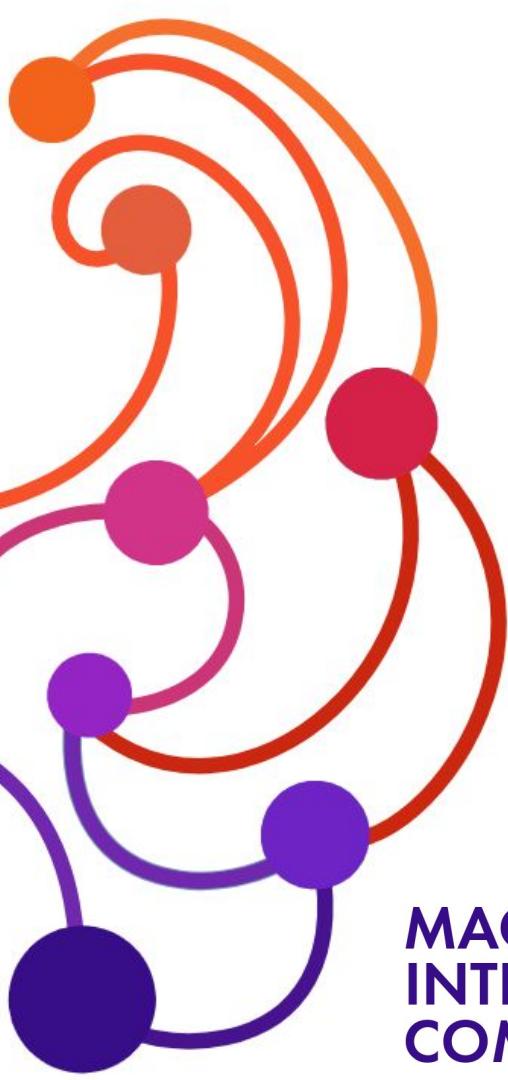


Intro to Natural Language Processing

[http://bit.ly/come join bumic slack](http://bit.ly/come_join_bumic_slack)

MACHINE
INTELLIGENCE
COMMUNITY



Intro to Natural Language Processing

Jennifer, Pauli, Duy

MACHINE
INTELLIGENCE
COMMUNITY

2/27/2019

Sunspring

[Screenplay](#)

[Article](#)



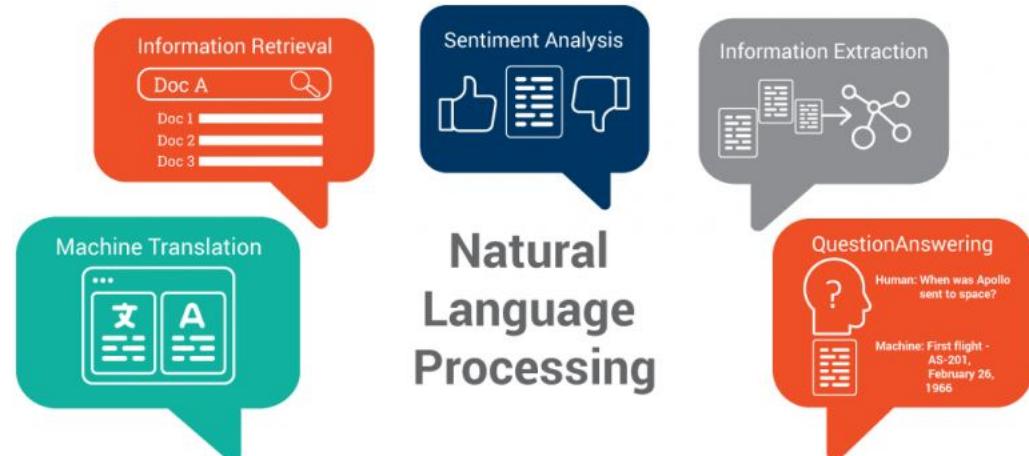


Natural Language Processing (NLP)

What is Natural Language Processing (NLP)?

Goal: To get computers to perform useful tasks involving human language

- enabling human-machine communication
- improving human-human communication
- processing of text or speech



Hardship of NLP

- Ambiguity

non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

segmentation issues

the New York-New Haven Railroad
the New York-New Haven Railroad

idioms

dark horse
get cold feet
lose face
throw in the towel

neologisms

unfriend
Retweet
bromance

world knowledge

Mary and Sue are sisters.
Mary and Sue are mothers.

tricky entity names

Where is *A Bug's Life* playing ...
Let It Be was recorded ...
... a mutation on the *for* gene ...

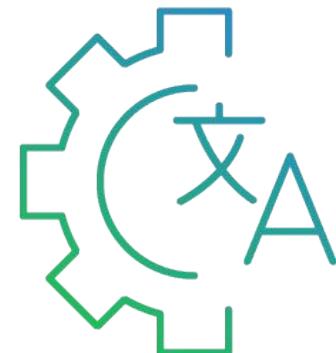




History

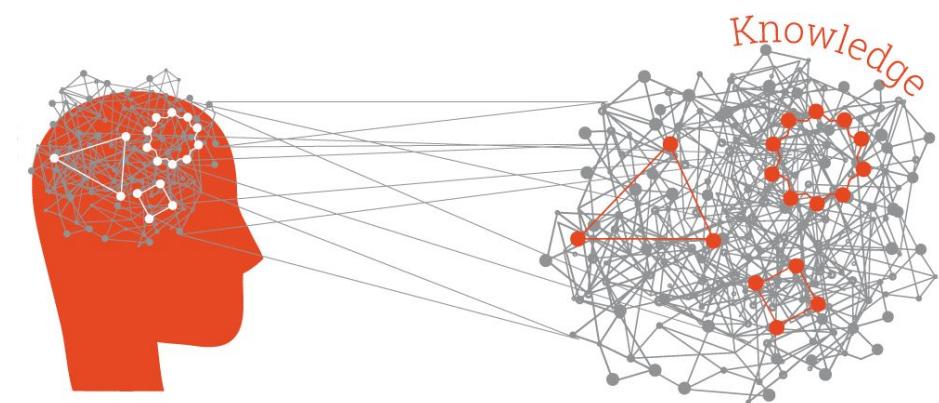
Phase 1: late 1940s to late 1960s

- **Machine Translation (MT)**
 - [1954] Automatic translation from Russian to English exhibited in the IBM-Georgetown Demonstration
 - [1950s] International conference on MT
 - [1961] The Teddington International Conference on Machine Translation of Languages and Applied Language Analysis
- Main accomplishment: Investigated many aspects of language
- Problem: Available computing resources were primitive



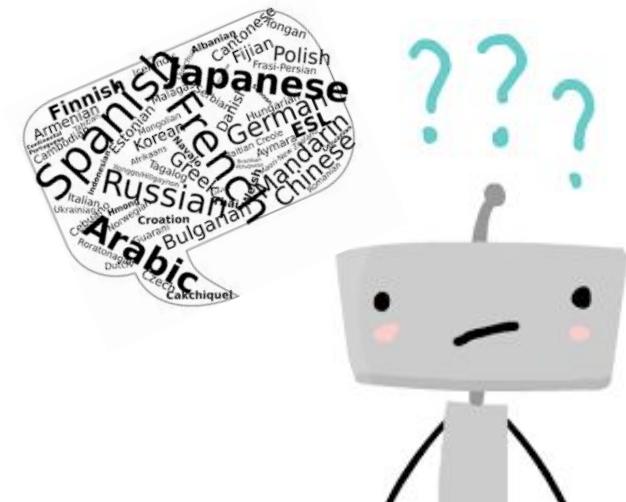
Phase 2: late 1960s to late 1970s

- More work done with **Artificial Intelligence (AI)**
 - [1961] BASEBALL Question-Answering system
- Developments in the field of **Semantics** (Meaning and Logic)
 - Delved into grammatical structures and frames or propositional units
 - Individual inputs to form a whole constituting generic model



Phase 2: late 1960s to late 1970s

- [1970s] Communicative function of language and the indirect meanings of linguistic expressions + Development of grammatical theory
- Emphasis on world knowledge and on its role in the construction and manipulation of meaning representations
- Problem: Failures of practical system building



Phase 3: late 1970s to late 1980s

- “Grammatico-Logical Phase”
 - Expansion of linguistic fields
 - Growth of logical process
- Computational grammar theory
 - Capturing the refinements of linguistics expression
 - [1992] SRI's Core Language Engine: powerful, general-purpose sentence processor
 - [1993] Discourse Representation Theory: Means of tackling more extended discourse within the framework



Phase 3: late 1970s to late 1980s

- Practical resources emerged
 - [1987] Alvey NATural Language Tools
 - More operational and commercial systems
- DARPA: Speech recognition and Information extraction



Phase 4: The 1990s

- Statistical language processing
 - Data analysis
 - Vast quantities of machine-readable data and machine power to handle it
 - Hidden Markov Modeling
- Speech Recognition (Transcription)

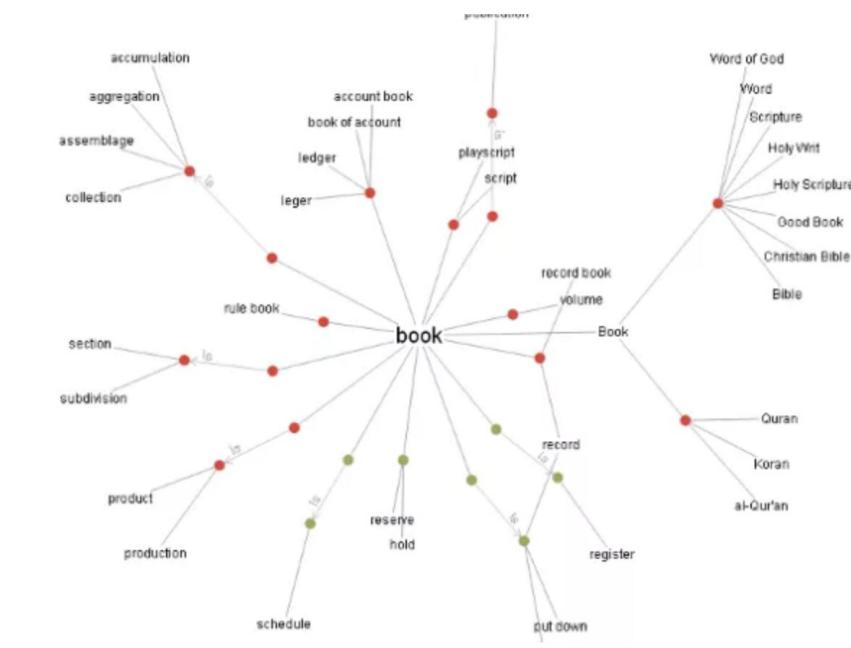


Phase 4: The 1990s

- **Evaluation**

- Global initiatives focused on multilingual tasks and issues
- [1999] WordNet
- British National Corpus

left context	KWIC	right context
the pipe . To e NN1/pipe /PUNC TO/to	find VVI/find	the actual cantilever AT/the JJ/actual NN1/cantilever
as help with II33/as NN1/help IW/with	finding VVG/find	housing , etc NN1/housing ,/PUNC RA/etc.
he could not be VM/could XX/not VBl/be	found VVN/find	distinctly , while RR/distinctly ./PUNC CS/while





Traditional NLP

Parsing Rule

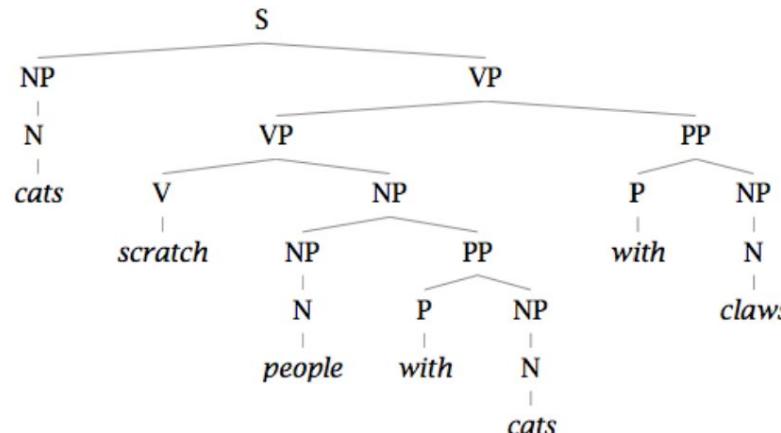
What is Parsing?

- First task for any NLP-based system: To read the text
- Depends on lexicon, categorization, grammar rules



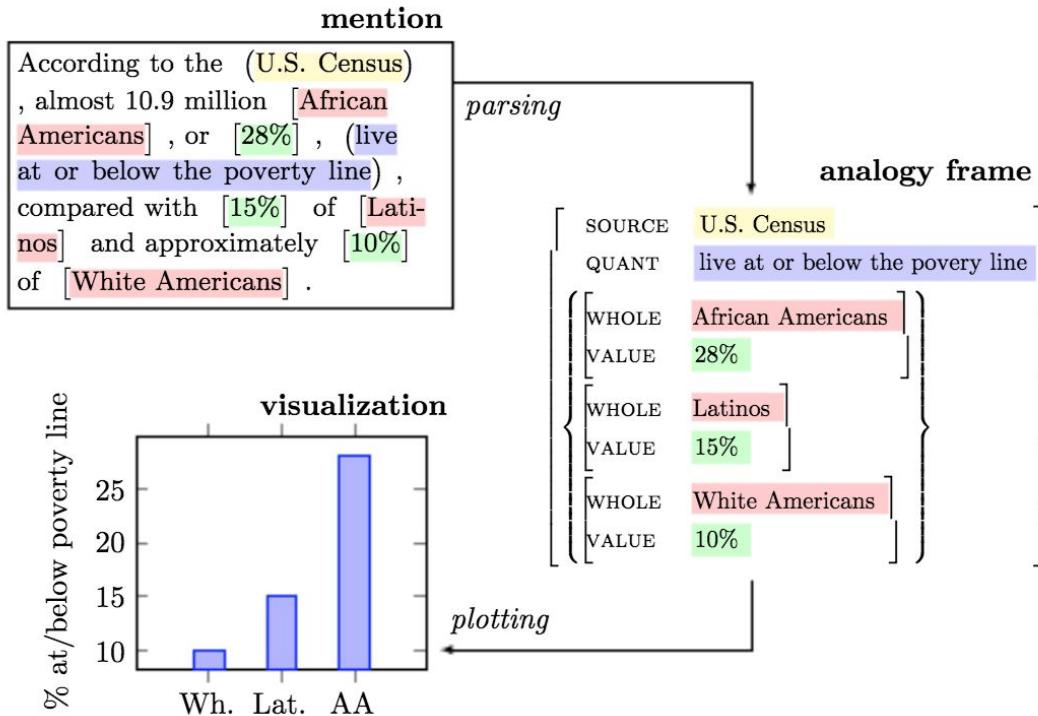
Pre 1990s: Classical NLP Parsing

- Method:
 - Wrote symbolic grammar and lexicon
 - Used grammar / proof systems to prove parse from words
- Annotated data: The Penn Treebank



Textual Analogy Parsing (TAP)

- Dan Jurafsky's Model



Statistical NLP

- Large sets of well established, reputable data collected

Rule based



Linguists constructs context based
on grammar

Noun

A.k.a. Kitten

Statistical

Model system with probability of
what normally happens in the
given context



30% chance that the word "Cat"
will appear in *word + [context]*

A decorative graphic in the background consists of several circular nodes in shades of orange, pink, and purple, connected by thin, curved lines that form a network-like structure.

Modern NLP

Integration of Deep Learning

[1960] Chatbot: Eliza

VS

[Now] AI Assistance

```
Welcome to
EEEEEE LL     IIII   ZZZZZZ  AAAA
EE    LL     II     ZZ  AA  AA
EEEEEE LL     II     ZZ  MAMAAA
EE    LL     II     ZZ  AA  AA
EEEEEE LLLLLL IIII   ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

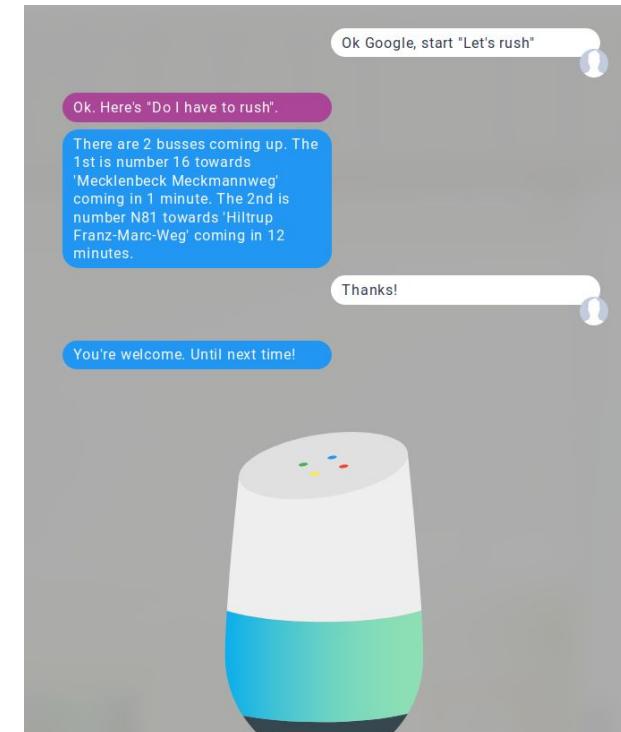
ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU: 
```

Person: "... I'm depressed all the time."

Eliza: I am sorry to hear that you are depressed.

Person: It's true. I am unhappy.

Eliza: Can you explain what made you unhappy?





Deep Learning in NLP

Word Embedding

- Basic idea of Word Embedding:

Mapping words or phrases to vectors of real number.

- “You shall know a word by the company it keeps. ”

--- J.R.Firth

- Walk, walking, walked



Word2Vec

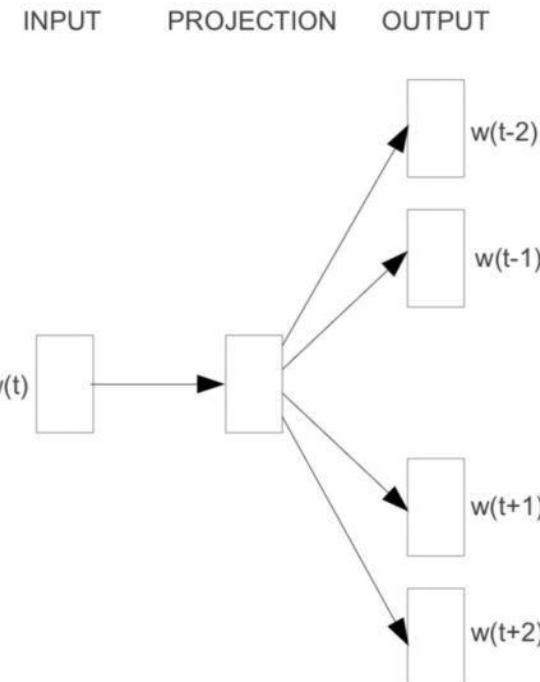
- Basic Idea:
 - Predict between every words and its context words
- Two Algorithms:
 - Skip-grams(SG):
Predict context words given target
 - Continuous Bag of Words(CBOW):
Predict target word from bag-of-words context



Skip-gram

For each step:

- $w(t)$: the input center word
- Window size: prediction range of each center words.



Skip-gram



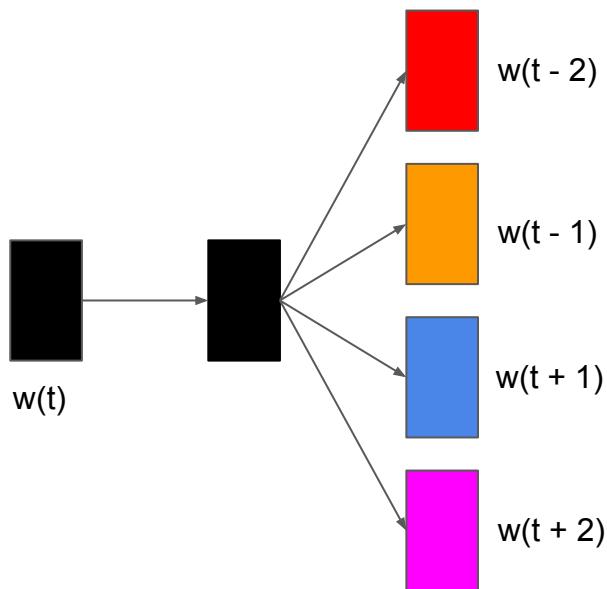
Summary of Word2Vec

- Go through each word of the whole corpus
- Predict surrounding words of each word
- Use Softmax to turn numbers into a **probability distribution**
- Choose vector of word that maximize the probability distribution.

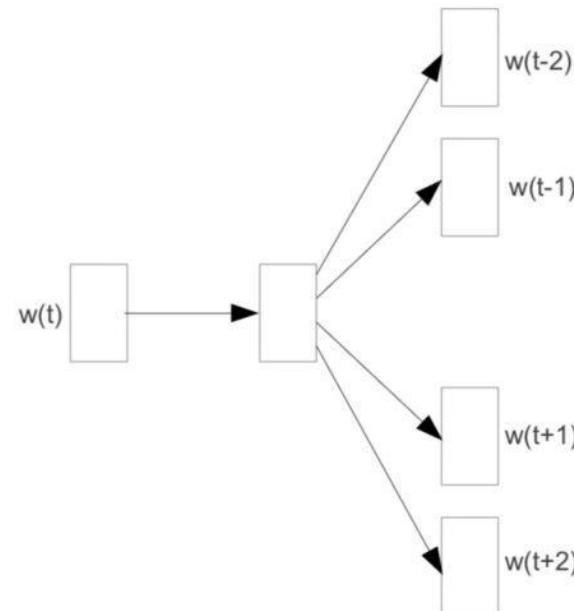


Example

“My biggest interest is machine learning.”



INPUT PROJECTION OUTPUT



Skip-gram

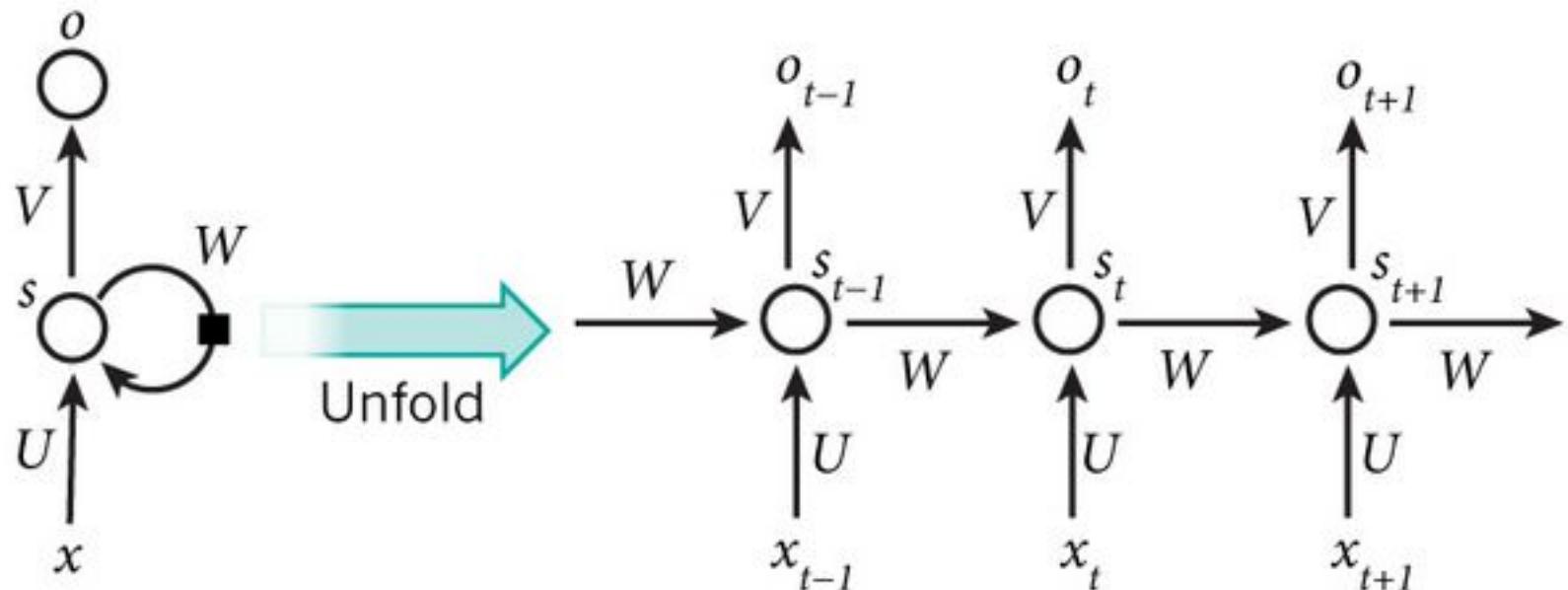
Recurrent Neural Network(RNN)

- **Hidden state** allows RNNs to store information about the past.
- **Non-linear dynamics** allows RNNs to update their hidden state.



Recurrent Neural Network

Given a list of word vectors: $X_1, X_2, X_3 \dots X_T$;



Problem of RNNs

The vanishing/exploding gradient problem:

As you go backwards through time, and you try
to send the air signal many time steps into
the past, signal will get too weak/strong

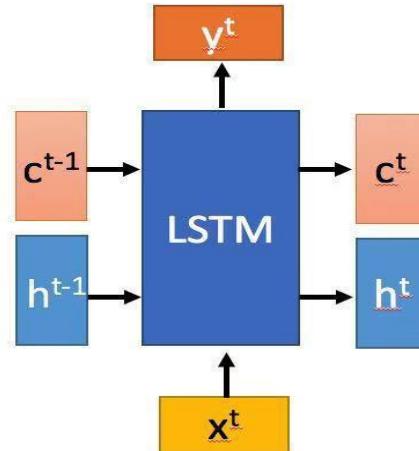
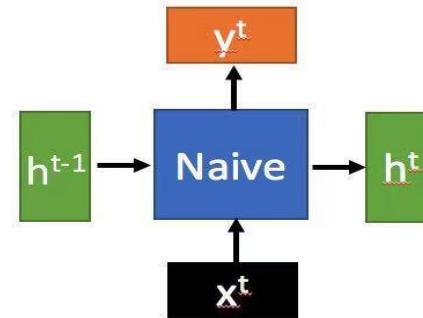


Long Short Term Memory(LSTM)

Cell State:

- Allow information to flow through unchanged
- Can be modified by LSTM

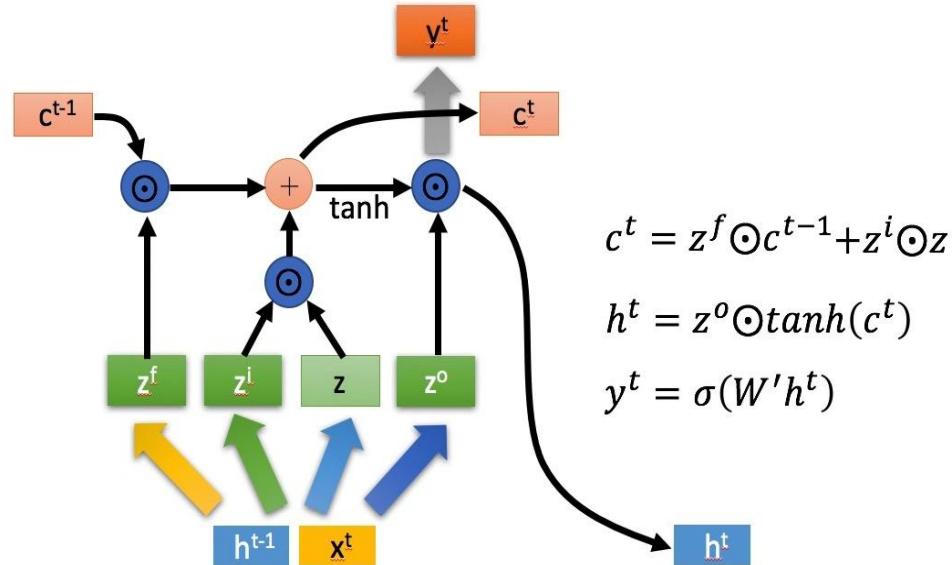
LSTM



Source: Hung-yi Lee
of NTU

Stages of LSTM

1. Using Z_f as a controller to decide which part of C_{t-1} need to be forgot
2. Using Z_i as a controller to decide which part of Z need to be remembered
3. Output



Source: Hung-yi Lee of NTU

Other Problems of RNNs

- Cannot capture phrases without prefix context
- Often capture too much of last words in the final vector



Convolutional Neural Network(CNN)

Basic Idea of the first layer of CNNs:

- Compute vectors for every possible phrases
- Regardless of whether phrase is grammatical
- Group them afterwards



Examples

- My biggest interest is machine learning.

My biggest/ biggest interest/ interest is/ is machine/ machine learning;

My biggest interest/ biggest interest is/ interest is machine/ is machine learning;

My biggest interest is/ biggest interest is machine/ interest is machine learning;

My biggest interest is machine/ biggest interest is machine learning.



Deficiency of CNNs

- Convenient for computer to do its work,
- **But** having inability to model long distance dependencies



Transformer



Attention Mechanism

Basic Idea:

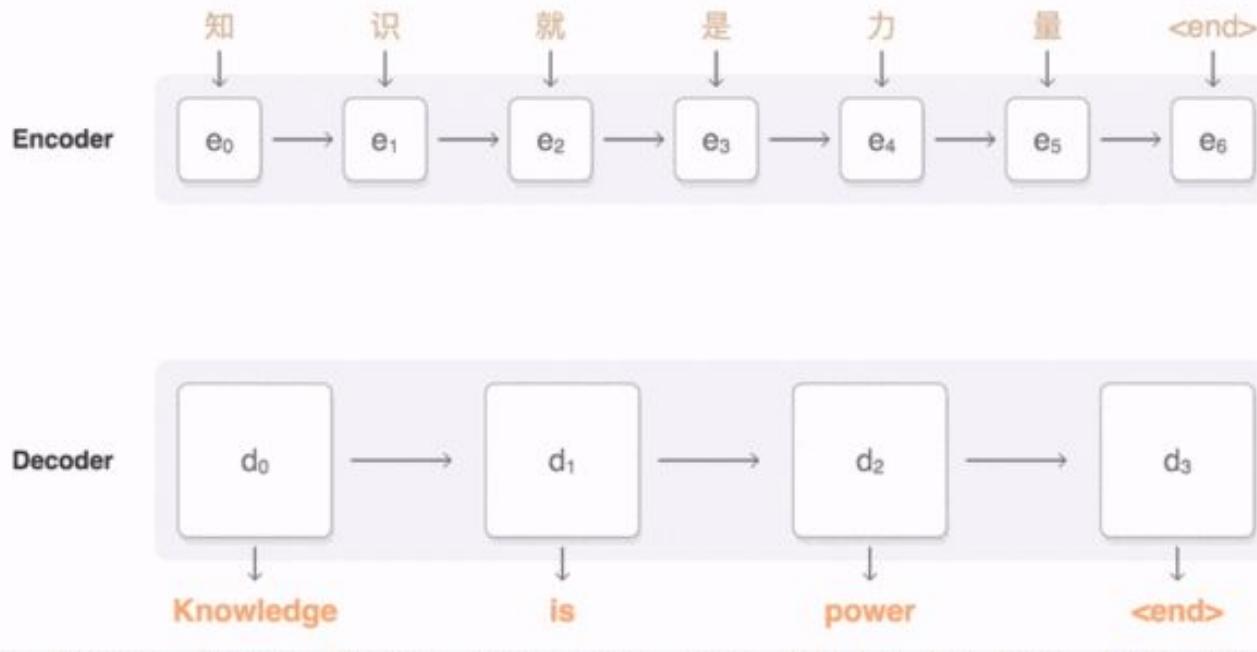
- A trick that enable NLP models to pay **attention** to relatively important information in a sources context.
- Weight each word
 - “I arrived at the bank after crossing the river.”

NOTICE:

Attention Mechanism is more of a trick rather than a complete model



Visualizing Attention Mechanism

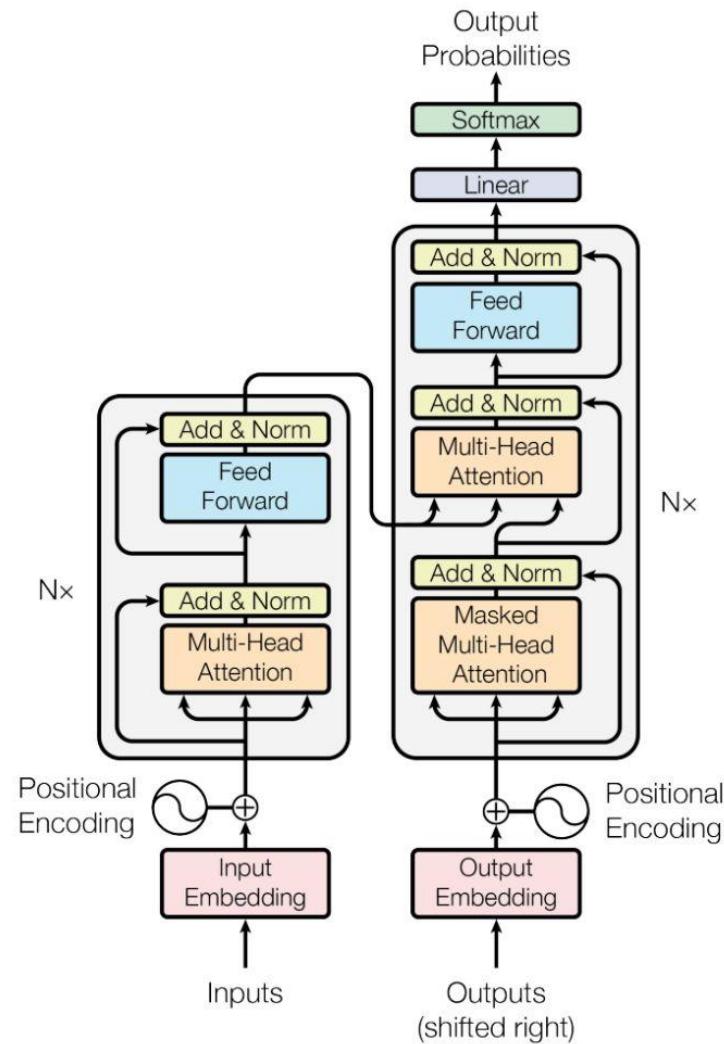


<https://google.github.io/seq2seq/>



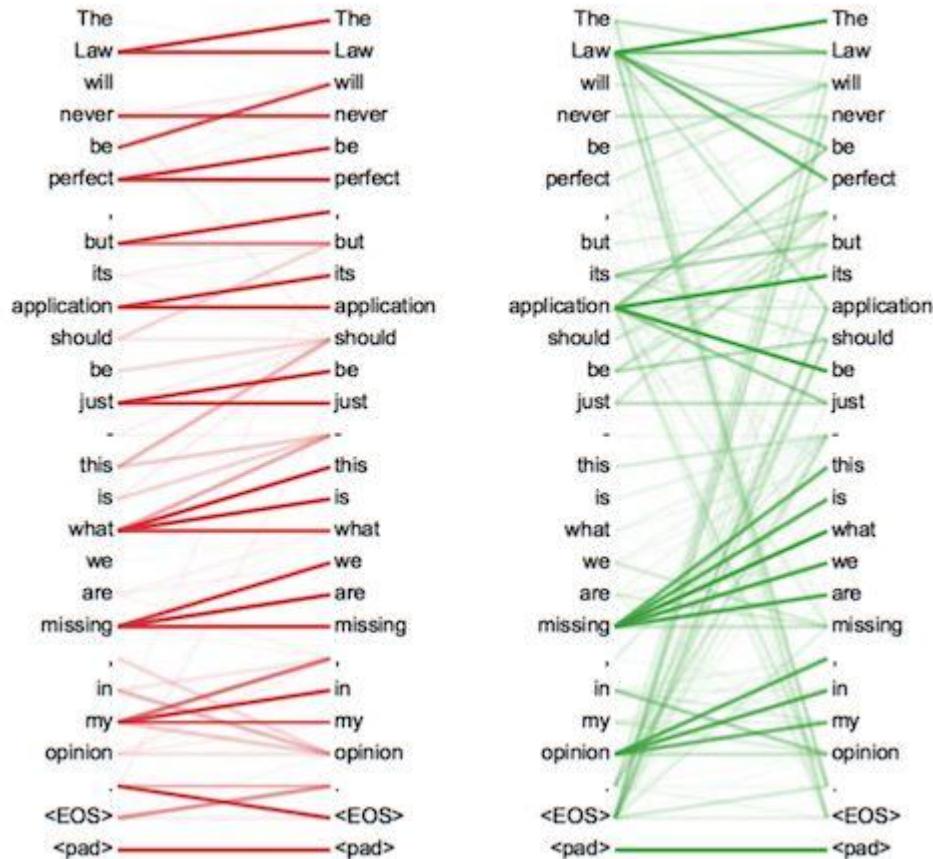
Transformer

- Multi-Head Attention
is multiple of self-attention



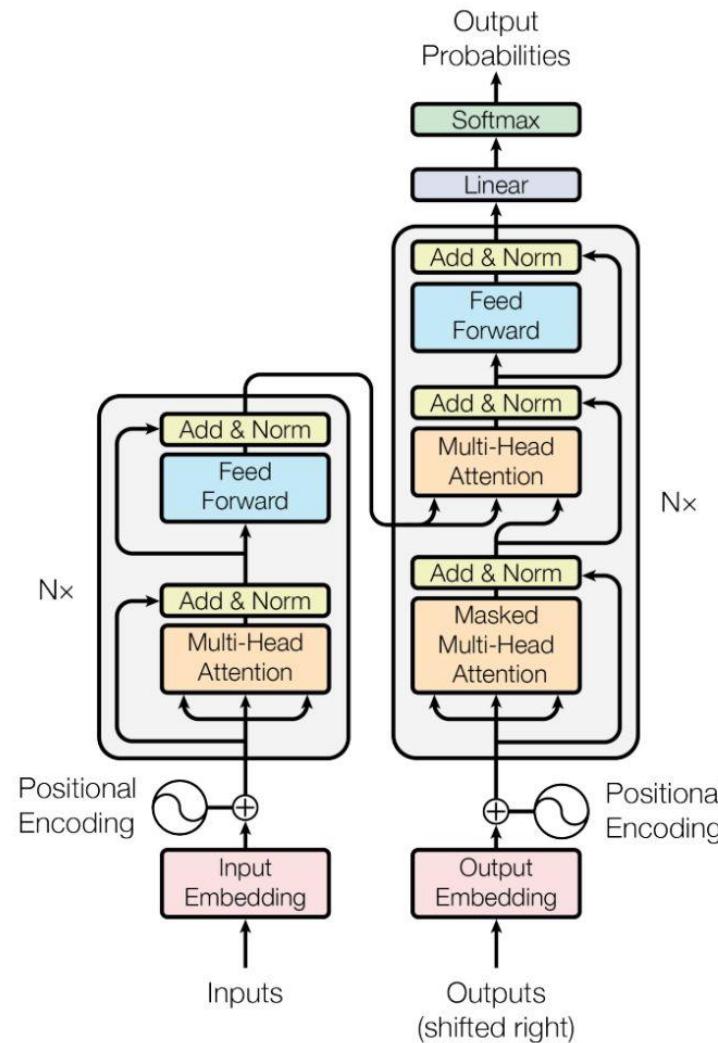
Multi-Head Attention

- Different head can have different interpretation of the source.
- Increasing the capacity of model



Transformer

- Masked Multi-Head Attention
 - When predict Tth word, cover words that didn't appear before and run Self-attention to Tth Word



Virtualizing transformer

<https://ai.googleblog.com>

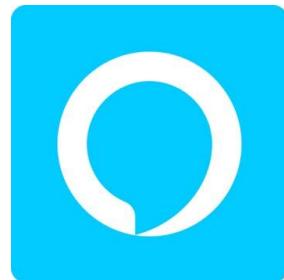
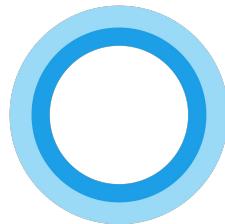




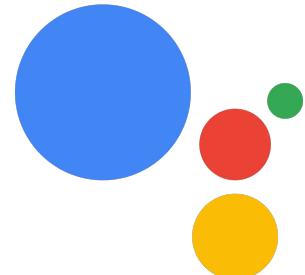
Applications

Virtual Assistants

- Every big company has one
- Main parts:
 - Speech recognition/synthesis
 - Conversational agent

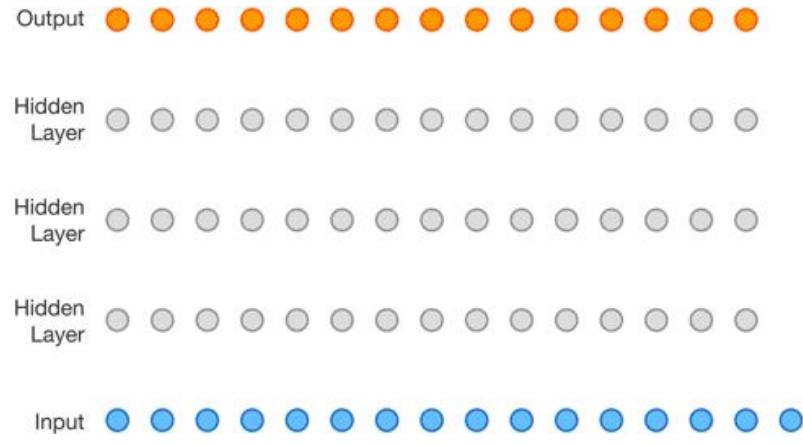


Bixby



Speech Recognition / Synthesis

- Traditionally, ASR is done by matching the spectrogram to a dictionary of phoneme.
- Better performance is done by running the spectrogram through CNN.



[WaveNet](#)

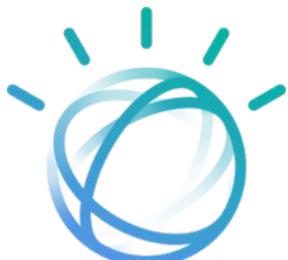
Conversational Agent

- Exists early in the history of NLP.
Chatbots like ELIZA
- Early chatbot uses pattern matching and rules to response to users.
- Deep learning allows building more natural and expressive conversational agents

```
Welcome to
EEEEE   LL      IIII    ZZZZZZ  AAAAA
EE      LL      II      ZZ      AA      AA
EEEEE   LL      II      ZZZ     AAAAAAAA
EE      LL      II      ZZ      AA      AA
EEEEE   LLLLLL  IIII    ZZZZZZ  AA      AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU: ■
```



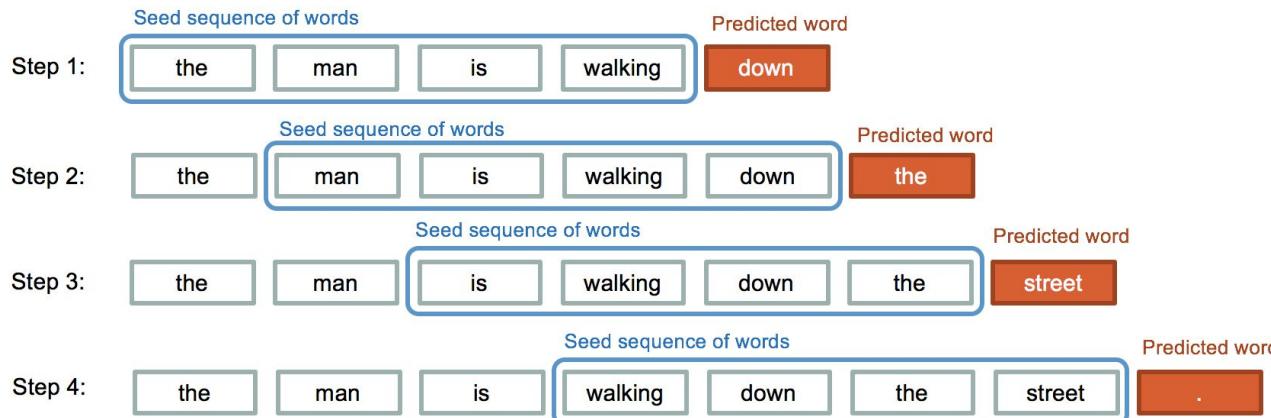
Machine Translation

- One of the core of NLP from the beginning
- Traditionally uses statistical machine translation
- More are using Neural Translation
- Still we are facing problems with low-resource languages.



Content generation

- Deep Learning algorithms enable AI to produce content with human level of linguistics fluency.
 - AI writing news articles
 - AI writing scripts



[David Campion](#)

References and Further Reading

1. Sunspring (movie): <https://www.youtube.com/watch?v=LY7x2lhqjmc>
2. Jones, Natural language processing: a historical review:
<https://www.cl.cam.ac.uk/archive/ksj21/histdw4.pdf>
3. Stanford University CS224n:
https://www.youtube.com/playlist?list=PL3FW7Lu3i5Jsnh1rnUwq_TcylNr7EkRe6
4. CMU Neural Nets for NLP: https://www.youtube.com/user/neubig/videos?disable_polymer=1
5. Attention is all you need: <https://arxiv.org/abs/1706.03762>

Upcoming Events

- Society and ML
 - Wednesday, March 6th. 7pm - 8pm
- Paper discussion
 - Tuesday, March 5th.
- ML Series Part 2: Intro to Clustering Algorithms
 - After Spring break

