

CS 533: Natural Language Processing

Introduction

Karl Stratos



Rutgers University

Modern Natural Language Processing (NLP)

NLP is everywhere



Other examples?

Short-Term Goals

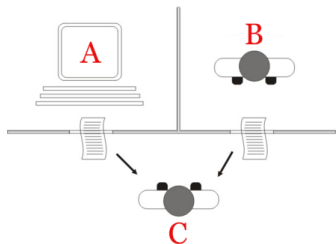
Make machines understand human language



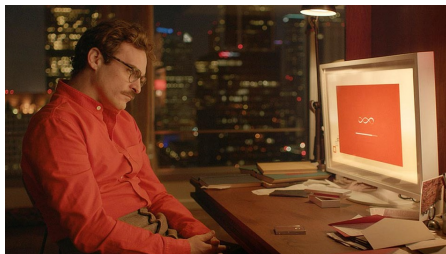
Countless applications: machine translation (MT), personal assistant, crucial component in any AI system (e.g., autonomous driving)

Long-Term Goals

Make machines as intelligent and conscious as humans (or more)



The Turing test



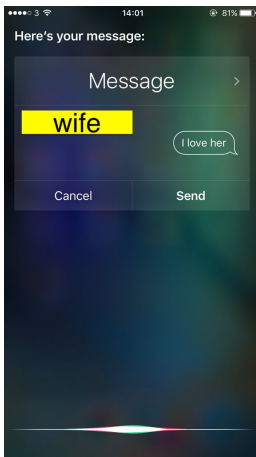
Her (2013)

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)

Reality

“Hey Siri, tell my wife I love her”

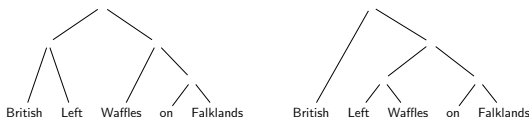


Why NLP is Hard: Ambiguity

Actual headline in *Guardian* (1982)

“British Left Waffles on Falklands”

- ▶ **Syntactic ambiguity**



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

- ▶ **Semantic ambiguity**



Why NLP is Hard: Nonsmoothness

A single word can completely change the meaning



Jack Black **Black Jack Black**



Black Jack Black playing **BlackJack**



Jack Black Playing **BlackJack**
with **BlackJack Black**

Why NLP is Hard: World Knowledge

Winograd (1972)

- ▶ The city councilmen refused the demonstrators a permit because **they** feared violence.

- ▶ The city councilmen refused the demonstrators a permit because **they** advocated violence.

Course Topics

- ▶ Language modeling
 - n -gram models, log-linear models, neural language models (feedforward, recurrent)
- ▶ Conditional language modeling
 - Attention-based models, translation, summarization
- ▶ Text classification
 - Naive Bayes classifier
- ▶ Structured prediction
 - Hidden Markov models, probabilistic context-free grammars
- ▶ Unsupervised learning
 - Expectation maximization algorithm, variational autoencoders
- ▶ Special topics on various applications
 - Information extraction, question answering, dialogue, grounding

(Subject to change)