Lectures on Natural Language Processing

1. General Introduction

Karl Stratos

Modern Natural Language Processing (NLP)

NLP is everywhere

Other examples?

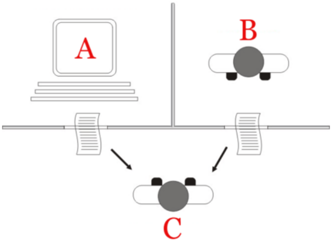
Short-Term Goals of NLP

Make machines understand human language to do useful tasks

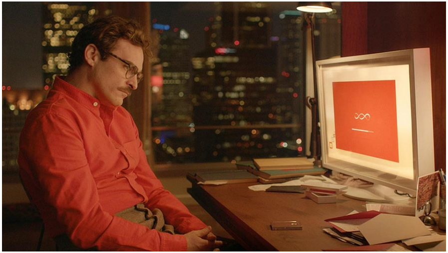
Countless applications: machine translation (MT), personal assistant, crucial component in any AI system (e.g., autonomous driving)

Long-Term Goals of NLP

Make machines human-level (or super-human?) intelligent



The Turing test



Her (2013)

Intelligence $\equiv$ Language?

Some History

- 1950: Alan Turing proposes the Turing test
- 1954: Georgetown–IBM experiment (rule-based MT)

  "Within three or five years, machine translation will be a solved problem"

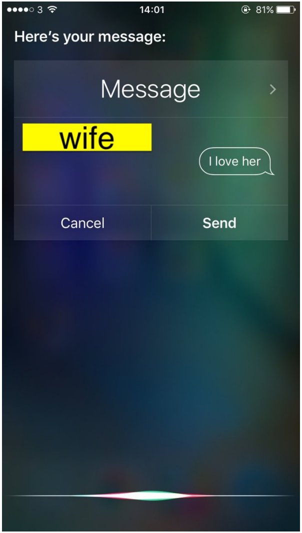
- 50-90s: focus on rule-based AI systems (e.g., SHRDLU)
- **From early 90s: Rise of statistical/data-driven NLP**
  - IBM: statistical MT and speech recognition

    "Every time I fire a linguist, the performance of the speech recognizer goes up" -Fred Jelinek

  - UPenn/AT&T: statistical techniques for tagging and parsing
- 2011: IBM Watson wins *Jeopardy!* against human champions
- **From early 2010s: Rise of deep learning for NLP**
  - "Human-level" MT: The Great A.I. Awakening (*NYT*, 2016)
  - "Human-level" conversation: Google Duplex (2018)
  - Language model considered "sentient" (*Insider*, 2022)

Reality

"Hey Siri, tell my wife I love her"

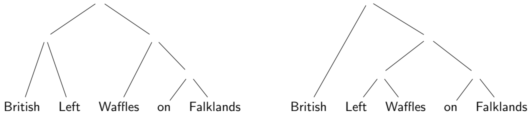


Why NLP is Hard: Ambiguity

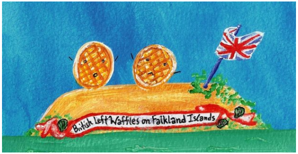
Actual headline in *Guardian* (1982)

"British Left Waffles on Falklands"

- **Syntactic ambiguity**

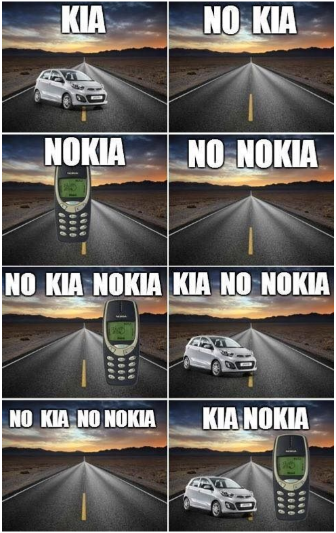


- **Lexical ambiguity**: Every single word
- **Semantic ambiguity**



Why NLP is Hard: Discrete Signals

Nonsmooth: A single word can completely change the meaning



Why NLP is Hard: Discrete Signals (Cont.)

Image

= **Tiger**

Language

negative

positive

“failed to not disappoint”

negative

Why NLP is Hard: World Knolwedge, Prejudice

Winograd (1972): Who does "they" refer to?

- The city councilmen refused the demonstrators a permit because they <u>feared</u> violence.
- The city councilmen refused the demonstrators a permit because they <u>advocated</u> violence.

Language Capabilities: Innate or Learned?

(Acknowledgement: This slide is adapted from Nathan Srebro's lecture.)

Noam Chomsky

The ability to learn grammars is **hard-wired** into the brain. It is not possible to "learn" linguistic ability—rather, we are born with a brain apparatus specific to language representation.

There exists some "universal" learning algorithm that can learn ***anything***: language, vision, speech, etc. The brain is based on it, and we're working on uncovering it. (Hint: the brain uses neural networks)

Geoffrey Hinton

Is Language All We Need for Intelligence?

Yann LeCun

A system trained on language alone will never approximate human intelligence... The knowledge underlying language understanding can only be acquired non-linguistically.

Language modeling alone should ultimately be sufficient for extracting understanding... If defining the true distribution of novels requires understanding, then fully solving the language modeling problem requires the ability to understand.

David McAllester

Core Topic 1: Language Models

Task: Predict a likely future text given a context

$\underbrace{\texttt{Alan Turing is the founder of}}_{\text{context}} \mapsto \underbrace{\begin{array}{l}\texttt{computer science}\ (0.98) \\ \texttt{medicine}\ (0.01) \\ \texttt{rock and roll (0.00)} \\ \ldots\end{array}}_{\text{possible completions}}$

Training requires *no special annotation* (e.g., can use all the web)

Modern NLP: recipe

1. Train a huge language model on huge text data
2. "Prompt"/finetune the model for any task of interest:
    - `Who founded information theory?` (question answering)
    - `Translate 동그라미 into English.` (translation)
    - `Write a Python program for computing Fibonacci numbers.` (coding)
    - `Draw an ascii art of a cat.` (image generation)

Demo

ChatGPT

Core Topic 2: Retrievers

Task: Given a query, retrieve the most relevant entries in a knowledge base

Canonical knowledge base: **Wikipedia**

(~ 6 million articles)

Also others

Most NLP problems can be approached with a combination of pretrained language models and retrievers.

Plan

Foundations

- Cross-entropy loss and gradient-based optimization
- Linear and nonlinear classifiers (i.e., neural networks)
- The transformer architecture

Core topics

- Pretrained language models
- Retrievers

Additional topics

- Latent-variable models
- Structured prediction

Applications: simple text classification, translation, summarization, entity linking, coreference resolution, and many others