CS 1571 Introduction to AI
Lecture 27

Appied AI topics

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

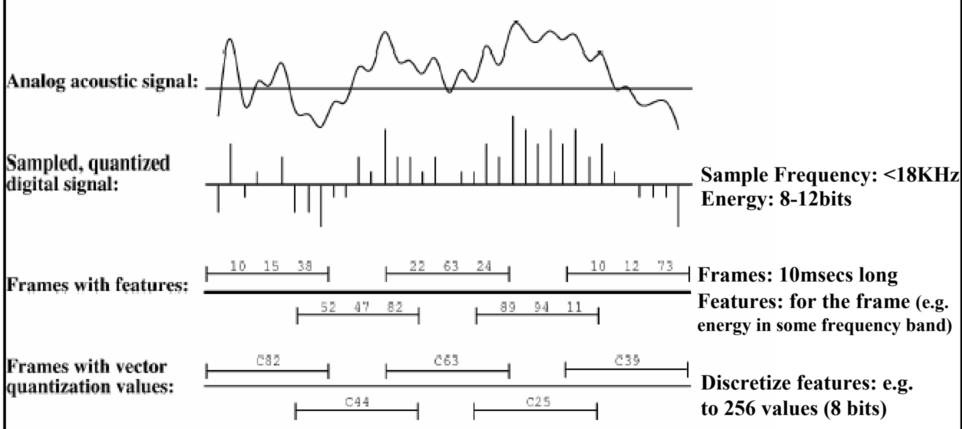
Topics in AI

Five main areas:
- **Problem solving and search**
- **Logic and knowledge representations**
- **Planning**
- **Uncertainty**
- **Learning**

Many other topics:
- **AI programming languages**
- **Speech recognition** ⬅
- **Natural language processing** ⬅ **Next**
- **Image understanding** ⬅
- **Robotics, ….**

Speech recognition

- **Objective**: take acoustic signal and convert it to text



Speech recognition

- **We want to determine the sequence of words that is most probable given the input signal**

  $P(wordseq = \mathbf{w} \mid signal = \mathbf{s})$
- It is easier to define an **acoustic model** that relates:

  $P(signal = \mathbf{s} \mid wordseq = \mathbf{w})$
- This is like a diagnosis problem , we can use the Bayes rule:

$$P(wordseq = \mathbf{w} \mid signal = \mathbf{s}) = \frac{P(signal = \mathbf{s} \mid wordseq = \mathbf{w}) P(wordseq = \mathbf{w})}{P(signal = \mathbf{s})}$$

- Assume we have multiple possible word sequences:

  $\mathbf{w}^1, \mathbf{w}^2, \ldots \mathbf{w}^k$
- **The best word sequence**:

$$\arg\max_{\mathbf{w}^i} P(signal = \mathbf{s} \mid wordseq = \mathbf{w}^i) P(wordseq = \mathbf{w}^i)$$

Speech recognition

- We need to define:

  $P(signal = \mathbf{s} \mid wordseq = \mathbf{w})$ and $P(wordseq = \mathbf{w})$

  for all possible word and signal sequences
- **Defining the probability:** $P(wordseq = \mathbf{w})$ $\quad \mathbf{w} = w_1 w_2 \ldots w_n$

  $P(wordseq = w_1 w_2 \ldots w_n) = P(w_1) P(w_2 \mid w_1) \ldots P(w_n \mid w_1 w_2 \ldots, w_{n-1})$
  - By the **chain rule**
- **Simplifications:**
  - **Unigram model:** a probability of each word is independent of the previous word

  $P(wordseq = w_1 w_2 \ldots w_n) = P(w_1) P(w_2) P(w_3) \ldots P(w_n)$
  - **Bigram model:** only the previous word matters

  $P(wordseq = w_1 w_2 \ldots w_n) = P(w_1) P(w_2 \mid w_1) P(w_3 \mid w_2) \ldots P(w_n \mid w_{n-1})$

Speech recognition

- **Defining the probability:** $P(signal = \mathbf{s} \mid wordseq = \mathbf{w})$

  $\mathbf{s} = s_1 s_2 s_3 \ldots s_m$ $\qquad \mathbf{w} = w_1 w_2 \ldots w_n$
- **Two simplifications:**
  1. **Define signal signatures for individual words**

     $P(\mathbf{s} = s_1 s_2 \ldots s_j \mid word = w_i)$
  2. **Divide the acoustic word models into a sequence of phones and define signal signature models for phones**

     $P(\mathbf{p} = p_1 p_2 \ldots p_u \mid word = w_i)$

     $P(\mathbf{s} = s_1 s_2 \ldots s_r \mid phone = p_q)$

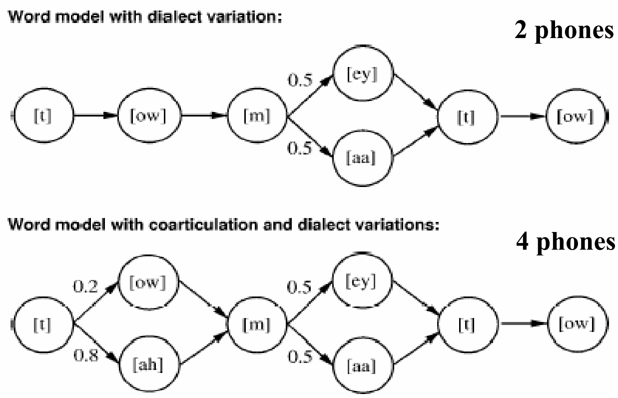
Conditional probabilities of sequences modeled most often as:

- **Hidden Markov Models (HMMs)**

Speech recognition

HMM models of words $P(\mathbf{p} = p_1 p_2 \ldots p_u \mid word = w_i)$

- **Example:** **word:** tomato



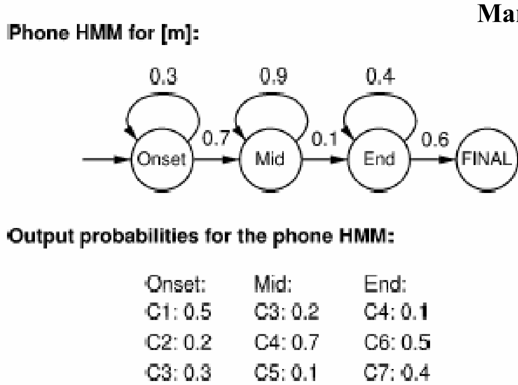
2 phones sequences

4 phones sequences

Speech recognition

HMM model of phones $P(\mathbf{s} = s_1 s_2 \ldots s_r \mid phone = p_q)$

Example:



Many possible feature sequences:

C1 C4 C6

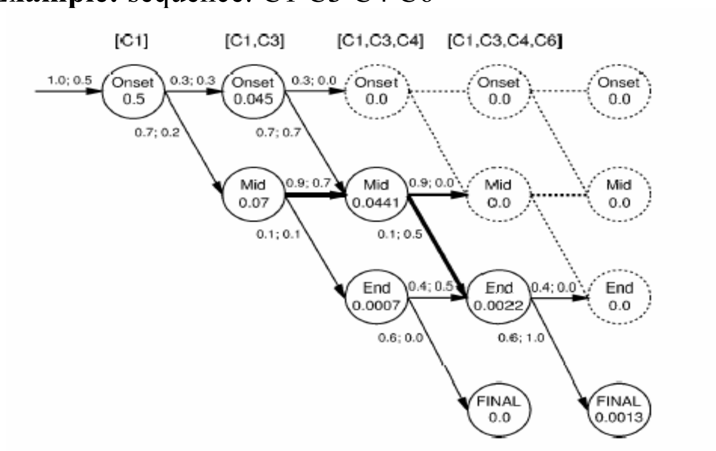
C1 C1 C4 C6

C1 C1 C5 C4 C6

…

Speech recognition

- **Finding the most probable path** through an HMM for [m]
- **Example:** sequence: C1 C3 C4 C6



Natural language processing

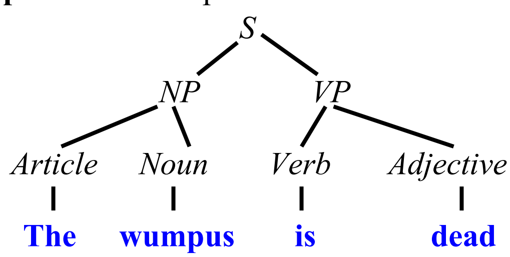
Goal: **Analyze and interpret the text in the natural language**

- **Input:** **text sentences**.
  - Speech recognition system
  - Optical character recognition (OCR)
  - Documents in the electronic form
- **Output:**
  - Knowledge extracted from the text that supports various inferences
- **Processing (multi-step process):**
  - Syntactic interpretation (parsing)
  - Semantic interpretation
  - Disambiguation & Incorporation

Natural language processing

Syntactic interpretation (parsing):

- **Input:** **a sentence**
- **Output:** **a parse tree**
- Uses grammar models for parsing the sentence to phrases and terminal symbols
- **Example:** ‘The wumpus is dead’



- Sometimes we have more than one possible parse. **Stochastic grammars** (quantify the goodness of possible parses)

Natural language processing

- **Semantic interpretation:**
  - **input:** **a parse tree**
  - **output:** **a set of meanings, e.g. in First order logic (FOL)**
- **Example:** ‘The wumpus is dead’
  - Gives two possible semantic interpretations:

$$\neg Alive(Wumpus, Now)$$
$$Tired(Wumpus, Now)$$

- **Disambiguation:**
  - chooses the most probable interpretation
- **Incorporation:**
  - The extracted knowledge is checked for consistency against other pieces of knowledge before it is incorporated into the KB

Image processing and vision

- **Classic image processing problem:**
  - **Analysis of image and extraction of information from the image**
  - **Can be used in many applications:**
    - Scene analysis
    - Manipulation and navigation tasks
    - Image retrieval
- **Other image processing problems:**
  - **Image enhancement:** degraded image should be improved to restore particular features
  - **Storage and Compression:** Large amounts of data need to be archived or transmitted
  - **Visualization**

Image processing

Image is defined by

- a **light intensity function** over the **image plane**

(Continuous) image is typically **discretized**

- **Image plane is discretized into:**
  - Pixels arranged on the rectangular grid
  - Resolution of the grid determines the spatial quality of the discretization
- **Light intensity values are discretized into:**
  - Integers values in some interval
- **Typical (black and white) image input:**
  - 512x512 pixels
  - Light intensity: 8 bits – 512 types of gray

Image processing

Analysis of image and extraction of information from the image

- **Segmentation:**
  - Division of the image to meaningful entities in the scene
  - Relies heavily on edge detection algorithms

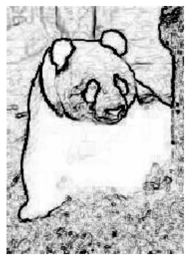




Image processing and vision

Analysis of image and extraction of information from the image

- To recognize (identify) the object from the image we need to compare it with the class pattern
- **Problem:** The position, orientation and the scale of the object in the scene may vary
- **Solution: Use a set of basic transformations:**
  - **scaling,**
  - **translation,**
  - **rotation of the object**
  - Transformations are relatively easy for 2D objects, much harder for 3-D objects
- **Other problems:** light sources and shadows

Image processing and vision

- **More complex task:** **analysis of a sequence of related images (videos)**
- **Image registration:** **the process of measuring visual motion between images.**
- **When this is useful:**
  - **Video - commercial skip**
  - **Detection and tracking of objects in the real world**