Lecture 01: Introduction

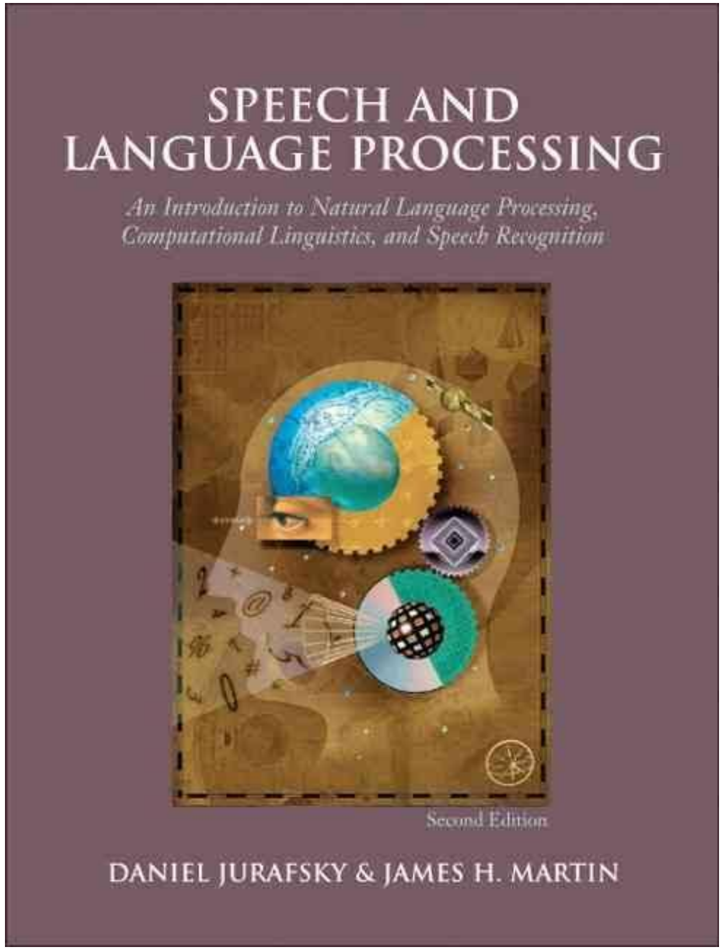
Instructor: Dr. Hossam Zawbaa

Image by
kirkh.deviantart.com

Course Syllabus

- **Introduction.**
- **Probability**
- **Language Modeling.**
- **Speech phonetics.**
- **Automatic speech recognition.**
- **Template matching.**
- **HMMs.**
- **Acoustic Modeling.**
- **Voice building.**
- **Multilingual Speech Processing.**

Textbook



An electronic copy is also available free online:

https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Grading

- **Lab activities and assignments: 10%**
- **Final project: 20%**
- **Mid-term exam: 20%**
- **Final exam: 50%**
- **Extra credit: 5% for students who participate actively on the lectures.**
- **Extra credit: 5% for students who obtain the best final project.**

Prerequisites

- **Good knowledge of digital signal processing**
- **Statistics and probability**
- **Basic knowledge of machine learning**
- **Basic knowledge of natural language processing**
- **Experience with Matlab will help**

Automatic Speech Recognition

- **Speech recognition** is the inter-disciplinary sub-field of **computational linguistics** that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers.
- **Speech recognition** is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format.
- It is used to identify the words a person has spoken or to authenticate the identity of the person speaking into the system.
- **Automatic speech recognition** is also known as **automatic voice recognition** (AVR), voice-to-text or simply **speech recognition**.

Definitions

- Speech Recognition
  - Speech-to-Text
    - Input: a wave file,
    - Output: string of words
- Speech Synthesis
  - Text-to-Speech
    - Input: a string of words
    - Output: a wave file

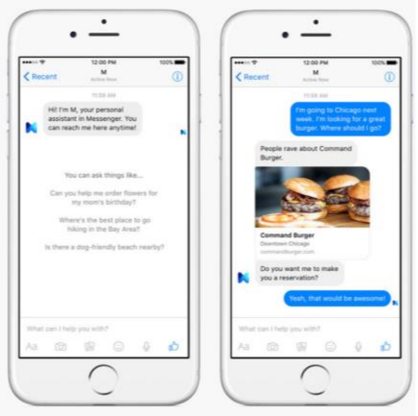
Automatic Speech Recognition (ASR) Automatic Speech Understanding (ASU)

- **Applications**
  - Voice Dictation
  - Telephone-based Information (directions, air travel, banking, etc)
  - Hands-free (in car)
  - Second language (accent reduction)
  - Audio archive searching
  - Linguistic research
    - Automatically computing word durations, etc.

An exciting time for spoken language processing

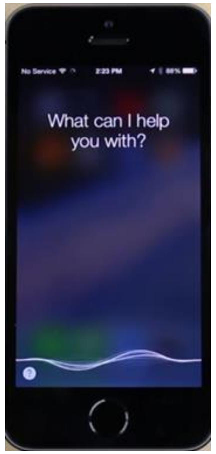
Amazon Echo
2015

Google Home
2016

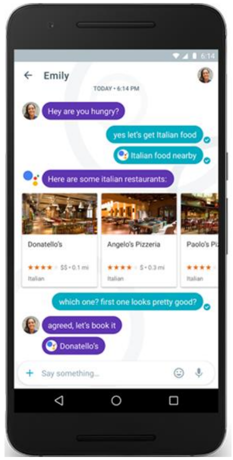


Facebook M
2015

Anki Cozmo
2016

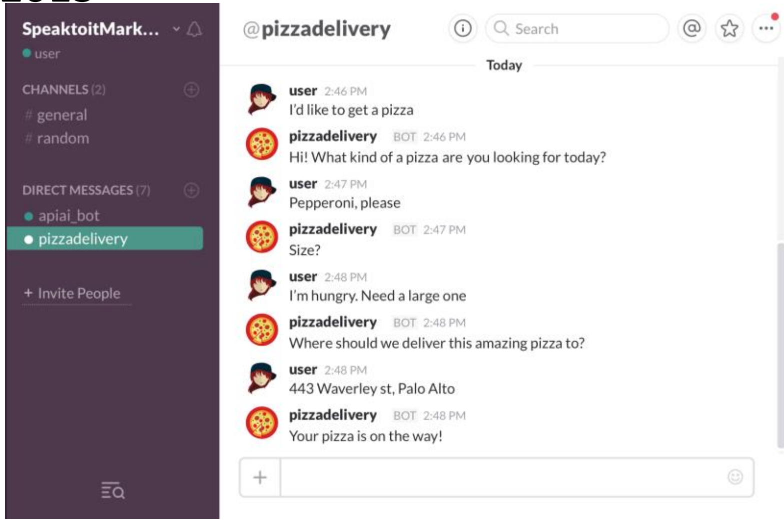


Apple
Siri
2011



Google
Assistant
2016

Microsoft
Cortana
2014



Slack Bot API
2015

Applications of Speech Synthesis/Text-to-Speech (TTS)

- Games
- Telephone-based Information (directions, air travel, banking, etc)
- Eyes-free (in car)
- Reading/speaking for disabled
- Education: Reading tutors

Applications of Speaker/Language Recognition

- Language recognition for call routing
- Speaker Recognition:
  - Speaker verification (binary decision)
    - Voice password, telephone assistant
  - Speaker identification (one of N)
    - Criminal investigation

History: foundational insights 1900s-1950s

- Automaton:
  - Markov 1911
  - Turing 1936
  - McCulloch-Pitts neuron (1943)
  - Shannon (1948) link between automata and Markov models
- Human speech processing
  - Fletcher at Bell Labs (1920’s)
- Probabilistic/Information-theoretic models
  - Shannon (1948)

History: early ASR systems

- 1950's: Early Speech recognizers
  - 1952: Bell Labs single-speaker digit recognizer
    - Measured energy from two bands (formants)
    - Built with analog electrical components
    - 2% error rate for single speaker, isolated digits
  - 1958: Dudley built classifier that used continuous spectrum rather than just formants
  - 1959: Denes ASR combining grammar and acoustic probability
- 1960's
  - FFT - Fast Fourier transform (Cooley and Tukey 1965)
  - LPC - linear prediction (1968)
  - 1969 John Pierce letter "Whither Speech Recognition?"
    - Random tuning of parameters,
    - Lack of scientific rigor, no evaluation metrics
    - Need to rely on higher level knowledge

ASR: 1970's and 1980's

- Hidden Markov Model 1972
  - Independent application of Baker (CMU) and Jelinek/Bahl/Mercer lab (IBM) following work of Baum and colleagues at IDA
- ARPA project 1971-1976
  - 5-year speech understanding project: 1000 word vocab, continuous speech, multi-speaker
- 1980's+
  - Large corpus collection
    - Resource Management
    - Wall Street Journal

State of the Art

- ASR
  - speaker-independent, continuous, no noise, world's best research systems:
    - Human-human speech: ~13-20% Word Error Rate (WER)
    - Human-machine speech: ~3-5% WER

LVCSR Overview

- Large Vocabulary Continuous (Speaker-Independent) Speech Recognition
  - ~64,000 words
  - Speaker independent (vs. speaker-dependent)
  - Continuous speech (vs isolated-word)
  - Build a statistical model of the speech-to-words process
  - Collect lots of speech and transcribe all the words
  - Train the model on the labeled speech
  - Paradigm: Supervised Machine Learning + Search

Introduction to Probability

- Experiment (trial)
  - Repeatable procedure with well-defined possible outcomes
- Sample Space (S)
    - the set of all possible outcomes
    - *finite or infinite*
  - Example
    - coin toss experiment
    - possible outcomes: S = {head, tail}
  - Example
    - die toss experiment
    - possible outcomes: S = {1,2,3,4,5,6}

Introduction to Probability

- Definition of sample space depends on what we are asking
  - Sample Space (S): the set of all possible outcomes
  - Example
    - die toss experiment for whether the number is even or odd
    - possible outcomes: {even, odd}
    - *not* {1,2,3,4,5,6}

More definitions

- Events
  - an ***event*** is any subset of outcomes from the ***sample space***
- Example
  - die toss experiment
  - let A represent the event such that the outcome of the die toss experiment is divisible by 3
  - A = {3, 6}
  - A is a subset of the sample space S= {1, 2, 3, 4, 5, 6}

Introduction to Probability

- Some definitions
  - Counting
    - suppose operation o_i can be performed in n_i ways, then
    - a sequence of k operations $o_1 o_2 ... o_k$
    - can be performed in $n_1 \times n_2 \times ... \times n_k$ ways
  - Example
    - die toss experiment, 6 possible outcomes
    - two dice are thrown at the same time
    - number of sample points in sample space = $6 \times 6 = 36$

Definition of Probability

- The probability law assigns to an event a nonnegative number Called P(A), also called the probability A
- That encodes our knowledge or belief about the collective likelihood of all the elements of A
- Probability law must satisfy certain properties

Probability Axioms

- Nonnegativity
  - P(A) >= 0, for every event A
- Additivity
  - If A and B are two disjoint events, then the probability of their union satisfies:
  - P(A U B) = P(A) + P(B)
- Normalization
  - The probability of the entire sample space S is equal to 1, I.e. P(S) = 1.

An example

- An experiment involving a single coin toss
- There are two possible outcomes, H and T
- Sample space S is {H, T}
- If coin is fair, should assign equal probabilities to 2 outcomes
- Since they have to sum to 1
- P({H}) = 0.5
- P({T}) = 0.5
- P({H, T}) = P({H}) + P({T}) = 1.0

Probability definitions

- In summary:

$$P(E) = \frac{\text{number of outcomes corresponding to event E}}{\text{total number of outcomes}}$$

Probability of drawing a heart from 52 well-shuffled playing cards:

$$\frac{13}{52} = \frac{1}{4} = 0.25$$

Probabilities of two events

- If two events A and B are independent
- Then
  - P(A and B) = P(A) x P(B)

- If flip a fair coin twice
  - What is the probability that they are both heads?

How about non-uniform probabilities? An example

- A biased coin,
  - twice as likely to come up tail as head,
  - is tossed twice
- What is the probability that at least one head occurs?
- Sample space = {hh, ht, th, tt} (h = head, t = tail)
- Sample points/probability for the event:
  - ht 1/3 x 2/3 = **2/9** hh 1/3 x 1/3 = **1/9**
  - th 2/3 x 1/3 = **2/9** tt 2/3 x 2/3 = 4/9
- Answer: 5/9 = ≈0.56 (*sum of weights in **red***)

Conditional Probability

- A way to reason about the outcome of an experiment based on partial information
  - In a word guessing game the first letter for the word is a “t”. What is the likelihood that the second letter is an “h”?
  - How likely is it that a person has a disease given that a medical test was negative?
  - A spot shows up on a radar screen. How likely is it that it corresponds to an aircraft?
- We need a new probability law that gives us the conditional probability of A given B
- **P(A|B)**

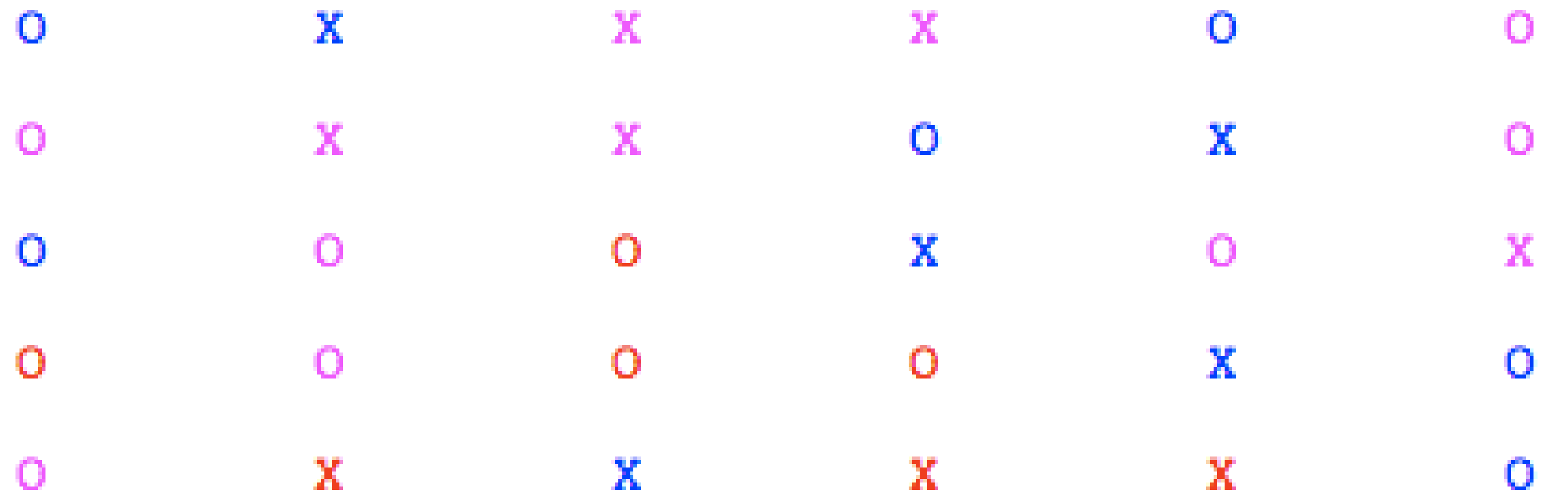
Conditional Probability

- let A and B be events
- p(B|A) = the *probability* of event B *occurring given* event A *occurs*
- ***definition:*** **p(B|A) = p(A ∩ B) / p(A)**

S

Conditional probability

- One of the following 30 items is chosen at random
- What is P(X), the probability that it is an X?
- What is P(X|red), the probability that it is an X given that it is red?



Bayes Theorem

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

- Swap the conditioning
- Sometimes easier to estimate one kind of dependence than the other

How many words?

- I do mainly business data processing
  - Fragments
  - Filled pauses
- Are **cat** and **cats** the same word?
- Some terminology
  - **Lemma**: a set of lexical forms having the same stem, major part of speech, and rough word sense
    - Cat and cats = **same lemma**
  - **Wordform**: the full inflected surface form.
    - Cat and cats = **different wordforms**

Moving toward language
Probability and part of speech tags

- What’s the probability of a random word (from a random dictionary page) being a verb?

- How to compute each of these:
  - ➢All words = just count all the words in the dictionary
  - ➢# of ways to get a verb: number of words which are verbs!
  - ➢If a dictionary has 50,000 entries, and 10,000 are verbs
    - ➢**P(V) is 10000/50000 = 0.2**