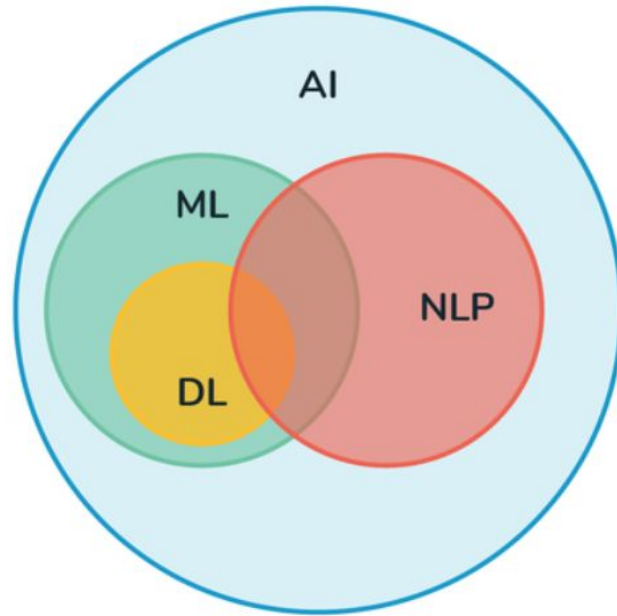


Introduction to Natural Language Processing

—

What is Natural Language Processing?

—



- Artificial intelligence
- Machine learning
- Language Processing
- Deep learning



Pragmatics

Semantics

Syntax

Morphology

NATURAL LANGUAGE
PROCESSING PYRAMID

How to change a text data into numerical data?

—

Bag of Words

good movie	good	movie	not	a	did	like
not a good movie	1	1	0	0	0	0
did not like	1	1	1	1	0	0
	0	0	1	0	1	1

Tf-Idf

$$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \log \left(\frac{N}{\text{df}_i} \right)$$

$\text{tf}_{i,j}$ = total number of occurrences of i in j

df_i = total number of documents (speeches) containing i

N = total number of documents (speeches)

Tf-Idf Data Format

	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	0.18	0.48	0.18							
Doc 2	0.18		0.18	0.48	0.18	0.18				
Doc 3					0.18	0.18	0.48	0.95	0.48	0.48

What are the 'data techniques' we can use in such data format?

—

Stop Words & Stemming & Lemmatization





1. TEXT
INFORMATION



2. SEGMENTATION
AND TOKENIZATION



3. TEXT
CLEANING



4. VECTORIZATION AND
FEATURE ENGINEERING



5. TEXT LEMMATIZATION
AND STEAMING



6. MACHINE LEARNING
ALGORITHMS



7. INTERPRETATION
OF THE RESULT

Artificial Neural Networks and NLP

Future Work?



Online ANN courses (Andrew Ng Online is a good course!)

Featurized representation: word embedding

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	<u>0.93</u>	<u>0.95</u>	-0.01	0.00
Age	0.03	0.02	0.7	0.69	0.03	-0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97
size	⋮	⋮				
cost						
alive						
verb						

I want a glass of orange juice.

I want a glass of apple juice.

e_{5391} e_{9853}

Andrew Ng

Do we train all these embeddings on our own?

Transfer Learning!

—

Latest State of the Art NLP Model!



Thank You!

