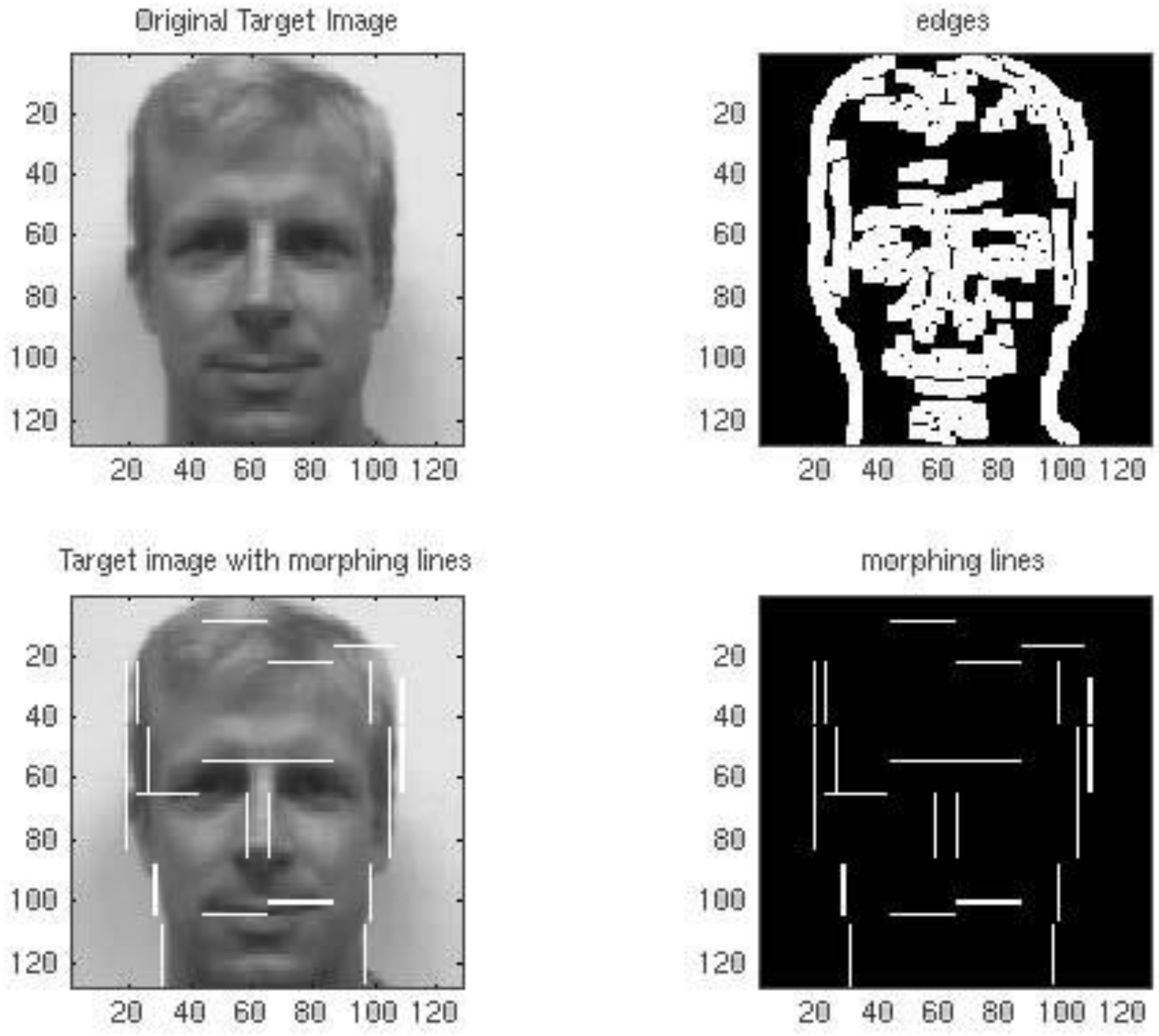
A sneak-peak into ELMo architecture..

- Consists of 3 Layers –

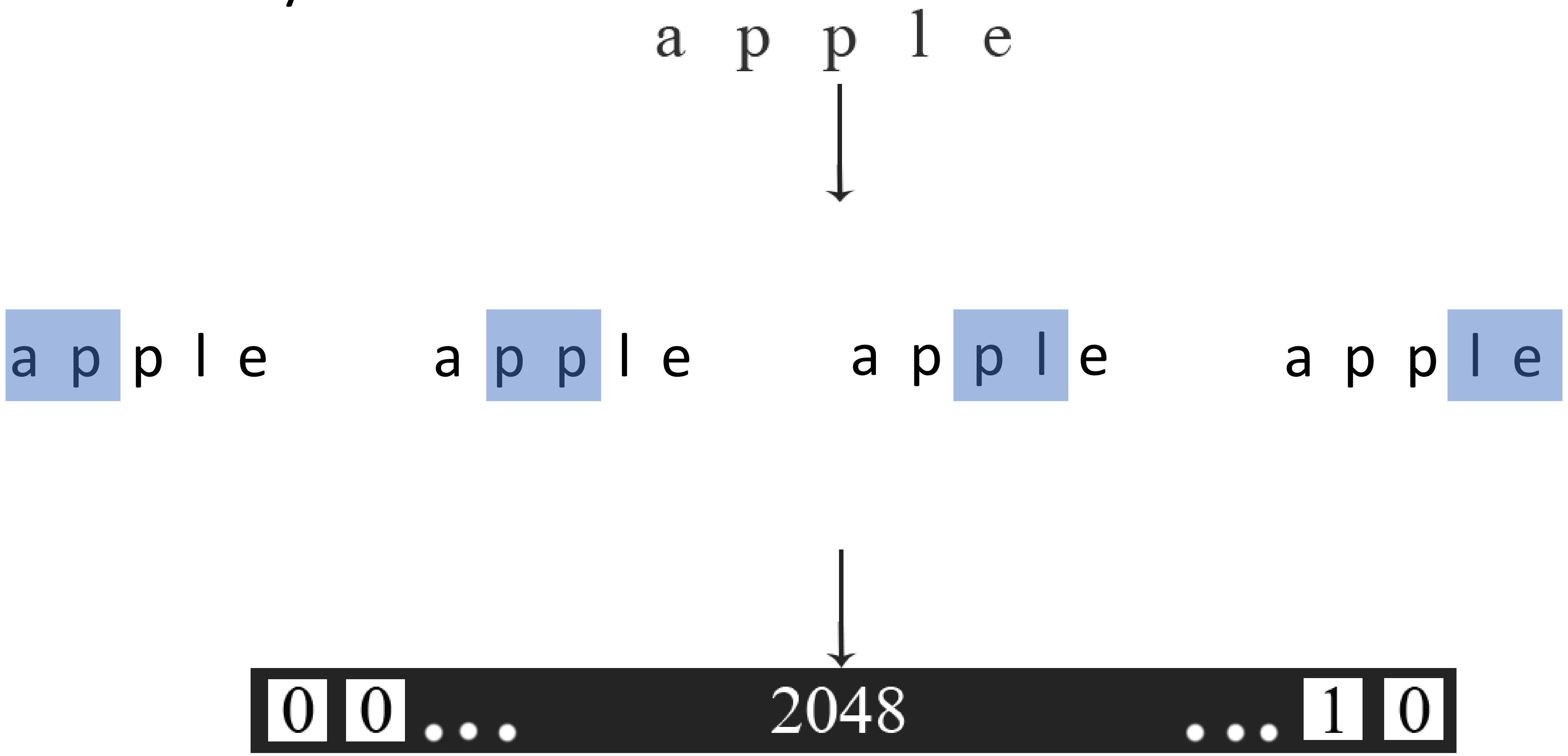


PyData MUMBAI

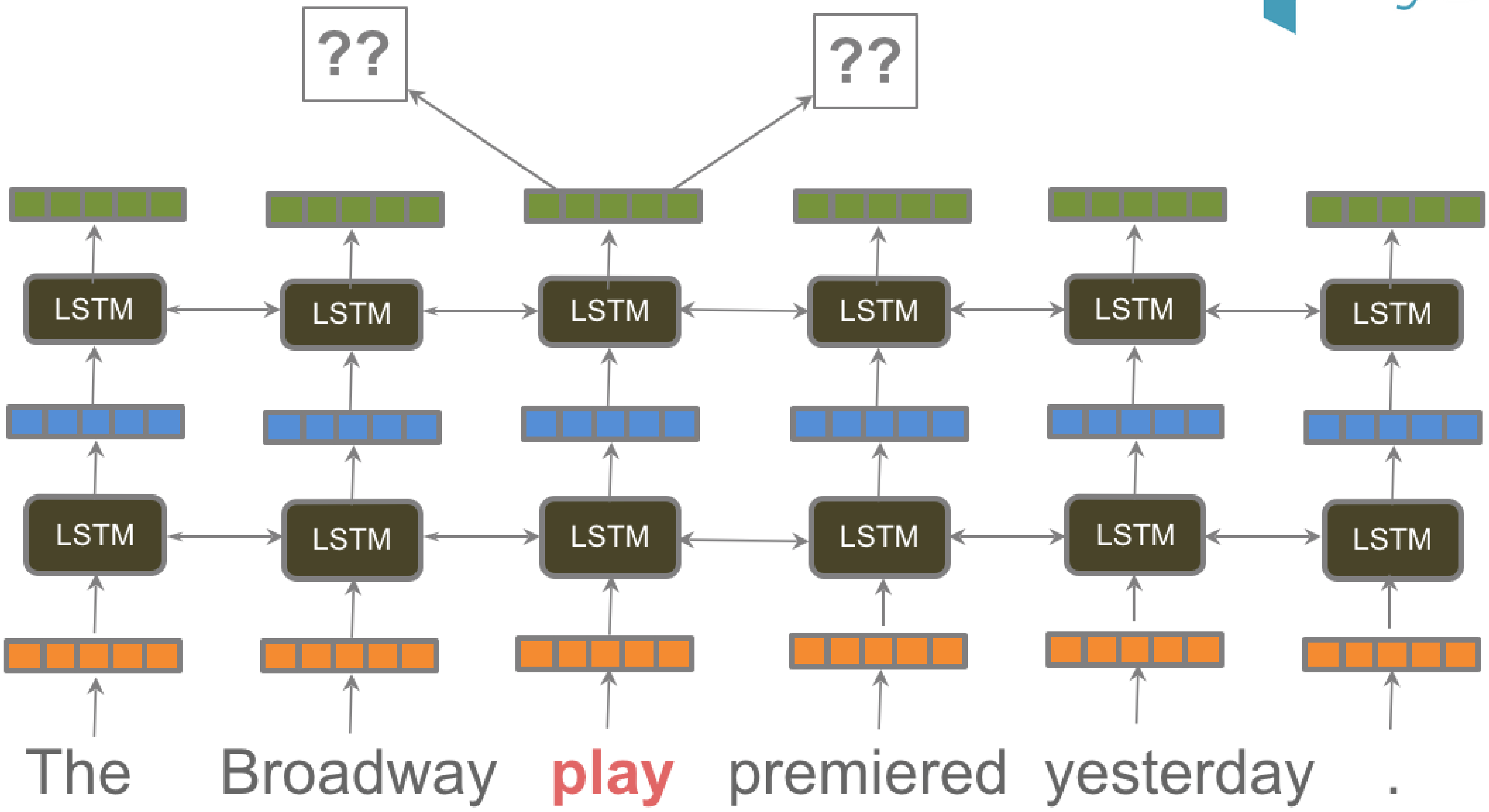
CNN Layer


Original Target Image
edges
Target image with morphing lines
morphing lines
20
40
60
80
100
120

CNN Layer



LSTM layers



??
??
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
The
Broadway
play
premiered
yesterday
.

Global benchmarks achieved

Task	Previous SOTA		Our baseline	ELMo + Baseline	Increase (Absolute/Relative)
SQuAD	SAN	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al (2017)	88.6	88.0	88.7 +/- 0.17	0.7 / 5.8%
SRL	He et al (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al (2017)	91.93 +/- 0.19	90.15	92.22 +/- 0.10	2.06 / 21%
Sentiment (5-class)	McCann et al (2017)	53.7	51.4	54.7 +/- 0.5	3.3 / 6.8%

Our Results and Observations

1. The **play** performed by the artists was very funny.

2. All work and no **play** makes everyone dull.

3. We must **play** everyday to be fit and fine.

1. The **play** performed by the artists was very funny.

Noun

2. All work and no **play** makes everyone dull.

Verb

Similarity = 0.79

2. All work and no **play** makes everyone dull.

Verb

3. We must **play** everyday to be fit and fine.

Verb

Similarity = 0.87

Conclusion

Similarity (Verb-Verb) > Similarity (Verb-Noun)

0.87 0.79

Our Results and Observations

1. The **cat** sat on the mat.
2. **Dog** came on the **cricket** field.
3. The **football** World Cup is held in Russia.

The **cat** sat on the mat.

A **Dog** came on the cricket field.

Similarity = 0.49

A Dog came on the **cricket** field.

The **football** World Cup is held in Russia.

Similarity = 0.68

Conclusion

The model gives decent similarity between words that are different but used in similar context.

Applications of ELMo

Feature	Current Model	Future Model
Summarization	GloVe	ELMo
Question Generation	GloVe	ELMo
Answer Generation	GloVe	ELMo
Stance Detection	GloVe	ELMo (Implemented)
Document Retriever	Concept Net	ELMo

Questions!