



Speech Processing 15-492/18-492

Speech Recognition

Grammars

Other ASR techniques

But not just acoustics

- But not all phones are equi-probable
- Find word sequences that maximizes

$$P(W | O)$$

- Using Bayes' Law

$$\frac{P(W)P(O|W)}{P(O)}$$

- Combine models

- Use HMMs to provide

$$P(O | W)$$

- Use language model to provide

$$P(W)$$

Beyond n-grams

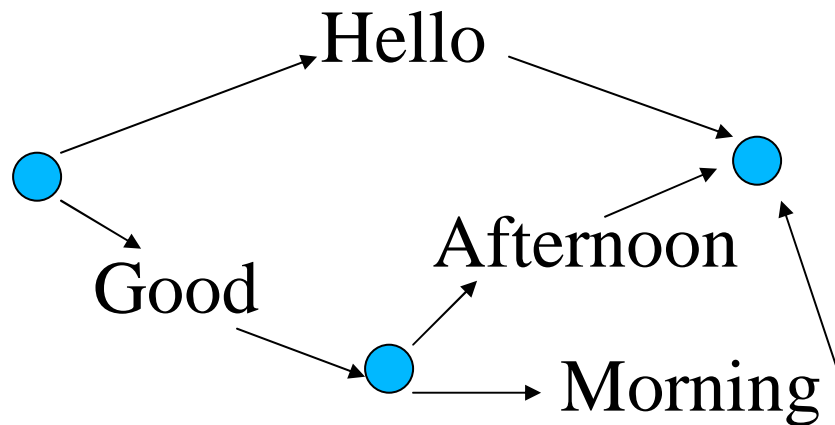
- ◆ *Tri-gram languages models*
 - *Good for general ASR*
- ◆ *More targeted models for dialog systems*
 - *Look for more structure*

Formal Language Theory

- ◆ *Chomsky Hierarchy*
 - *Finite State Machines*
 - *Context Free Grammars*
 - *Context Sensitive Grammars*
 - *Generalized Rewrite Rules/Turing machines*
- ◆ *As LM or as Understanding mechanism*
 - *Folded into the ASR or only ran on output*

Finite State Machines

- ◆ *Trigram is a word² FSM*
- ◆ *FSM for greeting*



Finite State Grammar

- ◆ *Sentences -> Start Greeting End*
- ◆ *Greeting -> "Hello"*
- ◆ *Greeting -> "Good" TOD*
- ◆ *TOD -> Morning*
- ◆ *TOD -> Afternoon*

Context Free Grammar

- ◆ $X \rightarrow YZ$
- ◆ $Y \rightarrow \text{"Terminal"}$
- ◆ $Y \rightarrow \text{NonTerminal NonTerminal}$

JSGF

- ◆ *Simple grammar formalism for ASR*
- ◆ *Standard for writing ASR grammars*
- ◆ *Actually finite state*
- ◆ <http://www.w3.org/TR/jsqf>

Finite State Machines

◆ *Finite State Machines:*

- *Deterministic*

- ⊗ *Each arc leaving a state has unique label*

- ⊗ *There always exists a Deterministic machine representing a non-Deterministic one*

- *Minimal*

- ⊗ *There exists an FSM with less (or equal) states that accepts the same language*

Probabilistic FSMs

- ◆ *Each arc has a label and a probability*
- ◆ *Collect probabilities from data*
 - *Can do smoothing like ngrams*

Natural Language Processing

- ◆ *Probably mildly context sensitive*
 - *i.e. you need context sensitive rules*
- ◆ *But if we only accept context free*
 - *Probably OK*
- ◆ *If we only accept finite state*
 - *Probably OK too*

Writing Grammars for Speech

- ◆ *What do people say?*
 - *No what do people *really* say!*
- ◆ *Write examples*
 - *Please, I'd like a flight to Boston*
 - *I want to fly to Boston*
 - *What do you have going to Boston*
 - *What about Boston*
 - *Boston*
- ◆ *Write rules grouping things together*

Ignore the unimportant things

- ◆ *I'm terribly sorry but I would greatly appreciate if you might be able to help me find an acceptable **flight to Boston**.*
- ◆ *I, I wanna want to go to ehm Boston.*

What do people really say

- ◆ *A: see who else will somebody else important all the {mumble} the whole school are out for a week*
- ◆ *B: oh really*
- ◆ *A: {lipsmack} {breath} yeah*
- ◆ *B: okay {breath} well when are you going to come up then*
- ◆ *A: um let's see well I guess I I could come up actually anytime*
- ◆ *B: okay well how about now*
- ◆ *A: now*
- ◆ *B: yeah*
- ◆ *A: have to work tonight –laugh-*

Class based language models

- ◆ *Conflate all words in same class*
 - *Cities, Names, numbers etc*
- ◆ *Can be automatic or designed*

Adaptive Language Models

- ◆ *Update with new News stories*
 - *Update your language model every day*
- ◆ *Update your language model with daily use*
 - *Using user generated data (if ASR is good)*

Combining models

- ◆ *Use “background” model*
 - *General tri-gram model*
- ◆ *Use specific model*
 - *Grammar based*
 - *Very localized*
- ◆ *Combine*
 - *Interpolated (just a weight factor)*
 - *More elaborate combinations*
 - ⊗ *Maximum entropy models*

Vocabulary size

- ◆ *Command and control*
 - *< 100 words, grammar based*
- ◆ *Simple dialog*
 - *< 1000 words, grammar/tri-gram*
- ◆ *Complex dialog*
 - *< 10K words, tri-gram (some grammar for control)*
- ◆ *Dictation*
 - *< 64K words, tri-gram*
- ◆ *Broadcast News*
 - *256K plus, tri-gram (and lots of other possibilities)*

Homework 1

- ◆ *Build a speech recognition system*
 - *An acoustic model*
 - *A pronunciation lexicon*
 - *A language model*
- ◆ *Note it takes time to build*
- ◆ *What is your initial WER*
 - *How did you improve it*
- ◆ *Submitted by 3:30pm Monday 29th Sep*

WFSTs