Introduction to Natural Language Processing

Rutu Mulkar
@rutumulkar
me@rutumulkar.com

About me:

Founder @ Hunchera - Text Understanding Simplified

- BS, MS, PhD in Computer Science
- Ph.D. thesis in Natural Language Processing

Previously:

- Founder of AI company - Ticary Solutions - sold to Sigmoidal in 2019
- Contributor to IBM Watson that defeated humans in Jeopardy!
- Worked at - Pricewaterhousecoopers, IBM, Moz and Precyse Healthcare

Outline

1. Syntax and Semantics of English Language
    - Structure and meaning of written text
2. Language Models
    - Converting text into a series of probabilities
3. Word Embeddings
    - Learning semantics of text as mathematical vectors

Outline

1. **Syntax and Semantics of English Language**
    - **Structure and meaning of written text**
2. Language Models
    - Converting text into a series of probabilities
3. Word Embeddings
    - Learning semantics of text as mathematical vectors

Syntax

Syntax makes the grammatical structure of a sentence

Syntax

S → NP VP

NP → DET ADJ* NOUN

VP → VERB ADV

DET → “the” | “a”

ADJ → “happy”

NOUN → “person”

VERB → “sleeps”

ADV → “well”

Syntax

S → NP VP

NP → DET ADJ* NOUN

VP → VERB ADV

DET → “the” | “a”

ADJ → “happy”

NOUN → “person”

VERB → “sleeps”

ADV → “well”

S

NP VP

DET ADJ* NOUN VP

The happy person VP

Semantics

What is semantics?

- Meaning of the sentence

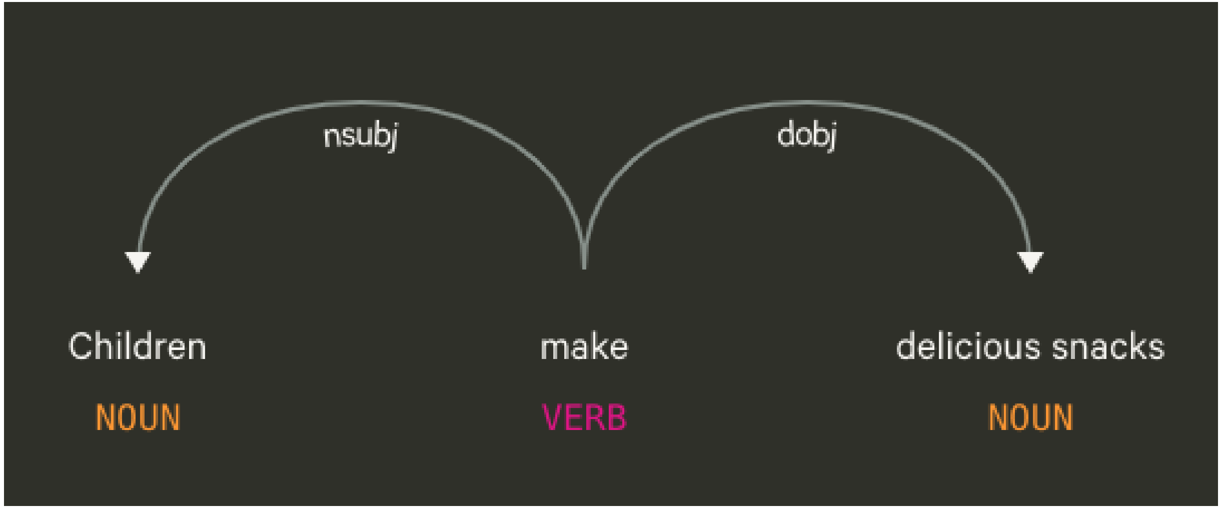
Why use syntax?

- Use syntax and syntactic dependencies to infer semantics (meaning) of the text

Syntax and Semantics of text is hard

Children **make** delicious snacks

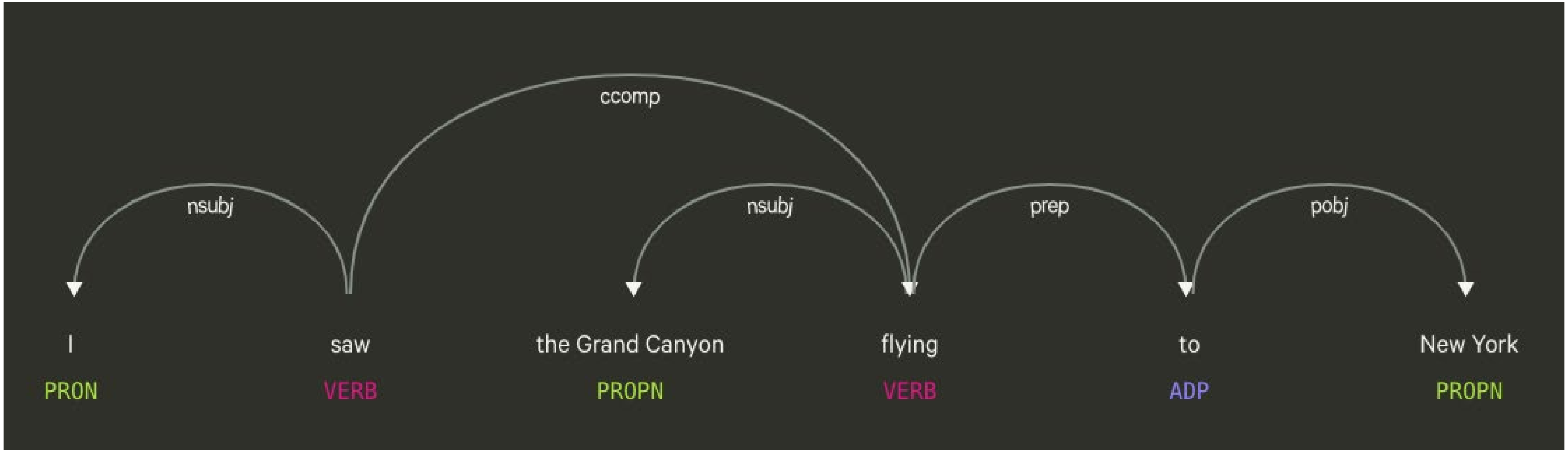
https://explosion.ai/demos/displacy?text=children%20are%20delicious%20snacks&model=en_core_web_sm&cpu=1&cph=1



Syntax and Semantics of text is hard

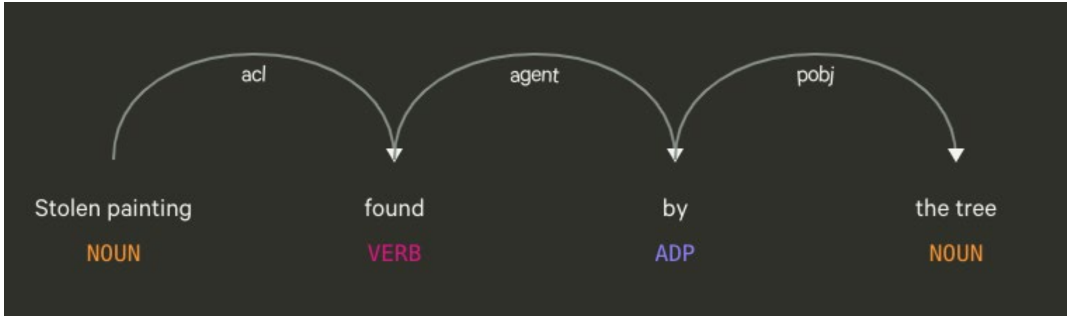
I saw the Grand Canyon **flying** to New York

https://explosion.ai/demos/displacy?text=I%20saw%20the%20Grand%20Canyon%20flying%20to%20New%20York&model=en_core_web_sm&cpu=1&cph=1

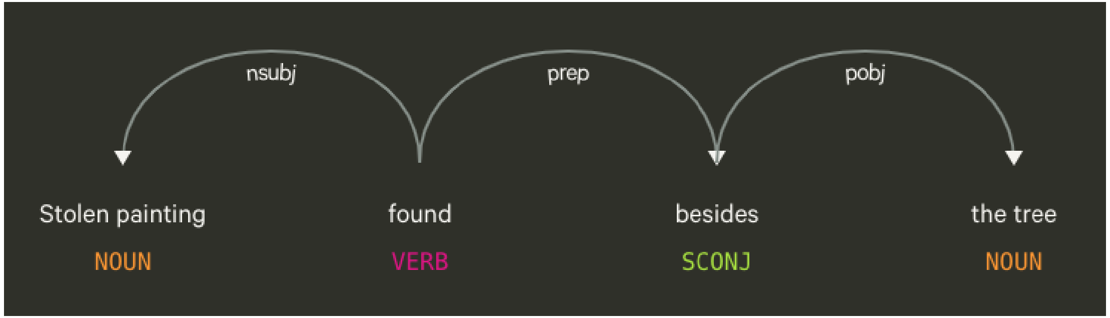


Syntax and Semantics of text is hard

Stolen painting **found** by the tree



Stolen painting **found** besides the tree

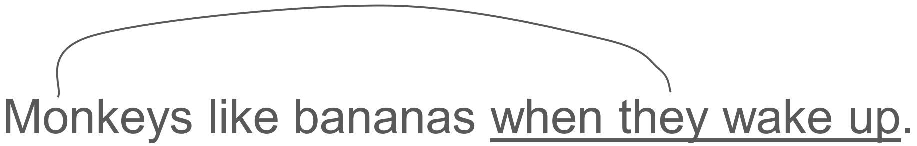


Context Matters - a lot!

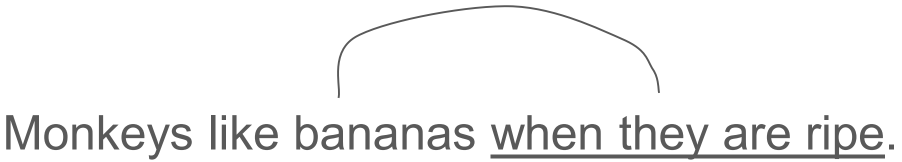
Monkeys like bananas when they wake up.

Monkeys like bananas when they are ripe.

Context Matters - a lot!



Monkeys like bananas when they wake up.



Monkeys like bananas when they are ripe.

Part of Speech tags - syntax of a sentence

All the wall-talls were idilicious.

It was callicial.

What are the underlined words?

- wall-talls
  - Noun
- idilicious
  - Noun or Adjective
- callical
  - Noun or Adjective

Part of Speech Tags

Nouns

Pronouns

Determiners

Adjectives

Verbs

Adverbs

Prepositions

Particles

Building your own tagger?

Corpora:

- Universal Dependencies Corpus:

  https://universaldependencies.org/

- Penn Treebank Corpus

  https://catalog.ldc.upenn.edu/LDC99T42

Parsing Resources

- SpaCy
  - python, high accuracy, fast
  - https://spacy.io/
- Stanford Core NLP
  - java, high accuracy, medium
  - http://nlp.stanford.edu/software /corenlp.shtml
- NLTK
  - python, low accuracy, fast
  - http://www.nltk.org/

What can we do with POS tags?

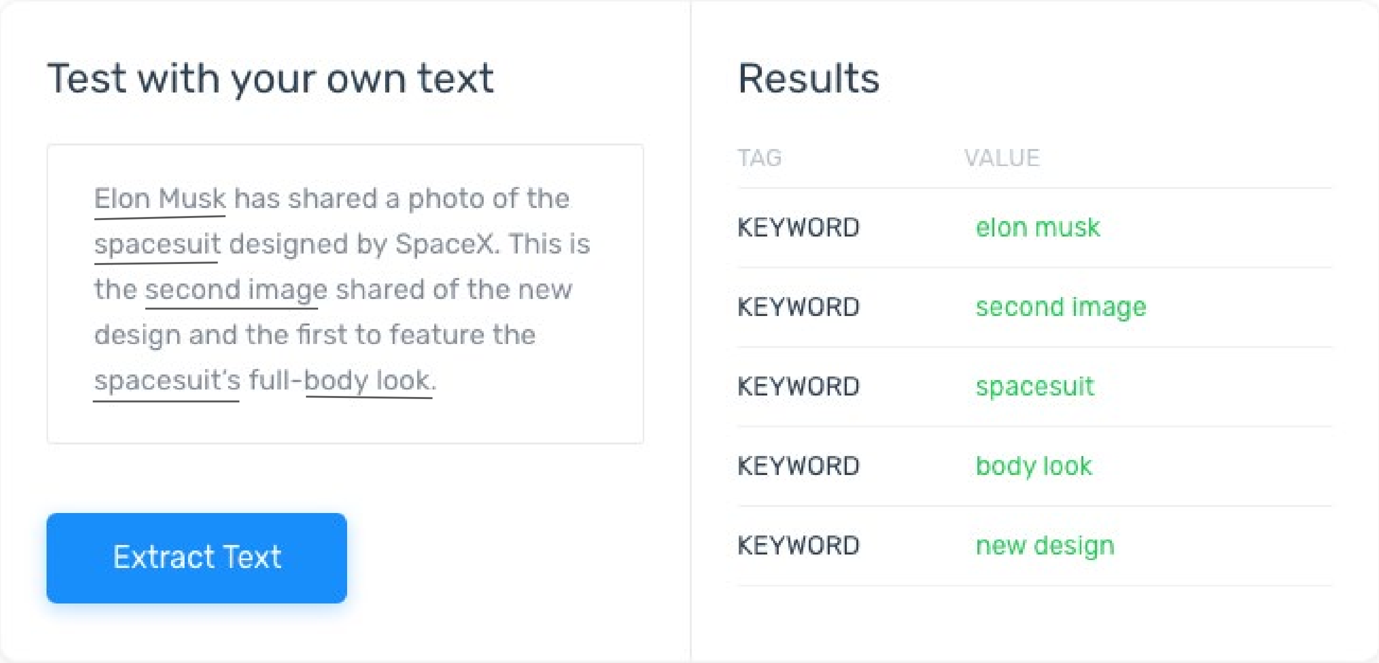
Keyword Extraction

- Nouns and Noun phrases are often the most significant pieces of information

Entity Extraction

- These are names of people, places etc

Example of Keyword Extraction



Test with your own text

Elon Musk has shared a photo of the spacesuit designed by SpaceX. This is the second image shared of the new design and the first to feature the spacesuit's full-body look.

Results

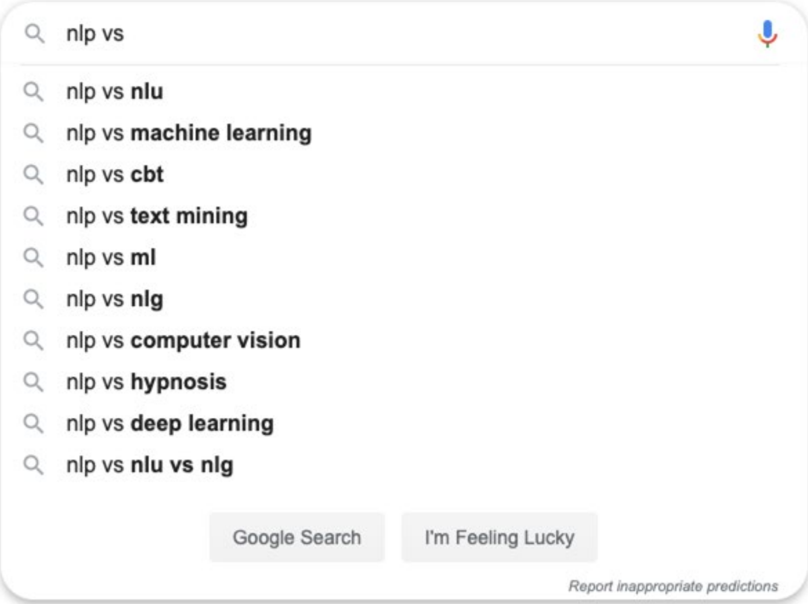
TAG	VALUE
KEYWORD	elon musk
KEYWORD	second image
KEYWORD	spacesuit
KEYWORD	body look
KEYWORD	new design

Process of Keyword Extraction

Extract candidate keywords, and rerank their relevance to the document based on a chosen (custom) metric.

Where is keyword extraction used?

SEO (Search Engine Optimization)



nlp vs

nlp vs **nlu**

nlp vs **machine learning**

nlp vs **cbt**

nlp vs **text mining**

nlp vs **ml**

nlp vs **nlg**

nlp vs **computer vision**

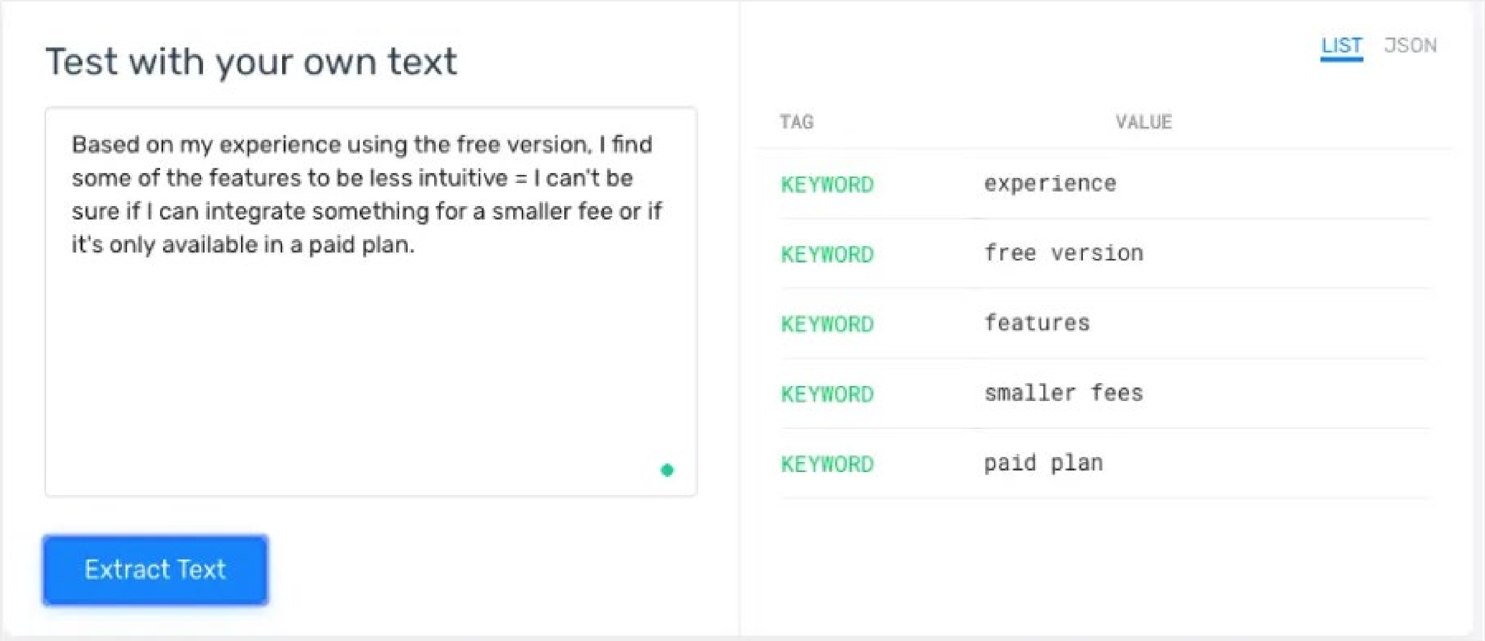
nlp vs **hypnosis**

nlp vs **deep learning**

nlp vs **nlu vs nlg**

Where is keyword extraction used?

Customer Feedback analysis



Based on my experience using the free version, I find some of the features to be less intuitive = I can't be sure if I can integrate something for a smaller fee or if it's only available in a paid plan.

TAG	VALUE
KEYWORD	experience
KEYWORD	free version
KEYWORD	features
KEYWORD	smaller fees
KEYWORD	paid plan

Outline

1. Syntax and Semantics of English Language
    - Structure and meaning of written text
2. **Language Models**
    - **Converting text into a series of probabilities**
3. Word Embeddings
    - Learning semantics of text as mathematical vectors

Syntax to Language Models

Using grammar and syntax provides just one representation of the text. But there can be more than one meaning of a sentence.

Language Models (LM) provides a list of probable explanations/representations of a given sentence.

Understanding Language Models

A model that can predict the probability of a given word given a list of previous words:

P(Natural Language Processing) =

p(Natural) * p(Language | Natural) * p(Processing | Natural Language)

But the chains of text can be very long!

E.g.

“Natural Language Processing is a very interesting field.”

```
p(w1...ws) = p(w1) * p(w2 | w1) * p(w3 | w1 w2) * p(w4 | w1 w2 w3) ..... p(wn | w1...wn-1)
```

Markov Assumption

Markov Assumption

Andrey Markov was a Russian mathematician who described a stochastic process with a property called Markov Property or Markov Assumption. This basically states that one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history, hence independently from such history.[2]

Based on this assumption we can rewrite conditional probability of 'pleasant' as:

$$P(pleasant \mid weather\, is) \approx P(pleasant \mid weather)$$

How do we compute these probabilities?

Bigram model

$$P(w_i \mid w_{i-1}) = \frac{count(w_{i-1}, w_i)}{count(w_{i-1})}$$

Other models:

Trigram, 4-gram, 5-gram etc.

Applications of Language Models

1. Spelling Correction

For spelling correction probability of incorrect sentence will be much smaller then correct sentence

$$P(weather\, is\, \mathbf{pleasent}) > P(weather\, is\, \mathbf{plesen})$$

Applications of Language Models

2. Speech Recognition

As words ‘weather’ & ‘whether’ may have similar phonetics, system may confuse among these but probability of ‘weather is pleasant’ will be higher than ‘whether is pleasant’

$$P(\mathbf{weather}\, is\, pleasant) > P(\mathbf{whether}\, is\, pleasant)$$

Applications of Language Models

3. Machine Translation

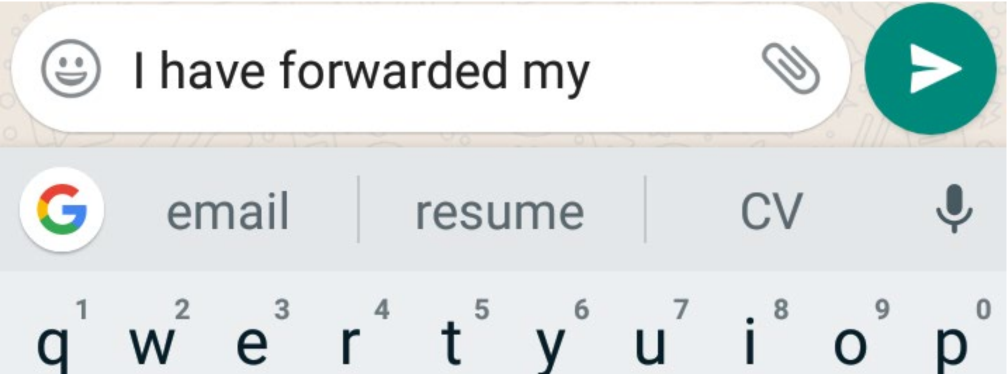
Selecting appropriate sequence while translating from one language to another can also use probability of sequence for providing more appropriate translations

$$P(\mathbf{high}\, winds\, tonight) > P(\mathbf{large}\, winds\, tonight)$$

Applications of Language Models

4. Predictive Text

By looking at previous sequence of words language model can predict next word, this feature was recently introduced in android phone keyboard by Google.



I have forwarded my

Language Models to check out

https://books.google.com/ngrams

Outline

1. Syntax and Semantics of English Language
   - Structure and meaning of written text
2. Language Models
   - Converting text into a series of probabilities
3. **Word Embeddings**
   - **Learning semantics of text as mathematical vectors**

What are word embeddings?

Word embeddings are vector representations of a given word to capture the semantics of the word.

What is a vector?

A list of numbers that can capture the meaning of a word

E.g. The

0	1	1	0	0.5

This is a vector representing the word “the” in 5 dimensions.

The value of the vector is arbitrary in this example.

Word Vectors

Consider 2 sentences:

Have a good day

Have a great day.

Total words: have, a, good, great, day

Vector Representation:

Have = [1,0,0,0,0]

a=[0,1,0,0,0]

good=[0,0,1,0,0]

great=[0,0,0,1,0]

day=[0,0,0,0,1]`

Examples of word embeddings

- Word2Vec
- Glove

These models represent the knowledge of the earth in over 300 dimensions

How Does Word2Vec work

Word2Vec is a method to construct such an embedding. It can be obtained using two methods (both involving Neural Networks): Skip Gram and Common Bag Of Words (CBOW)

CBOW

CBOW Model: This method takes the context of each word as the input and tries to predict the word corresponding to the context.

Skip Gram Model

Skip Gram: We take a target word and try to predict its context

Which approach is better?

Skip Gram works well with small amount of data and is found to represent rare words well.

CBOW is faster and has better representations for more frequent words.

Where can word embeddings be used?

Used as a thesaurus

topic identification

Summary

1. Syntax and Semantics of English Language
   - Structure and meaning of written text
2. Language Models
   - Converting text into a series of probabilities
3. Word Embeddings
   - Learning semantics of text as mathematical vectors

Thank you for listening!

@rutumulkar
me@rutumulkar.com