Text Summarization

Using seq2seq model

Steve Nouri

Head of Data Science at Nod.

9th May, 2019

Nod.

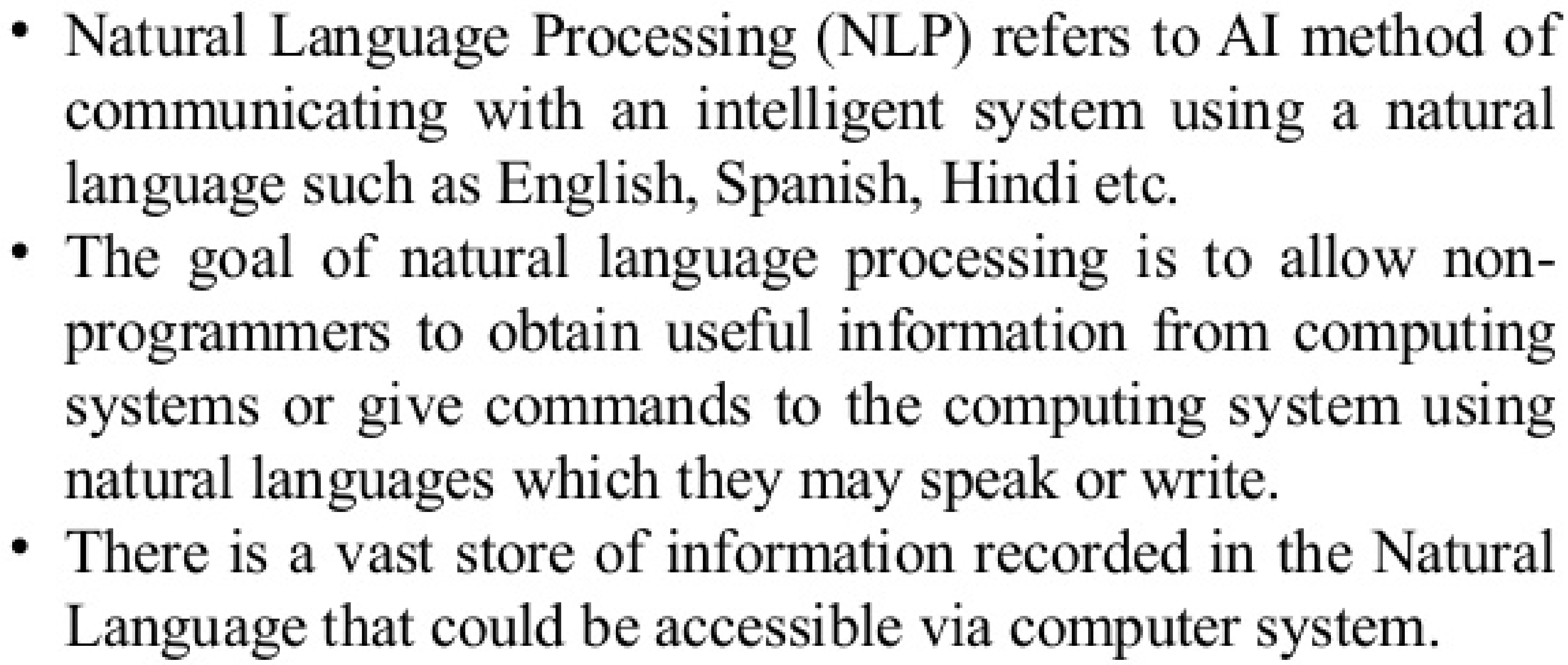
Introduction and Background

- Leading the data science team at Nod.
- Casual academic at UTS
- 10+ years of experience in different IT roles

- Executive Degree from MIT
- Master of Data science
- Bachelor of software Engineering

Nod.

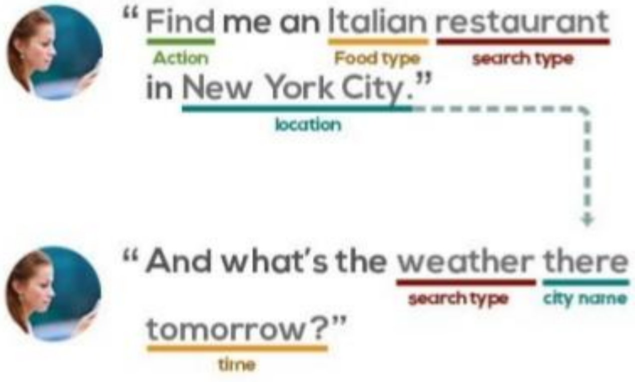
Natural Language Processing



- Natural Language Processing (NLP) refers to AI method of communicating with an intelligent system using a natural language such as English, Spanish, Hindi etc.
- The goal of natural language processing is to allow non-programmers to obtain useful information from computing systems or give commands to the computing system using natural languages which they may speak or write.
- There is a vast store of information recorded in the Natural Language that could be accessible via computer system.

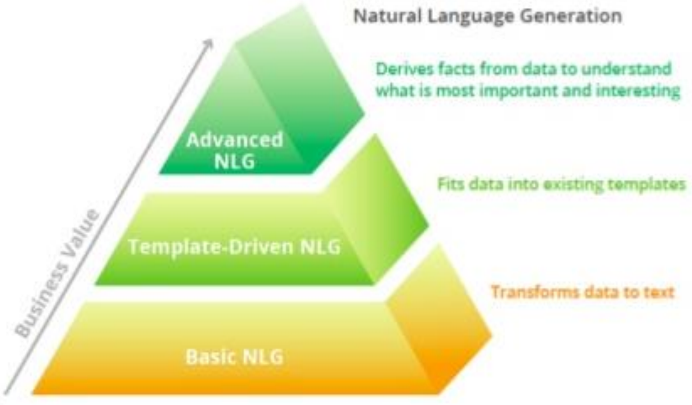
Components of NLP

- Natural Language Understanding
  - Mapping the given input in natural language into useful representations.
- Natural Language Generation
  - Producing meaningful phrases and sentences in the form of natural language from some internal representation



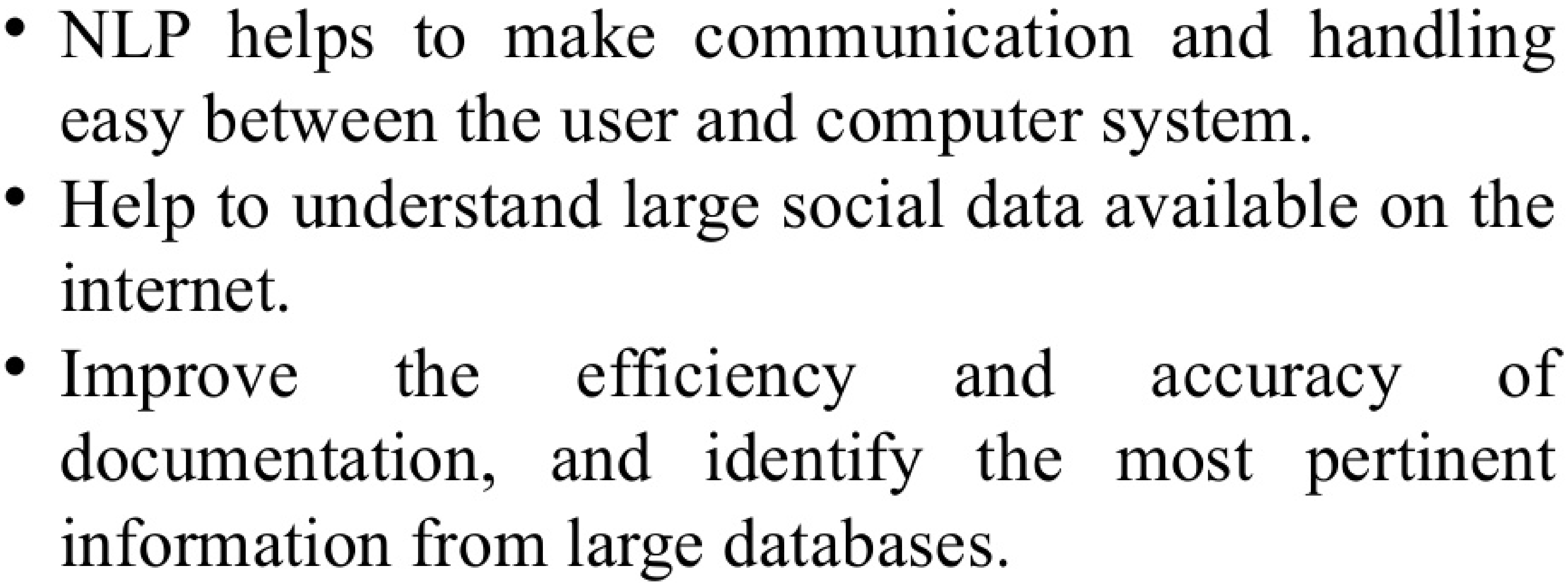
“Find me an Italian restaurant in New York City.”

“And what’s the weather there tomorrow?”



Natural Language Generation
Derives facts from data to understand what is most important and interesting
Advanced NLG
Fits data into existing templates
Template-Driven NLG
Transforms data to text
Basic NLG
Business Value

Importance



- NLP helps to make communication and handling easy between the user and computer system.
- Help to understand large social data available on the internet.
- Improve the efficiency and accuracy of documentation, and identify the most pertinent information from large databases.

Applications of Nat. Lang. Processing

- Machine Translation
- Database Access
- Information Retrieval
  - Selecting from a set of documents the ones that are relevant to a query
- Text Categorization
  - Sorting text into fixed topic categories
- Extracting data from text
  - Converting unstructured text into structure data
- Spoken language control systems
- Spelling and grammar checkers

Real world example

- Understand a Job Resume
- Match it with a Job Description
- Rank the resumes based on relevance
- Rank the resumes based on capability

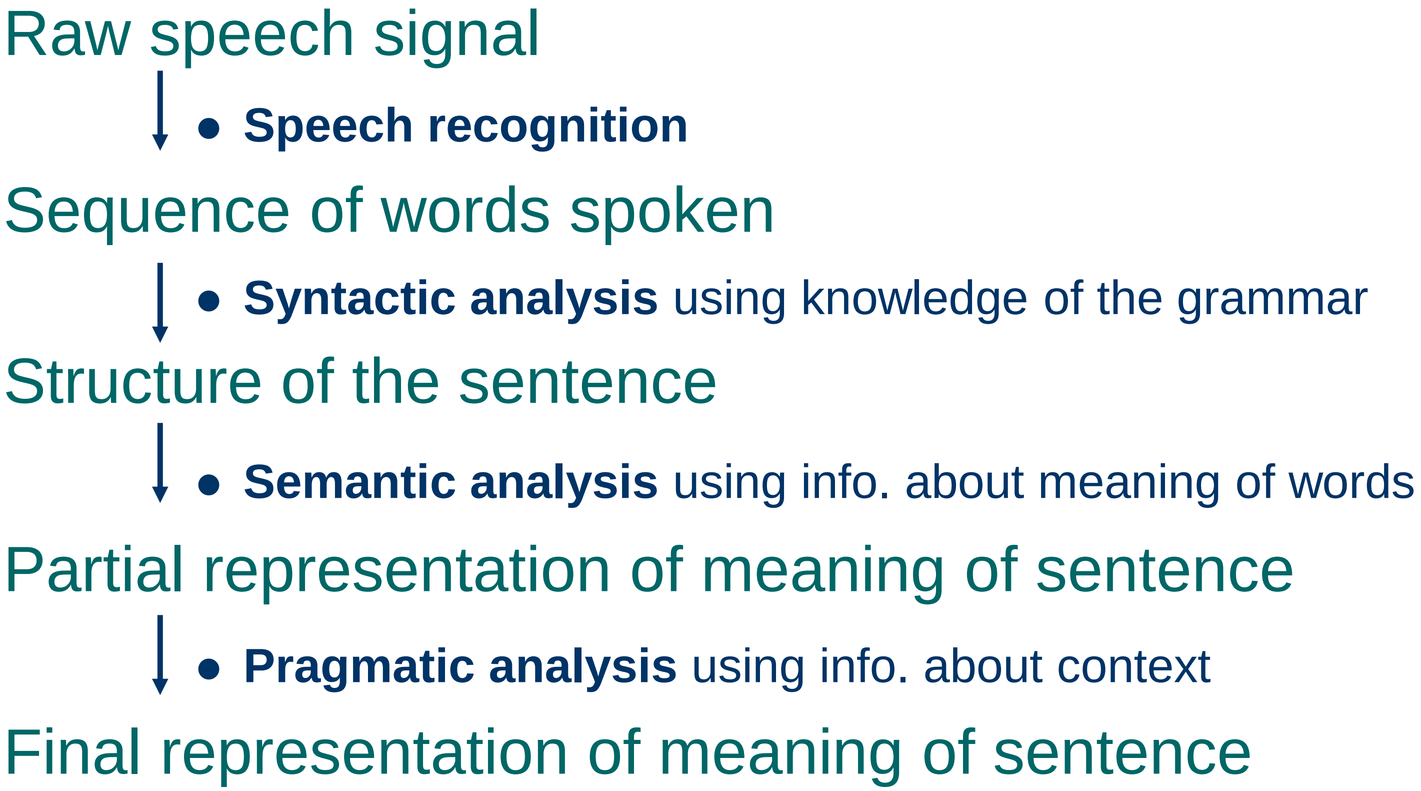
Ambiguity

- Ambiguity
  - Lexical ambiguity
    - Treating the word “board” as noun or verb?
  - Syntactical ambiguity
    - “He lifted the beetle with red cap”
    - Did he use cap to lift the beetle or he lifted a beetle that had red cap?
  - Referential ambiguity
    - Rima went to Gauri. She said, “I am tired.”
    - Exactly who is tired?

Challenges

- Phrases / Idioms
  - “A perfect storm”
    - The worst possible situation
- Connecting language and machine perception
- Sentence generation
- Text summarization
- Keyword extraction

Natural language understanding



Raw speech signal

- **Speech recognition**

Sequence of words spoken

- **Syntactic analysis** using knowledge of the grammar

Structure of the sentence

- **Semantic analysis** using info. about meaning of words

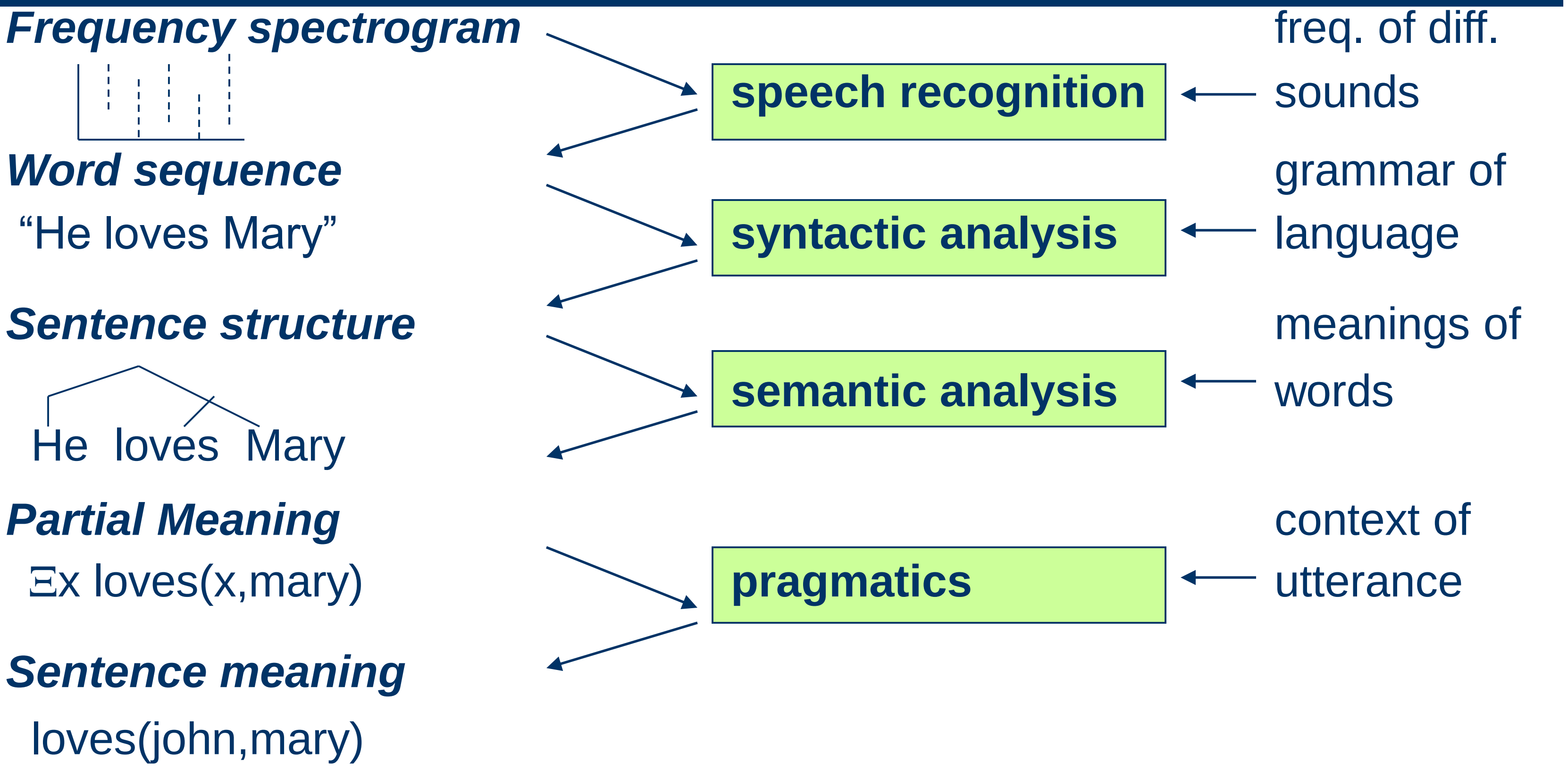
Partial representation of meaning of sentence

- **Pragmatic analysis** using info. about context

Final representation of meaning of sentence

Natural Language Understanding

- Input/Output data Processing stage Other data used



Frequency spectrogram

speech recognition ← freq. of diff. sounds

Word sequence

“He loves Mary”

syntactic analysis ← grammar of language

Sentence structure

He loves Mary

semantic analysis ← meanings of words

Partial Meaning

Ξx loves(x,mary)

pragmatics ← context of utterance

Sentence meaning

loves(john,mary)

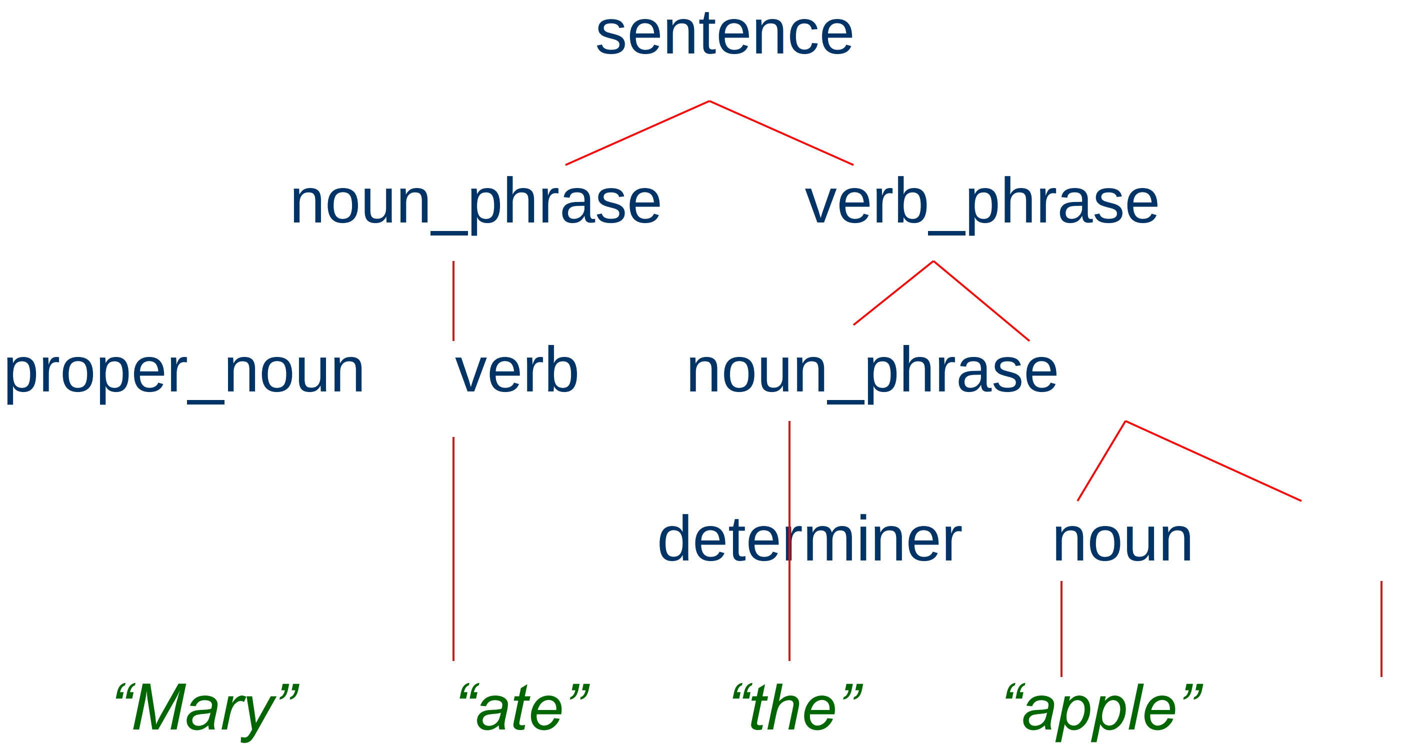
Syntactic Analysis

- Rules of syntax (grammar) specify the possible organization of words in sentences and allows us to determine sentence’s structure(s)
  - “John saw Mary with a telescope”
    - John saw (Mary with a telescope)
    - John (saw Mary with a telescope)
- Parsing: given a sentence and a grammar
  - Checks that the sentence is correct according with the grammar and if so returns a **parse tree** representing the structure of the sentence

Syntactic Analysis - Grammar

- sentence -> noun_phrase, verb_phrase
- noun_phrase -> proper_noun
- noun_phrase -> determiner, noun
- verb_phrase -> verb, noun_phrase
- proper_noun -> [mary]
- noun -> [apple]
- verb -> [ate]
- determiner -> [the]

Syntactic Analysis - Parsing


sentence

noun_phrase verb_phrase

proper_noun verb noun_phrase

determiner noun

“Mary” *“ate”* *“the”* *“apple”*

Syntactic Analysis – Complications (1)

- Number (singular vs. plural) and gender
  - `sentence-> noun_phrase(n),verb_phrase(n)`
  - `proper_noun(s) -> [mary]`
  - `noun(p) -> [apples]`
- Adjective
  - `noun_phrase-> determiner,adjectives,noun`
  - adjectives-> adjective, adjectives
  - adjective->[ferocious]
- Adverbs, …

Syntactic Analysis – Complications (2)

- Handling ambiguity
  - Syntactic ambiguity: “fruit flies like a banana”

- Having to parse syntactically incorrect sentences

Semantic Analysis – Complications

- Handling ambiguity
  - Semantic ambiguity: “I saw the prudential building flying into Boston”

Pragmatics

- Uses context of utterance
  - Where, by who, to whom, why, when it was said
  - Intentions: *inform, request, promise*, *criticize,* …
- Handling Pronouns
  - “Mary eats apples. She likes them.”
    - She=“Mary”, them=“apples”.
- Handling ambiguity
  - Pragmatic ambiguity: “you’re late”: What’s the speaker’s intention: informing or criticizing?

Natural Language Generation

- Talking back! ☺
- What to say or text planning
  - flight(AA,london,boston,$560,2pm),
  - flight(BA,london,boston,$640,10am),
- How to say it
  - "There are two flights from London to Boston. The first one is with American Airlines, leaves at 2 pm, and costs $560 …"
- Speech synthesis
  - Simple: Human recordings of basic templates
  - More complex: string together phonemes in phonetic spelling of each word
    - Difficult due to stress, intonation, timing, liaisons between words

SOA Generation

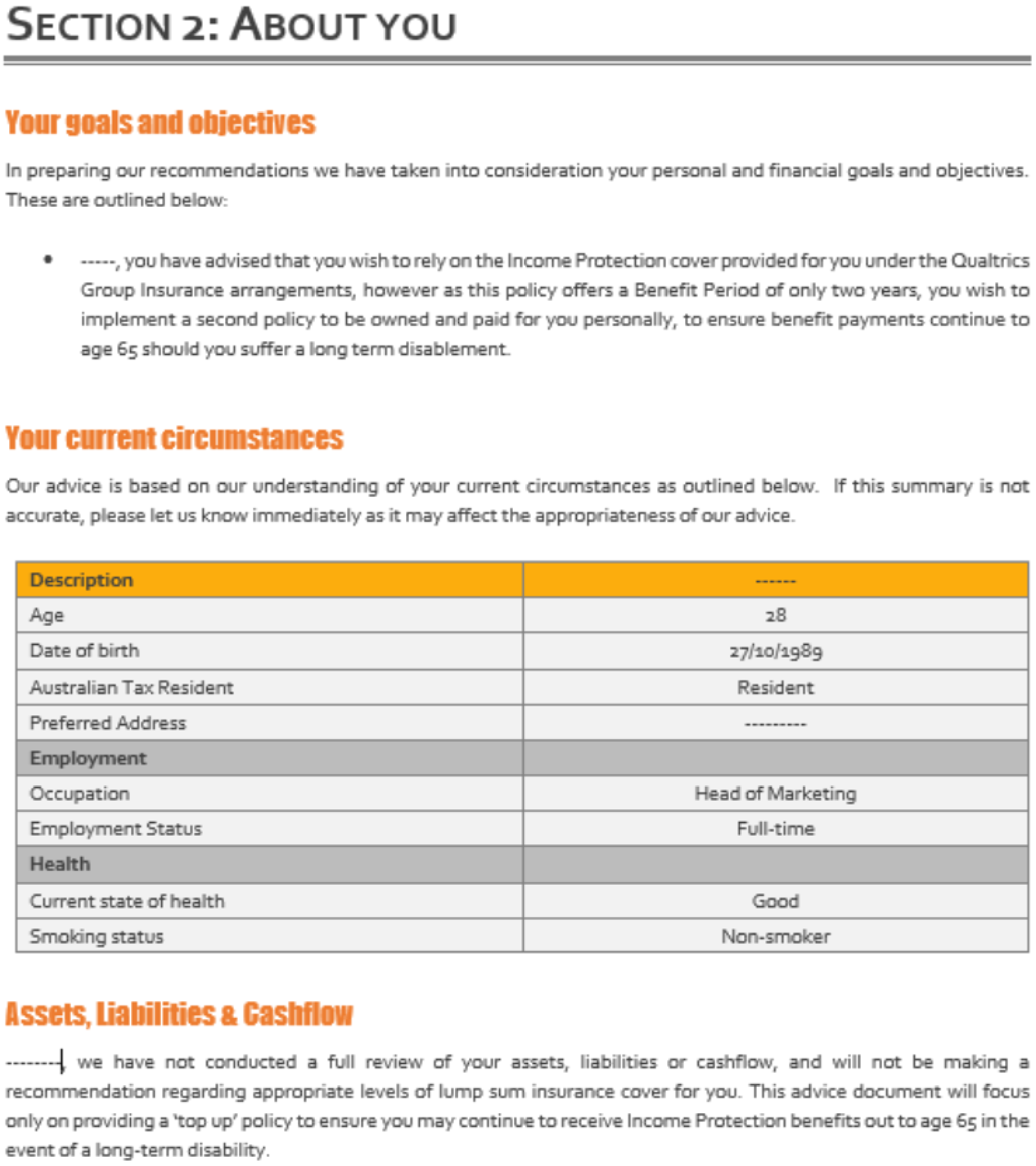
Producing complete advice documents is time consuming, costly and repetitive work across multiple industries.

Example from Financial Advice

5-7 days	per document waiting for paraplanners to produce Statements of Advice
$400 – $2000	per document on production costs**.**
5-7 days	per document going through compliance processes and checks

Structure of an SOA

- Headings
- Paragraphs
- Styles
- Tables
- Images



SECTION 2: ABOUT YOU

Your goals and objectives

In preparing our recommendations we have taken into consideration your personal and financial goals and objectives. These are outlined below:

- -----, you have advised that you wish to rely on the Income Protection cover provided for you under the Qualtrics Group Insurance arrangements, however as this policy offers a Benefit Period of only two years, you wish to implement a second policy to be owned and paid for you personally, to ensure benefit payments continue to age 65 should you suffer a long term disablement.

Your current circumstances

Our advice is based on our understanding of your current circumstances as outlined below. If this summary is not accurate, please let us know immediately as it may affect the appropriateness of our advice.

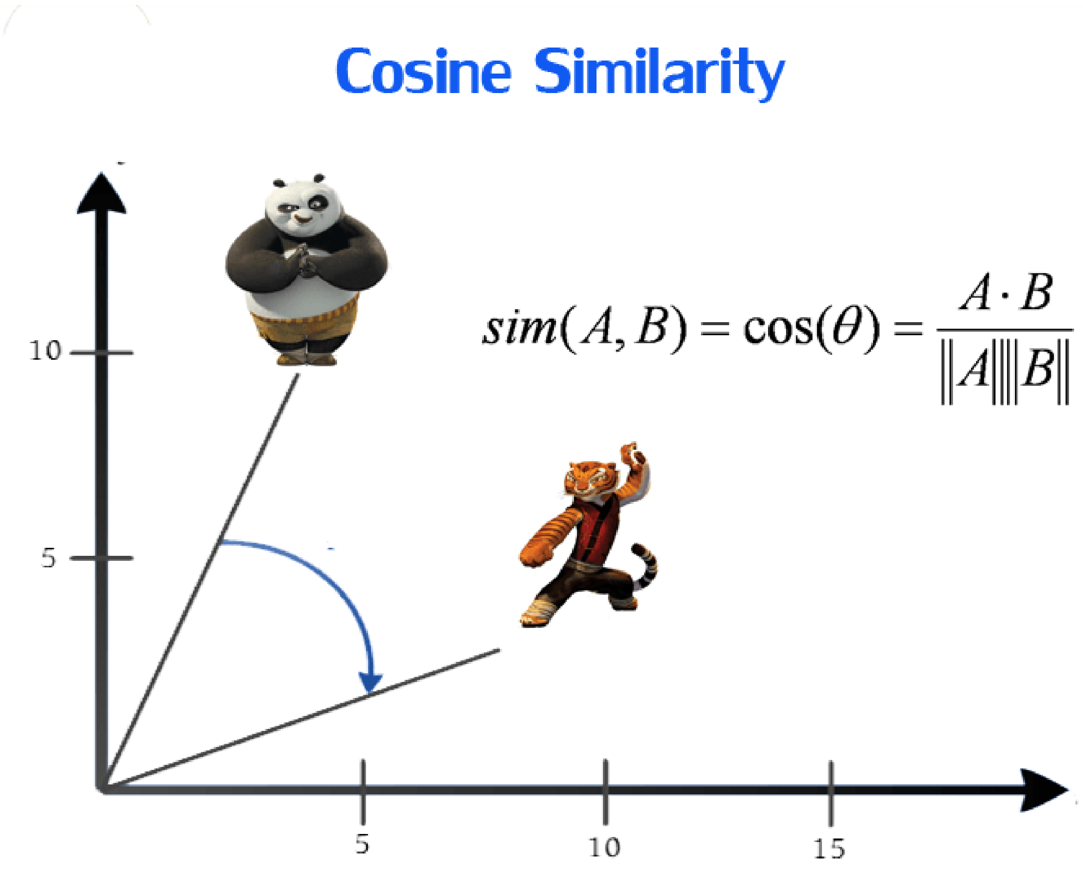
Description	------
Age	28
Date of birth	27/10/1989
Australian Tax Resident	Resident
Preferred Address	---------
Employment	
Occupation	Head of Marketing
Employment Status	Full-time
Health	
Current state of health	Good
Smoking status	Non-smoker

Assets, Liabilities & Cashflow

-------, we have not conducted a full review of your assets, liabilities or cashflow, and will not be making a recommendation regarding appropriate levels of lump sum insurance cover for you. This advice document will focus only on providing a 'top up' policy to ensure you may continue to receive Income Protection benefits out to age 65 in the event of a long-term disability.

Word Embedding

- **Skip-Gram model**
- ***CBOW Model***

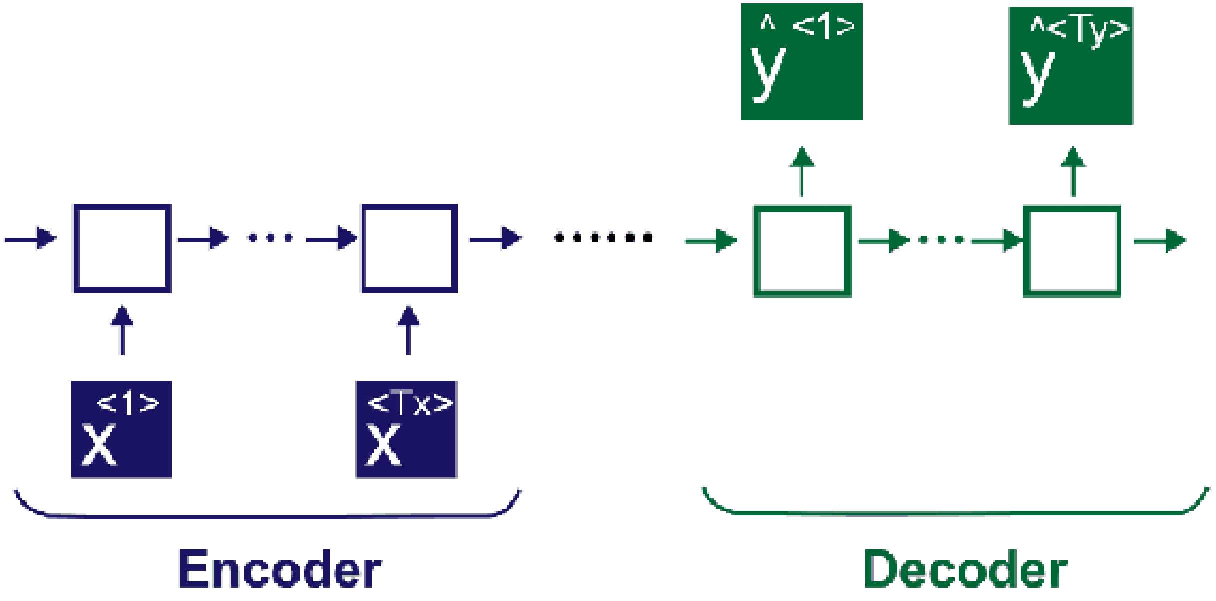


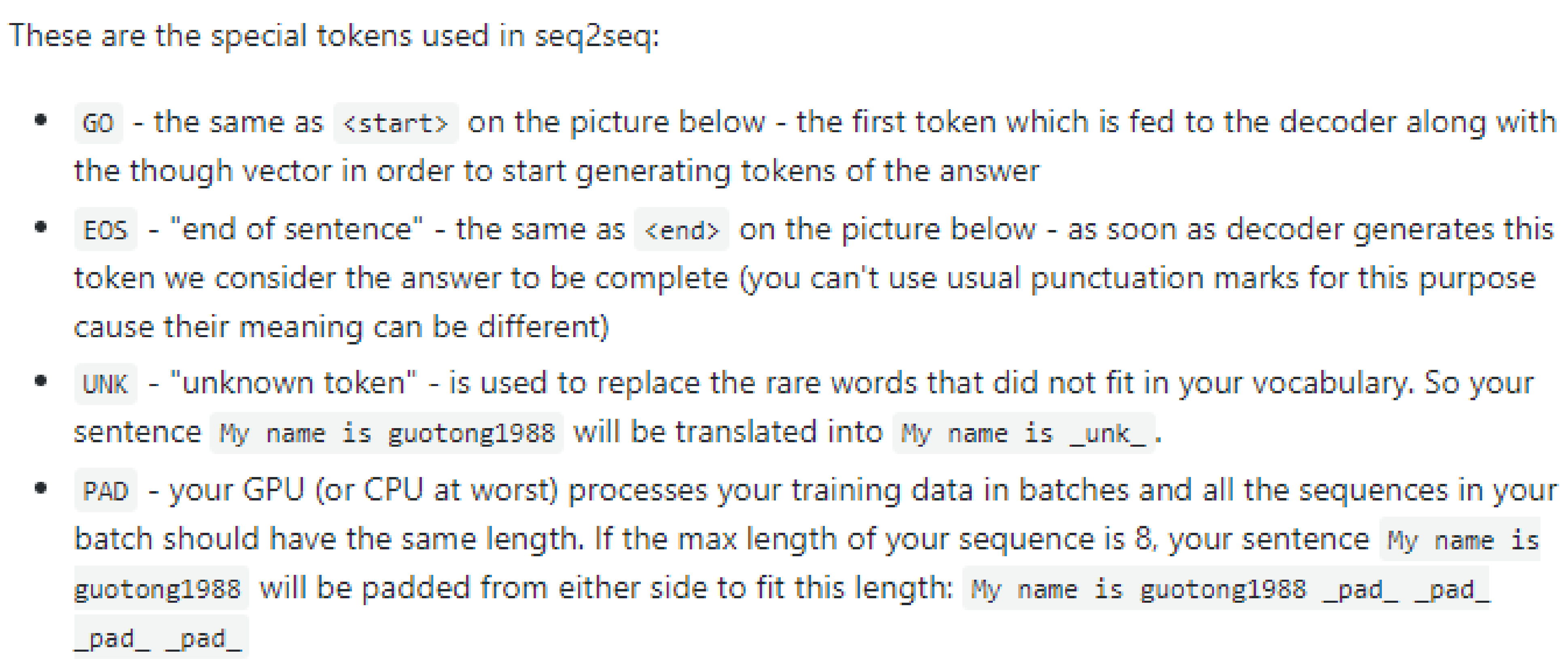
Cosine Similarity

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

Seq2Seq Model

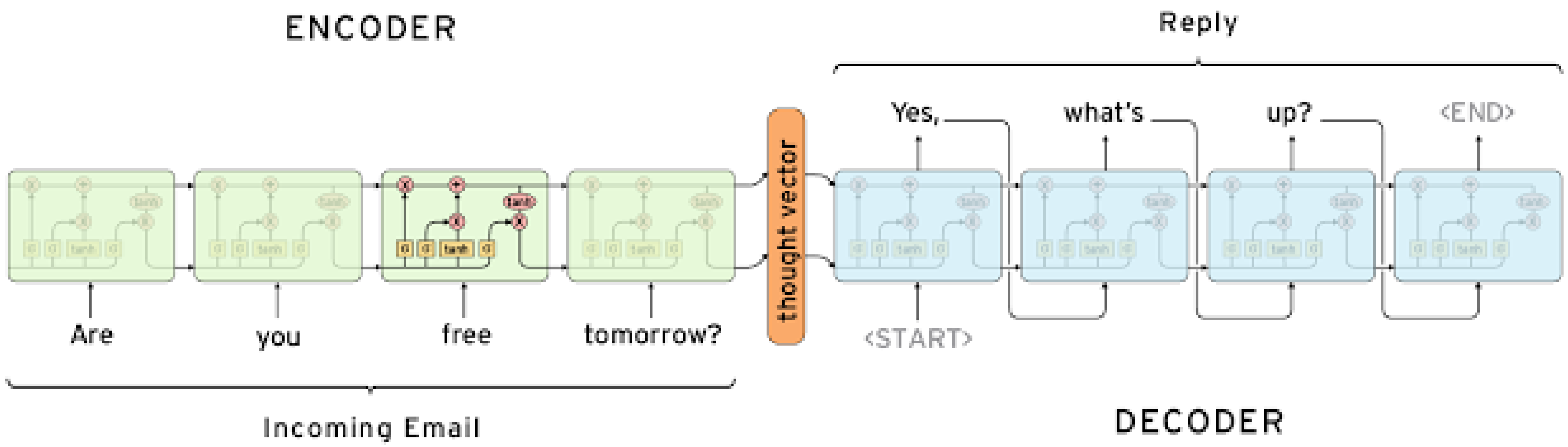
Both Encoder Decoder here are RNN network , but encoder uses input , and generates an output state that is then used as input to decoder stage





These are the special tokens used in seq2seq:

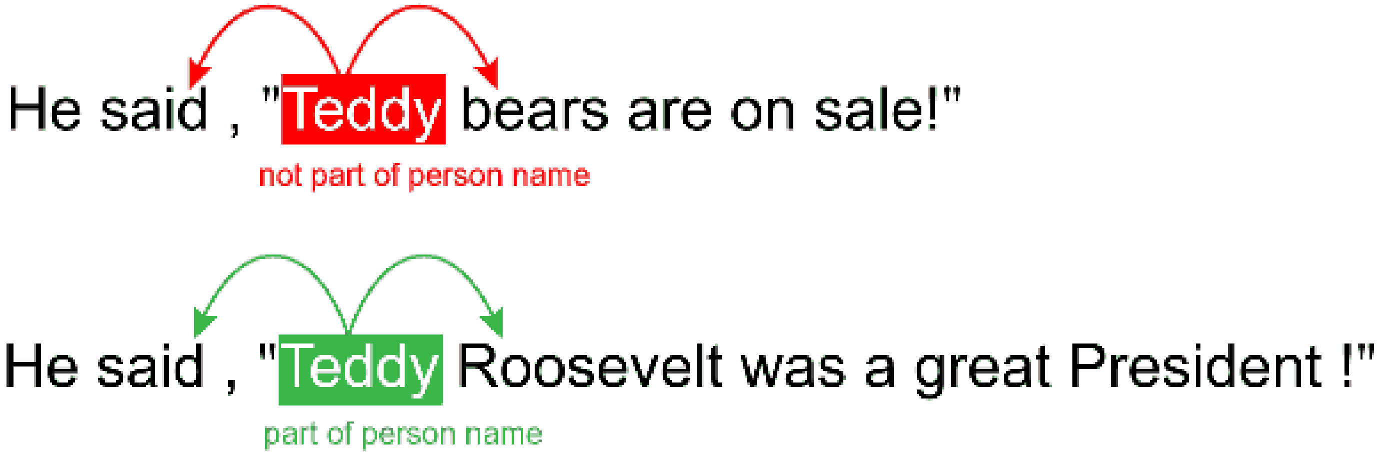
- `GO` - the same as `<start>` on the picture below - the first token which is fed to the decoder along with the though vector in order to start generating tokens of the answer
- `EOS` - "end of sentence" - the same as `<end>` on the picture below - as soon as decoder generates this token we consider the answer to be complete (you can't use usual punctuation marks for this purpose cause their meaning can be different)
- `UNK` - "unknown token" - is used to replace the rare words that did not fit in your vocabulary. So your sentence `My name is guotong1988` will be translated into `My name is _unk_`.
- `PAD` - your GPU (or CPU at worst) processes your training data in batches and all the sequences in your batch should have the same length. If the max length of your sequence is 8, your sentence `My name is guotong1988` will be padded from either side to fit this length: `My name is guotong1988 _pad_ _pad_ _pad_ _pad_`



Nod.

Bidirectional LSTM

as in nlp , sometimes to understand a word we need not just to the previous word , but also to the coming word , like in this example



He said , "Teddy bears are on sale!"

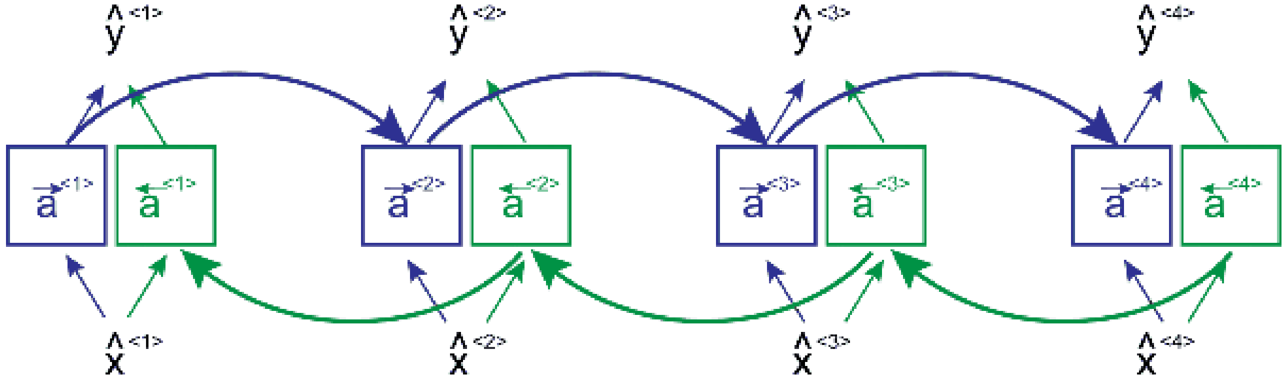
not part of person name

He said , "Teddy Roosevelt was a great President !"

part of person name

More Bidirectional

Bidirectional networks is a general architecture that can utilize any RNN model (normal RNN , GRU , LSTM)

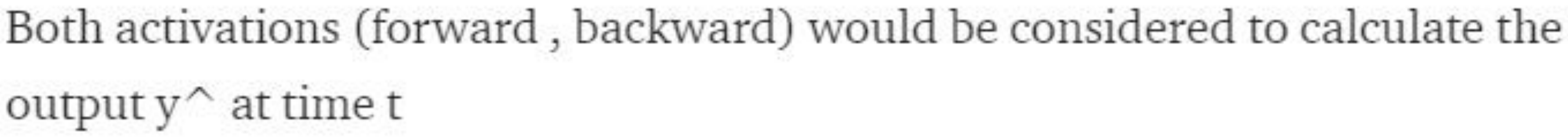


$\hat{y}^{<1>}$
$\hat{y}^{<2>}$
$\hat{y}^{<3>}$
$\hat{y}^{<4>}$
$\overrightarrow{a}^{<1>}$
$\overleftarrow{a}^{<1>}$
$\overrightarrow{a}^{<2>}$
$\overleftarrow{a}^{<2>}$
$\overrightarrow{a}^{<3>}$
$\overleftarrow{a}^{<3>}$
$\overrightarrow{a}^{<4>}$
$\overleftarrow{a}^{<4>}$
$\hat{x}^{<1>}$
$\hat{x}^{<2>}$
$\hat{x}^{<3>}$
$\hat{x}^{<4>}$

Connecting the backward cells

forward propagation for the 2 direction of cells

Even More Bidirectional



Both activations (forward , backward) would be considered to calculate the output y^ at time t

$$\hat{y}^{<t>} = g(W_y \, [\, \boxed{\overrightarrow{a}^{<t>}} \, , \boxed{\overleftarrow{a}^{<t>}} \,] + b_y)$$

Thank you!

Steve Nouri

Keep in touch and connect on